RANDOM WALKS

(lecture notes)

Luca Avena, Markus Heydenreich, Frank den Hollander, Evgeny Verbitskiy, Willem van Zuijlen

Mathematical Institute, Leiden University, The Netherlands

Preface

This syllabus contains the notes of a course on *Random Walks* offered at the Mathematical Institute of Leiden University. The course is aimed at second-year and third-year mathematics students who have completed an introductory course on probability theory. The goal of the course is to describe a number of topics from modern probability theory that are centred around random walks. Random walks are key examples of a random processes, and have been used to model a variety of different phenomena in physics, chemistry, biology and beyond. Along the way a number of key tools from probability theory are encountered and applied.

Chapter 1 lists basic properties of finite-length random walks, including space-time distributions, stopping times, the ruin problem, the reflection principle and the arcsine law. Chapter 2 describes basic limit theorems for infinite-length random walks, including the strong law of large numbers, the central limit theorem, the large deviation principle, and recurrence versus transience.

Chapter 3 shows how random walks can be used to describe electric flows on finite and infinite networks, which leads to estimates of effective resistances of such networks via the Dirichlet principle and the Thomson principle. Chapter 4 deals with self-avoiding walks, which are lattice paths constrained not to intersect themselves. The counting of such paths and the study of their spatial properties is quite challenging, especially in low dimensions. Chapter 5 focusses on random walks in the vicinity of an interface with which they interact, which serve as a model for polymer chains acting as surfactants. Each contact with the interface is either rewarded or penalised, and it is shown that a phase transition occurs from a desorbed phase to an adsorbed phase as the interaction strength increases.

Chapter 6 introduces Brownian motion, which is the space-time continuous analogue of random walk. Also Brownian motion is a key example of a random process. It arises as the scaling limit of random walk, has powerful scaling properties, and is the pillar of stochastic analysis, the area that deals with stochastic differential equations, i.e., differential equations with noise. Chapter 7, finally, treats a topic from finance, namely, the binomial asset pricing model for stock exchange. It introduces the notion of arbitrage, and uses random walks to compute the fair price of a certain financial derivative called option, which leads to the discrete version of the so-called Black-Scholes formula.

A rough indication of the pace at which the course can be taught is as follows (1 lecture = 2×45 minutes): Chapters 1+2: 3 or 4 lectures; Chapter 3: 3 lectures; Chapters 4+5: 3 lectures; Chapter 6: 1 or 2 lectures; Chapter 7: 2 lectures. Some sections in the lecture notes are marked with a $\star$. These sections contain more advanced material that provides important background but that is not required for the final examination.

The present notes were written up and polished during courses taught since the Spring of 2014. Comments by students continue to be welcome.

Luca Avena, Markus Heydenreich, Frank den Hollander, Evgeny Verbitskiy, Willem van Zuijlen
Februari 2016

Contents

Chapter 1

Finite-Length Random Walks

Random walks on discrete spaces are easy objects to study. Yet there is a certain complication coming from the uncountability of infinite-length paths. In order to circumvent this difficulty, we start in this chapter by considering finite-length random walks. The presentation in this chapter is based on unpublished notes of H. Föllmer.

We use this chapter to illustrate a number of useful concepts for *one-dimensional* random walk. In later chapters we will consider d-dimensional random walk as well. Section 1.1 provides the main definitions. Section 1.2 introduces the notion of stopping time, and looks at random walk from the perspective of a fair game between two players. Section 1.3 solves the classical problem of the "gambler's ruin". Section 1.4 proves the so-called reflection principe and shows how this can be used to derive laws of first hitting times. Section 1.5, finally, discusses the so-called arc sine law for last hitting times, which in a game setting records the last time the two players had the same capital.

1.1 Definition

Throughout the sequel we adopt the notation $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

We begin by considering *simple random walk* on $\mathbb{Z}$. Fix $N \in \mathbb{N}$. The configuration space is given by the binary sequences of length N, i.e.,

$$\Omega_N = \left\{ \omega = (\omega_1, \dots, \omega_N) \in \{-1, +1\}^N \right\}. \tag{1.1}$$

Write

$$X_k(\omega) = \omega_k, \qquad 1 \le k \le N,\ \omega \in \Omega_N, \tag{1.2}$$

to denote the projection on the k-th component of ω, which is to be thought of as the *step* of the random walk at time k. The *position* of the random walk after n steps (i.e., after n units of time) is given by

$$S_n(\omega) = \sum_{k=1}^{n} X_k(\omega), \qquad 1 \le n \le N, \qquad S_0(\omega) = 0. \tag{1.3}$$

In this way, for every $\omega \in \Omega_N$ we obtain a *trajectory* $(S_n)_{n=0}^N$, also called a path (see Fig. 1.1). As probability distribution on Ω_N we take the *uniform distribution*, i.e.,

$$P_N(A) = |A|\, 2^{-N}, \qquad A \subseteq \Omega_N. \tag{1.4}$$

This means that all binary sequences ω (equivalently, all trajectories) have the *same* probability.

Definition 1.1. The sequence of random variables $(S_n)_{n=0}^N$ on the finite probability space (Ω_N, P_N) is called a *simple random walk of length N starting at* 0. In what follows we suppress the index N from the notation.

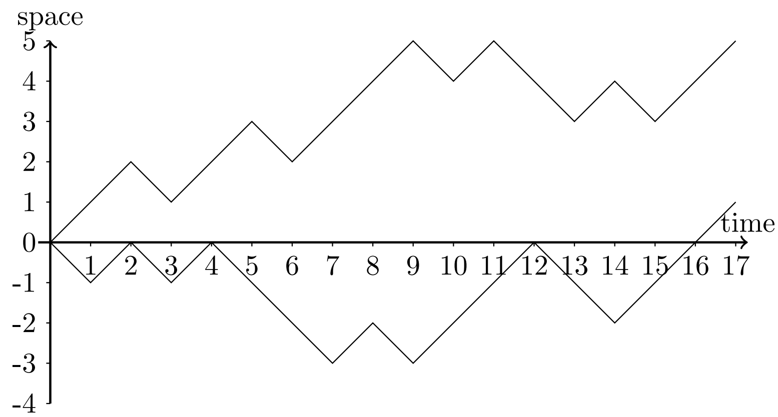

Figure 1.1: Two possible trajectories for $(S_n)_{n=0}^N$.

It follows from (1.4) that for $1 \le k_1 < \cdots < k_n \le N$ and $x_{k_i} \in \{-1,1\}$, $i = 1,\ldots,n$,

$$P(X_{k_1} = x_{k_1},\ldots,X_{k_n} = x_{k_n}) = 2^{N-n}2^{-N} = 2^{-n}. \tag{1.5}$$

Exercise 1.1. (a) Use this to conclude that $X_1,\ldots,X_N$ are independent and identically distributed with $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$.
(b) Prove that a simple random walks has *independent increments*, which means that for all $0 < k_1 < k_2, \cdots < k_n \le N$ the vectors $S_{k_1} - S_0, S_{k_2} - S_{k_1},\ldots,S_{k_n} - S_{k_{n-1}}$ are independent.
(c) Prove that an increment $S_m - S_k$ for $0 < k < m \le N$ has the same distribution as S_{m-k}, which means that $P(S_m - S_k = a) = P(S_{m-k} = a)$ for all $a \in \mathbb{Z}$.
(d) Prove that a simple random walk satisfies the Markov property: $P(S_n = a_n|S_{n-1} = a_{n-1},\ldots,S_1 = a_1) = P(S_n = a_n|S_{n-1} = a_{n-1})$ for $0 < n \le N$ and $a_1,\ldots,a_n \in \mathbb{Z}$ (such that $P(S_{n-1} = a_{n-1},\ldots,S_1 = a_1) > 0$).
(e) Prove that for $0 < k < m \le N$ one has $P(S_m = b|S_k = a) = P(S_{m-k} = b-a)$ for $a,b \in \mathbb{Z}$ (assuming $P(S_k = a) > 0$).

We obtain $E(X_k) = 0$, $E(X_k^2) = 1$, $k = 1,\ldots,N$ and $E(X_k X_L) = 0$ for $k \ne l$. This leads us to our first result:

Claim 1.2. $E(S_n) = 0$, $E(S_n^2) = n$.

Proof. Compute $E(S_n) = \sum_{k=1}^n E(X_k) = 0$ and $E(S_n^2) = \sum_{k,l=1}^n E(X_k X_l) = n$. □

It is in fact easy to determine the distribution of S_n:

Claim 1.3. *For $x \in \{-n,-n+2,\ldots,n-2,n\}$, the probability that simple random walk is in x after n steps equals (see Fig. 1.2)*

$$P(S_n = x) = \binom{n}{\frac{n+x}{2}} 2^{-n}. \tag{1.6}$$

Proof. Observe that $S_n = x$ if and only if the first n components of ω take precisely $k = \frac{n+x}{2}$ times the value +1. Indeed, then $S_n(\omega) = k(+1) + (n-k)(-1) = 2k - n = x$. Hence $|\{\omega \in \Omega\colon\ S_n(\omega) = x\}| = \binom{n}{k}2^{N-n}$ and

$$P(S_n = x) = |\{\omega \in \Omega\colon\ S_n(\omega) = x\}|\,2^{-N} = \binom{n}{k} 2^{-n}. \tag{1.7}$$

□

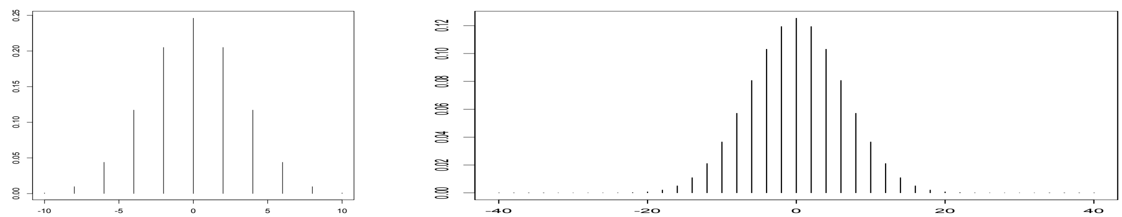

Figure 1.2: Plot of $P(S_n = x)$, $x \in \mathbb{Z}$, for $n = 10$ (left) and $n = 40$ (right).

Remark 1.4. Several observations are in order.

(i) The distribution of S_n is *symmetric around 0*:

$$P(S_n = x) = \frac{n!}{(\frac{n-x}{2})!\,(\frac{n+x}{2})!}\, 2^{-n} = P(S_n = -x). \tag{1.8}$$

(ii) The *maximal weight* of the distribution (also called mode) is achieved in the middle:

$$P(S_{2n} = 0) = P(S_{2n-1} = 1) = \binom{2n}{n} 2^{-2n}. \tag{1.9}$$

(The identity $2\binom{2n-1}{n} = \binom{2n}{n}$ is used.)

(iii) Stirling's formula says that $n! \sim n^n e^{-n} \sqrt{2\pi n}$ as $n \to \infty$, where $\sim$ means that the quotient of the left-hand side and the right-hand side tends to 1 (i.e., $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$). From this one deduces (see Exercise 1.2)

$$P(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}, \qquad n \to \infty. \tag{1.10}$$

In particular,

$$\lim_{n\to\infty} P(a \le S_n \le b) = 0 \quad \text{for any finite interval } [a, b], \tag{1.11}$$

because (ii) and (1.10) imply that

$$P(a \le S_n \le b) \le (b - a + 1)\, P(S_n \in \{0, 1\}) \to 0, \qquad n \to \infty. \tag{1.12}$$

Exercise 1.2. (a) Deduce (1.10) from (1.8) with the help of Stirling's formula.
(b) In Chapter 5 we will need the finer asymptotics $P(S_{2n} = 0) - P(S_{2n+2} = 0) \sim 1/(2n\sqrt{\pi n})$, $n \to \infty$. Deduce this from (1.10). *Hint*: Show that $P(S_{2n} = 0) - P(S_{2n+2} = 0) = \frac{1}{2(n+1)} P(S_{2n} = 0)$.

Note: In order to be able to pass to the limit $n \to \infty$, we need to let $N \to \infty$ as well. In Chapter 2 we will deal with the situation where $N = \infty$.

1.2 Stopping times and games

We can interpret a simple random walk as a game, where in round k a player wins the amount X_k. Then S_n is the "capital" of the player after n rounds. We have seen that $E(S_n) = 0$ for any $0 \le n \le N$. In other words, the expected "gain" after n rounds is 0. An interesting question reads: "Is it possible to *stop* the game in a favorite moment, i.e., can clever stopping lead to a positive expected gain?" Of course, the decision to stop may only depend on the trajectory until that time: no "insider knowledge" about the future of the trajectory is permitted. To answer this question we need to formalise the setting.

Definition 1.5. An event $A \subseteq \Omega$ is *observable until time n* when it can be written as a union of basic events of the form

$$\{\omega \in \Omega : \omega_1 = o_1, \dots, \omega_n = o_n\}, \qquad o_1, \dots, o_n \in \{-1, +1\}. \tag{1.13}$$

Write $\mathscr{A}_n$ to denote the class of events A that are observable until time n (which includes the empty union).

We further define the *indicator function* $\mathbb{1}_A$ for an event $A \subseteq \Omega$ to be the random variable

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases} \tag{1.14}$$

Note that $\{\emptyset, \Omega\} = \mathscr{A}_0 \subseteq \mathscr{A}_1 \subseteq \cdots \subseteq \mathscr{A}_N = \{\text{the set of all subsets of } \Omega\}$. Further note that $\mathscr{A}_n$ is closed with respect to taking union, intersection and complement. A sequence $(\mathscr{A}_n)_{n=0}^N$ with these properties is called a *filtration*. Our filtration has the following properties:

Lemma 1.6. *For $n = 0, 1, \dots, N-1$ and $A_n \in \mathscr{A}_n$,*

$$P\big(A_n \cap \{X_{n+1} = +1\}\big) = \tfrac{1}{2}\, P(A_n), \qquad E\big(X_{n+1} \mathbb{1}_{A_n}\big) = 0. \tag{1.15}$$

Exercise 1.3. Prove Lemma 1.6.

For a random variable $Y : \Omega \to \mathbb{Z}$ we use the following notation for $a \in \mathbb{Z}$

$$\{Y = a\} = \{\omega \in \Omega : Y(\omega) = a\}, \tag{1.16}$$

similarly we write $\{Y \geq a\} = \{\omega \in \Omega : Y(\omega) \geq a\}$, analogously $\{Y \leq a\}$ and $\{Y \in A\} = \{\omega \in \Omega : Y(\omega) \in A\}$ for $A \subset \mathbb{Z}$.

Definition 1.7. A map $T\colon\ \Omega \to \{0, \dots, N\} \cup \{\infty\}$ is called a *stopping time* if

$$\{T = n\} = \{\omega \in \Omega\colon\ T(\omega) = n\} \in \mathscr{A}_n, \qquad n = 0, \dots, N. \tag{1.17}$$

Note: Since $n \mapsto \mathscr{A}_n$ is non-decreasing, it follows that $T\colon\ \Omega \to \{0, \dots, N\}$ is a stopping time if and only if $\{T \leq n\} \in \mathscr{A}_n$, $n = 0, \dots, N$.

Example. For $a \in \mathbb{Z}$, let

$$\sigma_a(\omega) = \inf\{n \in \mathbb{N}\colon\ S_n(\omega) = a\} \tag{1.18}$$

denote the *first hitting time* of a after time 0 (with $\inf \emptyset = \infty$), which is the *first return time* to a when $S_0(\omega) = a$. Since $\{\sigma_a = n\} \in \mathscr{A}_n$, we have that $\min\{\sigma_a, N\}$ is a stopping time.

We are now ready to give an answer to the question posed at the beginning of this section. The answer is somewhat unsatisfying for those who like a good gamble.

Theorem 1.8 (Impossibility of profitable stopping)**.** *For any stopping time $T\colon\ \Omega_N \to \{0, \dots, N\}$,*

$$E[S_T] = 0, \tag{1.19}$$

where $S_T = S_{T(\omega)}(\omega)$ is the outcome of the trajectory ω at the stopping time $T(\omega)$.

Proof. For every $k = 0,\dots,N$ we have $\{T \ge k\} \in \mathscr{A}_{k-1}$, because

$$\{T \ge k\}^c = \bigcup_{\ell=0}^{k-1} \{T = \ell\} \in \mathscr{A}_{k-1}. \tag{1.20}$$

Since

$$S_T = \sum_{k=1}^{N} X_k \mathbb{1}_{\{T \ge k\}}, \tag{1.21}$$

it follows from (1.15) and (1.20) that

$$E(S_T) = \sum_{k=1}^{N} E\big(X_k \mathbb{1}_{\{T \ge k\}}\big) = 0. \tag{1.22}$$

□

Theorem 1.8 is a special case of general theorem about the *impossibility of profitable games*, which we state next. A *game system* is a sequence of $\mathbb{R}$-valued random variables $V_1, V_2, \dots, V_N\colon\ \Omega_N \to \mathbb{R}$ such that

$$\{V_k = c\} \in \mathscr{A}_{k-1}, \qquad c \in \mathbb{R},\ k = 1,2,\dots,N. \tag{1.23}$$

The interpretation is that in the k-th round you bet the amount V_k, so that the result of the k-th round is $V_k X_k$. Mind that V_k can be positive, zero or negative. The total gain of the game is

$$S_N^V(\omega) = \sum_{k=1}^{N} V_k(\omega)\, X_k(\omega). \tag{1.24}$$

Theorem 1.9 (Impossibility of profitable games)**.** *For any game system $V_1, V_2, \dots, V_N$, the expected gain vanishes: $E(S_N^V) = 0$.*

Proof. It is sufficient to show that $E(V_k\, X_k) = 0$ for all $1 \le k \le N$ (because of the linearity of expectation). Now, V_k can be written as $V_k = \sum_{i=1}^{M} c_i \mathbb{1}_{\{V_k = c_i\}}$ for certain $M \in \mathbb{N}$ and $c_1, c_2, \dots, c_M \in \mathbb{R}$. Furthermore, $\{V_k = c_i\} \in \mathscr{A}_{k-1}$ for all $i = 1,\dots,M$, so that (1.15) implies

$$E(V_k\, X_k) = \sum_{i=1}^{M} c_i\, E\big(X_k \mathbb{1}_{\{V_k = c_i\}}\big) = 0. \tag{1.25}$$

□

Note that any stopping time T can be written as a game system by putting $V_k = \mathbb{1}_{\{T \ge k\}}$, so that

$$S_N^V = \sum_{k=1}^{N} X_k\, \mathbb{1}_{\{T \ge k\}} = \sum_{k=1}^{T} X_k = S_T. \tag{1.26}$$

Thus, Theorem 1.8 is indeed a special case of Theorem 1.9.

Theorem 1.9 is quite powerful. We can use it to derive a link between the expected value of stopping times and the variance of payoffs. To this end, let T be a stopping time, and consider the game system

$$V_k = S_{k-1}\, \mathbb{1}_{\{T \ge k\}}, \qquad k = 1,\dots,N. \tag{1.27}$$

Since

$$S_k^2 = (S_{k-1} + X_k)^2 = S_{k-1}^2 + 2S_{k-1}X_k + 1, \tag{1.28}$$

we have

$$V_k X_k = \tfrac{1}{2}\left(S_k^2 - S_{k-1}^2 - 1\right) \mathbb{1}_{\{T \ge k\}}. \tag{1.29}$$

Summing over $k = 1, \dots, N$ and recalling (1.24), we get

$$S_N^V = \tfrac{1}{2}\left(S_T^2 - T\right). \tag{1.30}$$

Theorem 1.9 implies that the expected value of (1.30) vanishes, and so we arrive at the following identity for the variance of S_T.

Corollary 1.10. *For any stopping time T,*

$$\operatorname{Var}(S_T) = E\big(S_T^2\big) = E(T). \tag{1.31}$$

1.3 The ruin problem

- *"Millionaires should always gamble, poor men never."* [J. M. Keynes]

We continue the interpretation in the previous section of the N-step simple random walk as an N-round game between two players, A and B, where in each round player A wins 1 Euro from player B with probability $\frac{1}{2}$, or loses 1 Euro with probability $\frac{1}{2}$. In this setting, S_n expresses the gain of player A after n rounds (and $-S_n$ the loss of player B after n rounds).

Let $a, b \in \mathbb{N}$. We interpret a and b as the *initial capital* of players A and B. Denote by σ_{-a} and σ_b the first hitting time of the states $-a$ and b, respectively. The event

$$\{\sigma_{-a} < \sigma_b,\ \sigma_{-a} \le N\} \tag{1.32}$$

expresses *ruin of player A* after N rounds (all his capital is lost). We are interested in the *ruin probabilities* of players A and B:

$$r_N^A = P\big(\sigma_{-a} < \sigma_b,\ \sigma_{-a} \le N\big), \qquad r_N^B = P\big(\sigma_b < \sigma_{-a},\ \sigma_b \le N\big), \tag{1.33}$$

in particular, the limits

$$r^A = \lim_{N\to\infty} r_N^A, \qquad r^B = \lim_{N\to\infty} r_N^B, \tag{1.34}$$

which exist by monotone convergence.

We can calculate r^A and r^B as follows. Since the game is *fair* (= expected gain is 0), r^A and r^B only depend on the initial capital a and b.

Exercise 1.4. Prove that the minimum of two stopping times is again a stopping time. *Hint*: One of the two equivalent definitions of a stopping time in Definition 1.7 is more useful here than the other.

Hence $T_N = \min\{\sigma_{-a}, \sigma_b, N\}$ is a stopping time. Therefore, by Theorem 1.8,

$$0 = E(S_{T_N}) = -a r_N^A + b r_N^B + E\big(S_N \mathbb{1}_{\{\min\{\sigma_{-a},\sigma_b\} > N\}}\big). \tag{1.35}$$

The last term is bounded from above by $\max\{a,b\}\, P(-a \le S_N \le b)$, and therefore vanishes in the limit as $N \to \infty$ by (1.11). Consequently,

$$-a r^A + b r^B = 0. \tag{1.36}$$

On the other hand,

$$1 - (r_N^A + r_N^B) = P\big(\min\{\sigma_{-a}, \sigma_b\} > N\big) \le P(-a \le S_N \le b) \to 0, \qquad N \to \infty, \tag{1.37}$$

and hence

$$r^A + r^B = 1. \tag{1.38}$$

We thus have two linear equations for r^A, r^B, which we can solve as

$$r^A = \frac{b}{a+b}, \qquad r^B = \frac{a}{a+b}. \tag{1.39}$$

This is the solution of the classical *gambler's ruin* problem.

How long will we typically have to wait until one of the players is ruined? The *expected waiting time for a ruin* can be computed with the help of Corollary 1.10:

$$E(T_N) = E\big(S_{T_N}^2\big) = a^2 r_N^A + b^2 r_N^B + E\big(S_N^2\,\mathbb{1}_{\{\min\{\sigma_{-a},\sigma_b\}>N\}}\big) \to a^2 r^A + b^2 r^B, \qquad N \to \infty. \tag{1.40}$$

Filling in the values for r^A and r^B, we get

$$\lim_{N\to\infty} E(T_N) = ab. \tag{1.41}$$

1.4 The reflection principle

Let $a \in \mathbb{N}$, and recall from (1.18) that $\sigma_a = \min\{n \in \mathbb{N}\colon\ S_n = a\}$ is the first hitting time of a after time 0.

Lemma 1.11 (Reflection Principle)**.** *For $a, c \in \mathbb{N}$,*

$$P\big(S_n = a - c, \sigma_a \le n\big) = P\big(S_n = a + c\big). \tag{1.42}$$

Proof. The proof follows from the observation that the number of n-step paths that first visit a and afterwards end in $a - c$ is equal to the number of n-step paths that end in $a + c$ (see Fig. 1.3). Recall that all n-step paths have the same probability 2^{-n}. □

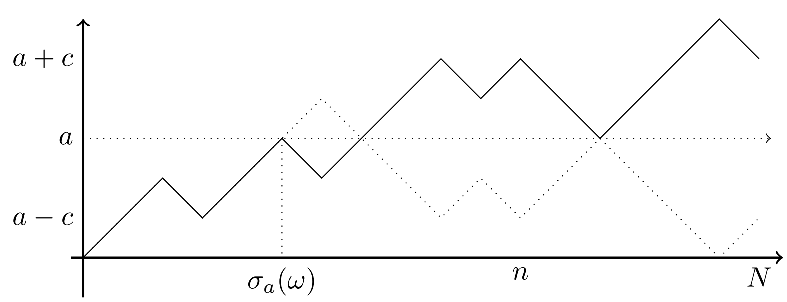

Figure 1.3: Illustration of the Reflection Principle.

Lemma 1.11 enables us to compute the law of the random variable σ_a.

Theorem 1.12. *For $a, n \in \mathbb{N}$,*

$$P\big(\sigma_a \le n\big) = P\big(S_n \notin [-a, a-1]\big), \tag{1.43}$$

$$P\big(\sigma_a = n\big) = \tfrac{1}{2}\big[P(S_{n-1} = a-1) - P(S_{n-1} = a+1)\big]. \tag{1.44}$$

Proof. The reflection principle is at the very heart of the proof. Write

$$P\big(\sigma_a \le n\big) = \sum_{b\in\mathbb{Z}} P\big(S_n = b, \sigma_a \le n\big) \tag{1.45}$$

$$= \sum_{\substack{b\in\mathbb{Z}\\ b\ge a}} P\big(S_n = b\big) + \sum_{\substack{b\in\mathbb{Z}\\ b<a}} P\big(S_n = b, \sigma_a \le n\big). \tag{1.46}$$

By the reflection principle in Lemma 1.11, the right-hand side equals $P(S_n \geq a) + P(S_n > a)$. By symmetry, this is equal to $P(S_n \geq a) + P(S_n < -a)$, which proves (1.43).

To get (1.44), write $P(\sigma_a = n) = P(\sigma_a \leq n) - P(\sigma_a \leq n-1)$, and apply (1.43).

Exercise 1.5. Write out the computation.

We expose a different route, one that introduces the useful concept of *time reversal*. Namely, note that (cf. Figure 1.4)

$$\begin{aligned}
&\#n\text{-step paths from } 0 \text{ to } a \text{ with } \sigma_a = n \\
&\quad = \#(n-1)\text{-step paths from } 0 \text{ to } a-1 \text{ without visit to } a \\
&\quad = \#(n-1)\text{-step paths from } a-1 \text{ to } 0 \text{ without visit to } a \\
&\quad = \#(n-1)\text{-step paths from } 0 \text{ to } 1-a \text{ without visit to } 1. \qquad (1.47)
\end{aligned}$$

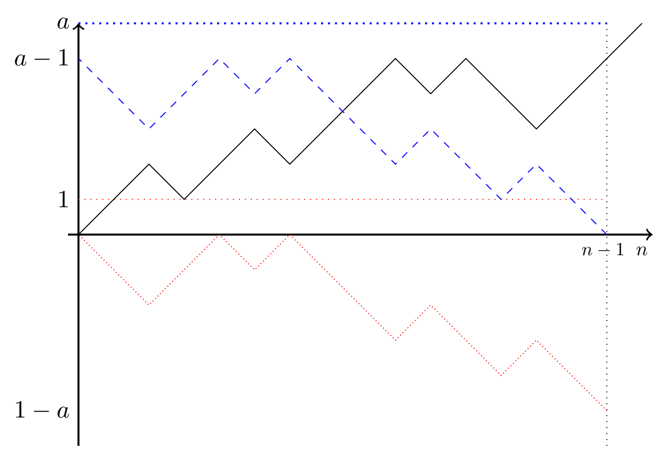

Figure 1.4: Illustration of (1.47).

Consequently,

$$\begin{aligned}
P(\sigma_a = n) &= P(S_{n-1} = 1-a, \sigma_1 > n-1, X_n = 1) \\
&= \tfrac{1}{2} P(S_{n-1} = 1-a, \sigma_1 > n-1) \\
&= \tfrac{1}{2} \Big[P(S_{n-1} = 1-a) - P(S_{n-1} = 1-a, \sigma_1 \leq n-1) \Big], \qquad (1.48)
\end{aligned}$$

and the last line, by symmetry and Lemma 1.11, equals $\frac{1}{2}[P(S_{n-1} = a-1) - P(S_{n-1} = 1+a)]$. □

Theorem 1.12 is very useful when it comes to *computing* distributions of certain observables of the random walk, as we show in the three corollaries below.

Corollary 1.13 (Hitting time distribution)**.** *For $a, n \in \mathbb{N}$,*

$$P\big(\sigma_a = n\big) = \frac{a}{n} P\big(S_n = a\big). \qquad (1.49)$$

Proof. Let $k(a,n) = \frac{a+n}{2}$, and note that $k(a,n) = k(a+1,n-1) = k(a-1,n-1)+1$. Note further that

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \qquad \binom{n-1}{k} = \frac{n-k}{n}\binom{n}{k}. \qquad (1.50)$$

Hence, by (1.6),

$$P(S_{n-1}=a-1)=2^{-(n-1)}\binom{n-1}{k(a,n)-1}=2\,\frac{k(a,n)}{n}\,P(S_n=a) \tag{1.51}$$

and

$$P(S_{n-1}=a+1)=2^{-(n-1)}\binom{n-1}{k(a+1,n-1)}=2\,\frac{n-k(a,n)}{n}\,P(S_n=a). \tag{1.52}$$

Inserting these identities into (1.44), we get the claim. □

Corollary 1.13 combined with (1.6) gives us a way to compute the law of hitting times.

Corollary 1.14 (Escape time distribution)**.** *For $n\in\mathbb{N}$,*

$$P(\sigma_0>2n)=P(S_{2n}=0). \tag{1.53}$$

Proof. Theorem 1.12 is not immediately applicable, since it requires that $a\in\mathbb{N}$ while here $a=0$. In order to apply the theorem, we first have to bring the problem into a suitable form. We start by writing

$$\begin{aligned}P(\sigma_0>2n)&=P(S_1\neq 0,\dots,S_{2n}\neq 0)\\&=2\,P(S_1>0,\dots,S_{2n}>0)\\&=2\,2^{-2n}\,\#\,(2n-1)\text{-step paths that start in 1 without visits to 0}\\&=2\,2^{-2n}\,\#\,(2n-1)\text{-step paths that start in 0 without visits to }-1\\&=P(\sigma_{-1}>2n-1)=P(\sigma_1>2n-1).\end{aligned} \tag{1.54}$$

Now we can apply Theorem 1.12. Indeed, using (1.43) with $a=1$ and (1.9), we can rewrite the right-hand side of (1.54) as

$$P(S_{2n-1}\in\{-1,0\})=P(S_{2n-1}=-1)=P(S_{2n}=0). \tag{1.55}$$

□

Remark 1.15. We see from (1.10) and (1.53) that

$$\lim_{n\to\infty}P(\sigma_0>2n)=0. \tag{1.56}$$

However, Corollary 1.14 tells us that

$$E(\sigma_0)=\sum_{n\in\mathbb{N}_0}P(\sigma_0>n)=2\sum_{n\in\mathbb{N}_0}P(\sigma_0>2n)=2\sum_{n\in\mathbb{N}_0}P(S_{2n}=0)=\infty, \tag{1.57}$$

where the last equality follows from (1.10). Hence, one-dimensional simple random walk is *recurrent*, i.e., with probability 1 returns to 0 eventually, but with a large probability we have to wait *a very long time* until its first return to 0 happens. Random walks with the property in (1.56)–(1.57) are called *null-recurrent.*

As another application of Theorem 1.12, we compute the probability that the random walk reaches a certain level $a\in\mathbb{N}$ *without returning to the origin.*

Corollary 1.16 (Crossing probabilities)**.** *For $a,n\in\mathbb{N}$,*

$$P(S_n=a,\sigma_0>n)=\frac{a}{n}\,P(S_n=a). \tag{1.58}$$

Proof. The proof is yet another example of the *time inversion* technique. Note that

$$\begin{aligned}&\#n\text{-step paths from 0 to } a \text{ without return to 0}\\&\quad=\#n\text{-step paths from } a \text{ to 0 without visit to 0 in between}\\&\quad=\#n\text{-step paths from 0 to } a \text{ with } \sigma_a=n\end{aligned} \tag{1.59}$$

and use Corollary 1.13. □

1.5 ⋆he arc sine law for the last visit to the origin

So far we have computed the probability distributions of *first* visits to certain states. We will now consider the *last* visit to the origin before time $2N$:

$$L = \max\{0 \le n \le 2N : S_n = 0\}. \tag{1.60}$$

Note that L is *not* a stopping time. In the game interpretation of Section 1.3, L is the time when one of the two players takes the lead for the rest of the game. In view of (1.10), we might guess that $L/2N \approx 1$. However, the answer is a little more complex.

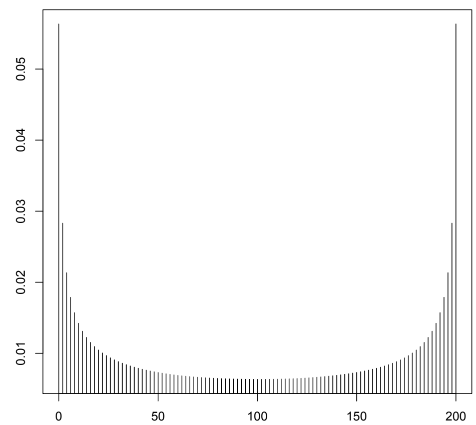

Figure 1.5: Plot of $P(L = n)$ for $n = 0, \dots, 200$ and $N = 200$.

Theorem 1.17. *For $n \in \mathbb{N}_0$ with $n \le N$,*

$$P(L = 2n) = P(S_{2n} = 0)\, P(S_{2N-2n} = 0) = 2^{-2N} \binom{2n}{n} \binom{2N-2n}{N-n}. \tag{1.61}$$

Proof. The number of paths of length $2N$ with $L = 2n$ is equal to the number of paths of length $2n$ with $S_{2n} = 0$ times the number of paths of length $2N - 2n$ with $\sigma_0 > 2N - 2n$. Hence

$$P(L = 2n) = P(S_{2n} = 0)\; P(\sigma_0 > 2N - 2n) = P(S_{2n} = 0)\; P(S_{2N-2n} = 0) \tag{1.62}$$

by Corollary 1.14. □

The distribution in Theorem 1.17 is called the *discrete arc sine distribution* (see Fig. 1.5). Note that it is symmetric around N, and has peaks for small and for large values of n. Apparently, either the winner takes the lead early on, or the game is tight until the very end.

What is the reason for the name *arc sine distribution*? We again use Stirling's formula to approximate the binomial coefficients in (1.62), to write

$$P(L = 2n) \sim \frac{1}{\pi\sqrt{n(N-n)}} \sim \frac{1}{N} f\left(\frac{n}{N}\right), \qquad n, N, N-n \to \infty, \tag{1.63}$$

with f the function $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$, $x \in [0,1]$. Hence

$$P\left(\frac{L}{2N} \le x\right) \sim \sum_{\substack{n\in\mathbb{N}_0 \\ \frac{n}{N} \le x}} \frac{1}{N} f\left(\frac{n}{N}\right) \sim \int_0^x f(y)\, dy = \frac{2}{\pi} \arcsin\sqrt{x}, \qquad N \to \infty. \tag{1.64}$$

Chapter 2

Infinite-Length Random Walks

We extend the set-up of Chapter 1 in two ways: (1) we consider simple random walk on the d-dimensional lattice $\mathbb{Z}^d$, $d \in \mathbb{N}$, rather than on the integers $\mathbb{Z}$; (2) we extend the probability space to deal with an infinite time horizon, abandoning the finite time horizon N that was used in Chapter 1. These extensions are made in Section 2.1. After that we state three basic limit theorems, formulated in Sections 2.2–2.4. In Section 2.5 we address the question of recurrence versus transience.

2.1 Definitions

2.1.1 Higher dimension

Recall the definitions given at the beginning of Chapter 1. Fix $d \in \mathbb{N}$. For $x \in \mathbb{Z}^d$, write

$$|x| = \left(\sum_{j=1}^{d} x_j^2\right)^{1/2}, \tag{2.1}$$

where x_j is the j-th component of the vector x. For given $N \in \mathbb{N}$, we have

$$\Omega_N = \left\{\omega = (\omega_1, \ldots, \omega_N) \colon\ \omega_k \in \mathbb{Z}^d,\ |\omega_k| = 1\ \forall\, 1 \le k \le N\right\} \tag{2.2}$$

as the configuration space, replacing (1.1). As before, $X_k(\omega) = \omega_k$, $1 \le k \le N$, and $S_n(\omega) = \sum_{k=1}^{n} X_k(\omega)$, $1 \le n \le N$, and $S_0(\omega) = 0$. Furthermore, we still have the uniform distribution on all paths of length N:

$$P_N(A) = |A|\,(2d)^{-N}, \qquad A \subseteq \Omega_N, \tag{2.3}$$

which replaces (1.4). Note that now S_n is a d-dimensional random vector

$$S_n = \begin{pmatrix} S_n^{(1)} \\ S_n^{(2)} \\ \vdots \\ S_n^{(d)} \end{pmatrix}, \qquad 0 \le n \le N, \quad S_n^{(j)} \in \mathbb{Z},\ j = 1, \ldots, d. \tag{2.4}$$

2.1.2 Infinite time horizon

To extend the probability space to infinite trajectories, i.e., $N = \infty$, requires a more argument subtle. However, there is a standard way to deal with this problem.

Let $0 < N < M$, and denote by $\pi_N\colon \Omega_M \to \Omega_N$ the projection

$$\pi_N(\omega_1,\dots,\omega_N,\omega_{N+1},\dots\omega_M) = (\omega_1,\dots,\omega_N). \tag{2.5}$$

Then the sequence of probability spaces $(\Omega_1,P_1),(\Omega_2,P_2),\dots$ satisfies the *compatibility condition*

$$P_M\big(\{\omega\in\Omega_M\colon\ \pi_N\omega=\bar\omega\}\big) = \frac{(2d)^{M-N}}{(2d)^M} = \frac{1}{(2d)^N} = P_N\big(\{\bar\omega\}\big), \qquad 0<N<M, \quad \bar\omega\in\Omega_N, \tag{2.6}$$

sometimes also referred to as the *consistency condition*. The so-called *Kolmogorov extension theorem* (see Fig. 2.1) states that for any sequence $(\Omega_1,P_1),(\Omega_2,P_2),\dots$ satisfying the above compatibility condition there exists a unique probability measure P on the space of infinite sequences $\Omega=\Omega_\infty$, i.e.,

$$\Omega_\infty = \{\omega=(\omega_1,\omega_2,\dots)\colon\ \omega_k\in\mathbb{Z}^d,\ |\omega_k|=1\ \forall k\in\mathbb{N}\}, \tag{2.7}$$

such that (with π_N defined on Ω_∞ analogously as on Ω_M as in (2.5))

$$P\big(\{\omega\in\Omega\colon\ \pi_N\omega=\bar\omega\}\big) = \frac{1}{(2d)^N} = P_N\big(\{\bar\omega\}\big), \qquad N\in\mathbb{N}, \quad \bar\omega\in\Omega_N. \tag{2.8}$$

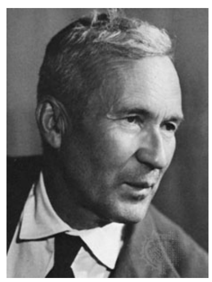

Figure 2.1: Andrey Kolmogorov, the founding father of probability theory.

Definition 2.1. The sequence of random variables $(S_n)_{n\in\mathbb{N}_0}$ on the infinite probability space (Ω,P) is called *simple random walk starting at* 0.

We will also define random walks starting at x, for some $x\in\mathbb{Z}^d$. Whenever $(S_n)_{n\in\mathbb{N}_0}$ is a random walk on (Ω,P) starting at 0 we could consider $(S_n+x)_{n\in\mathbb{N}}$. Usually instead we will define another probability P_x on Ω such that the law of $(S_n)_{n\in\mathbb{N}}$ under P_x is the same as the law of $(S_n+x)_{n\in\mathbb{N}}$ under P, so that for all $N\in\mathbb{N}_0$

$$\begin{aligned} &P_x(S_1=a_1,\dots,S_N=a_N) = P(S_1+x=a_1,\dots,S_N+x=a_N)\\ &(\text{and } E_x(f(S_1,\dots,S_N)) = E(f(S_1+x,\dots,S_N+x))). \end{aligned} \tag{2.9}$$

Whenever $m\in\mathbb{N}$ is such that $P(S_m=x)>0$, then one also has the equality $P_x(S_1=a_1,\dots,S_N=a_N) = P(S_{m+1}=a_1,\dots,S_{m+N}=a_N|S_m=x)$. We will also say P_x is the law of S given $S_0=x$.

In the literature, the term *random walk* refers to any stochastic process of independent and identically distributed jumps drawn from $\mathbb{Z}^d$. For the purpose of this course we limit our attention to the special case of simple random walk considered here. Random walks are the simplest examples of discrete-time stochastic processes. In Chapter 4 we will encounter random objects derived from random walks that cannot be interpreted as stochastic processes, namely, self-avoiding walks.

Definition 2.2. Similar to Definition 1.5 one defines (but with $\Omega = \Omega_\infty$, the space of infinite sequences) an event $A \subseteq \Omega$ to be *observable until time n* when it can be written as a union of basic events of the form

$$\{\omega \in \Omega : \omega_1 = o_1, \dots, \omega_n = o_n\}, \qquad o_1, \dots, o_n \in \mathbb{Z}^d, |o_i| = 1. \tag{2.10}$$

Write $\mathscr{A}_n$ to denote the class of events A that are observable until time n.
A map $T\colon\ \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called a *stopping time* if

$$\{T = n\} = \{\omega \in \Omega\colon\ T(\omega) = n\} \in \mathscr{A}_n, \qquad n \in \mathbb{N}_0. \tag{2.11}$$

For $A \subset \mathbb{Z}^d$ we define $\tau_A : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ by

$$\tau_A(\omega) = \inf\{n \in \mathbb{N}_0 : S_n(\omega) \in A\}. \tag{2.12}$$

For $a \in \mathbb{Z}^d$ we will write $\tau_a = \tau_{\{a\}}$.

Exercise 2.1. Show that τ_A is a stopping time.

One has the following facts (these can be proved, but the proof is rather administrative); for $z \in \mathbb{Z}^d$, $o \in \mathbb{Z}^d$, $|o| = 1$, $A, B \subset \mathbb{Z}^d$ with $z \notin A$ and $z \notin B$ and for $k \in \mathbb{N}_0$

$$P_z(\tau_A = k | S_1 = z + o) = P_{z+o}(\tau_A = k - 1), \tag{2.13}$$

$$P_z(\tau_A < \tau_B | S_1 = z + o) = P_{z+o}(\tau_A < \tau_B). \tag{2.14}$$

Exercise 2.2 (Roulette)**.** Roulette is one of fairest games that is offered in a casino. If you bet on the colour *red* (or *black*), then your bet will be doubled if the ball lands in one of the 18 red (black) holes. If the ball lands in a hole with the other colour, then the bet is lost. An exception is the special hole "0" (color green): if the ball lands in that hole, then half the bet is lost (see Fig. 2.2). In order to include the latter, you assume that you double your bet with probability $p = 18.25/37$ and lose your bet with probability $q = 1 - p = 18.75/37$.

Figure 2.2: Roulette wheel.

Suppose that initially you have \$500, and your ambition is to double that amount. Since you loose on average in every bet, your best strategy is to achieve your goal in as few steps as possible, betting all the \$500 at once. In that case the probability of achieving your goal is p. However, the casino imposed the special rule that you are allowed to bet at most \$10 at a time. What is your chance of accumulating the desired \$1000 (without bankruptcy)?

Solve this problem via the following steps. Define a simple random walk $(S_n)_{n=1}^\infty$ with probability to go up p and go down q (one does this by letting $X_1, X_2, \dots$ be independent random variables with $P(X_k = +1) = p$, $P(X_k = -1) = q$). Consider the quantity

$$A(z) = P_z(\tau_M < \tau_0), \tag{2.15}$$

(note that $A(z)$ equils $P(\tau_{M-z} < \tau_{-z})$) for $z, M \in \mathbb{N}$ with $z \le M$. So $A(z)$ is the probability that the random walk S starting in z reaches M before 0.
(a) Prove that $A(z)$ satisfies the relation

$$A(z) = q\,A(z-1) + p\,A(z+1) \tag{2.16}$$

with boundary values $A(0) = 0$ and $A(M) = 1$.
(b) Let $B(z) = A(z+1) - A(z)$ for $z = 0, \ldots, M-1$. Derive a recursive relation for $B(z)$ (i.e., $B(z+1)$ expressed as a function of $B(z)$) and give a solution of $B(z)$ in terms of $B(0)$. Derive the solution for A.
(c) What is your chance of accumulating the desired \$1000 (without bankruptcy)?

2.2 The strong law of large numbers

Limit theorems form the essence of any introductory probability course. There are two versions of the law of large numbers, the strong one and the weak one. In our situation the strong law of large numbers (SLLN) applies: Since $X_1, X_2, X_3, \ldots$ form an i.i.d. (= independent and identically distributed) sequence of random variables with $E(|X_1|) = \sum_{x \in \mathbb{Z}^d} |x|\, P(X_1 = x) < \infty$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E(X_1) \qquad P-a.s., \tag{2.17}$$

which means $P(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E(X_1)) = 1$.

Mind that the X_k's are d-dimensional vectors $X_k = (X_k^{(1)}, \ldots, X_k^{(d)})$, and so also $E(X_1)$ is a d-dimensional vector. By symmetry we have $E(X_1^{(j)}) = \sum_{x \in \mathbb{Z}^d} x_j\, P(X_1 = x) = 0$ for all $j = 1, \ldots, d$. In other words, $E(X_1)$ is the null-vector, and we arrive at our first limit result for random walks.

Theorem 2.3 (SLLN)**.** *For simple random walk on $\mathbb{Z}^d$,*

$$\lim_{n\to\infty} \frac{S_n}{n} = 0 \qquad P-a.s. \tag{2.18}$$

This theorem may be interpreted as saying that simple random walk on $\mathbb{Z}^d$ grows sublinearly, also called *subballistically*.

2.3 The central limit theorem

Like for the SLLN, we can readily feed our simple random walk into the standard central limit theorem (CLT) for i.i.d. random variables. The only (slight) complication is that we need a multi-dimensional version of the CLT, in which (co-)variances are given by a matrix rather than by a number.

For $\mu \in \mathbb{R}^d$ and a positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, the d-dimensional normal distribution $\mathcal{N}_d(\mu, \Sigma)$ has probability density function (see Fig. 2.3)

$$\phi_d(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\Big\}, \qquad x \in \mathbb{R}^d. \tag{2.19}$$

The probability of a set $A \subset \mathbb{R}^d$ is given by $\int_A \phi_d(x)\, dx$. The CLT for sums of i.i.d. d-dimensional vectors states that

$$\frac{1}{\sqrt{n}} \left(\sum_{k=1}^{n} [X_k - E(X_1)] \right) \tag{2.20}$$

converges *in distribution* to a d-dimensional centered normal distribution, i.e., $\mu = 0$ and the covariance matrix Σ has elements

$$\Sigma_{ij} = \mathrm{Cov}\big(X_1^{(i)}, X_1^{(j)}\big), \qquad 1 \le i, j \le d. \tag{2.21}$$

We have already shown that $E(X_1)$ is the d-dimensional null-vector. In our situation,

$$\mathrm{Cov}\big(X_1^{(i)}, X_1^{(j)}\big) = E\big(X_1^{(i)} X_1^{(j)}\big) = \sum_{x \in \mathbb{Z}^d} x_i x_j P(X_1 = x) = \begin{cases} 1/d, & i = j, \\ 0, & i \neq j. \end{cases} \tag{2.22}$$

Thus, Σ is the unit matrix Id scaled by the factor $1/d$.

Theorem 2.4 (CLT)**.** *For simple random walk on $\mathbb{Z}^d$, $n^{-1/2} S_n$ converges in distribution to $\mathcal{N}_d(0, d^{-1}\mathrm{Id})$.*

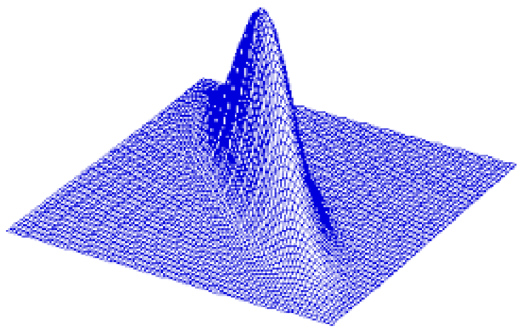

Figure 2.3: The bivariate normal distribution $\mathcal{N}_2(\mu, \Sigma)$.

2.4 The large deviation principle

The CLT implies that

$$\lim_{n\to\infty} P\big(|S_n| \ge a\sqrt{n}\big) = \int_{|x| \ge a} \phi_d(x)\,dx, \qquad a \in [0, \infty). \tag{2.23}$$

Thus, simple random walk *typically* deviates from 0 by an amount of order $\sqrt{n}$. In what follows we consider events of the form $\{|S_n| \ge an\}$, $a \in [0, \infty)$. These are *rare* events, in the sense that their probability tends to zero as $n \to \infty$. Large deviation theory analyses how fast, namely, *exponentially* fast in n. In order to simplify the presentation, we only consider the case of one dimension ($d = 1$).

Figure 2.4: Harald Cramér.

The following theorem is a special version of a theorem proved by Cramér (see Fig. 2.4) in 1938 for large deviations of the empirical average of i.i.d. random variables.

Theorem 2.5 (LDP)**.** *For $a > 0$,*

$$\lim_{n\to\infty} \tfrac{1}{n} \log P(S_n \geq an) = \lim_{n\to\infty} \tfrac{1}{n} \log P(S_n \leq -an) = -I(a) \tag{2.24}$$

with

$$I(a) = \begin{cases} \log 2 + \frac{1+a}{2} \log \frac{1+a}{2} + \frac{1-a}{2} \log \frac{1-a}{2}, & a \in [-1,1], \\ \infty, & \textit{otherwise}, \end{cases} \tag{2.25}$$

where $0 \log 0 = 0$.

Proof. The claim is trivial for $a > 1$. Let $a \in (0,1]$. Then

$$P(S_n \geq an) = 2^{-n} \sum_{k \geq \frac{1+a}{2} n} \binom{n}{k}. \tag{2.26}$$

This yields the estimate

$$2^{-n} Q_n(a) \leq P(S_n \geq an) \leq n 2^{-n} Q_n(a) \tag{2.27}$$

with

$$Q_n(a) = \max_{k \geq \frac{1+a}{2} n} \binom{n}{k}. \tag{2.28}$$

The maximum is attained at $k = \lceil (1+a)n/2 \rceil$, the smallest integer larger than or equal to $(1+a)n/2$. Stirling's formula, as quoted above (1.10), therefore allows us to infer

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(a) = -\tfrac{1+a}{2} \log \tfrac{1+a}{2} - \tfrac{1-a}{2} \log \tfrac{1-a}{2}. \tag{2.29}$$

Combining (2.27)–(2.29), we get the statement for $a \in (0,1]$. The case $a = 0$ corresponds to the typical behaviour. □

The function $z \mapsto I(z)$ is called the *rate function*. Note that I is infinite outside $[-1,1]$, finite and strictly convex inside $[-1,1]$, and has a unique zero at 0. The latter corresponds to the SLLN (see Fig. 2.5).

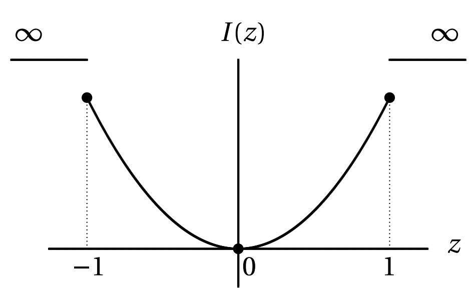

Figure 2.5: The rate function for one-dimensional simple random walk.

Exercise 2.3. The goal of this exercise is to show that I is a convex function.
Let $J \subset \mathbb{R}$ be an interval in $\mathbb{R}$ (which may equal $\mathbb{R}$ itself). A function $f : J \to [-\infty,\infty]$ is called convex (on J) if

$$f(\lambda x + (1-\lambda) y \leq \lambda f(x) + (1-\lambda) f(y) \qquad \text{for all } x, y \in J \text{ and } \lambda \in [0,1]. \tag{2.30}$$

(a) Show that whenever J_g, J_h are intervals in $\mathbb{R}$, $g : J_g \to [-\infty,\infty]$, $h : J_h \to [-\infty,\infty]$ are convex functions, then so is $g + h$ on $J_g \cap J_h$. Show that if $g(J_g) \subset J_h$ then $h \circ g$ is convex.
(b) *[Bonus]* Show that $f : J \to \mathbb{R}$ is convex if and only

$$\frac{f(z)-f(y)}{z-y} \geq \frac{f(y)-f(x)}{y-x} \qquad \text{for all } x,y,z \in J \text{ with } x<y<z. \tag{2.31}$$

Use this to show the following: (1) If J is an open interval and f is differentiable, then f is convex if and only if $f'(y) \geq f'(x)$ for all $x, y \in J$ with $y > x$. (2) If J is an open interval and f is twice differentiable, then f is convex if and only if $f'' \geq 0$.
(c) Use (b) to show that $x \mapsto x \log x$ is convex on $(0,1)$.
(d) Use (a) and (c) to show that $x \mapsto I(x)$ is convex on $(-1,1)$.
(e) Show that $\lim_{x \downarrow -1} I(x) = I(-1)$ and $\lim_{x \uparrow 1} I(x) = I(1)$. Use this to show I is convex on $[-1,1]$.
(f) Show that I is convex (on $\mathbb{R}$).

Exercise 2.4. The goal of this exercise is to show that the LDP in Theorem 2.5, together with the fact that $a \mapsto I(a)$ has a unique zero at $a = 0$, implies the SLLN in Theorem 2.3.
(a) Show that $P(|n^{-1}S_n| \geq \epsilon) = 2P(n^{-1}S_n \geq \epsilon)$ for $\epsilon > 0$.
(b) Fix $\epsilon > 0$. Show that for every $\delta > 0$ there exists an $N \in \mathbb{N}$ (depending on ϵ and δ) such that $P(n^{-1}S_n \geq \epsilon) \leq e^{-n[I(\epsilon)-\delta]}$ for all $n \geq N$.
(c) Use the Borel-Cantelli lemma, a special case of which states the following: If $\sum_{n \in \mathbb{N}_0} P(|X_n - X| \geq \epsilon) < \infty$ for all $\epsilon > 0$, then $\lim_{n\to\infty} X_n = X$ P-a.s.

2.5 Recurrence versus transience

One of the most natural questions to ask about a random walk is whether it is certain to return to its starting point or not.

Definition 2.6. The random walk $(S_n)_{n\in\mathbb{N}_0}$ is called *recurrent* when $P(\sigma_0 < \infty) = 1$. Otherwise, it is called *transient*.

We have seen in (1.56) that one-dimensional simple random walk is recurrent. What about higher dimensions? It turns out that the answer, which was given by George Polya (see Fig. 2.6) in 1921, depends on the dimension.

Theorem 2.7. *Simple random walk is recurrent in dimension $d = 1,2$ and transient in dimension $d \geq 3$.*

Figure 2.6: George Pólya.

2.5.1 Proof of Polya's theorem

In order to prove Theorem 2.7, we introduce a new object that is important in its own right.

Definition 2.8. The random walk *Green function* is defined by

$$G(x;1) = \sum_{n\in\mathbb{N}_0} P(S_n = x), \qquad x \in \mathbb{Z}^d. \tag{2.32}$$

Note that

$$G(x;1) = \sum_{n\in\mathbb{N}_0} E(\mathbb{1}_{\{S_n=x\}}) = E\Bigg(\sum_{n\in\mathbb{N}_0} \mathbb{1}_{\{S_n=x\}}\Bigg) \tag{2.33}$$

and so $G(x;1)$ equals the *expected number of visits to* x. There is an intricate relationship between the Green function and recurrence.

Theorem 2.9 (Criterion for recurrence)**.** *The random walk* $(S_n)_{n\in\mathbb{N}_0}$ *is recurrent if and only if* $G(0;1) = \infty$.

Proof. We start by claiming

$$P(S_n = 0) = \sum_{i=1}^{n} P(\sigma_0 = i)\, P(S_{n-i} = 0), \qquad n \in \mathbb{N}. \tag{2.34}$$

Indeed, i accounts for the *first* return time to 0, and then there are $n-i$ steps left, after which the walk must be at the origin again. Let $z \in [0,1]$, and consider the *generating functions*

$$G(0;z) = \sum_{n\in\mathbb{N}_0} z^n P(S_n = 0), \qquad F(0;z) = \sum_{n\in\mathbb{N}} z^n P(\sigma_0 = n). \tag{2.35}$$

Mind that $\{\sigma_0 = 0\} = \emptyset$ by the definition of σ_0 (recall (1.18)), and hence $P(\sigma_0 = 0) = 0$. Multiplying both sides of (2.34) by z^n, summing over $n \in \mathbb{N}$ and using that $P(S_0 = 0) = 1$, we get

$$\begin{aligned} G(0;z) &= 1 + \sum_{n\in\mathbb{N}} z^n P(S_n = 0) = 1 + \sum_{n\in\mathbb{N}} \sum_{i=1}^{n} z^{i+(n-i)} P(\sigma_0 = i)\, P(S_{n-i} = 0) \\ &= 1 + \sum_{i\in\mathbb{N}} z^i P(\sigma_0 = i) \sum_{j\in\mathbb{N}_0} z^j P(S_j = 0) = 1 + F(0;z)\, G(0;z), \end{aligned} \tag{2.36}$$

where we take $z \in [0,1)$ to make sure that the sums converge. This relation can also be written as $F(0;z) = 1 - G(0;z)^{-1}$. Thus, we have

$$P(\sigma_0 < \infty) = \sum_{n\in\mathbb{N}} P(\sigma_0 = n) = F(0;1) = \lim_{z\uparrow 1} F(0;z) = 1 - \lim_{z\uparrow 1} \frac{1}{G(0;z)}, \tag{2.37}$$

where the last two equalities use monotone convergence.

If $G(0;1) = \sum_{n\in\mathbb{N}_0} P(S_n = 0) < \infty$, then

$$\lim_{z\uparrow 1} G(0;z) = G(0;1) = \sum_{n\in\mathbb{N}_0} P(S_n = 0) < \infty \tag{2.38}$$

and

$$F(0;1) = 1 - \lim_{z\uparrow 1} \frac{1}{G(0;z)} = 1 - \frac{1}{\sum_{n\in\mathbb{N}_0} P(S_n = 0)} < 1, \tag{2.39}$$

which means that the random walk is transient. On the other hand, if $G(0;1) = \sum_{n\in\mathbb{N}_0} P(S_n = 0) = \infty$, then we have $\lim_{z\uparrow 1} G(0;z)^{-1} = 0$. Indeed, fix $\varepsilon > 0$, and find $N = N(\varepsilon)$ such that $\sum_{n=0}^{N} P(S_n = 0) \geq 2/\varepsilon$. Then, for z sufficiently close to 1, we have $\sum_{n=0}^{N} z^n P(S_n = 0) \geq 1/\varepsilon$. Consequently, for such z,

$$\frac{1}{G(0;z)} \leq \frac{1}{\sum_{n=0}^{N} z^n P(S_n = 0)} \leq \varepsilon. \tag{2.40}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{z\uparrow 1} G(0;z)^{-1} = 0$. Thus, we find that $F(0;1) = 1$, and so the random walk is recurrent. □

Exercise 2.5. Prove that $G(0;1) = \infty$ if and only if $G(x;1) = \infty$ for all $x \in \mathbb{Z}^d$. *Hint*: Show that for every $x \in \mathbb{Z}^d$ there exists an event A_x such that $P(A_x) > 0$ and $G(x;1) \geq P(A_x)G(0;1)$.

2.5.2 ⋆ Fourier analysis

In view of Theorem 2.9, it remains to show that $G(0;1) = \infty$ for $d = 1,2$ and $G(0;1) < \infty$ for $d \geq 3$. There are many ways to achieve this goal. For example, a *local central limit theorem* yields the asymptotics of $P(S_n = 0)$ as $n \to \infty$, which in turn yields the desired result. Below we pursue a different line of argument, namely, one that involves characteristic functions. In Chapter 3 we encounter yet another argument.

Let us start by recalling a few basic facts about characteristic functions. A finite measure μ on $\mathbb{Z}^d$ uniquely determines its *characteristic function*

$$\phi_\mu(k) = \sum_{x\in\mathbb{Z}^d} e^{ik\cdot x}\mu(\{x\}), \qquad k \in [-\pi,\pi)^d, \tag{2.41}$$

where $k \cdot x = \sum_{j=1}^d k_j x_j$ is the standard d-dimensional inner product, with k_j the j-th component of the vector k. Conversely, given the characteristic function ϕ_μ, the measure μ can be retrieved through the so-called *Fourier inversion formula*

$$\mu(\{x\}) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} e^{-ik\cdot x}\phi_\mu(k)\,dk, \qquad x \in \mathbb{Z}^d. \tag{2.42}$$

An extremely handy property of characteristic functions is the *convolution rule*, which says that the characteristic function of the *sum* of two independent random variables is equal to the *product* of the characteristic functions of the two random variables.

Let us now see how this formalism can be used to solve our problem. In our setting, μ is the distribution of the random variable X_1, i.e.,

$$\mu(x) = \frac{1}{2d}\mathbb{1}_{\{|x|=1\}}, \qquad x \in \mathbb{Z}^d. \tag{2.43}$$

Via symmetry we can check that the characteristic function of μ, which is $\sum_{x\in\mathbb{Z}^d} e^{ik\cdot x}P(X_1 = x)$, has the form

$$\phi(k) = \frac{1}{d}\sum_{j=1}^d \cos\left(k_j\right), \qquad k \in [-\pi,\pi)^d. \tag{2.44}$$

Exercise 2.6. Prove (2.44).

Since S_n is the sum of the i.i.d. random variables $X_1,\ldots,X_n$, its characteristic function is the n-th power $\phi(k)^n$, and the Fourier inversion formula in (2.42) gives

$$P(S_n = x) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} e^{-ik\cdot x}\phi(k)^n\,dk, \qquad x \in \mathbb{Z}^d. \tag{2.45}$$

Hence, for $z \in [0,1)$,

$$G(0;z) = \sum_{n\in\mathbb{N}_0} z^n P(S_n = 0) = \sum_{n\in\mathbb{N}_0}\frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} z^n\phi(k)^n\,dk = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d}\frac{1}{1-z\phi(k)}\,dk. \tag{2.46}$$

Since $\lim_{z\uparrow 1} G(0;z) = G(0;1)$ by monotone convergence, we see that

$$G(0;1) < \infty \qquad \Longleftrightarrow \qquad \int_{[-\pi,\pi)^d}\frac{1}{1-\phi(k)}\,dk < \infty. \tag{2.47}$$

Thus, we have found an integral test for transience.

Exercise 2.7. Use the expression in (2.44) to prove that the right-hand side of (2.47) holds if and only if $d \geq 3$. This goes in two steps:
(a) Show that $\frac{2}{\pi^2} t^2 \leq 1 - \cos t \leq \frac{1}{2} t^2$ for $t \in [-\pi, \pi)$. *Hint for the upper bound:* First conclude from $\cos(x) \leq 1$ that $\sin(x) \leq x$ for $x \in [0, \pi)$. *Hint for the lower bound:* Show $f(t) = \frac{1}{\pi} t - \sin(\frac{t}{2})$ is convex (on $[0, \pi)$ for example) (see Exercise 2.3). Show that this implies $f \leq 0$. Then use $2\sin^2(\frac{t}{2}) = 1 - \cos(t)$.
(b) Conclude from (a) that

$$\frac{2}{\pi^2 d} \sum_{j=1}^{d} k_j^2 \leq 1 - \phi(k) \leq \frac{1}{2d} \sum_{j=1}^{d} k_j^2, \qquad k \in [-\pi, \pi)^d, \tag{2.48}$$

and integrate the reciprocal of these bounds using polar coordinates to finish the proof. *Hint:* Use that for any non-increasing function $f\colon (0,\infty) \to [0,\infty)$ and any $\lambda > 0$,

$$\int_{0<|k|\leq\lambda} f(|k|)\, dx = V_d \int_0^\lambda f(r) r^{d-1}\, dr, \tag{2.49}$$

where V_d is the volume of the d-dimensional unit sphere (the integral on the left-hand side is over $\mathbb{R}^d$).

This finishes the proof of Theorem 2.7.

Combining the results of Exercises 2.5 and 2.7, we see that $G(x;1) < \infty$ for all $x \in \mathbb{Z}^d$ when $d \geq 3$. Since $G(x;1)$ is the expected number of visits to x, it must tend to zero as $|x| \to \infty$: a transient random walk is unlikely to hit a far away point. With the help of the same type of computation as above it is possible to derive the asymptotic formula

$$G(x;1) \asymp \frac{1}{|x|^{d-2}}, \qquad |x| \to \infty, \tag{2.50}$$

where $\asymp$ means that the ratio of the two sides is bounded away from 0 and ∞. Indeed, as in (2.45)–(2.46) we have

$$\begin{aligned} G(x;1) &= \sum_{n\in\mathbb{N}_0} P(S_n = x) = \sum_{n\in\mathbb{N}_0} \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} e^{-ik\cdot x} \phi(k)^n\, dk \\ &= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} \frac{\cos(k\cdot x)}{1-\phi(k)}\, dk. \end{aligned} \tag{2.51}$$

Standard Fourier theory shows that for $|x| \to \infty$ the integral is dominated by the small values of $|k|$ (because for large k the cosine oscillates fast), for which we can approximate $\cos(k\cdot x) = \exp[-\frac{1}{2}(k\cdot x)^2 + O(|k|^3)]$ and $1 - \phi(k) = \frac{1}{2d}|k|^2 + O(|k|^3)$. Substitute these relations into the right-hand side of (2.51) and go to polar coordinates (as in Exercise 2.7), to obtain the claim in (2.50).

Chapter 3

Random Walks and Electric Networks

In this chapter we look at random walks and flows on *electric networks*. We show that these processes are intimately related. In Section 3.1 we consider finite networks, beginning with linear and planar networks, and afterwards building up a general theory. In Section 3.2 we turn to infinite networks. The exposition is based on P.G. Doyle and J.L. Snell, *Random Walks and Electric Networks*, Carus Mathematical Monograph 22, Mathematical Association of America, 1984.

Figure 3.1: Georg Ohm, Gustav Kirchhoff.

Before we start we recall three basic facts about electric networks:

(I) The rule for the composition of *resistances* is shown in in Fig. 3.2.

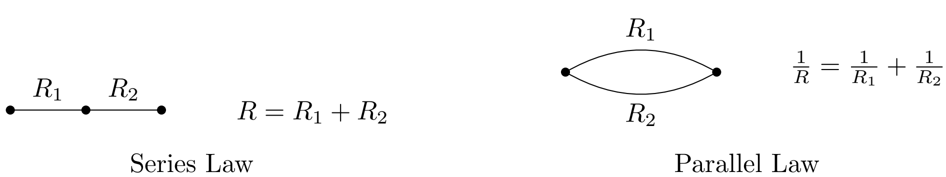

Figure 3.2: The Series Law and the Parallel Law.

The inverse of resistance is called *conductance*. Thus, in series resistances add up, while in parallel conductances add up. By iteration this leads to formulas for the composition of $n \in \mathbb{N}$ resistances: $\sum_{i=1}^{n} R_i$ for resistances in series and $[\sum_{i=1}^{n} (1/R_i)]^{-1}$ for resistances in parallel.

(II) *Ohm's law* says that

$$V = I \times R \qquad \text{potential difference} = \text{current} \times \text{resistance}. \tag{3.1}$$

(III) *Kirchhoff's law* says that the total current in and out of an isolated vertex is zero: current can only flow in at a source vertex and flow out of a sink vertex (e.g. via a battery that is connected to the source vertex and the sink vertex).

3.1 Finite networks

3.1.1 A linear network

The network we start with is a finite piece of $\mathbb{Z}$, say, the set of vertices $\mathcal{V} = \{0, 1, \dots, N-1, N\}$, $N \in \mathbb{N}$, with edges between neighbouring vertices (see Fig. 3.3).

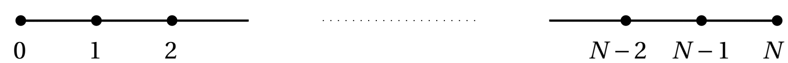

Figure 3.3: A linear network.

Random Walk. Consider a simple random walk $S = (S_n)_{n\in\mathbb{N}_0}$ on $\mathcal{V}$, i.e., at x the walk has probability $\frac{1}{2}$ to jump to $x-1$ and probability $\frac{1}{2}$ to jump to $x+1$, with the convention that a jump out of $\mathcal{V}$ is replaced by a pause, i.e., $P_0(S_{n+1} = 0 \mid S_n = 0) = \frac{1}{2}$ and $P_N(S_{n+1} = N \mid S_n = N) = \frac{1}{2}$ for all $n \in \mathbb{N}_0$. Let

$$p_x = P_x(\tau_N < \tau_0), \qquad x \in \mathcal{V}, \tag{3.2}$$

where, as before, P_x stands for the law of S given $S_0 = x$, and

$$\tau_y = \inf\{n \in \mathbb{N}_0 \colon S_n = y\} \tag{3.3}$$

is the first hitting time of $y \in \mathcal{V}$. (Note that τ_y differs from σ_y used in Chapter 1 in that it allows the hitting time to be zero, i.e., $\tau_y = 0$ when $S_0 = y$.) The probability in (3.2) has the following properties:

$$p_0 = 0, \qquad p_N = 1, \qquad p_x = \tfrac{1}{2} p_{x-1} + \tfrac{1}{2} p_{x+1}, \quad x \notin \partial\mathcal{V} = \{0, N\}. \tag{3.4}$$

The third line follows by recording the first step of S and using that S is a Markov process (see also (2.14))

$$\begin{aligned} p_x &= P_x(\tau_N < \tau_0, S_1 = x+1) + P_x(\tau_N < \tau_0, S_1 = x-1) \\ &= \tfrac{1}{2} P_x(\tau_N < \tau_0 \mid S_1 = x+1) + \tfrac{1}{2} P_x(\tau_N < \tau_0 \mid S_1 = x-1) \\ &= \tfrac{1}{2} P_{x+1}(\tau_N < \tau_0) + \tfrac{1}{2} P_{x-1}(\tau_N < \tau_0) \\ &= \tfrac{1}{2} p_{x+1} + \tfrac{1}{2} p_{x-1}. \end{aligned} \tag{3.5}$$

Lemma 3.1. *(Recall Exercise* 2.2.*) The difference equation with boundary condition in* (3.4) *has a unique solution.*

Proof. We already know that there is a solution, namely, (3.2). Our task is to show that there is at most one solution. The proof comes in two steps, which will be generalised later by the Maximum Principle (Lemma 3.2) and the Uniqueness Principle (Lemma 3.3).

Step 1: Put $M = \max_{x\in\mathcal{V}} p_x$. Since p is bounded, the maximum is attained at some $x_0 \in \mathcal{V}$. Suppose that $x_0 \notin \partial\mathcal{V}$. Then, by the third equation of (3.4), we must have $p_{x_0-1} = M$ and $p_{x_0+1} = M$. Iteration gives that $p_x = M$ for all $x \in \mathcal{V}$, which contradicts the boundary condition. Hence $x_0 \in \partial\mathcal{V}$. The same argument shows that also $m = \min_{x\in\mathcal{V}} p_x$ is attained at $\partial\mathcal{V}$.

Step 2: Let p, p' be two solutions. Put $q = p - p'$. This function has the properties $q_0 = q_N = 0$ and $q_x = \frac{1}{2}q_{x-1} + \frac{1}{2}q_{x+1}$. As shown in Step 1, the latter ensures that q attains both its maximum and its minimum at $\partial\mathcal{V}$. The former ensures that both the maximum and the minimum are zero. Hence $q \equiv 0$ on $\mathcal{V}$. □

Electric Flow. Suppose that the edges in $\mathcal{V}$ are wires with an electric resistance of 1 Ohm (= unit resistance). Place a 1-Volt battery across $\partial\mathcal{V}$, fixing the potential at $x = 0$ to be 0 and the potential $x = N$ to be 1. Let

$$v_x = \text{ voltage at } x, \qquad x \in \mathcal{V}. \tag{3.6}$$

By Ohm's Law, the potential in (3.6) has the following properties:

$$v_0 = 0, \qquad v_N = 1, \qquad v_x = \tfrac{1}{2}v_{x-1} + \tfrac{1}{2}v_{x+1}, \quad x \notin \partial\mathcal{V} = \{0, N\}. \tag{3.7}$$

The third equation follows from Kirchhoff's Law: the current into x, which is $(v_x - v_{x-1}) + (v_x - v_{x+1})$, must be equal to 0 when $x \notin \partial\mathcal{V}$. Now, (3.7) is the precise same equation as (3.4), and so by Lemma 3.1 we have

$$p_x = v_x, \qquad x \in \mathcal{V}. \tag{3.8}$$

Thus, we see that there is a *direct link* between the Random-Walk-problem and the Electric-Flow-problem.

It is not hard to guess the solution of (3.4) and (3.7):

$$p_x = v_x = \frac{x}{N}, \qquad x \in \mathcal{V}. \tag{3.9}$$

Compare this formula and the following exercise with what was found in Section 1.3 for the ruin problem.

Exercise 3.1. Let $m_x = E_x(\tau_{\{0,N\}})$ be the average number of steps until S hits the set $\{0, N\}$ given $S_0 = x$. (Here, $\tau_{\{0,N\}} = \inf\{n \in \mathbb{N}_0 \colon S_n \in \{0, N\}\}$ and E_x is expectation over S given $S_0 = x$.) Write down a difference equation with boundary condition for m_x. Derive the solution for m_x (similar as in Exercise 2.2).

3.1.2 A planar network

Take for $\mathcal{V}$ a finite piece of $\mathbb{Z}^2$, say, $\mathcal{V} = \{0, 1, \ldots, N-1, N\}^2$, $N \in \mathbb{N}$ (see Fig. 3.4).

Random Walk. Consider a simple random walk $S = (S_n)_{n\in\mathbb{N}_0}$ on $\mathcal{V}$, i.e., at x the walk has probability $\frac{1}{4}$ to jump along one of the 4 edges going out of x, with the convention that a jump out of $\mathcal{V}$ is replaced by a pause. Let

$$p_x = P_x(\tau_{(N,N)} < \tau_{(0,0)}), \qquad x \in \mathcal{V}. \tag{3.10}$$

Then (3.4) becomes

$$p_{(0,0)} = 0, \qquad p_{(N,N)} = 1, \qquad p_x = \tfrac{1}{4}\sum_{y\sim x} p_y, \quad x \notin \{(0,0), (N,N)\}, \tag{3.11}$$

where $y \sim x$ means that y is a neighbour of x, with the convention that if y falls outside $\mathcal{V}$, then y is replaced by x. As in Lemma 3.1, this difference equation with boundary condition has a unique solution (see also the uniqueness principle in Lemma 3.3 below).

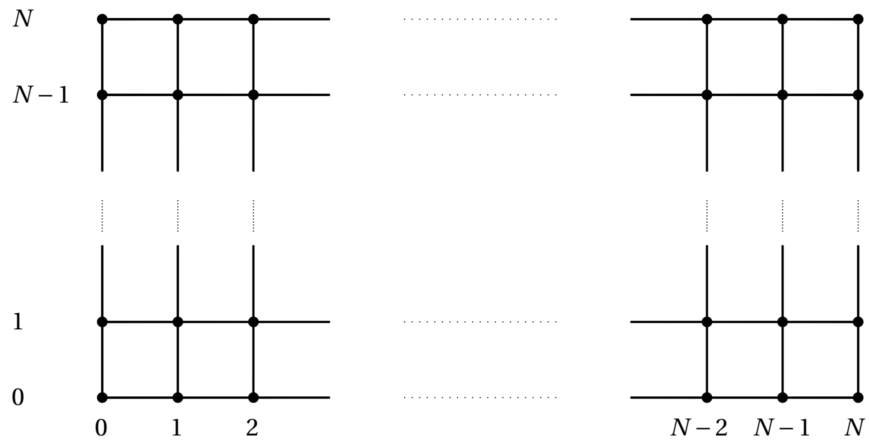

Figure 3.4: A planar network.

Electric Flow. Again, suppose that the edges are wires with an electric resistance of 1 Ohm. Place a 1-Volt battery across $\{(0,0),(N,N)\}$, fixing the potential at $x=(0,0)$ to be 0 and the potential $x=(N,N)$ to be 1. Then

$$v_{(0,0)} = 0, \qquad v_{(N,N)} = 1, \qquad v_x = \tfrac{1}{4}\sum_{y\sim x} v_y, \quad x \notin \{(0,0),(N,N)\}, \tag{3.12}$$

where the third line again follows from Kirchhoff's Law: the current into x, which is $\sum_{y\sim x}(v_x - v_y)$, must be equal to 0 when $x \notin \{(0,0),(N,N)\}$. Here the same convention is used: if y falls outside $\mathscr{V}$, then y is replaced by x. By the uniqueness of the solution we get the analogue of (3.8):

$$p_x = v_x, \qquad x \in \mathscr{V}. \tag{3.13}$$

However, this time it is *not* easy to guess the explicit form of the solution.

Exercise 3.2. Check that $p_{(x_1,x_2)} = x_1 x_2 / N^2$ is not a solution.

3.1.3 General networks

Definitions. We generalise the above two examples to an *arbitrary finite network.* We think of the network as a *graph* $\mathscr{G} = (\mathscr{V},\mathscr{E})$, where $\mathscr{V}$ is a set of *vertices* and $\mathscr{E} \subset \mathscr{V}\times\mathscr{V}$ is a set of *directed edges.* An edge from x to y is denoted by xy. We assume that

$$xy \in \mathscr{E} \iff yx \in \mathscr{E}, \tag{3.14}$$

i.e., each pair of sites is either not connected by an edge or is connected by two edges in opposite directions. (Notationally it will be convenient not to use a single undirected edge.) With each edge $xy \in \mathscr{E}$ we associate a *conductance* $C_{xy} \in (0,\infty)$ in such a way that

$$C_{xy} = C_{yx} \qquad \forall\, xy \in \mathscr{E}. \tag{3.15}$$

We assume that $\mathscr{G}$ is connected, and consider a Markov Chain $S = (S_n)_{n\in\mathbb{N}_0}$ on $\mathscr{G}$ with *transition matrix* $P = (P_{xy})_{x,y\in\mathscr{V}}$, i.e., P_{xy} is the probability to go from x to y, given by

$$P_{xy} = \frac{C_{xy}}{C_x}, \qquad C_x = \sum_{y\sim x} C_{xy}, \qquad x,y \in \mathscr{V}, \tag{3.16}$$

where $y \sim x$ means that $xy \in \mathscr{E}$, and C_x is the conductance out of x. For this choice, S has equilibrium distribution

$$\mu_x = \frac{C_x}{C}, \qquad x \in \mathscr{V}, \tag{3.17}$$

where $C = \sum_{x\in\mathscr{V}} C_x$ is the normalisation constant. Indeed, (3.16–3.17) imply that

$$\mu_x P_{xy} = \mu_y P_{yx} \qquad \forall\, x, y \in \mathscr{V}, \tag{3.18}$$

from which it follows that $\sum_{x\in\mathscr{V}} \mu_x P_{xy} = \mu_y$ for all $y \in \mathscr{V}$. In fact, (3.18) says that μ is the *reversible equilibrium* of S. In physics terminology, the latter property is referred to as "detailed balance": in equilibrium the probability that S moves from x to y is the same as the probability that S moves from y to x. Note that transitions can only occur along the directed edges in $\mathscr{G}$. Alternatively, we may put

$$C_{xy} = 0 \qquad \forall, xy \notin \mathscr{E}. \tag{3.19}$$

Given $\mathscr{C} = (C_{xy})_{xy\in\mathscr{E}}$, the *Laplacian* Δ associated with the Markov chain, acting on the set of functions $\{f\colon \mathscr{V} \to \mathbb{R}\}$, is defined as

$$(\Delta f)_x = \sum_{y\sim x} P_{xy}(f_y - f_x), \qquad x \in \mathscr{V}. \tag{3.20}$$

Pick $a, b \in \mathscr{V}$, $a \neq b$. A function f is called *harmonic* on $\mathscr{V}\backslash\{a,b\}$ when $\Delta f \equiv 0$ on $\mathscr{V}\backslash\{a,b\}$. The following two lemmas generalise Steps 1 and 2 in the proof of Lemma 3.1.

Lemma 3.2 (Maximum Principle)**.** *Let f be a harmonic function on $\mathscr{V}\backslash\{a,b\}$ with $f_a \geq f_b$. Then f attains its maximal value M at a and its minimal value m at b.*

Lemma 3.3 (Uniqueness Principle)**.** *Let f, g be two harmonic functions on $\mathscr{V}\backslash\{a,b\}$ such that $f = g$ on $\{a,b\}$. Then $f = g$ on $\mathscr{V}$.*

Exercise 3.3. Give the proof of both lemmas.

Dirichlet problem. Again, think of edge xy as a wire with conductance C_{xy}. Pick $a, b \in \mathscr{V}$, $a \neq b$, and place a 1-Volt battery across $\{a,b\}$, fixing the potential at $x = b$ to be 0 and the potential at $x = a$ to be 1 (see Fig. 3.5). Let v_x denote the *potential* at x, and i_{xy} the *current* from x to y. Note that $i_{xy} = -i_{yx}$.

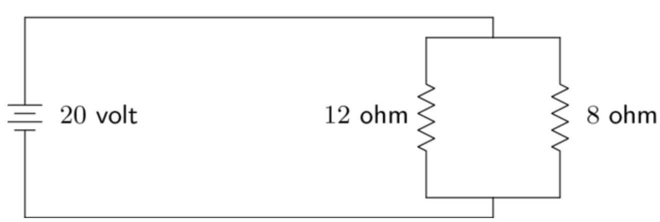

Figure 3.5: Example of an electric networks

By *Ohm's Law*, we have [1]

$$i_{xy} = (v_x - v_y)C_{xy} = (v_x - v_y)\,\frac{1}{R_{xy}}, \tag{3.21}$$

[1] Infinite resistance = zero conductance = no current. Zero resistance = infinite conductance = no potential difference.

where $R_{xy} = 1/C_{xy}$ is the *resistance* of with edge xy. (Note that the current flows in the direction where the potential decreases.) *Kirchhoff's Law* requires that the total current flowing out of $x \in \mathcal{V}\backslash\{a,b\}$ is 0, i.e.,

$$\sum_{y\in\mathcal{V}} i_{xy} = 0, \qquad x \in \mathcal{V}\backslash\{a,b\}. \tag{3.22}$$

Substitution of (3.21) into (3.22) gives

$$v_x C_x = \sum_{y\in\mathcal{V}} C_{xy} v_y, \qquad x \in \mathcal{V}\backslash\{a,b\}, \tag{3.23}$$

which in terms of the transition matrix reads

$$v_x = \sum_{y\in\mathcal{V}} P_{xy} v_y, \qquad x \in \mathcal{V}\backslash\{a,b\}. \tag{3.24}$$

Thus, recalling (3.20), we see that the voltage solves the so-called *Dirichlet problem*

$$v_a = 1, \qquad v_b = 0, \qquad \Delta v = 0 \quad \text{on} \quad \mathcal{V}\backslash\{a,b\}. \tag{3.25}$$

Putting

$$p_x = P_x(\tau_a < \tau_b), \qquad x \in \mathcal{V}, \tag{3.26}$$

as in (3.2), we see that also p solves the Dirichlet problem. This not only tells us that the solution of the Dirichlet problem exists, via Lemmas 3.2–3.3 it also tells us that the solution is unique, and so we once again have

$$p_x = v_x \quad \forall\, x \in \mathcal{V}. \tag{3.27}$$

Thus, the link between the Markov-Chain-problem and the Electric-Flow-problem is valid in full generality.

Exercise 3.4. Compute v for the planar network in Fig. 3.4 with $N = 1$ and $N = 2$ when all edges have unit resistance (i.e., unit conductance). Recall that $a = (N,N)$ and $b = (0,0)$. (*Hint:* Exploit symmetry to simplify the computation.)

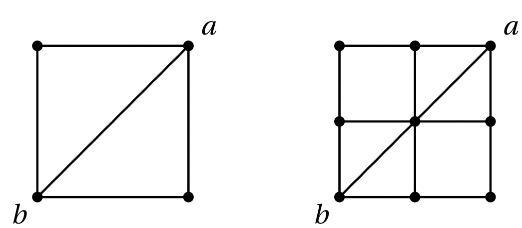

Figure 3.6: Two planar networks.

Exercise 3.5. Compute v for the planar networks in Fig. 3.6 with unit resistances. For the first network, compute the effective resistance between a and b with the help of the Series Law and the Parallel Law (recall Fig. 3.2). Can this be done for the second network too?

Note that (3.12) coincides with (3.25) when $\mathcal{G}$ is the planar network in Fig. 3.4, and $a = (N,N)$ and $b = (0,0)$.

Effective resistance. The total current that flows through the network from a to b equals the current *out of* a, $i_a = \sum_{y\in\mathcal{V}} i_{ay}$, and also equals the current *into* b, $-i_b = \sum_{y\in\mathcal{V}}[-i_{by}] = \sum_{y\in\mathcal{V}} i_{yb}$. Indeed,

$$\sum_{x,y\in\mathcal{V}} i_{xy} = -\sum_{x,y\in\mathcal{V}} i_{yx} \quad\Longrightarrow\quad 0 = \sum_{x,y\in\mathcal{V}} i_{xy} = \sum_{y\in\mathcal{V}} i_{ay} + \sum_{y\in\mathcal{V}} i_{by} = i_a + i_b. \tag{3.28}$$

The battery effectively sees the network as a single edge connecting a and b through which the total current flows (see Fig. 3.7). The *effective resistance* of the network between a and b therefore is equal to

$$R_{\text{eff}} = \frac{v_a - v_b}{i_a} = \frac{1}{i_a} = -\frac{1}{i_b} = \frac{v_b - v_a}{i_b}. \tag{3.29}$$

Since $p \equiv v$ by (3.27), we can use (3.16) and (3.20) to write

$$i_a = \sum_{y\in\mathcal{V}} i_{ay} = \sum_{y\in\mathcal{V}} (v_a - v_y) C_{ay} = C_a \sum_{y\in\mathcal{V}} (v_a - v_y)\frac{C_{ay}}{C_a} = C_a\left[1 - \sum_{y\in\mathcal{V}} P_{ay} p_y\right] = C_a p_a^{\text{esc}} \tag{3.30}$$

with $p_a^{\text{esc}} = P_a(\sigma_b < \sigma_a)$, where we recall that $\sigma_x = \inf\{n\in\mathbb{N}\colon S_n = x\}$ is the first hitting time of $x\in\mathcal{V}$ *after* time 0. The latter is the *escape probability* at a, i.e., the probability that the Markov chain starting at a reaches b before returning to a. Hence (3.29) gives us the relation

$$R_{\text{eff}} = \frac{1}{C_a p_a^{\text{esc}}}. \tag{3.31}$$

This shows yet another aspect of the general link between the Markov-Chain-problem and the Electric-Flow-problem.

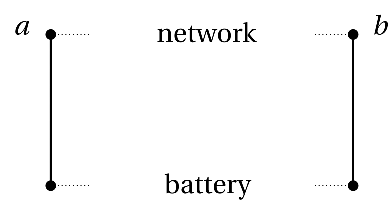

Figure 3.7: Effective resistance between two vertices of a network.

Exercise 3.6. Interchange the role of a and b to deduce from (3.31) that $C_a p_a^{\text{esc}} = C_b p_b^{\text{esc}}$. Explain this simple relation with the help of path reversal. Motivate all the details of the argument.

Energy dissipation. The energy dissipation along edge xy is the product of the current $i_{x,y}$ from x to y and the voltage difference $v_x - v_y$. Writing $R_{xy} = 1/C_{xy}$ for the *resistance* of edge xy, we see that the *total energy dissipation* of the network equals

$$E = \sum_{xy\in\mathcal{E}} i_{xy}(v_x - v_y) = \tfrac{1}{2}\sum_{x,y\in\mathcal{V}} i_{xy}(v_x - v_y) = \tfrac{1}{2}\sum_{x,y\in\mathcal{V}} i_{xy}^2 R_{xy} = \tfrac{1}{2}\sum_{x,y\in\mathcal{V}} (v_x - v_y)^2 C_{xy}, \tag{3.32}$$

where we use (3.21). Since i_a is the total current from a to b and $v_b = 0$, we have

$$E = \sum_{x\in\mathcal{V}} v_x \sum_{y\in\mathcal{V}} i_{xy} = i_a v_a = i_a^2 R_{\text{eff}} = \frac{v_a^2}{R_{\text{eff}}}. \tag{3.33}$$

Since $v_a = 1$, this says that *E is the reciprocal of the effective resistance.*

Definition 3.4. A *flow* $j = (j_{xy})_{x,y\in\mathcal{V}}$ from a to b on $\mathcal{G}$ is an assignment of real numbers j_{xy} to all pairs x, y in $\mathcal{V}$ such that:

(1) $j_{xy} = -j_{yx}$ for all $x, y \in \mathcal{V}$.

(2) $\sum_{y\in\mathcal{V}} j_{xy} = 0$ for all $x \in \mathcal{V}\setminus\{a,b\}$.

(3) $j_{xy} = 0$ when $xy \notin \mathcal{E}$.

Think of j as a "permissible flow" through the network, not the actual flow. The flow *out of* a, $j_a = \sum_{y\in\mathcal{V}} j_{ay}$, equals the flow *into* b, $-j_b = \sum_{y\in\mathcal{V}}[-j_{by}]$. If $j_a = -j_b = 1$, then we say that j is a *unit flow*. The set of unit flows is denoted by $\mathcal{UF}$.

An important property of a flow is that the energy supplied by the battery must balance against the energy dissipated by the network.

Lemma 3.5 (Conservation of Energy)**.** *For any $w\colon \mathcal{V} \to \mathbb{R}$ and any flow j from a to b,*

$$(w_a - w_b)\, j_a = \tfrac{1}{2} \sum_{x,y\in\mathcal{V}} (w_x - w_y)\, j_{xy}. \tag{3.34}$$

Proof. Write, using properties (1) and (2) in Definition 3.4,

$$\begin{aligned}\sum_{x,y\in\mathcal{V}} (w_x - w_y)\, j_{xy} &= \sum_{x\in\mathcal{V}} w_x \sum_{y\in\mathcal{V}} j_{xy} - \sum_{y\in\mathcal{V}} w_y \sum_{x\in\mathcal{V}} j_{xy}\\ &= w_a \sum_{y\in\mathcal{V}} j_{ay} + w_b \sum_{y\in\mathcal{V}} j_{by} - w_a \sum_{x\in\mathcal{V}} j_{xa} - w_b \sum_{x\in\mathcal{V}} j_{xb}\\ &= w_a j_a + w_b j_b - w_a(-j_a) - w_b(-j_b) = 2(w_a - w_b)\, j_a,\end{aligned} \tag{3.35}$$

where the last line use that $j_b = -j_a$. □

Variational principles. The total energy dissipation associated with j equals

$$\hat{E}(j) = \tfrac{1}{2} \sum_{x,y\in\mathcal{V}} j_{xy}^2 R_{xy}. \tag{3.36}$$

The following theorem shows that E in (3.32–3.33) is the solution of a variational principle after a small but crucial modification. In order to state this theorem properly, we must adjust the battery so that $i_a = 1$ instead of $v_a = 1$, i.e.,

- *The voltage of the battery must be tuned in such a way that the flow i determined by Ohm's Law becomes a unit flow.*

This can be done because any flow multiplied by any constant is again a flow.

Theorem 3.6 (Thomson Principle)**.** *Let i be the unit flow from a to b determined by Ohm's Law. Then*

$$\hat{E}(i) = \min_{j\in\mathcal{UF}} \hat{E}(j) \tag{3.37}$$

with the minimum uniquely attained at $j = i$.

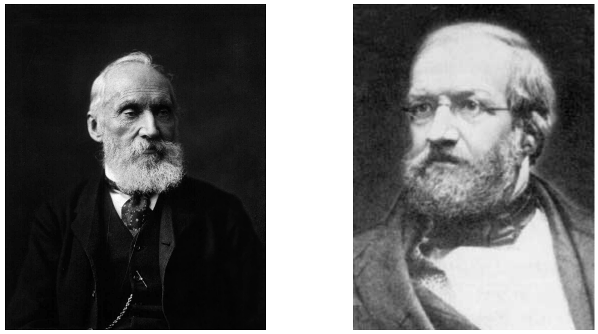

Figure 3.8: William Thomson, Gustav Dirichlet.

Proof. Fix $j \in \mathscr{UF}$. Put $\delta_{xy} = j_{xy} - i_{xy}$. Then δ is a flow from a to b with $\delta_a = \sum_{y\in\mathscr{V}} \delta_{ay} = j_a - i_a = 1 - 1 = 0$. Write

$$\begin{aligned}\hat{E}(j) &= \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} (i_{xy} + \delta_{xy})^2 R_{xy} \\ &= \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} i_{xy}^2 R_{xy} + \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} 2 i_{xy} \delta_{xy} R_{xy} + \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} \delta_{xy}^2 R_{xy} \\ &= \hat{E}(i) + \sum_{x,y\in\mathscr{V}} (v_x - v_y)\delta_{xy} + \hat{E}(\delta).\end{aligned} \tag{3.38}$$

Picking $w = v$ and $j = \delta$ in Lemma 3.5, we see that the middle term equals $2(v_a - v_b)\delta_a = 0$, and so we have

$$\hat{E}(j) = \hat{E}(i) + \hat{E}(\delta). \tag{3.39}$$

Since $\hat{E}(\delta) \geq 0$ with equality if and only if $\delta \equiv 0$, we get the claim. □

Thus, the *true current* through the network $\mathscr{G}$ is the one that *minimises the total energy dissipation* among all currents with the *same total flow*.

There is a dual variational principle, which is of equal interest. Let $\mathscr{UP}$ be the set of unit potentials, i.e., $u\colon \mathscr{V} \to \mathbb{R}$ with $u_a = 1$ and $u_b = 0$. The total energy dissipation associated with u equals

$$E(u) = \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} \frac{(u_x - u_y)^2}{R_{xy}}. \tag{3.40}$$

Theorem 3.7 (Dirichlet Principle)**.** *Let v be the unit potential determined by Kirchhoff's Law. Then*

$$E(v) = \min_{u\in\mathscr{UP}} E(u) \tag{3.41}$$

with the minimum uniquely attained at $u = v$.

Exercise 3.7. Give the proof.

$E(v)$ is the same as E in (3.32). When we adjust the battery from $v_a = 1$ to $i_a = 1$, we also change the total energy dissipation. Since $E(v)$ corresponds to $v_a = 1$ and $\hat{E}(i)$ corresponds to $i_a = 1$, we see from (3.33) that

$$E(v) = \frac{1}{R_{\text{eff}}}, \qquad \hat{E}(i) = R_{\text{eff}}. \tag{3.42}$$

Combining Theorems 3.6–3.7 with (3.42), we get

$$\max_{u\in\mathscr{U}\mathscr{P}} \frac{1}{E(u)} = R_{\text{eff}} = \min_{j\in\mathscr{U}\mathscr{F}} \hat{E}(j). \tag{3.43}$$

This formula shows that we can obtain upper and lower bounds on R_{eff} by inserting test unit flows and test unit potentials, and hence provides a powerful way to *estimate effective resistances* for networks that are so complex that an exact computation is not feasible.

Exercise 3.8. With the help of (3.43) it is possible to estimate the effective resistance of the planar network in Fig. 3.4 when each edge has unit resistance. As a *test unit potential*, take $u_{(x_1,x_2)} = x_1x_2/N^2$, $0 \le x_1, x_2 \le N$, and compute $E(u)$ (recall Exercise 3.2). As a *test unit flow*, normalise the flow given in Figure 3.9, and compute $\hat{E}(j)$. Explain why this is indeed a flow. *Hint*: Use that $\sum_{k=1}^{N} k^2 = \frac{1}{6}N(N+1)(2N+1)$.

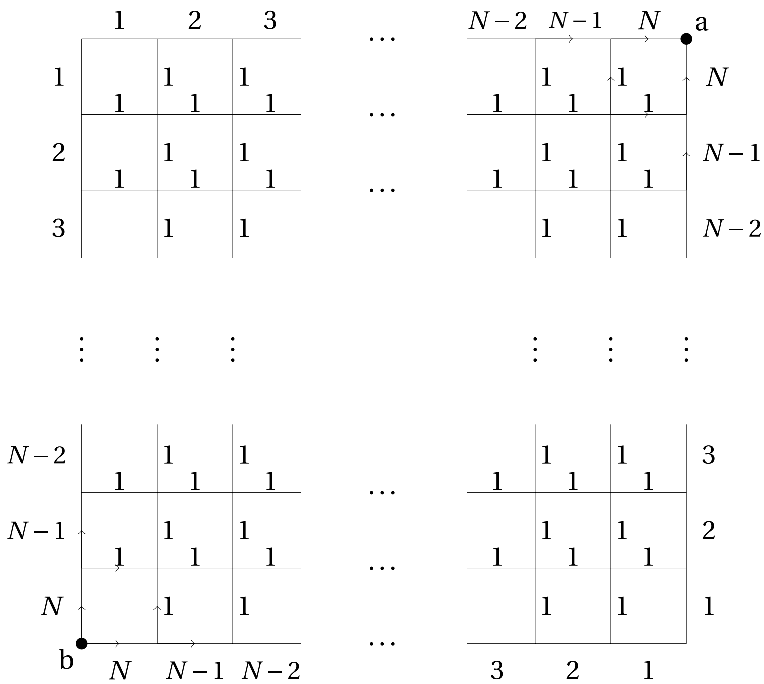

Figure 3.9: The test flow j.

Rayleigh Monotonicity Law. The following result is physically obvious but mathematically non-trival.

Theorem 3.8 (Rayleigh Monotonicity Law)**.** *If the resistances of a network are increased (decreased), then the effective resistance between any pair of points increases (decreases).*

Proof. Consider two copies of the network, one with resistances $\mathscr{R} = (R_{xy})_{xy\in\mathscr{E}}$ and one with resistances $\mathscr{R}' = (R'_{xy})_{xy\in\mathscr{E}}$, such that $R'_{xy} \ge R_{xy}$ for all $xy \in \mathscr{E}$. Let i be the unit flow from a to b in the copy with resistances $\mathscr{R}$ and j the unit flow from a to b in the copy with resistances $\mathscr{R}'$. Then, by (3.36) and (3.42),

$$R'_{\text{eff}} = \hat{E}_{\mathscr{R}'}(j) = \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} j_{xy}^2 R'_{xy} \ge \tfrac{1}{2} \sum_{x,y\in\mathscr{V}} j_{xy}^2 R_{xy} = \hat{E}_{\mathscr{R}}(j). \tag{3.44}$$

By the Thomson Principle we have $\hat{E}_{\mathscr{R}}(j) \ge \hat{E}_{\mathscr{R}}(i) = R_{\text{eff}}$, which proves the increasing part of the claim. The decreasing part of the claim follows the same argument. □

Note that the monotonicity in Theorem 3.8 is *not* automatically strict: the effective resistance of the network strictly increases when the resistance of single wire strictly increases *only* when there is a non-zero current through that wire. For most networks all wires have a non-zero current, but not all.

Returning to (3.31), Theorem 3.8 has the following interpretation in terms of the Markov Chain:

- *If the conductances of a network are increased (decreased), then the escape probability between any pair of points increases (decreases).*

In this form the monotonicity property is *far from obvious.* Indeed, by increasing the conductance of the edges $xy, yx \in \mathscr{E}$, we increase both P_{xy} and P_{yx}, but at the same time we decrease P_{xz}, $z \neq y$, and P_{yz}, $z \neq x$ (recall (3.16)). Apparently, the combined effect is that it is easier to escape from a to b. Doyle and Snell offer a nice probabilistic explanation (see Section 4.2 of their book).

3.2 Infinite networks

The theory that was built up in Section 3.1 applies to *finite* networks. For infinite networks we need to be more careful, namely, the quantities we work with are not necessarily well-defined, e.g. the total energy dissipation. In this section we show that it is possible to describe electric flows on infinite networks and that such flows have interesting properties. For random walks it was already shown in Chapter 2 that there is no problem to deal with infinite lattices like $\mathbb{Z}^d$, $d \in \mathbb{N}$. Our focus will be on such lattices as well.

Effective resistance. Nothing in Section 3.1 prevents us from replacing a and b by disjoint non-empty subsets of $\mathscr{V}$, say, $\mathscr{A}$ and $\mathscr{B}$. All we have to do is impose the boundary condition $v \equiv 1$ on $\mathscr{A}$ and $v \equiv 0$ on $\mathscr{B}$, which amounts to connecting all the vertices in $\mathscr{A}$ and all the vertices in $\mathscr{B}$ by wires with infinite conductance (or zero resistance). After that the same formulas can be used when we treat $\mathscr{A}$ and $\mathscr{B}$ as single vertices.

Armed with this idea, we consider the infinite network

$$\mathscr{G}_d = (\mathbb{Z}^d, \mathbb{Z}^d_*), \quad d \in \mathbb{N}, \tag{3.45}$$

where $\mathbb{Z}^d_*$ denotes the set of edges between neighbouring vertices of $\mathbb{Z}^d$. Each edge represents a wire with resistance 1 Ohm. We fix an $N \in \mathbb{N}$, look at the finite block $B_N = \{-N, \dots, N\}^d \subset \mathbb{Z}^d$, and pick $\mathscr{A} = \{0\}$ and $\mathscr{B} = \partial B_N = B_N \backslash B_{N-1}$. According to (3.31), the effective resistance $R_{\text{eff}}(0, \partial B_N)$ between 0 and ∂B_N (see Fig. 3.10), which we abbreviate as $R_N(\mathscr{G}_d)$, equals

$$R_N(\mathscr{G}_d) = \frac{1}{C_0 p_0^{\text{esc}}(0, \partial B_N)}, \tag{3.46}$$

where $C_0 = 2d$ and $p_0^{\text{esc}}(0, \partial B_N)$ is the probability that a simple random walk on $\mathbb{Z}^d$ starting at 0 hits ∂B_N before returning to 0. By the Rayleigh Monotonicity Law, $N \mapsto R_N(\mathscr{G}_d)$ is non-decreasing. Therefore

$$R_\infty(\mathscr{G}_d) = \lim_{N\to\infty} R_N(\mathscr{G}_d) \qquad \text{exists.} \tag{3.47}$$

What can we say about $R_\infty(\mathscr{G}_d)$?

We may think of $R_\infty(\mathscr{G}_d)$ as the effective resistance of $\mathscr{G}_d$ between 0 and infinity. Now,

$$p^{\text{esc}}(d) = \lim_{N\to\infty} p_0^{\text{esc}}(0, \partial B_N) = \lim_{N\to\infty} P_0(\sigma_{\partial B_N} < \sigma_0) = P_0(\sigma_0 = \infty) = P_0(S_n \neq 0 \ \forall\, n \in \mathbb{N}) \tag{3.48}$$

is the escape probability from 0 of simple random walk on $\mathbb{Z}^d$. As shown in Chapter 2, we have

$$p^{\text{esc}}(d) \begin{cases} = 0, & \text{for } d = 1, 2, \\ > 0, & \text{for } d \geq 3. \end{cases} \tag{3.49}$$

Therefore we find that

$$R_\infty(\mathscr{G}_d) = \frac{1}{2d\,p^{\text{esc}}(d)} \begin{cases} = \infty, & \text{for } d = 1,2, \\ < \infty, & \text{for } d \geq 3. \end{cases} \tag{3.50}$$

Thus, if we connect our 1-volt battery to any vertex of $\mathscr{G}_d$ and imagine that infinity has potential 0, then no current will flow when $d = 1,2$, while a positive current will flow when $d \geq 3$. Interestingly, *networks on which simple random walk is recurrent have infinite resistance, while networks on which simple random walk is transient have finite resistance.*

Exercise 3.9. Show that $d \mapsto R_\infty(\mathscr{G}_d)$ is non-increasing. (*Hint:* Rayleigh monotonicity law.)

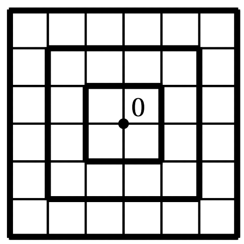

Figure 3.10: The layers ∂B_M for $M = 1,2,3$. In $\mathscr{G}_2$, all edges have resistance 1. In $\mathscr{G}_2'$, only the thin edges have resistance 1, while the thick edges have resistance 0. All the vertices connected by the thick edges have the same potential and therefore can be thought of as being merged into a single vertex.

Short-cut method. Can we verify (3.50) by direct computation? By the series and parallel law for resistances it is clear that $R_N(\mathscr{G}_1) = (N^{-1} + N^{-1})^{-1} = \frac{1}{2}N$, and so the fact that $R_\infty(\mathscr{G}_1) = \infty$ is trivial. With the help of (3.43) it is possible to show that $R_N(\mathscr{G}_2) \geq \frac{1}{8}\log(2N+1)$. This provides an alternative route to $R_\infty(\mathscr{G}_2) = \infty$ and goes as follows.

For each $M \in \mathbb{N}$, replace all the wires between the vertices in ∂B_M by wires with resistance 0. We then get a new network, say $\mathscr{G}_2'$, in which only the edges between ∂B_{M-1} and ∂B_M have resistance 1 (see Fig. 3.10). Moreover, by the Rayleigh Monotonicity Law, we have

$$R_N(\mathscr{G}_2) \geq R_N(\mathscr{G}_2'), \qquad N \in \mathbb{N}. \tag{3.51}$$

When we connect our 1-Volt battery to 0 in $\mathscr{G}_2'$, the resulting potential is constant on each ∂B_M (because the edges in ∂B_M have resistance 0). The computation of $R_N(\mathscr{G}_2')$ therefore reduces to finding the effective resistance of the linear network in Fig. 3.3 in which the edge between $M-1$ and M has conductance $|\partial B_{M-1}|+4$, where the extra 4 comes from the fact that each of the 4 corner vertices of ∂B_{M-1} has 2 edges towards ∂B_M rather than 1. Since $|\partial B_{M-1}| = 8(M-1)$, we find

$$R_N(\mathscr{G}_2') = \sum_{M=1}^{N} \frac{1}{8(M-1)+4} \geq \int_0^N \frac{dm}{8m+4} = \frac{1}{8}\log\left(\frac{8N+4}{4}\right) = \frac{1}{8}\log(2N+1), \tag{3.52}$$

which proves the claim.

Exercise 3.10. Compute the effective resistance of the rooted binary tree in Fig. 3.11 when the edges have unit resistance. *Hint:* Use symmetry to reduce the problem to the computation of the effective resistance of a linear network.

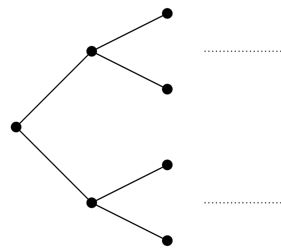

Figure 3.11: Two generations of the rooted binary tree.

Embedding method. In the previous paragraph we used that the resistance decreases when wires are replaced by *perfect conductors*, which led to a lower bound on the effective resistance. We can take the opposite route and use that the resistance increases when wires are replaced by *perfect insulators*, which leads to an upper bound on the effective resistance. In what follows we will show that $R_\infty(\mathscr{G}_3) \le 1$ by removing edges from $\mathscr{G}_3$ until we are left with a tree-like network for which the effective resistance can be computed by hand.

The idea is to build a tree inside the positive octant of $\mathscr{G}_3$ as follows. Let $\mathscr{P}_n = \{(x,y,z) \in \mathbb{N}_0^3\colon\ x+y+z = 2^n - 1\}$, $n \in \mathbb{N}_0$. We place the root of the tree at the origin $\mathscr{P}_0 = \{(0,0,0)\}$. Next, we pick the three *rays* emanating from the origin in the three positive directions, and we add all the edges and the vertices on these rays until they hit $\mathscr{P}_1$. When a ray hits $\mathscr{P}_1$, it splits into three further rays, again emanating in the three positive directions, and we add all the edges and the vertices on these rays until $\mathscr{P}_2$ is hit. After that a further splitting into three rays occurs, etc.

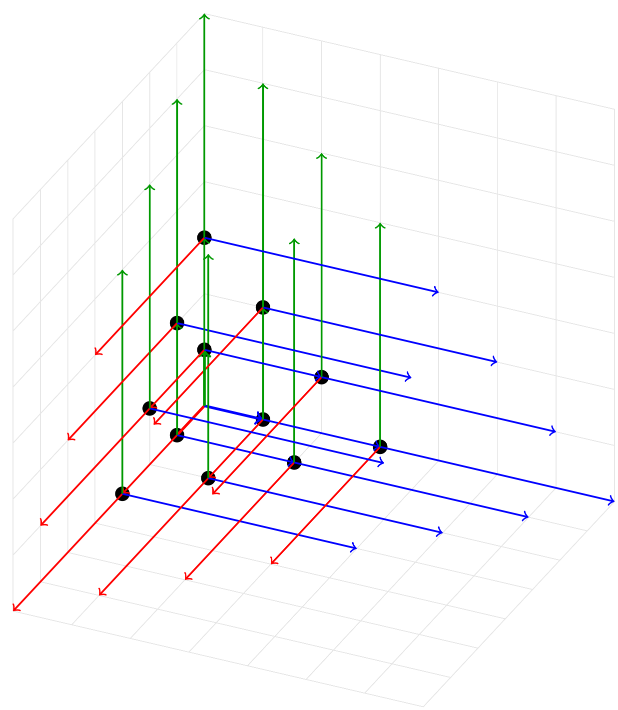

Figure 3.12: Picture of $\tilde{\mathscr{G}}_3$.

Let us call the resulting infinite network $\tilde{\mathscr{G}}_3$ (see Fig. 3.12). All the wires in $\tilde{\mathscr{G}}_3$ have resistance 1, and since $\tilde{\mathscr{G}}_3 \subset \mathscr{G}_3$ the Rayleigh Monotonicity Law gives [2]

$$R_\infty(\mathscr{G}_3) \le R_\infty(\tilde{\mathscr{G}}_3). \tag{3.53}$$

[2] The Rayleigh Monotonicity Law was derived for finite networks, but carries over to infinite networks via a straightforward limiting argument.

Now, we can lift $\tilde{\mathscr{G}}_3$ out of $\mathscr{G}_3$ and draw it as a tree $\tilde{\mathscr{T}}_3$ in the plane. In doing so, we *separate vertices* where rays intersect (we make multiple vertices with zero resitances between the edges between them), which coincide in $\tilde{\mathscr{G}}_3$ but do not coincide in $\tilde{\mathscr{T}}_3$. This does not affect the effective resistance:

$$R_\infty(\tilde{\mathscr{G}}_3) = R_\infty(\tilde{\mathscr{T}}_3). \tag{3.54}$$

Let $\mathscr{T}_3$ be the tree $\tilde{\mathscr{T}}_3$ where we erase the edges (of zero resistance) between the separated vertices (as though we put infinite resistance on these edges). Then

$$R_\infty(\tilde{\mathscr{T}}_3) \le R_\infty(\mathscr{T}_3). \tag{3.55}$$

Exercise 3.11. Draw pictures of $\tilde{\mathscr{G}}_2$ and $\tilde{\mathscr{T}}_2$, i.e., let $\mathscr{P}_n = \{(x,y) \in \mathbb{N}_0^2 \colon\ x+y = 2^n - 1\}$, $n \in \mathbb{N}_0$, pick the two rays emanating from the origin $\mathscr{P}_n = \{(0,0)\}$ in the two positive directions, etc.

The effective resistance of $\mathscr{T}_3$ can be easily computed. There are 3^n vertices at distance $2^n - 1$ from the root. Each of these vertices is connected to 3 vertices at distance $2^{n+1} - 1$ from the root. All the vertices at the same distance from the root have the same potential. Hence $R_\infty(\mathscr{T}_3)$ is the same as the effective resistance of the linear network with vertices $\mathbb{N}_0$ such that the conductance between vertex n and vertex $n+1$ is

$$\frac{3^{n+1}}{(2^{n+1}-1)-(2^n-1)} = 3\left(\frac{3}{2}\right)^n. \tag{3.56}$$

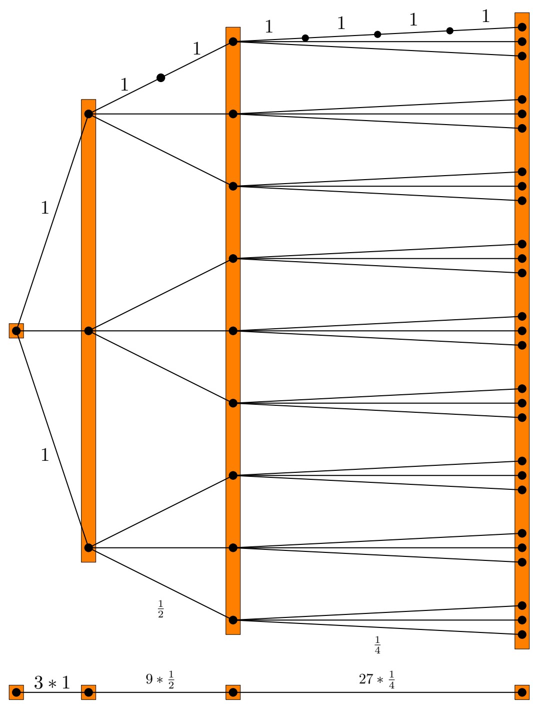

Figure 3.13: $\mathscr{T}_3$ and the corresponding linear network with conductances between vertices.

Hence

$$R_\infty(\mathscr{T}_3) = \frac{1}{3}\sum_{n\in\mathbb{N}_0}\left(\frac{2}{3}\right)^n = 1, \tag{3.57}$$

which proves the claim.

The bound $R_\infty(\mathscr{G}_3) \le R_\infty(\tilde{\mathscr{G}}_3) \le 1$ implies that $p^{\text{esc}}(3) \ge \frac{1}{6}$ (recall (3.50)).

Exercise 3.12. Improve the lower bound to $p^{\text{esc}}(3) \geq \frac{1}{3}$ by adding a mirror image of the embedded tree in the negative octant.

The improved lower bound in Exercise 3.12 turns out to be quite reasonable: exact computations show that in fact $p^{\text{esc}}(3) \approx 0.66 \approx \frac{2}{3}$.

Chapter 4

Self-Avoiding Walks and Configurations of Polymer Chains

In this chapter we describe *self-avoiding walks*, i.e., lattice paths that do not intersect themselves. We borrow from F. Caravenna, F. den Hollander and N. Pétrélis, *Lectures on Random Polymers* and R. Bauerschmidt, H. Duminil-Copin, J. Goodman and G. Slade, *Lectures on Self-Avoiding Walks*, both of which appeared in: *Probability and Statistical Physics in Two and More Dimensions*, Clay Mathematics Proceedings 15, American Mathematical Society, Providence, RI, USA, 2012, pp. 319–393 and 395–467. In Section 4.1 we count how many self-avoiding walks of a given length there are on $\mathbb{Z}^d$. In Section 4.2 we look at the spatial scaling properties of a *random* self-avoiding walk.

Self-avoiding walks are models for *polymer chains*. A polymer is a large molecule consisting of *monomers* that are tied together by *chemical bonds*. The monomers can be either small units (such as CH_2 in polyethylene) or larger units with an internal structure (such as the adenine-thymine and cytosine-guanine base pairs in the DNA double helix). Because two monomers cannot occupy the same space, the polymer chain cannot intersect itself. Paul Flory (1974 Nobel Prize in Chemistry) was the first to use self-avoiding walks to analyse the configurational properties of polymer chains. John Hammersley was the first to develop a mathematical theory of self-avoiding walks (see Fig. 4.1).

Polymers occur everywhere in nature because of the multivalency of atoms like carbon, oxygen, nitrogen and sulfur, which are capable of forming long concatenated structures. The chemical bonds in a polymer are flexible, so that the polymer can arrange itself in *many different shapes*. The longer the chain, the more involved these shapes tend to be. For instance, the polymer may wind around itself to form a knot.

Figure 4.1: Paul Flory and John Hammersley.

4.1 Counting self-avoiding walks

A self-avoiding walk (SAW) is a lattice path that does not self-intersect itself. On $\mathbb{Z}^d$, for $n \in \mathbb{N}_0$ an n-step SAW is an element of the path space ($\|\cdot\|$ is the lattice norm)

$$\mathscr{W}_n^{\neq} = \Big\{ w = (w_i)_{i=0}^n \in (\mathbb{Z}^d)^{n+1} \colon\ w_0 = 0,\ \|w_{i+1} - w_i\| = 1\ \forall\, 0 \le i < n,\ w_i \neq w_j\ \forall\, 0 \le i < j \le n \Big\}. \tag{4.1}$$

In this section we look at the problem of counting the number of n-step SAWs, i.e., $c_n = |\mathscr{W}_n^{\neq}|$. This is a hard combinatorial problem, with a history of at least 60 years.

Exercise 4.1. An easy computation shows that $c_0 = 1$, $c_1 = 2d$, $c_2 = 2d(2d-1)$ and $c_3 = 2d(2d-1)^2$. Compute c_4 and c_5.

Exact enumeration. For $d = 1$ the problem is trivial: $c_0 = 1$ and $c_n = 2$, $n \in \mathbb{N}$. For $d \ge 2$, however, no closed form expression is known for c_n. Exact enumeration methods allow for the computation of c_n up to moderate values of n. The current record is: $n = 71$ for $d = 2$, $n = 36$ for $d = 3$, $n = 24$ for $d \ge 4$. Larger n can be handled either via numerical simulation (presently up to $n = 2^{25} \approx 3.3 \times 10^7$ in $d = 3$) or with the help of extrapolation techniques.

On the homepage of Iwan Jensen (Melbourne) [http://www.ms.unimelb.edu.au/~iwan/] exact enumerations of SAWs for $d = 2$ up to $n = 71$ can be found. The first 15 entries read:

0	1
1	4
2	12
3	36
4	100
5	284
6	780
7	2172
8	5916
9	16268
10	44100
11	120292
12	324932
13	881500
14	2374444
15	6416596

Exercise 4.2. What is c_n, $n \in \mathbb{N}_0$, for the binary tree in Fig. 3.11? (The root of the tree is the origin.)

Asymptotics. What can we say about c_n for $n \to \infty$? We begin with the observation that $n \mapsto c_n$ is submultiplicative:

$$c_{m+n} \le c_m c_n, \qquad m, n \in \mathbb{N}_0. \tag{4.2}$$

The reason is that when we concatenate an m-step SAW and n-step SAW, we get an $(m+n)$-step path that may or may not be self-avoiding. Taking logarithms, we see that $n \mapsto \log c_n$ is subadditive:

$$\log c_{m+n} \le \log c_m + \log c_n, \qquad m, n \in \mathbb{N}_0. \tag{4.3}$$

Since also $c_n \ge 1$, Fekete's Lemma 4.1 implies that the sequence $(\frac{1}{n} \log c_n)_{n\in\mathbb{N}}$ converges in $\mathbb{R}$ and

$$\lim_{n\to\infty} \frac{1}{n} \log c_n = \inf_{n\in\mathbb{N}} \frac{1}{n} \log c_n. \tag{4.4}$$

Lemma 4.1 (Fekete's Lemma). *Let* $(a_n)_{n\in\mathbb{N}}$ *be a subadditive sequence in* $\mathbb{R}$ *for which* $\inf_{n\in\mathbb{N}} \frac{a_n}{n}$ *exists in* $\mathbb{R}$*. Then the sequence* $(\frac{a_n}{n})_{n\in\mathbb{N}}$ *converges and* $\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n}$.

Exercise 4.3. Give the proof of Lemma 4.1.

The quantity (which exists by (4.4))

$$\mu = \lim_{n\to\infty} (c_n)^{1/n} \tag{4.5}$$

is called the *connective constant* and depends on the dimension of the lattice: $\mu = \mu(d)$ (the limit in (4.4) equals $\log\mu(d)$). Since

$$d^n \le c_n \le 2d(2d-1)^{n-1}, \qquad n \in \mathbb{N}, \tag{4.6}$$

we have $\mu(d) \in [d, 2d-1]$. For $d = 1$ this yields the value $\mu(1) = 1$, but for $d \ge 2$ it only provides bounds. Numerical simulation leads to the estimate

$$\mu(2) = 2.63815853031\ldots, \tag{4.7}$$

but no exact value is known for $\mu(2)$. Similar estimates have been obtained for $\mu(d)$, $d \ge 3$, but with less accuracy. It is known that $\mu(d)$ has an asymptotic expansion in powers of $1/2d$, namely,

$$\mu(d) = 2d - \sum_{k\in\mathbb{N}_0} \frac{a_k}{(2d)^k} \tag{4.8}$$

with a_k integer coefficients. Up to now 13 coefficients have been identified, e.g. $a_0 = 1$, $a_1 = 1$, $a_2 = 3$, $a_3 = 16$, $a_4 = 102$. The a_k's appear to grow so rapidly with k in absolute value that the expansion in (4.8) is believed to be non-summable, i.e., its radius of convergence is believed to be zero. Moreover, for large values of k the signs of a_k appear to change in an irregular manner.

The series in (4.8) is believed to be *Borel summable*. Try to find out what Borel summability is.

The connective constant only gives the rough asymptotics for the growth of c_n, namely,

$$c_n = \mu^{n+o(n)}, \qquad n\to\infty, \tag{4.9}$$

where $o(n)$ means any function of n that grows slower than n. It is predicted that

$$c_n \sim \begin{cases} A\mu^n n^{\gamma-1}, & d \ne 4, \\ A\mu^n (\log n)^{1/4}, & d = 4, \end{cases} \qquad n\to\infty, \tag{4.10}$$

with A an amplitude and γ an exponent. The value of γ is predicted to be

$$\gamma = 1\ (d=1), \quad \tfrac{43}{32}\ (d=2), \quad 1.16\ldots\ (d=3), \quad 1\ (d\ge 5). \tag{4.11}$$

For $d \ge 5$ a proof of (4.10–4.11) has been given with the help of a combinatorial expansion technique called the lace expansion. For $d = 2,3,4$ the claim is open.

Exercise 4.4. (1) Consider two independent simple random walks S and $\bar S$ on $\mathbb{Z}^d$, both starting at 0. Let $I = \sum_{i,j\in\mathbb{N}_0} 1_{\{S_i=\bar S_j\}}$ denote their total intersection local time. In this exercise we show that $E(I) < \infty$ if and only if $d \ge 5$. This will be done with the use of the following

$$\begin{aligned} E(I) &= \sum_{i,j\in\mathbb{N}_0} P(S_i = \bar S_j) = \sum_{x\in\mathbb{Z}^d} \sum_{i,j\in\mathbb{N}_0} P(S_i = \bar S_j = x) \\ &= \sum_{x\in\mathbb{Z}^d} \sum_{i,j\in\mathbb{N}_0} P(S_i = x) P(\bar S_j = x) = \sum_{x\in\mathbb{Z}^d} G(x;1)^2, \end{aligned} \tag{4.12}$$

where $G(x;1) = \sum_{i\in\mathbb{N}_0} P(S_i = x)$ denotes the *Green function* at x. We know that $G(x;1) = \infty$ for all $x \in \mathbb{Z}^d$ when $d = 1,2$ (because SRW is recurrent), while for $d \geq 3$ we found with the help of the Fourier analysis explained in Section 2.5.2 that

$$G(x;1) \asymp \|x\|^{-(d-2)}, \qquad \|x\| \to \infty. \tag{4.13}$$

Here $f(x) \asymp h(x)$, $\|x\| \to \infty$ means that there exists $\alpha, \beta > 0$ and $K > 0$ such that

$$\alpha|h(x)| \leq |f(x)| \leq \beta|h(x)| \qquad \text{for all } x \in \mathbb{Z}^d \text{ with } \|x\| > K. \tag{4.14}$$

Prove (1) using the following steps:
(a) Conclude from (4.12) that $E(I) = \infty$ if and only if

$$\sum_{x\in\mathbb{Z}^d\setminus\{0\}} \|x\|^{-2(d-2)} = \infty.$$

(b) Show that $\|x\|_\infty \leq \|x\| \leq \sqrt{d}\|x\|_\infty$ for all $x \in \mathbb{Z}^d$, where $\|x\|_\infty = \max_{i\in\{1,\dots,d\}} |x_i|$. Conclude from this that $E(I) = \infty$ if and only if

$$\sum_{x\in\mathbb{Z}^d\setminus\{0\}} \|x\|_\infty^{-2(d-2)} = \infty.$$

(c) Show that

$$\sum_{x\in\mathbb{Z}^d\setminus\{0\}} \|x\|_\infty^{-2(d-2)} = \sum_{n\in\mathbb{N}} \Big((2n+1)^d - (2n-1)^d\Big)\, n^{-2(d-2)}.$$

(*Hint:* Compute $\#\{x \in \mathbb{Z}^d : \|x\|_\infty \leq n\}$.)
(d) Use the Binomium of Newton on $(2n+1)^d$ and $(2n-1)^d$ to conclude that there exists a $a_{d-1} > 0$ and $a_{d-2}, a_{d-3}, \dots, a_0 \geq 0$ for which

$$\begin{aligned}\sum_{x\in\mathbb{Z}^d\setminus\{0\}} \|x\|_\infty^{-2(d-2)} &= a_{d-1} \sum_{n\in\mathbb{N}} n^{d-1} n^{-2(d-2)} \\ &\quad + a_{d-2} \sum_{n\in\mathbb{N}} n^{d-2} n^{-2(d-2)} + \cdots a_1 \sum_{n\in\mathbb{N}} n^1 n^{-2(d-2)} + a_0 \sum_{n\in\mathbb{N}} n^{-2(d-2)}..\end{aligned}$$

(e) Show that $\sum_{n\in\mathbb{N}} n^{d-1} n^{-2(d-2)} < \infty$ if and only if $d \geq 5$ and conclude that $E(I) < \infty$ if and only if $d \geq 5$.
(2) *Bonus:* Why does the result in (1) provide intuitive support for the fact that there is a crossover at $d = 4$ in (4.10)?

What the crossover at $d = 4$ shows is that in low dimension the effect of the self-avoidance constraint in SAW is long-ranged, whereas in high dimension it is short-ranged. Phrased differently, since SRW in dimension $d \geq 2$ has "Hausdorff dimension 2", it tends to intersect itself frequently for $d < 4$ and not so frequently for $d > 4$. Consequently, the self-avoidance constraint in SAW changes the qualitative behaviour of the path for $d < 4$ but not for $d > 4$.

4.2 Spatial extension of a random self-avoiding walk

What can we say about the spatial extension of a self-avoiding walk that is drawn randomly from $\mathcal{W}_n^{\neq}$? In order to better appreciate the effect of the self-avoidence restriction (which in physics terminology is referred to as "excluded-volume", i.e., no two monomers in the polymer chain can occupy the same space), we recall a few facts about simple random walk that were already mentioned in Chapter 1. Simple random walk models a polymer without excluded volume.

4.2.1 Simple random walk

SRW has path space

$$\mathcal{W}_n = \Big\{ w = (w_i)_{i=0}^n \in (\mathbb{Z}^d)^{n+1} \colon\ w_0 = 0,\ \|w_{i+1} - w_i\| = 1\ \forall\, 0 \le i < n \Big\}. \tag{4.15}$$

The law P_n of SRW up to time n is the uniform distribution on $\mathcal{W}_n$ (recall (2.3)). Counting is easy: $|\mathcal{W}_n| = (2d)^n$.

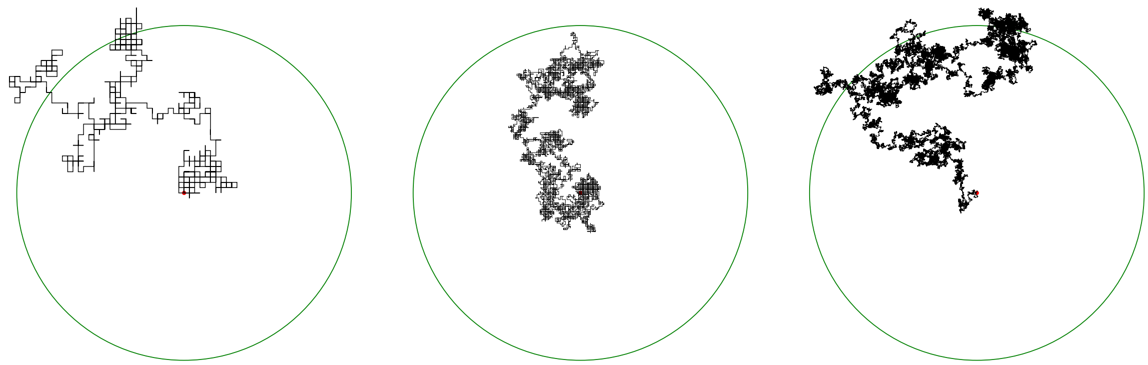

Figure 4.2: Simulation of SRW on $\mathbb{Z}^2$ with $n = 10^3$, 10^4 and 10^5 steps. The circles have radius $n^{1/2}$ in units of the step size.

Let S_n be the random variable whose law is that of the end-point w_n under P_n, i.e., S_n is the position of SRW at time n. A distinctive feature of SRW is that it exhibits *diffusive behavior*, i.e.,

$$E_n(S_n) = 0 \quad \text{and} \quad E_n(\|S_n\|^2) = n \qquad \forall\, n \in \mathbb{N}_0 \tag{4.16}$$

as shown in Claim 1.2 for $d = 1$, and

$$\left(\frac{1}{n^{1/2}} S_{\lfloor nt \rfloor} \right)_{0 \le t \le 1} \quad \Longrightarrow \quad (B_t)_{0 \le t \le 1}, \qquad n \to \infty, \tag{4.17}$$

where the right-hand side is Brownian motion on $\mathbb{R}^d$ (the definition of Brownian motion or Wiener process for higher dimensions will be introduced in Section 6.7) , and $\Longrightarrow$ denotes convergence in distribution on the space of càdlàg paths endowed with the Skorohod topology (see Fig. 4.2).

4.2.2 Self-avoiding walk

A random SAW is described by the uniform distribution $P_n^{\neq}$ on $\mathcal{W}_n^{\neq}$ in (4.1), i.e., $P_n^{\neq}(w) = 1/|\mathcal{W}_n^{\neq}|$ for all $w \in \mathcal{W}_n^{\neq}$. Let S_n be the random variable whose law (= probability distribution) is that of the end-point w_n under $P_n^{\neq}$.

Exercise 4.5. The evolutions in time of SRW and SAW have different characteristics.
(1) Show that $(P_n)_{n\in\mathbb{N}}$ is a consistent family, i.e., for every $n \in \mathbb{N}_0$ under the law P_{n+1} the first n steps of the path have law P_n (i.e. the events of the form $\{S_1 = s_1, \dots, S_n = s_n\}$ has te same probability under P_{n+1} as under P_n).
(2) Show that $(P_n^{\neq})_{n\in\mathbb{N}}$ is not a consistent family. *Hint:* Find a path in $\mathcal{W}_n^{\neq}$ that cannot be extended to a path in $\mathcal{W}_{n+1}^{\neq}$.

The consistence property in (1) is the same as the compatibility condition in (2.6). What (2) says is that SAW cannot be seen as an "evolving random process”. In other words, there is no definition of an infinite SAW (at least not in a straightforward sense).

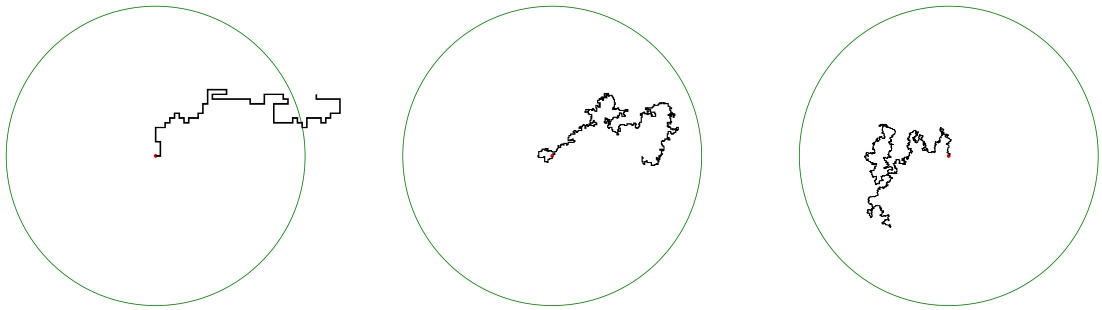

Figure 4.3: Simulation of SAW on $\mathbb{Z}^2$ with $n = 10^2$, 10^3 and 10^4 steps. The circles have radius $n^{3/4}$ in units of the step size.

The *mean-square displacement* is predicted to scale like

$$E_n^{\neq}(\|S_n\|^2) \sim \begin{cases} D\,n^{2\nu}, & d \neq 4, \\ D\,n(\log n)^{\frac{1}{4}}, & d = 4, \end{cases} \qquad n \to \infty, \tag{4.18}$$

with D a diffusion constant and ν an exponent. The value of ν is predicted to be

$$\nu = 1\ (d = 1), \quad \tfrac{3}{4}\ (d = 2), \quad 0.588\ldots\ (d = 3), \quad \tfrac{1}{2}\ (d \geq 5). \tag{4.19}$$

What (4.18–4.19) say is that

- SAW is *ballistic* in $d = 1$, *subballistic and superdiffusive* in $d = 2,3,4$, and *diffusive* in $d \geq 5$

(ballistic means $S_n \asymp n$, diffusive means $S_n \asymp \sqrt{n}$). For $d \geq 5$ a proof has been given via the lace expansion. For $d = 2,3,4$ the claim is open. Iwan Jensen's homepage lists $E_n(\|S_n\|^2)$ for $d = 2$ up to $n = 59$.

For $d \geq 5$, SAW scales to Brownian motion:

$$\left(\frac{1}{Dn^{1/2}} S_{\lfloor nt \rfloor}\right)_{0 \leq t \leq 1} \Longrightarrow (B_t)_{0 \leq t \leq 1}, \qquad n \to \infty. \tag{4.20}$$

In physics terminology this is expressed by saying that:

- For $d \geq 5$, SAW is in the *same universality class* as SRW.

For $d = 2$ the scaling limit is predicted to be $\mathrm{SLE}_{8/3}$, the so-called Schramm Loewner Evolution with parameter 8/3 (see Fig. 4.3). This is part of an elaborate theory describing two-dimensional random paths with self-interaction.

Chapter 5

Random Walks and Adsorption of Polymer Chains

A polymer is a large molecule consisting of monomers that are tied together by chemical bonds. In Chapter 4 we looked at the effect of *excluded volume*, i.e., two monomers cannot occupy the same space. In the present chapter we study a polymer in the vicinity of a linear substrate. Each monomer that touches the substrate feels a *binding energy*, resulting in an attractive or a repulsive interaction between the polymer and the substrate (depending on whether the binding energy is positive or negative). We will consider two situations:

(1) Section 5.1: the substrate is *penetrable* ("pinning").

(2) Section 5.2: the substrate is *impenetrable* ("wetting").

We will show that, in the limit as the length of the polymer tends to infinity, as the binding energy is varied there is a crossover between a *localized phase* where the polymer stays close to the substrate and a *delocalized phase* where it wanders away from the substrate (see Fig. 5.2). This crossover is referred to as a *phase transition*. Michel Fisher (see Fig. 5.1) was the first to study this phase transition in detail.

Figure 5.1: Michael Fisher.

In Section 5.3 we show that the wetting situation can be used to describe the denaturation transition of DNA. Other applications, not discussed here, include the study of chemical surfactants, i.e., chemical materials that coat a surface (like paint).

For background see: G. Giacomin,*Random Polymer Models*, Imperial College Press, London, 2007 and F. den Hollander, *Random Polymers*, Lecture Notes in Mathematics 1974, Springer, 2009.

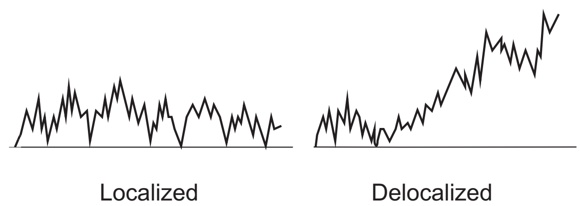

Figure 5.2: Typical paths in the localized phase and the delocalized phase in the wetting situation.

5.1 Pinning

Definitions. Let

$$\mathscr{W}_n = \Big\{ w = (i, w_i)_{i=0}^n \colon\ w_0 = 0,\ w_{i+1} - w_i \in \{-1, +1\}\ \forall\, 0 \le i < n \Big\} \tag{5.1}$$

denote the set of n-step *directed* paths starting at the origin. Here, steps move up-right ($\nearrow$) or down-right ($\searrow$). For $w \in \mathscr{W}_n$, let

$$K_n(w) = \sum_{i=1}^{n} 1_{\{w_i = 0\}}, \tag{5.2}$$

be the number of times w hits the origin (where the visit at time $i = 0$ is not counted). Fix $\zeta \in \mathbb{R}$ and define a *path measure* on $\mathscr{W}_n$ by putting

$$\bar{P}_n^{\zeta}(w) = \frac{1}{Z_n^{\zeta}}\, e^{\zeta K_n(w)}\, \bar{P}_n(w), \qquad w \in \mathscr{W}_n. \tag{5.3}$$

Here, $\bar{P}_n$ is the projection onto $\mathscr{W}_n$ of the path measure $\bar{P}$ of a *directed* SRW (i.e., under $\bar{P}$ the vertical increments $w_{i+1} - w_i$, $i \in \mathbb{N}_0$, are i.i.d. and take values ± 1 with probability $\frac{1}{2}$ each, just like SRW), ζ is the *interaction strength*, and Z_n^{ζ} is the normalisation constant

$$Z_n^{\zeta} = \sum_{w \in \mathscr{W}_n} e^{\zeta K_n(w)}\, \bar{P}_n(w). \tag{5.4}$$

Note that $|\mathscr{W}_n| = 2^n$ and that $\bar{P}_n$ is the uniform distribution on $\mathscr{W}_n$. In physics terminology, $\bar{P}_n^{\zeta}$ in (5.3) is called the *Gibbs measure* of the polymer, Z_n^{ζ} is called the *partition sum*, and ζ is "the binding energy divided by the Boltzmann constant times the absolute temperature". The Gibbs measure $\bar{P}_n^{\zeta}$ models a two-dimensional directed polymer in $\mathbb{N}_0 \times \mathbb{Z}$ where each visit to the substrate $\mathbb{N} \times \{0\}$ carries a weight e^{ζ}, which is a reward when $\zeta > 0$ and a penalty when $\zeta < 0$ (see Fig. 5.3).

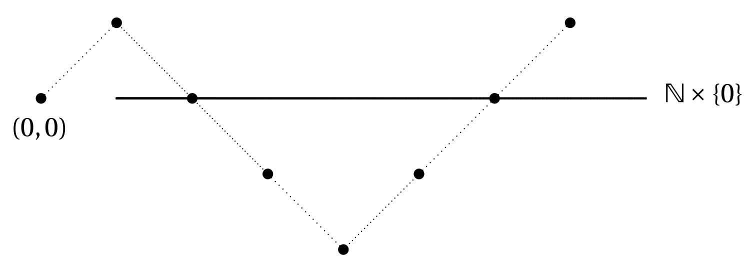

Figure 5.3: A 7-step path in $\mathbb{N}_0 \times \mathbb{Z}$ that makes 2 visits to $\mathbb{N} \times \{0\}$.

Let $S = (S_i)_{i\in\mathbb{N}_0}$ denote SRW starting from $S_0 = 0$. Write P to denote the law of S. Let $\sigma = \inf\{n \in \mathbb{N}\colon\ S_n = 0\}$ denote the first return time to 0. For $k \in \mathbb{N}$, define

$$a(k) = P(\sigma > k), \qquad b(k) = P(\sigma = k), \qquad k \in \mathbb{N}_0. \tag{5.5}$$

Note that $a(0) = 1$, $b(0) = 0$ and that the support of b is $2\mathbb{N}$, because SRW can only return to the origin after an even number of steps. Since $b(k) = a(k-1) - a(k)$, $k \in \mathbb{N}$, this implies that $a(k-1) = a(k)$, $k \in \mathbb{N}$ odd. We know from (1.10), Corollary 1.14 and Exercise 1.2 that

$$a(2k) \sim \frac{1}{\sqrt{\pi k}}, \qquad b(2k) \sim \frac{1}{2k\sqrt{\pi k}}, \qquad k \to \infty. \tag{5.6}$$

(See also F. Spitzer, *Principles of Random Walk*, 2nd. ed., Springer, 1976.)

Free energy. The *free energy* of the polymer is defined as

$$f(\zeta) = \lim_{n\to\infty} \frac{1}{n} \log Z_n^\zeta. \tag{5.7}$$

In Theorem 5.1 below we will show that the limit exists in $\mathbb{R}$. Before doing so, we explain why the function $\zeta \mapsto f(\zeta)$ is important.

Exercise 5.1. Let $f_n(\zeta) = \frac{1}{n}\log Z_n^\zeta$, $\zeta \in \mathbb{R}$, $n \in \mathbb{N}$.
(1) Show that f_n is convex for every $n \in \mathbb{N}$, i.e., $f_n(\lambda\zeta' + (1-\lambda)\zeta'') \le \lambda f_n(\zeta') + (1-\lambda) f_n(\zeta'')$ for all $\zeta', \zeta'' \in \mathbb{R}$ and $\lambda \in [0,1]$. *Hint:* Use the Hölder inequality, which says that

$$\sum_{i=1}^n g(i)h(i)p_i \le \left(\sum_{i=1}^n g(i)^p p_i\right)^{\frac{1}{p}} \left(\sum_{i=1}^n g(i)^q p_i\right)^{\frac{1}{q}} \tag{5.8}$$

for all $g,h\colon \{1,\dots,n\} \to [0,\infty)$ and all $p_1,\dots,p_n \in [0,1]$ such that $\sum_{i=1}^n p_i = 1$, where $p,q > 0$ satisfy the equation $\frac{1}{p} + \frac{1}{q} = 1$. Check that the conditions for (5.8) are satisfied.
(2) Show that if $f = \lim_{n\to\infty} f_n$ exists, then also f is convex.

An important consequence of (1) and (2), not proved here, is that $f'(\zeta) = \lim_{n\to\infty} f_n'(\zeta)$ whenever $f'(\zeta)$ exists. By (5.3–5.4), we have

$$f_n'(\zeta) = \left[\frac{1}{n}\log Z_n(\zeta)\right]' = \frac{1}{n} \sum_{w\in\mathcal{W}_n} [K_n(w)]\, \bar{P}_n^\zeta(w) = \frac{1}{n} E^{\bar{P}_n^\zeta}(K_n). \tag{5.9}$$

What this says is that $f_n'(\zeta)$ is the *average fraction of adsorbed monomers* under the law $\bar{P}_n^\zeta$ of the pinned polymer. Letting $n \to \infty$ in (5.9), we find that $f'(\zeta)$ is the limiting *average fraction of adsorbed monomers* whenever $f'(\zeta)$ exists. Consequently, at those values of ζ where the free energy fails to be differentiable this fraction is discontinuous, signalling the occurrence of what is called a *phase transition*, i.e., a drastic change in the behaviour of the typical path under the Gibbs measure. See Figs. 5.8–5.9 below for an illustration.

A similar computation as in (5.9) shows that

$$f_n''(\zeta) = \frac{1}{n} \mathrm{Var}^{\bar{P}_n^\zeta}(K_n), \tag{5.10}$$

i.e., $f_n''(\zeta)$ is the variance of $K_n/\sqrt{n}$ under the law $\bar{P}_n^\zeta$ of the pinned polymer. Thus, also the higher derivatives of the free energy have an interpretation (as long as they exist and are equal to the limit of the derivatives for finite n as $n \to \infty$, which is often the case).

Phase transition. Our first theorem settles the existence of the free energy.

Theorem 5.1. *$f(\zeta)$ exist for all $\zeta \in \mathbb{R}$ and is non-negative.*

Proof. As $Z_n^\zeta = Z_{n+1}^\zeta$ for even n, i.e., $n \in 2\mathbb{N}$, it is sufficient to show that $\lim_{n\to\infty, n\in 2\mathbb{N}} \frac{1}{n} \log Z_n^\zeta$ exists. Let (E denotes expectation w.r.t. SRW, i.e., with respect to $\bar{P}_n$)

$$Z_n^{\zeta,0} = E\left(\exp\left[\zeta \sum_{i=1}^n 1_{\{S_i=0\}}\right] 1_{\{S_n=0\}}\right), \qquad n \in 2\mathbb{N}, \tag{5.11}$$

be the partition sum for the polymer *constrained to end at the substrate*. We begin by showing that there exists a $0 < C < \infty$ such that

$$Z_n^{\zeta,0} \leq Z_n^\zeta \leq (1 + Cn) Z_n^{\zeta,0} \qquad \forall\, n \in 2\mathbb{N}. \tag{5.12}$$

The lower bound is obvious. The upper bound is proved as follows. By splitting the expectation in (5.11) according to the *last* hitting time of 0 *prior* to time n, we may write (see Fig. 5.4)

$$Z_n^\zeta = Z_n^{\zeta,0} + \sum_{\substack{k=1\\ k\in 2\mathbb{N}}}^n Z_{n-k}^{\zeta,0}\, a(k) = Z_n^{\zeta,0} + \sum_{\substack{k=1\\ k\in 2\mathbb{N}}}^n Z_{n-k}^{\zeta,0}\, b(k)\, \frac{a(k)}{b(k)}. \tag{5.13}$$

By (5.6) we know that $a(k)/b(k) \leq Ck$ for all $k \in 2\mathbb{N}$ and some $0 < C < \infty$. However, without the factor $a(k)/b(k)$, the last sum in (5.13) is precisely $Z_n^{\zeta,0}$, and so we get the upper bound in (5.12).

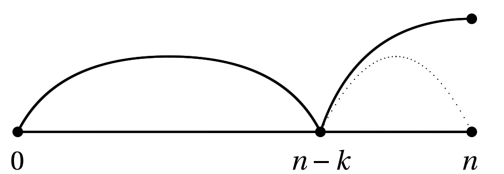

Figure 5.4: Illustration of (5.13).

To prove the existence of the free energy, note that (see Fig. 5.5)

$$Z_{m+n}^{\zeta,0} \geq Z_m^{\zeta,0}\, Z_n^{\zeta,0} \qquad \forall\, m, n \in 2\mathbb{N}, \tag{5.14}$$

which follows by inserting an extra indicator $1_{\{S_m=0\}}$ into (5.11) and using the Markov property of S at time m. This inequality says that $n \mapsto n f_n^0$ with $f_n^0(\zeta) = \frac{1}{n} \log Z_n^{\zeta,0}$ is superadditive. This implies (see Exercise 5.2) the existence of

$$f^0(\zeta) = \lim_{\substack{n\to\infty\\ n\in 2\mathbb{N}}} f_n^0(\zeta). \tag{5.15}$$

But $f^0 = f$ because of (5.12), and so f exists.

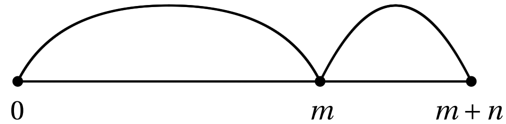

Figure 5.5: Illustration of (5.14).

Finally, $Z_n^\zeta \geq E(\exp[\zeta K_n] \mathbb{1}_{\{\sigma > n\}}) = a(n)$ implies that $f(\zeta) \geq 0$ because $\lim_{n\to\infty} \frac{1}{n} \log a(n) = 0$. □

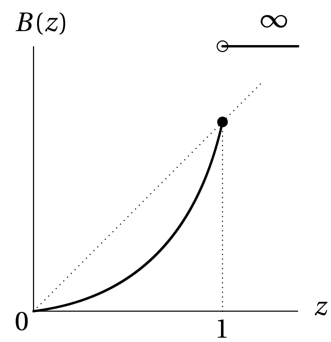

Figure 5.6: Qualitative picture of $z \mapsto B(z)$.

Exercise 5.2. Use Fekete's Lemma 4.1 to show that (5.14) implies existence of the limit in (5.15) with $f^0(\zeta) = \sup_{n \in 2\mathbb{N}} f_n^0(\zeta)$. Do not forget to show that the conditions for the application of the lemma are satisfied.

Our second theorem relates the free energy to the generating function for the length of the *excursions away from the interface*. Let (see Fig. 5.6)

$$B(z) = \sum_{k \in \mathbb{N}} z^k \, b(k), \qquad z \in [0, \infty). \tag{5.16}$$

Note that $B(z)$ is the same as $F(0; z)$ in Section 2.5.1.

Theorem 5.2. *The free energy is given by*

$$f(\zeta) = \begin{cases} 0, & \text{if } \zeta \le 0, \\ r(\zeta), & \text{if } \zeta > 0, \end{cases} \tag{5.17}$$

where $r(\zeta)$ is the unique solution of the equation

$$B(e^{-r}) = e^{-\zeta}, \qquad \zeta > 0. \tag{5.18}$$

Proof. For $\zeta \le 0$, we have the trivial bounds

$$a(n) \le Z_n^\zeta \le 1 \qquad \forall \, n \in \mathbb{N}, \tag{5.19}$$

which imply that $f(\zeta) = 0$. For $\zeta > 0$, we look at the constrained partition sum $Z_n^{\zeta,0}$ and write this out as follows (see Fig. 5.7)

$$Z_n^{\zeta,0} = \sum_{m=1}^{n} \sum_{\substack{j_1,\ldots,j_m \in \mathbb{N} \\ j_1+\cdots+j_m=n}} \prod_{i=1}^{m} e^{\zeta} \, b(j_i). \tag{5.20}$$

This expression counts the possible excursions away from the interface up to length n with their weight under the polymer measure.

Let

$$b^\zeta(k) = e^{\zeta - r(\zeta)k} \, b(k), \qquad k \in \mathbb{N}. \tag{5.21}$$

By (5.18), this is a probability distribution on $\mathbb{N}$. Moreover, because $r(\zeta) > 0$, we have

$$M_{b^\zeta} = \sum_{k \in \mathbb{N}} k b^\zeta(k) < \infty. \tag{5.22}$$

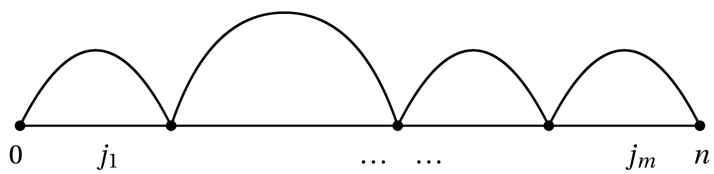

Figure 5.7: Illustration of (5.20).

Now, let $T = (T_l)_{l\in\mathbb{N}_0}$ denote the sequence of random times obtained from b^ζ as follows: $T_0 = 0$, and

$$\begin{aligned} &(1)\quad T_{l+1} - T_l,\ l \in \mathbb{N}_0, \text{ are i.i.d.,} \\ &(2)\quad P^\zeta(T_{l+1} - T_l = k) = P^\zeta(T_1 = k) = b^\zeta(2k), \qquad k \in \mathbb{N},\ l \in \mathbb{N}, \end{aligned} \tag{5.23}$$

where P^ζ denotes the law of T. A random process with property (1) is called a *renewal process*. The successive times T_l, $l \in \mathbb{N}$, are called *renewal times*. Because $b^\zeta(0) = 0$, $P^\zeta(T_{l+1} - T_l \geq 1) = 1$ for all $l \in \mathbb{N}_0$ and thus $P^\zeta(T_m \leq n) = 0$ for $n < m$. Hence $P^\zeta(\exists\, m \in \mathbb{N}_0\colon\ T_m = n) = P^\zeta(\exists\, m \in \{0,\dots,n\}\colon\ T_m = n)$.

With the help of (5.21), we may rewrite (5.20) as

$$\begin{aligned} Z_n^{\zeta,0} &= e^{r(\zeta)n}\, P^\zeta\big(\exists\, m \in \mathbb{N}_0\colon\ T_m = n\big), \\ P^\zeta\big(\exists\, m \in \mathbb{N}_0\colon\ T_m = n\big) &= P^\zeta\big(\exists\, m \in \{0,\dots,n\}\colon\ T_m = n\big) = \sum_{m=1}^{n} \sum_{\substack{j_1,\dots,j_m\in\mathbb{N} \\ j_1+\dots+j_m=n}} \prod_{i=1}^{m} b^\zeta(j_i). \end{aligned} \tag{5.24}$$

By the so-called *renewal theorem* (see Theorem 5.3 below), we have

$$\lim_{n\to\infty} P^\zeta(n \in T) = \frac{1}{M_{b^\zeta}}. \tag{5.25}$$

Intuitively, what (5.25) says is that a far away integer is hit by the renewal process with a probability that is equal to one over the average spacing between the renewal times. Combining (5.15) and (5.24–5.25), we find that $f^0(\zeta) = \lim_{n\to\infty} \frac{1}{n} \log Z_n^{\zeta,0} = r(\zeta)$. Since $f^0 = f$, as shown in the proof of Theorem 5.1, we get $f(\zeta) = r(\zeta)$. □

The proof of the following theorem can be found in standard textbooks on stochastic processes.

Theorem 5.3 (Renewal Theorem)**.** *Let $(\tau_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. $\mathbb{N}$-valued random variables with $0 < \mathbb{E}(\tau_1) < \infty$ and $P(\tau_1 = k) > 0$ for all $k \in \mathbb{N}$. Let $T_n = \sum_{i=1}^n \tau_i$. Then*

$$\lim_{n\to\infty} P\big(\exists\, m \in \mathbb{N}\colon\ T_m = n\big) = \frac{1}{E(\tau_1)}. \tag{5.26}$$

Phase transition. Theorem 5.2 shows that $\zeta \mapsto f(\zeta)$ is non-analytic at $\zeta_c = 0$. Since $x \mapsto B(x)$ is strictly increasing and analytic on $(0,1)$, it follows from (5.18) and the *implicit function theorem* that $\zeta \mapsto f(\zeta)$ is strictly increasing and analytic on $(0,\infty)$. Consequently, $\zeta_c = 0$ is the only point of non-analyticity of f, and corresponds to the phase transition.

Exercise 5.3. Give a reference for the implicit function theorem (i.e., give the Author, Title, Volume, Publisher(book)/Journal(article), Year (and place(book)) of publication, Pages(article)).

For SRW we have

$$B(z) = 1 - \sqrt{1 - z^2}, \qquad z \in [0,1]. \tag{5.27}$$

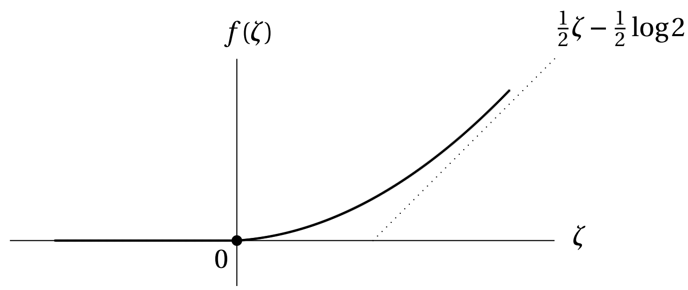

Figure 5.8: Plot of the pinned free energy.

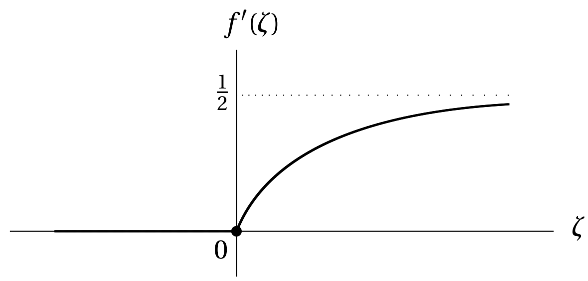

Figure 5.9: Plot of the average fraction of adsorbed monomers. In view of the relation of f' with the fraction of absorbed monomers (see (5.9)) one can see that $\zeta > 0$ corresponds to the localised phase and $\zeta \le 0$ to the delocalised phase.

Exercise 5.4. The goal of this exercise is to derive (5.27). Recall the definition of $F(0;z)$ and $G(0;z)$ in (2.35) and the definition of $a(k)$ and $b(k)$ in (5.5).
(1) Prove that $G(0;z) = \sum_{k\in\mathbb{N}_0} a(2k)z^{2k}$.
(2) Prove that $B(z) = F(0;z) = (z^2-1)G(0;z)+1$.
(3) Derive (5.27) *Hint:* use (2.36).

It follows from (5.18) and (5.27) that

$$r(\zeta) = \tfrac{1}{2}\zeta - \tfrac{1}{2}\log(2-e^{-\zeta}), \qquad \zeta > 0. \tag{5.28}$$

Figs. 5.8–5.9 show plots of the free energy and its derivative based on (5.28). Note that $\zeta \mapsto f(\zeta)$ is quadratic in a right-neighbourhood of $\zeta_c = 0$, namely, $f(\zeta) \sim \frac{1}{2}\zeta^2$, $\zeta \downarrow 0$. Therefore f' is continuous at $\zeta_c = 0$, while f'' is not. In physics terminology this is expressed by saying that the phase transition is *second order*.

5.2 Wetting

Next we investigate in what way the results in Section 5.1 are to be modified when the substrate is impenetrable, i.e., when the set of paths $\mathcal{W}_n$ in (5.1) is replaced by (see Fig. 5.10)

$$\mathcal{W}_n^+ = \Big\{w = (i, w_i)_{i=0}^n \colon w_0 = 0,\ w_{i+1} - w_i \in \{-1,+1\}\ \forall\, 0 \le i < n,\ w_i \in \mathbb{N}_0\ \forall\, 0 \le i \le n\Big\}. \tag{5.29}$$

Let $\bar{P}_n^+$ be the uniform distribution on $\mathcal{W}_n^+$. Analoguesly as in Section 5.1 let

$$\bar{P}_n^{\zeta,+}(w) = \frac{1}{Z_n^{\zeta,+}}\, e^{\zeta K_n(w)}\, \bar{P}_n^+(w), \qquad w \in \mathcal{W}_n, \qquad Z_n^{\zeta,+} = \sum_{w\in\mathcal{W}_n^+} e^{\zeta K_n(w)}\, \bar{P}_n^+(w), \qquad f^+(\zeta) = \lim_{n\to\infty} \frac{1}{n}\log Z_n^{\zeta,+} \tag{5.30}$$

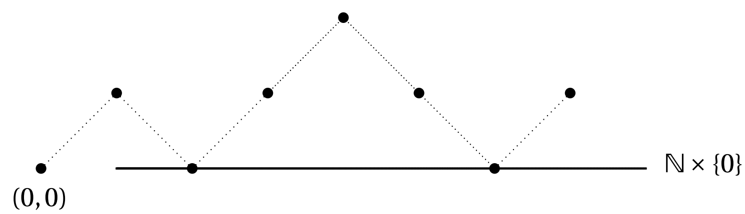

Figure 5.10: A 7-step path in $\mathbb{N}_0 \times \mathbb{N}_0$ that makes 2 visits to $\mathbb{N} \times \{0\}$.

be the path measure, partition sum and free energy in this version of the model.

The free energy can be computed using the same excursion approach as in Section 5.1. Theorem 5.1 therefore carries over immediately. The analogue of Theorem 5.2 reads as follows.

Theorem 5.4. *The free energy is given by*

$$f^+(\zeta) = \begin{cases} 0, & \text{if } \zeta \le \zeta_c^+, \\ r^+(\zeta), & \text{if } \zeta > \zeta_c^+, \end{cases} \tag{5.31}$$

where $r^+(\zeta)$ is the unique solution of the equation

$$B(e^{-r}) = e^{-(\zeta-\zeta_c^+)}, \qquad \zeta > \zeta_c^+, \tag{5.32}$$

and $\zeta_c^+ = \log 2$.

Proof. The proof uses a comparison with the pinned polymer. For $n \in \mathbb{N}$, let $\mathscr{A}_n = \{A \subset 2\mathbb{N}\colon\ A \subset (0,n]\}$. For $A \in \mathscr{A}_n$, let

$$\begin{aligned} \mathscr{N}_n(A) &= \Big\{w \in \mathscr{W}_n\colon\ w_a = 0 \text{ if and only if } a \in A \cup \{0\}\Big\}, \\ \mathscr{N}_n^+(A) &= \Big\{w \in \mathscr{W}_n^+\colon\ w_a = 0 \text{ if and only if } a \in A \cup \{0\}\Big\}. \end{aligned} \tag{5.33}$$

Then

$$|\mathscr{N}_n(A)| = \begin{cases} 2^{|A|+1}|\mathscr{N}_n^+(A)|, & \text{if } n \notin A, \\ 2^{|A|}|\mathscr{N}_n^+(A)|, & \text{if } n \in A. \end{cases} \tag{5.34}$$

Since

$$\begin{aligned} Z_n^\zeta &= 2^{-n} \sum_{A \in \mathscr{A}_n} |\mathscr{N}_n(A)|\, e^{\zeta|A|}, \\ Z_n^{\zeta,+} &= \frac{1}{|\mathscr{W}_n^+|} \sum_{A \in \mathscr{A}_n} |\mathscr{N}_n^+(A)|\, e^{\zeta|A|}, \end{aligned} \tag{5.35}$$

it follows that

$$Z_n^{\zeta,+} \le \frac{2^n}{|\mathscr{W}_n^+|} Z_n^{\zeta-\log 2} \le 2\, Z_n^{\zeta,+}. \tag{5.36}$$

One has $|\mathscr{W}_{2n}^+| \ge |\mathscr{N}_{2n}^+(\{2n\})| = C_n$, where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number, whence

$$0 \le \tfrac{1}{2n} \log \frac{2^{2n}}{|\mathscr{W}_{2n}^+|} \le -\tfrac{1}{n}\log(n+1) - \tfrac{1}{n}\log(\binom{2n}{n} 2^{-2n}). \tag{5.37}$$

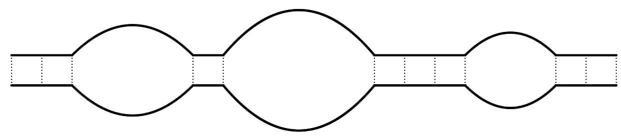

Figure 5.11: Schematic representation of the two strands of DNA in the Poland-Sheraga model. The dotted lines are the interacting base pairs, the loops are the denaturated segments without interaction.

As $|\mathscr{W}^+_{2n+1}| \geq |\mathscr{N}^+_{2n+1}(\{2n\})| = |\mathscr{N}^+_{2n}(\{2n\})|$ we have (see (1.10))

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{2^n}{|\mathscr{W}^+_n|} = 0. \tag{5.38}$$

Therefore $f^+(\zeta) = f(\zeta - \log 2)$ and $\zeta^+_c = \zeta_c + \log 2$. □

Note that $\zeta^+_c > 0$ in Theorem 5.4 whereas $\zeta_c = 0$ in Theorem 5.2. Apparently, *localization on an impenetrable substrate is harder than on a penetrable substrate.* Note that the wetted free energy is a shift of the pinned free energy.

Exercise 5.5. Prove (5.34)–(5.36).

In physics terminology, the polymer suffers a loss of entropy when it localizes in the wetting situation.

5.3 ⋆ Poland-Sheraga model

The wetting version of the polymer adsorption model can be used to describe the so-called *denaturation transition* of DNA. DNA is a string of adenine-thymine (A-T) and cytosine-guanine (C-G) base pairs forming a double helix. A and T share two hydrogen bonds, C and G share three. If we think of the two strands as performing random walks in three-dimensional space subject to the restriction that they do not cross each other, then the distance between the two strands is a random walk in the presence of a wall. This representation of DNA is called the Poland-Sheraga model (see Fig. 5.11). The localized phase corresponds to the *bounded* phase of DNA where the two strands are attached, the delocalized phase corresponds to the *denaturated phase* where the two strands are detached. Upon heating, the hydrogen bonds that keep the base pairs together can break and the two strands can separate, either partially or completely.

The Poland-Sheraga model is not entirely realistic. Since the order of the base pairs in DNA is irregular and their binding energies are different, we should actually think of DNA as a wetted polymer with binary disorder, i.e., we must modify the model in such a way that the binding energy that is picked up location i of the substrate is not a fixed number ζ but is given by a random variable ζ_i taking two different values. It is also not realistic to presume that the two strands can be modelled as SRWs. However, the theory of wetting allows for a general excursion length law (as long as it satisfies certain regularity properties). Hence, we may attempt to pick an excursion law that approximates the true spatial behaviour of DNA strands, one that takes into account for instance the self-avoidance *within* the denaturated segments. There is an extended literature on this subject.

Chapter 6

Random Walk and Brownian Motion

Brownian motion has already appeared in Chapter 4 as the scaling limit of simple random walk (in any dimension) and self-avoiding walk (in high enough dimension). It is time to give a formal definition of Brownian motion and identify its main properties.

The presentation below is based on the following literature:

(1) P. Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Statistics, 2008.

(2) P. Mörters, Y. Peres, *Brownian Motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 30, 2010. `http://www.stat.berkeley.edu/~peres/bmbook.pdf`

(3) S. Lalley, *Lecture Notes on Brownian Motion*, University of Chicago, 2012. `galton.uchicago.edu/~lalley/Courses/390/Lecture5.pdf`

6.1 Historical perspective

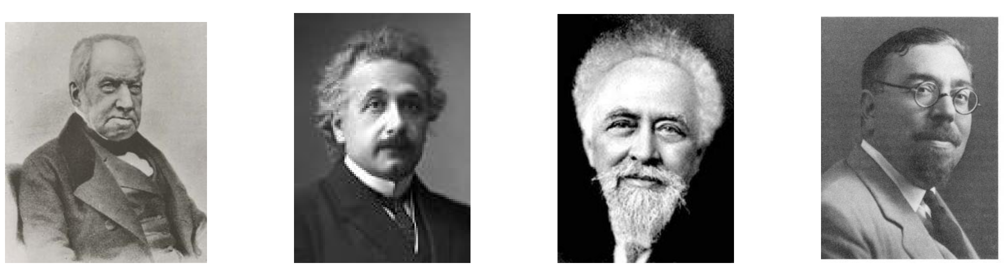

Figure 6.1: Robert Brown, Albert Einstein, Jean Perrin, Norbert Wiener.

Brownian motion is named after the Scottish botanist Robert Brown, who in 1827 observed the irregular motion of pollen particles floating in water and attributed this to an interaction between the pollen particles and the water. Later, in 1905, the Swiss physicist Albert Einstein showed that Brownian motion is the result of the erratic collision of solvent particles against solution particles, and used this to support the atomic view of matter put forward by the Austrian physicist Ludwig Boltzmann in the 1870's. His computations were confirmed in a series of experiments by French physicist Jean Perrin, who in 1926 received the Nobel Prize for his work.

Perhaps the discovery of Brownian motion is older. In 1785 the Dutch physiologist, biologist and chemist Jan Ingenhousz (1730-1799) described the irregular movement of coal dust on the surface of alcohol and therefore has a claim as the discoverer of what later came to be known as Brownian motion. The Dutch scientist Antoni van Leeuwenhoek (17th century), who invented the microscope, must have seen a similar motion also when observing mesoscopic samples from nature, but Brown was the first to understand that it was not caused organically, i.e., by the pollen particles themselves.

The American mathematician Norbert Wiener was the first to give a rigorous mathematical description of Brownian motion. He constructed a probability measure on the space of continuous paths with the appropriate properties.

6.2 Random graphs from simple random walks

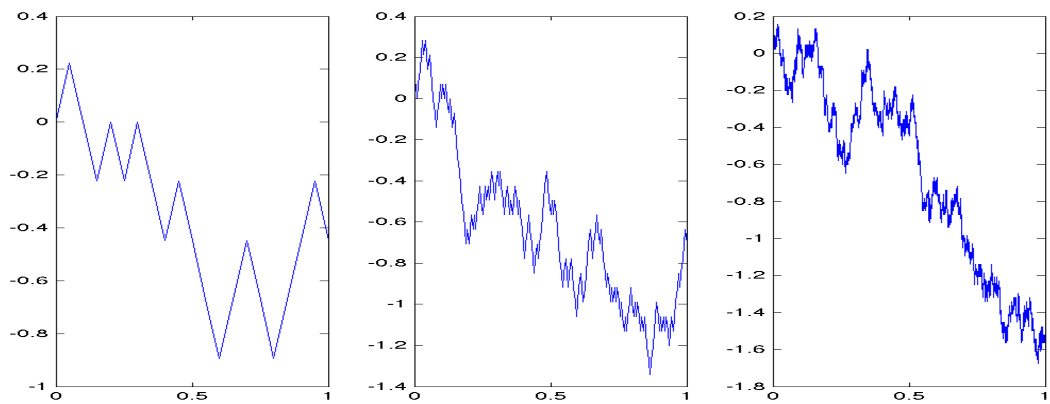

Figure 6.2: Simulations of $t \mapsto W_n(t)$ in (6.1–6.2) obtained from scaled simple random walks of length $n = 20, 200, 2000$.

Consider a one-dimensional simple random walk $S_n = \sum_{k=1}^n X_k$. The graphs in Fig. 6.2 are obtained after connecting consecutive points $\{(k, S_k)\}_{k=0}^n$ by straight lines and rescaling the horizontal axis by a factor $1/n$ and the vertical axis by a factor $1/\sqrt{n}$. More precisely, the random graphs $W_n(t)$, $t \in [0,1]$, are obtained as follows:

(1) For $t \in [0,1]$ such that nt is an integer, we put

$$W_n(t) = \frac{1}{\sqrt{n}} S_{nt}. \tag{6.1}$$

(2) For all other $t \in [0,1]$ there is a unique integer k such that $k/n < t < (k+1)/n$, namely, $k = \lfloor nt \rfloor$ with $\lfloor \cdot \rfloor$ denoting the lower integer part, and we put

$$W_n(t) = W_n\left(\frac{k}{n}\right)[1-(nt-k)] + W_n\left(\frac{k+1}{n}\right)(nt-k), \tag{6.2}$$

i.e., the linear interpolation between the points in (6.1) as drawn in Fig. 6.2.

Taking into account that

$$W_n\left(\frac{k+1}{n}\right) = W_n\left(\frac{k}{n}\right) + \frac{X_{k+1}}{\sqrt{n}}, \tag{6.3}$$

we can write (6.2) as

$$W_n(t) = W_n\left(\frac{k}{n}\right) + \frac{X_{k+1}}{\sqrt{n}}(nt-k), \tag{6.4}$$

where the last term is uniformly small, namely,

$$\left|\frac{X_{\lfloor nt\rfloor+1}}{\sqrt{n}}(nt-\lfloor nt\rfloor)\right| \le \frac{1}{\sqrt{n}}. \tag{6.5}$$

Thus, the increments across the interpolation vanish in the limit as $n\to\infty$.

The random function $W_n\colon [0,1]\to\mathbb{R}$ has the following four properties:

(i) $W_n(0)=0$.

(ii) $t\mapsto W_n(t)$ is a continuous function on $[0,1]$.

(iii) For all $t\in[0,1]$, $W_n(t)$ converges in distribution to the normal distribution with mean 0 and variance t:

$$W_n(t) \Longrightarrow \sqrt{t}Z, \qquad n\to\infty, \tag{6.6}$$

where $Z\sim\mathcal{N}(0,1)$.

(iv) For all m integer and all $0\le t_1<\ldots<t_m\le 1$ such that $nt_1,\ldots,nt_m$ are integer, the increments

$$W_n(t_j)-W_n(t_{j-1}), \qquad j=1,\ldots,m, \tag{6.7}$$

are independent random variables.

Property (iii) follows by writing (6.4) as

$$W_n(t)=\sqrt{\frac{\lfloor nt\rfloor}{n}}\frac{1}{\sqrt{\lfloor nt\rfloor}}S_{\lfloor nt\rfloor}+\frac{X_{\lfloor nt\rfloor+1}}{\sqrt{n}}(nt-\lfloor nt\rfloor) \tag{6.8}$$

and using (6.5) in combination with the CLT. Lemma 6.1 and Exercise 6.1 below explain the details.

Lemma 6.1. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ that converges to $a\in(0,\infty)$. Suppose that $(X_n)_{n\in\mathbb{N}}$ and X are random variables such that $X_n\Longrightarrow X$, and that $(Y_n)_{n\in\mathbb{N}}$ are random variables with $|Y_n|\le b_n$ for some sequence $(b_n)_{n\in\mathbb{N}}$ in $[0,\infty)$ such that $b_n\to 0$. Then*

$$a_nX_n+Y_n\Longrightarrow aX. \tag{6.9}$$

Proof. ⋆ Because $a_n\neq 0$ for all $n\in\mathbb{N}$, we have

$$P(a_nX_n+Y_n\le x)=P\big(X_n\le \tfrac{1}{a_n}(x-Y_n)\big), \tag{6.10}$$

$$P\big(X_n\le \tfrac{1}{a_n}(x-b_n)\big)\le P\big(X_n\le \tfrac{1}{a_n}(x-Y_n)\big)\le P\big(X_n\le \tfrac{1}{a_n}(x+b_n)\big). \tag{6.11}$$

Let $F(x)=P(X\le x)$, $x\in\mathbb{R}$. Let $x\in\mathbb{R}$ be such that $\frac{1}{a}x$ is a continuity point of F. Let $\epsilon>0$ and let $\delta>0$ be such that $|F(\frac{1}{a}x)-F(y)|<\epsilon$ for $y\in\mathbb{R}$ with $|\frac{1}{a}x-y|\le\delta$. Let $N\in\mathbb{N}$ be such that, for all $n\ge N$ (with $F_n(x)=P(X_n\le x)$),

$$|F_n(\tfrac{1}{a}x-\delta)-F(\tfrac{1}{a}x-\delta)|<\epsilon, \qquad |F_n(\tfrac{1}{a}x+\delta)-F(\tfrac{1}{a}x+\delta)|<\epsilon, \tag{6.12}$$

$$|\tfrac{1}{a_n}(x-b_n))-\tfrac{1}{a}x|<\delta, \qquad |\tfrac{1}{a_n}(x+b_n))-\tfrac{1}{a}x|<\delta. \tag{6.13}$$

Then, for $n\ge N$,

$$P(X_n\le\tfrac{1}{a_n}(x+b_n))\le P(X_n\le\tfrac{1}{a}x+\delta)\le F(\tfrac{1}{a}x+\delta)+\epsilon\le F(\tfrac{1}{a}x)+2\epsilon, \tag{6.14}$$

$$P(X_n\le\tfrac{1}{a_n}(x-b_n))\ge P(X_n\le\tfrac{1}{a}x-\delta)\ge F(\tfrac{1}{a}x-\delta)-\epsilon\ge F(\tfrac{1}{a}x)-2\epsilon. \tag{6.15}$$

Thus $|P(a_nX_n+Y_n\le x)-F(\frac{1}{a}x)|<2\epsilon$ for $n\ge N$. Since $P(aX\le x)=F(\frac{1}{a}x)$, this implies (6.9). □

Exercise 6.1. Show that for all $0\le t_1<t_2\le 1$ the increment $W_n(t_2)-W_n(t_1)$ converges in distribution to the normal distribution with mean 0 and variance t_2-t_1.

6.3 Wiener process

Definition 6.2. A standard Wiener process (also known as standard Brownian motion) is a stochastic process $W = (W(t))_{t\geq 0}$ assuming real values with the following four properties:

(i) $W(0) = 0$.

(ii) With probability 1, the function $t \mapsto W(t)$ is continuous.

(iii) For all $t, s > 0$, the increment $W(t+s) - W(t)$ is $\mathcal{N}(0, s)$-distributed.

(iv) For all $t_1 < t_2 < \ldots < t_m$, the increments

$$W(t_2) - W(t_1), \quad W(t_3) - W(t_2), \quad \ldots, \quad W(t_m) - W(t_{m-1}) \tag{6.16}$$

are independent.

A realisation of the Wiener process is drawn in Fig. 6.3.

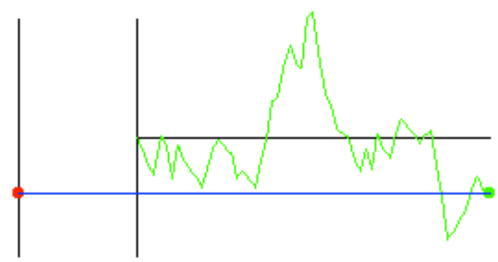

Figure 6.3: A realisation of the Wiener process, viewed from its starting point and from its ending point at a given time.

Remark 6.3. The following observations are in order.

(a) Property (iii) implies that W has stationary increments, i.e., the distribution of $W(t+s) - W(t)$ is independent of t.

(b) W is an example of a stochastic process with stationary and independent increments.

(c) $\widetilde{W} = (\widetilde{W}(t))_{t\geq 0}$ with $\widetilde{W}(t) = \sigma W(t)$, $\sigma > 0$, is a stochastic process with the same properties, except that

$$\widetilde{W}(t+s) - \widetilde{W}(t) \sim \mathcal{N}(0, s\sigma^2). \tag{6.17}$$

A priori it is *not* clear that there exists a family of random variables $W(t)$ indexed by $t \geq 0$ satisfying all four properties in Definition 6.2. However, comparing the four properties of the scaled simple random walk process $W_n(t)$ with those of the standard Wiener process $W(t)$, we are inclined to believe that the random function $W_n = W_n(\cdot)$ converges to the random function $W = W(\cdot)$ in some appropriate sense. In Sections 6.4–6.5 we will deal with the *existence* and the *uniqueness* of W.

Exercise 6.2. Assume that Wiener processes exist on $[0,1]$ (this will be proved in §6.4 and in §6.5). Assume that $W_1, W_2, \ldots$ are independent Wiener processes on $[0,1]$. Define $\overline{W}$ on $[0,\infty)$ by

$$\overline{W}(t) = W^1(1) + \cdots W^{\lfloor t\rfloor -1}(1) + W^{\lfloor t\rfloor}(t - \lfloor t\rfloor) \tag{6.18}$$

Show that $\overline{W}$ satisfies (i),(ii) and (iii) of Definition 6.2. *Hints:* For (ii): First show that for a countable number of events E_n with $P(E_n) = 1$ for all $n \in \mathbb{N}$ it is true that $P(\bigcap_{n\in\mathbb{N}} E_n) = 1$. For (iii): If $\lfloor s+t \rfloor > \lfloor t \rfloor$, write $\overline{W}(s+t) - \overline{W}(t)$ as a sum of independent normal random variables. Use that independence between the $W_1, W_2, \ldots$ implies independence between their increments. *Bonus:* Show that $\overline{W}$ satisfies (iv) of Definition 6.2.

Exercise 6.3. Prove that the following stochastics processes are also standard Wiener processes on $[0,\infty)$:

(1) $(-W(t))_{t\ge 0}$.

(2) $(W(t+s) - W(s))_{t\ge 0}$.

(3) $(aW(t/a^2))_{t\ge 0}$ for any $a \neq 0$.

6.4 ⋆ Existence of the Wiener process [intermezzo]

Let $C = C([0,1])$ be the space of continuous functions on $[0,1]$. We equip C with the *uniform norm* $\|\cdot\|\colon C \to \mathbb{R}_+$ defined by

$$\|f\| = \sup_{x\in[0,1]} |f(x)|. \tag{6.19}$$

Consider also the corresponding metric $\rho(f,g) = \|f-g\|$. It is easy to show that with this metric $C([0,1])$ becomes a complete separable metric space. The *Wiener space* $\mathscr{W} = \mathscr{W}([0,1])$ is defined to be the space of those $f \in C$ satisfying $f(0) = 0$.

Remark 6.4. A real-valued continuous-time stochastic process $X = (X(t))_{t\in[0,1]}$ with continuous sample paths satisfying $X(\omega,0) = 0$ can be thought of as a random variable defined on some probability space (Ω, P) taking values in $\mathscr{W}$:

$$X\colon \Omega \to \mathscr{W}. \tag{6.20}$$

Suppose for the moment that the Wiener process exists. This process induces the *Wiener measure*, a probability distribution on $\mathscr{W}$ defined for, say, open sets $A \subset \mathscr{W}$ by

$$P^W(A) = P\big(\omega \in \Omega\colon\ X(\omega,\cdot) \in A\big). \tag{6.21}$$

Conversely, if we construct a Wiener measure P^W on $\mathscr{W}$, then we establish existence of the Wiener process.

Let P_n^W be the probability measure induced on $\mathscr{W}$ but for the rescaled random walk path W_n:

$$W_n(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt\rfloor} + \frac{X_{\lfloor nt\rfloor+1}}{\sqrt{n}}(nt - \lfloor nt\rfloor), \qquad t \in [0,1]. \tag{6.22}$$

We claim that the sequence $(P_n^W)_{n\in\mathbb{N}}$ converges *weakly* to some probability measure P^W on $\mathscr{W}$, written $P_n^W \Longrightarrow P^W$. By definition, *weak convergence* means that for any bounded continuous function $f\colon \mathscr{W} \to \mathbb{R}$ the corresponding expected values converge, i.e.,

$$\lim_{n\to\infty} \mathbb{E}\big(f(W_n)\big) = \mathbb{E}\big(f(W)\big), \tag{6.23}$$

where $W_n \sim P_n^W$ and $W \sim P^W$. Moreover, the expected values $\mathbb{E}(f(W))$, where $W \sim P^W$ and f ranges over the set of all bounded continuous functions on $\mathscr{W}$, completely determine P^W. This definition of the weak convergence generalises the notion of *convergence in distribution* of probability distributions on the real line.

The desired convergence will be achieved in two steps (henceforth we drop the superscript W):

Step 1: Convergence of the finite-dimensional marginal distributions: for any $t_1 < t_2 < \ldots < t_k$,

$$(W_n(t_1),\ldots,W_n(t_k)) \Longrightarrow (W(t_1),\ldots,W(t_k)). \tag{6.24}$$

Therefore, if $f\colon \mathscr{W} \to \mathbb{R}$ is such that $f(W)$, $W \in \mathscr{W}$, depends only on the values of W at $t_1,\ldots,t_k \in [0,1]$, then we have established convergence

$$\lim_{n\to\infty} \mathbb{E}\big(f(W_n)\big) = \mathbb{E}\big(f(W)\big). \tag{6.25}$$

Step 2: In general, the weak convergence $P_n \Longrightarrow P$ does not follow from the weak convergence of the finite-dimensional distributions. We need to impose additional conditions to capture what is happening at intermediate times. In particular, we need to show *tightness*: a sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ on C is tight when for every $\epsilon > 0$ there exists a compact set $K = K(\epsilon) \subset C$ such that

$$P_n(K) \geq 1-\epsilon \qquad \forall\, n \in \mathbb{N}. \tag{6.26}$$

Exercise 6.4. Prove (6.24). *Hint*: Note the following facts: If $X_n^1,\ldots,X_n^d$ are independent for all $n \in \mathbb{N}$, then $(X_n^1,\ldots,X_n^d) \Longrightarrow (X^1,\ldots,X^d)$ when $X_n^i \Longrightarrow X^i$ for $i \in \{1,\ldots,d\}$. For a $d\times d$ invertable matrix A and a d-dimensional random vector X we have $P(AX \leq x) = P(X \leq A^{-1}x)$, and hence $X_n \Longrightarrow X$ if and only if $AX_n \Longrightarrow AX$.

On the space C the tightness of a sequence $(P_n)_{n\in\mathbb{N}}$ can be checked effectively.

Theorem 6.5. *The sequence of measures $(P_n)_{n\in\mathbb{N}}$ on C is tight if and only if the following two conditions hold:*

(a) For every $\eta > 0$ there exists an $a > 0$ such that

$$P_n\big(w \in C\colon\ |w(0)| > a\big) \leq \eta \qquad \forall\, n \in \mathbb{N}. \tag{6.27}$$

(b) For every $\epsilon,\eta > 0$ there exist $0 < \delta < 1$ and $n_0 \in \mathbb{N}$ such that

$$P_n\Big(w \in C\colon \sup_{t\leq s\leq t+\delta} |w(s)-w(t)| \geq \epsilon\Big) \leq \eta\delta, \qquad \forall\, n \geq n_0,\ t \geq 0. \tag{6.28}$$

1. For our random functions

$$W_n(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt\rfloor} + \frac{X_{\lfloor nt\rfloor+1}}{\sqrt{n}}(nt - \lfloor nt\rfloor), \tag{6.29}$$

we have $W_n(0) = 0$ and so condition (a) is automatic. To prove condition (b) we need to check that the functions $W_n(\cdot)$ do not oscillate too wildly. For simplicity, let us assume that $t = k/n$ and $\delta = m/n$, and consider the event

$$A = \left\{ \sup_{s\in\{\frac{k}{n},\ldots,\frac{k+m}{n}\}} |W_n(s) - W_n(t)| \geq \epsilon \right\} = \left\{ \sup_{j\in\{1,\ldots,m\}} |S_{k+j} - S_k| \geq \lambda\sqrt{m} \right\} \tag{6.30}$$

with $\lambda = \epsilon/\sqrt{\delta}$. In Step 3 below will prove that

$$P_n(A) \leq 2\,P_n\left(|S_{k+m} - S_k| \geq (\lambda - \sqrt{2})\sqrt{m}\right). \tag{6.31}$$

2. The inequality in (6.31) is trivial when $\lambda < \sqrt{2}$. Therefore suppose that $\lambda > 2\sqrt{2}$. Then $\lambda - \sqrt{2} > \frac{1}{2}\lambda$, and hence

$$P_n\left(\sup_{j\in\{1,\ldots,m\}} |S_{k+j} - S_k| \geq \lambda\sqrt{m} \right) \leq 2\,P_n\left(|S_{k+m} - S_k| \geq \tfrac{1}{2}\lambda\sqrt{m}\right). \tag{6.32}$$

However, by the Central Limit Theorem,

$$P_n\left(\frac{1}{\sqrt{m}}|S_{k+m}-S_k| \geq \tfrac{1}{2}\lambda\right) \to P\left(|Z| > \tfrac{1}{2}\lambda\right) = P\left(|Z|^3 > \tfrac{1}{8}\lambda^3\right) \leq \frac{8}{\lambda^3}E[|Z|^3]. \tag{6.33}$$

Therefore

$$\limsup_{n\to\infty} P_n\left(\sup_{s\in\{\frac{k}{n},\ldots,\frac{k+m}{n}\}} |W_n(s)-W_n(t)| \geq \epsilon\right) \leq \frac{C\delta^{3/2}}{\epsilon^3} = \left(\frac{C\delta^{1/2}}{\epsilon^3}\right)\delta \leq \eta\delta \tag{6.34}$$

for δ sufficiently small, uniformly in k, m. Hence condition (b) holds.

3. Suppose that $\lambda \geq \sqrt{2}$. Let $S'_j = \sum_{i=k+1}^{k+j} X_i$, and consider the events

$$E_j = \left\{|S'_i| < \lambda\sqrt{m}\ \forall\, i = 1,\ldots,j-1,\ |S'_j| \geq \lambda\sqrt{m}\right\}, \qquad j = 1,\ldots,m. \tag{6.35}$$

These events form a partition of A. Consider the event

$$B = \left\{|S'_m| \geq (\lambda-\sqrt{2})\sqrt{m}\right\}. \tag{6.36}$$

Then

$$A = \left(\bigcup_{i=1}^{m}(E_j \cap B)\right) \cup \left(\bigcup_{j=1}^{m}(E_j \cap B^c)\right). \tag{6.37}$$

Hence

$$P_n(A) \leq P_n(B) + \sum_{j=1}^{m-1} P_n(E_j \cap B^c). \tag{6.38}$$

Since the conditions $|S'_j| \geq \lambda\sqrt{m}$, $|S'_m| < (\lambda-\sqrt{2})\sqrt{m}$ imply that $|S'_m - S'_j| \geq \sqrt{2m}$, we obtain

$$\sum_{j=1}^{m-1} P_n(E_j \cap B^c) \leq \sum_{j=1}^{m-1} P_n\left(E_j \cap \left\{|S'_m - S'_j| \geq \sqrt{2m}\right\}\right) = \sum_{j=1}^{m-1} P_n(E_j)\,P_n\left(|S'_m - S'_j| \geq \sqrt{2m}\right), \tag{6.39}$$

where the last equality follow from the independence of E_j and $\{|S'_m - S'_j\}$ for all j. By Chebyshev's inequality,

$$P_n\left(|S'_m - S'_j| \geq \sqrt{2m}\right) \leq \frac{\operatorname{var}(S'_m - S'_j)}{2m} = \frac{m-j}{2m} \leq \frac{1}{2}. \tag{6.40}$$

Hence

$$P_n(A) \leq P_n(B) + \sum_{j=1}^{m-1} \tfrac{1}{2}P_n(E_j) \leq P_n(B) + \tfrac{1}{2}\sum_{j=1}^{m} P_n(E_j) = P_n(B) + \tfrac{1}{2}P_n(A). \tag{6.41}$$

Thus, $P_n(A) \leq 2P_n(B)$, and hence (6.31) holds.

6.5 ⋆ Explicit construction of the Wiener process [intermezzo]

In this section we show how the Wiener process can be constructed as the limit of piecewise linear random graphs, without taking recourse to a comparison with scaled simple random walk. The construction will be in the line of the following remark.

Remark 6.6. Let $(W(t))_{t\geq 0}$ denote the standard Wiener process. Suppose $(X(t))_{t\geq 0}$ is a stochastic process for which $X(0) = 0$, $t \mapsto X(t)$ is continuous, and the distribution of

$$(X(t_1), X(t_2), \ldots, X(t_m)) \tag{6.42}$$

coincides with $(W(t_1), W(t_2), \ldots, W(t_m))$ for all $t_1 < t_2 < \cdots < t_m$. Then $(X(t))_{t\geq 0}$ is a standard Wiener process, since this implies that the distribution of $(X(t_1), X(t_2) - X(t_1), \ldots, X(t_m) - X(t_{m-1}))$ and $(W(t_1), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1}))$ coincides for all $t_1 < t_2 < \cdots < t_m$.

Exercise 6.5. ⋆ Check that the last statement holds and that it implies properties (iii) and (iv) of Definition 6.2.

For $m \in \mathbb{N}_0$, let

$$D_m = \left\{ \frac{k}{2^m}, \quad k = 0, \dots, 2^m \right\} \tag{6.43}$$

be the set of m-dyadic rationals in $[0,1]$.

Step 1: $m = 0$. Let $W_0(0) = 0$, and $W_0(1) = \zeta_{0,1}$ with $\zeta_{0,1} \sim \mathcal{N}(0,1)$. Define $W_0(t)$ as the linear function on $[0,1]$ with values $W_0(0) = 0$ and $W_0(1) = \zeta_{0,1}$:

$$W_0(t) = t\zeta_{0,1}. \tag{6.44}$$

Note that distributions of $(W_0(0), W_0(1))$ and $(W(0), W(1))$ coincide.

Step 2: $m = 1$. Put

$$W_1(0) = W_0(0), \quad W_1(1) = W_0(1). \tag{6.45}$$

We select an appropriate random value for $W_1(\frac{1}{2})$, and choose $W_1(t)$ to be linear on $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, interpolating between the values $W_1(0)$, $W_1(\frac{1}{2})$ and $W_1(1)$. The choice of $W_1(\frac{1}{2})$ is motivated by the following observation. Suppose that W is the standard Wiener process. Then

$$X = W(\tfrac{1}{2}) - W(0), \quad Y = W(1) - W(\tfrac{1}{2}), \tag{6.46}$$

are independent $\mathcal{N}(0, \frac{1}{2})$-random variables. What is the distribution of $W(\frac{1}{2})$ given that we know $W(0) = 0$ and $W(1)$? The answer is given by the following lemma.

Lemma 6.7. *Suppose that X and Y are independent random variables with $X \sim \mathcal{N}(0,s)$ and $Y \sim \mathcal{N}(0,t)$. Then the conditional distribution of X given that $X + Y = z$ is also normal, with mean $\frac{zs}{s+t}$ and variance $\frac{st}{s+t}$, i.e.,*

$$\textit{distibution}\big(X \mid X + Y = z\big) = \mathcal{N}\left(\frac{zs}{s+t}, \frac{st}{s+t}\right). \tag{6.47}$$

Exercise 6.6. ⋆ Prove Lemma 6.7. *Hint:* Let $Z = X + Y$. Then $Z \sim \mathcal{N}(0, s+t)$. Show that

$$(X, Z) \sim \mathcal{N}(0, \Sigma), \quad \text{with} \quad \Sigma = \begin{pmatrix} s & s \\ s & s+t \end{pmatrix}. \tag{6.48}$$

Use

$$f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_Z(z)}, \tag{6.49}$$

to prove the claim.

Applying Lemma 6.7, we find that

$$\text{distibution}\big(W(\tfrac{1}{2}) \mid W(1) = \zeta_{0,1}\big) = \mathcal{N}\left(\tfrac{1}{2}\zeta_{0,1}, \tfrac{1}{4}\right). \tag{6.50}$$

Thus, if we let

$$W_1(\tfrac{1}{2}) = \tfrac{1}{2}W_1(1) + \tfrac{1}{2}\zeta_{1,1} = W_0(\tfrac{1}{2}) + \tfrac{1}{2}\zeta_{1,1} \tag{6.51}$$

with $\zeta_{1,1} \sim N(0,1)$ and independent of $\zeta_{0,1}$, then we find that the joint distributions of

$$(W_1(0), W_1(\tfrac{1}{2}), W_1(1)), \quad (W(0), W(\tfrac{1}{2}), W(1)), \tag{6.52}$$

coincide.

Step 3: $m \in \mathbb{N}$. Suppose that we have constructed random variables

$$\left(W_m(0), W_m\left(\frac{1}{2^m}\right), \ldots, W_m\left(\frac{2^m-1}{2^m}\right), W_m(1)\right), \tag{6.53}$$

indexed by the dyadic rationals D_m, such that their joint distribution coincides with the corresponding distribution for the Wiener process

$$\left(W(0), W\left(\frac{1}{2^m}\right), \ldots, W\left(\frac{2^m-1}{2^m}\right), W(1)\right). \tag{6.54}$$

If we want to extend this collection to the next level, then we need to define

$$W_{m+1}\left(\frac{k}{2^{m+1}}\right) \tag{6.55}$$

for *odd* k. The trick is to choose $W_{m+1}(k2^{-m-1})$ conditional on $W_{m+1}((k-1)2^{-m-1})$ and $W_{m+1}((k+1)2^{-m-1})$, which we already know. Indeed, an application of Lemma 1 gives

$$\text{distribution}\left(W_{m+1}\left(\frac{k}{2^{m+1}}\right) \,\middle|\, W_{m+1}\left(\frac{k-1}{2^{m+1}}\right) = a, W_{m+1}\left(\frac{k-1}{2^{m+1}}\right) = b\right) = \mathcal{N}\left(\frac{a+b}{2}, \frac{1}{2^{m+2}}\right). \tag{6.56}$$

Hence

$$W_{m+1}(t) = W_m(t) + \sum_{k=1}^{2^m} \frac{1}{2^{(m+2)/2}} \zeta_{m+1,k} G_{m,2k-1}(t), \tag{6.57}$$

where $\zeta_{m+1,k}$ are independent $N(0,1)$-random variables, and the Schrauder functions $G_{m,k}$ are given by

$$G_{m,2k-1}(t) = \begin{cases} 2^{m+1}t - (2k-2), & \text{for } \frac{2k-2}{2^{m+1}} \le t \le \frac{2k-1}{2^{m+1}}, \\ 2k - 2^{m+1}t, & \text{for } \frac{2k-1}{2^{m+1}} \le t \le \frac{2k}{2^{m+1}}, \\ 0, & \text{otherwise.} \end{cases} \tag{6.58}$$

Theorem 6.8 (P. Lévy)**.** *If $\zeta_{m,k}$ are independent $N(0,1)$-random variables, then with probability one the infinite series*

$$W(t) = \zeta_{0,1} t + \sum_{m\in\mathbb{N}_0} \sum_{k=1}^{2^m} \zeta_{m+1,k} G_{m,2k-1}(t) \frac{1}{2^{(m+2)/2}} \tag{6.59}$$

converges uniformly on $[0,1]$*. The limit function $W(t)$ is a standard Wiener process.*

Proof. The idea is if a sequence $(f_n)_{n\in\mathbb{N}}$ of continuous functions on $[0,1]$ converges uniformly to f, then f is continuous.

We have

$$W(t) = \zeta_{0,1} t + \sum_{m\in\mathbb{N}_0} Y_m(t), \quad Y_m(t) = \sum_{k=1}^{2^m} \zeta_{m+1,k} G_{m,2k-1}(t) \frac{1}{2^{(m+2)/2}}. \tag{6.60}$$

Let $H_m = \max_{t\in[0,1]} |Y_m(t)| = 1/2^{(m+2)/2} \max_k |\zeta_{m,k}|$. Hence, for every $c_m > 0$,

$$P(H_m \ge 2^{-(m+2)/2} c_m) \le P\left(\max_{k=1,\ldots,2^m} |\zeta_{m,k}| \ge c_m\right) \le \sum_{k=1}^{2^m} P(|\zeta_{m,k}| \ge c_m) \le 2^m \frac{A}{c_m} e^{-\frac{1}{2}c_m^2} \tag{6.61}$$

for some constant A. Let $c_m = B\sqrt{m}$ for some $B > \sqrt{2\log 2}$. Then

$$P(H_m \ge 2^{-(m+2)/2} c_m) \le \exp\left[m\left(\log 2 - \frac{B^2}{2}\right)\right] \frac{A'}{\sqrt{m}}, \tag{6.62}$$

and hence

$$\sum_{m\in\mathbb{N}_0} P(H_m \ge 2^{-(m+2)/2} c_m) < \infty. \tag{6.63}$$

Therefore, by the Borel-Cantelli lemma,

$$P\big(H_m \ge 2^{-(m+2)/2} c_m \text{ for infinitely many } m\big) = 0. \tag{6.64}$$

This means that, for almost all ω, there exists an $M = M(\omega)$ such that $H_m(\omega) < c_m$ for all $m \ge M$. Hence the series in (6.60) converges absolutely and uniformly on $[0,1]$:

$$\left|\sum_{m=N}^{\infty}\sum_{k=1}^{2^m} \zeta_{m+1,k} G_{m,2k-1}(t) \frac{1}{2^{(m+2)/2}}\right| \le \sum_{m=N}^{\infty} 2^{-(m+2)/2} c_m \to 0, \qquad N \to \infty. \tag{6.65}$$

□

Exercise 6.7. ⋆ Check that the process constructed in Theorem 6.8 is indeed the standard Wiener process: Use the preceding to conclude that for $s_1 < \cdots < s_n$ in D_m the distribution of the vector

$$(W(s_1), \dots, W(s_n)) \tag{6.66}$$

coincides with the distribution of the vector in $s_1 < \cdots < s_n$ of the Wiener process. Then prove that for general $t_1 < \cdots < t_n$ the vector $(W(t_1), \dots, W(t_n))$ also has the desired properties by a limiting argument: Almost sure convergence implies convergence in law, i.e., convergence in distribution, also called weak convergence (no proof is needed for this implication).

6.6 ⋆ Path properties of the Wiener process [intermezzo]

There are continuous functions that are nowhere differentiable. It is a good exercise to construct an example of such a function, but this is not so easy. Interestingly, it turns out that the path of a Brownian motion has this property almost surely. On the other hand, it does have some form of "smoothness". Below we list a few key properties without proof.

For $f\colon [0,1) \to \mathbb{R}$, define the upper and lower right derivatives as

$$D^* f(t) = \limsup_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}, \quad D_* f(t) = \liminf_{h\downarrow 0} \frac{f(t+h) - f(t)}{h}. \tag{6.67}$$

Theorem 6.9 (Paley, Wiener and Zygmund 1933)**.** *Almost surely, Brownian motion is nowhere differentiable and, for all t, either $D^* W(t) = \infty$ or $D_* W(t) = -\infty$ or both.*

Definition 6.10 (Hölder continuity)**.** A function f is said to be locally α-Hölder continuous at x when there exist $\epsilon, c > 0$ such that $|f(x) - f(y)| \le c|x-y|^\alpha$ for all y with $|y - x| < \epsilon$.

Lemma 6.11. *There exists a constant $C > 0$ such that, almost surely, for $h > 0$ sufficiently small and $0 \le t \le 1-h$,*

$$|W(t+h) - W(t)| \le C\sqrt{h \log(1/h)}. \tag{6.68}$$

Corollary 6.12. *For every $\alpha \in (0, \frac{1}{2})$, almost surely Brownian motion is everywhere locally α-Hölder continuous.*

Consider the set of zeroes of the Wiener process:

$$Z = \{t \ge 0\colon\ W(t) = 0\}. \tag{6.69}$$

Theorem 6.13. *Almost surely, Z is a perfect set, i.e., Z is closed and for every $t \in Z$ there exists a sequence $(t_n)_{n\in\mathbb{N}}$ of distinct elements in Z such that $\lim_{n\to\infty} t_n = t$.*

6.7 The higher-dimensional Wiener process

Definition 6.14. A d-dimensional Wiener process is a stochastic process $W = (W(t))_{t\geq 0}$ with

$$W(t) = (W_1(t), \ldots, W_d(t)) \tag{6.70}$$

assuming values in $\mathbb{R}^d$ with the following four properties:

(i) $W(0) = 0$.

(ii) With probability 1, the function $t \mapsto W(t)$ is continuous.

(iii) For $t, s > 0$, the increment $W(t+s) - W(t)$ is $\mathcal{N}(0, s\mathrm{Id}_d)$-distributed, where Id_d is the $d \times d$ identity matrix.

(iv) For all $t_1 < t_2 < \ldots < t_m$, the increments

$$W(t_2) - W(t_1), \quad W(t_3) - W(t_2), \quad \ldots, \quad W(t_m) - W(t_{m-1}), \tag{6.71}$$

are independent.

Corollary 6.15. *If W is the d-dimensional Wiener process, then each component $W_j = (W_j(t))_{t\geq 0}$ is the standard one-dimensional Wiener process.*

The Wiener process in d dimensions has interesting invariance properties that we will not discuss, e.g. its distribution is *isotropic*. For $d = 2$ it is even conformally invariant, i.e., its distribution is invariant under conformal mappings of $\mathbb{R}^2$ (see Fig. 6.4).

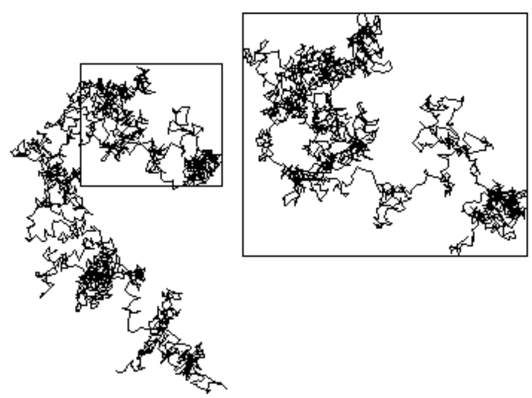

Figure 6.4: A realisation of the two-dimensional Wiener process.

6.8 Diffusion equations

Given a bounded continuous function $f\colon \mathbb{R} \to \mathbb{R}$, define

$$u(x,t) = E\big(f(x + W(t))\big). \tag{6.72}$$

Theorem 6.16 (Feynman-Kac formula)**.** *The function $u(x,t)$ defined in* (6.72) *is the unique solution of the partial differential equation (PDE)*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}, \qquad u(x,0) = f(x). \tag{6.73}$$

which is called the heat equation.

Proof. Note that

$$\begin{aligned} u(x,t+s) &= E(f(x+W(t+s))) = E\Big(f\big(x+[W(t+s)-W(t)]+W(t)\big)\Big) \\ &= E\big(u(x+[W(t+s)-W(t)],t)\big) = E\big(u(x+W(s),t)\big), \end{aligned} \tag{6.74}$$

where we use the independence of $W(t+s)-W(t)$ and $W(t)$, as well as the fact that $W(t+s)-W(t)$ and $W(s)$ have the same distribution. Consider the Taylor series expansion

$$u(x+W(s),t) = u(x,t) + \left[\frac{\partial}{\partial x}u(x,t)\right]W(s) + \left[\frac{1}{2}\frac{\partial^2}{\partial x^2}u(x,t)\right]W^2(s) + o(W^2(s)). \tag{6.75}$$

After taking expectation and using the fact that $W(s) \sim \mathcal{N}(0,s)$, in particular, $E(W(s)) = 0$ and $E(W^2(s)) = s$, we obtain that

$$u(x,t+s) = \mathbb{E}(u(x+W(s),t)) = u(x,t) + \left[\frac{1}{2}\frac{\partial^2}{\partial x^2}u(x,t)\right]s + o(s). \tag{6.76}$$

Therefore

$$\frac{\partial}{\partial t}u(x,t) = \lim_{s\downarrow 0}\frac{u(x,t+s)-u(x,t)}{s} = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(x,t). \tag{6.77}$$

The initial condition $u(x,0) = E(f(x+W_0)) = E(f(x)) = f(x)$ is clearly satisfied. □

The importance of the Feynman-Kac formula in (6.72) is that it provides the solution to the heat equation in terms of a simple formula involving Brownian motion. We think of $u(x,t)$ as the amount of heat at site x at time t when $f(x)$ is the amount of heat at site x at time $t = 0$.

The Feynman-Kac formula admits many generalisations. For example, if

$$v(x,t) = E(f(x+W(t))) + \int_0^t E(g(x+W(s)))\,ds, \tag{6.78}$$

then $v(x,t)$ satisfies the heat equation with a source term:

$$\frac{\partial v}{\partial t} = \frac{1}{2}\frac{\partial^2 v}{\partial x^2} + g(x), \qquad v(x,0) = f(x). \tag{6.79}$$

Similarly, if

$$w(x,t) = E\Big(f(x+W(t))\exp\Big[\int_0^t g(x+W(s))\,ds\Big]\Big), \tag{6.80}$$

then $w(x,t)$ satisfies

$$\frac{\partial w}{\partial t} = \frac{1}{2}\frac{\partial^2 w}{\partial x^2} + g(x)w, \qquad w(x,0) = f(x). \tag{6.81}$$

Exercise 6.8. ⋆ Given a bounded continuous function $f\colon \mathbb{R}\to\mathbb{R}$ and a closed interval $[a,b]\subset\mathbb{R}$, show that a solution of the PDE

$$\frac{1}{2}\frac{\partial^2\chi(x)}{\partial x^2} = 0, \quad x \in (a,b), \qquad \chi(x) = f(x), \quad x \in \{a,b\}, \tag{6.82}$$

is given by

$$\chi(x) = E_x\left(f\left(W_{\tau_{\{a,b\}}}\right)\right), \quad x \in [a,b], \tag{6.83}$$

where $\tau_{\{a,b\}}$ is the first hitting time of the boundary $\{a,b\}$ and E_x is expectation w.r.t. the one-dimensional Wiener process starting at x.

The result in Exercise 6.8 is in some sense the continuous analogue of what we found in Chapter 3 for the linear network.

Chapter 7

Random Walk and the Binomial Asset Pricing Model

In this chapter we consider an application of random walks in finance. In this area the "random walk hypothesis" states that stock market prices evolve according to a random walk and hence cannot be predicted. This concept dates back to the 19th century, and was developed by the French mathematician Louis Bachelier (see Fig. 7.1).

Wikipedia:

> "Louis Jean-Baptiste Alphonse Bachelier (1870–1946) was a French mathematician at the turn of the 20th century. He is credited with being the first person to model the stochastic process now called Brownian motion, which was part of his PhD thesis *The Theory of Speculation*, published in 1900. This thesis, which discussed the use of Brownian motion to evaluate stock options, is historically the first paper to use advanced mathematics in the study of finance. Thus, Bachelier is considered a pioneer in the study of financial mathematics and stochastic processes. Also notable is that Bachelier's work on random walks was more mathematical and predated Einstein's celebrated study of Brownian motion by five years."

Figure 7.1: Picture of Louis Bachelier at an early age.

Our exposition uses the following sources:

(1) J.C. Cox, S.A. Ross, M. Rubinstein, Option pricing: A simplified approach, Journal of Financial Economics 7, 229 (1979).
`http://www.dms.umontreal.ca/~morales/docs/cox_rubinstein_ross.pdf`

(2) S.E. Shreve, *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*, Springer Finance, 2004.

(3) T. Tao
http://terrytao.wordpress.com/2008/07/01/the-black-scholes-equation/

(4) Wikipedia `http://en.wikipedia.org/wiki/Binomial_options_pricing_model`

In Section 7.1 we formulate a model for stock pricing, look at the role of the money market and of interest rates, and introduce the notion of arbitrage. In Section 7.2 we look at a financial derivative called option and look at its pricing. In Section 7.3 we derive the Black-Scholes formula for the price of an option.

7.1 Stock pricing

7.1.1 The Binomial Asset Pricing Model

Stock prices are modelled in discrete time. Initially, the stock price is $S_0 > 0$. At each time step, the stock price changes to one of two possible values, dS_0 or uS_0, where d, u satisfy

$$0 < d < 1 < u. \tag{7.1}$$

The change from S_0 to dS_0 represents a *downward* movement, while the change from S_0 to uS_0 represents an *upward* movement. Suppose that a coin is tossed. When the outcome is "Head" the stock price moves up, when the outcome is "Tail" the stock price moves down:

$$S_1(\omega) = \begin{cases} uS_0, & \text{if } \omega = H, \\ dS_0, & \text{if } \omega = T. \end{cases} \tag{7.2}$$

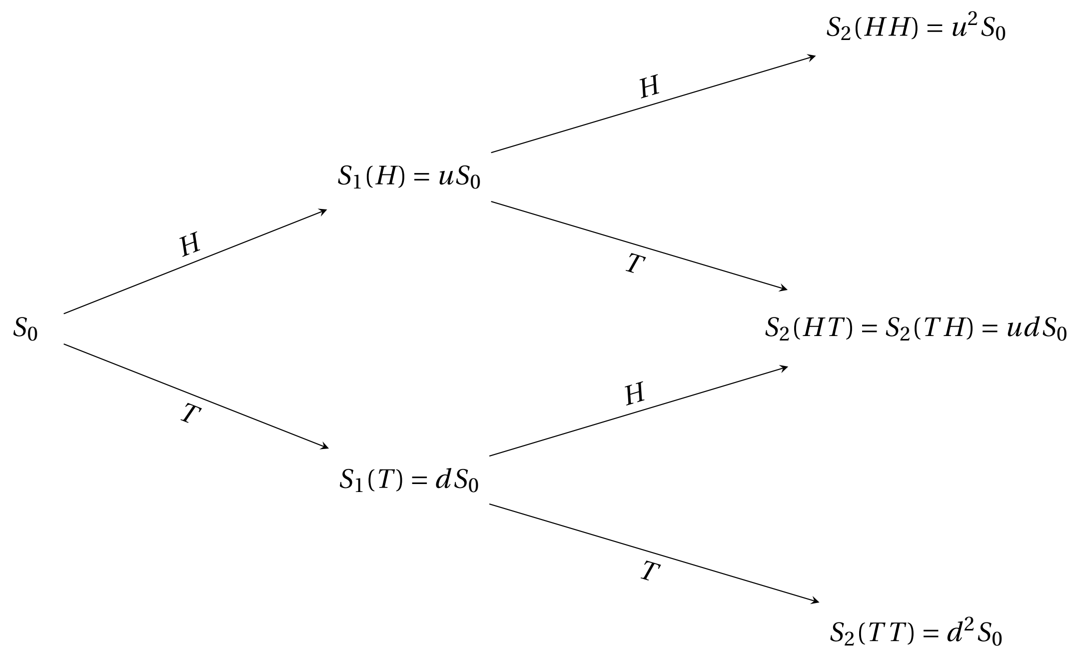

Figure 7.2: Price dynamics over two periods.

Using a series of N coin tosses, we can define an N-period model (see Fig. 7.2 for $N = 2$). The sample space of the N-period model is

$$\Omega = \{\omega = (\omega_1, \dots, \omega_N) : \ \omega_k \in \{T, H\}\}. \tag{7.3}$$

The link with random walks is obvious: we can equally well represent the sample space as

$$\Omega = \{\omega = (\omega_1, \dots, \omega_N) : \omega_k \in \{-1, +1\}\}, \tag{7.4}$$

with $X_k(\omega) = \omega_k$ being the outcome of the k-th coin flip. Then the price after N-periods is given by

$$S_N(\omega) = S_0 u^{|\{1\le k\le N:\ X_k(\omega)=+1\}|} d^{|\{1\le k\le N:\ X_k(\omega)=-1\}|} = S_0 (ud)^{\frac{N}{2}} \left(\frac{u}{d}\right)^{\frac{1}{2}\sum_{k=1}^N X_k(\omega)}. \tag{7.5}$$

In particular, if the up-factor u and the down-factor d satisfy

$$ud = 1, \tag{7.6}$$

then

$$S_N(\omega) = S_0 u^{\sum_{k=1}^N X_k(\omega)}. \tag{7.7}$$

Here, $\sum_{k=1}^N X_k(\omega)$, which is the position of the random walk at time N, models the stock price after N periods.

7.1.2 Money market, interest rate and portfolio

Apart from buying and selling stocks, an investor may lend or borrow money, and receive or pay rent. One euro invested in the money market at time 0 will yield $(1+r)$ euro at time 1. Similarly, one euro borrowed from the money market at time 0 will result in a debt of $1+r$ euro at time 1. Thus, we assume that the interest rate for lending and borrowing is the same. This assumption is definitely not realistic for university students, but it is nearly true for large financial institutions.

The interest rate r does not have to be non-negative. For example, in May 2013 the government of The Netherlands was able to borrow money with interest rate $r = -0.039$ percent, while in July 2014 the European Central Bank (ECB) set the interest rate for deposits at -0.10 percent.

Exercise 7.1. Think of at least two economical reasons for negative interest rates.

However, throughout the sequel we always assume that $1 + r > 0$.

Definition 7.1. A *portfolio* is a collection of assets $\mathbb{P} = (M, \Delta)$ consisting of M euros capital and Δ shares of stock at price S.

7.1.3 Arbitrage

Definition 7.2. Arbitrage is the possibility of a risk-free profit (see also Exercise 7.3).

More specifically, arbitrage is defined as a trading strategy that starts with no money, has zero probability of loss, and has a positive probability of making profit. In practice, arbitrage is the practice of taking advantage of a price difference between two or more markets by constructing a combination of deals that capitalise upon the imbalance, the profit being the difference between the market prices.

The theory of option pricing to be described below assumes *no arbitrage*, i.e., the financial market does not allow for a risk-free profit. This assumption is quite natural, for otherwise we could all become rich without running a risk. We show that this assumption places a restriction on our parameters, namely,

$$d < 1 + r < u. \tag{7.8}$$

Consider the Binomial Asset Pricing Model with parameters u, d and the money market with interest rate r. Suppose that

$$1 + r \le d < u. \tag{7.9}$$

In this case we can borrow money at interest rate r and invest in stock. The stock price increases at least as fast as the debt that is used to buy it and so profit is made.

Exercise 7.2. Demonstrate with the help "Short Selling" that there is also arbitrage when $d < u < 1 + r$.

Wikipedia:

> Short Selling (also known as shorting or going short) is the sale of a security that is not owned by the seller, or that the seller has borrowed. For the purposes of this chapter, the short seller receives a current price for the asset, but at the same time is obliged to deliver the asset to the buyer at the end of a period. If the price of the asset decreases, then the short seller profits: the cost of repurchase in the next period is less than the proceeds that were received upon the initial short sale. Conversely, if the price of the asset increases, then the short seller endures a loss because the proceeds are not sufficient to cover the price of the asset in the next period.

In the following we will assume that there is no arbitrage in the stock and money markets, i.e., (7.8) holds.

Exercise 7.3. Show that under (7.8) the financial market has no arbitrage. *Hint:* Suppose that $d < 1 + r < u$ and that arbitrage is possible. The latter means that at time zero we can form a portfolio $\mathbb{P} = (M_0, \Delta_0)$, consisting of M_0 euros and Δ_0 shares at price S_0, such that the value of the portfolio at time 0 is $V_0 = M_0 + \Delta_0 S_0 = 0$, while the value V_1 at time 1 satisfies $P(V_1 \geq 0) = 1$ and $P(V_1 > 0) > 0$ for any choice of the probabilities $p = P(\omega = H)$, $1 - p = P(\omega = T)$, $p \in (0, 1)$. Arrive at a contradiction by *identifying a probability distribution Q* on $\{H, T\}$, i.e., a $q \in (0, 1)$ with $q = Q(\omega = H)$, $1 - q = Q(\omega = T)$, such that under this distribution of upward and downward moves the expected value of V_1 is zero.

The probability distribution Q is called the *risk-neutral measure*. Note that Q is "equivalent to" the probability distribution P describing the true market moves, i.e., there exist two constants $\underline{c}, \overline{c} \in (0, \infty)$ such that

$$\underline{c}\, Q(\omega) \leq P(\omega) \leq \overline{c}\, Q(\omega) \qquad \forall\, \omega \in \Omega. \tag{7.10}$$

Exercise 7.3 is a particular case of the following more general statement, for which we refer to the literature.

Theorem 7.3 (First Fundamental Theorem of Asset Pricing)**.** *A discrete market on a discrete probability space (Ω, P) is arbitrage-free if and only if there exists at least one risk-neutral probability measure Q that is equivalent to the original probability measure P., i.e.,*

$$E_Q(V_1) = V_0, \tag{7.11}$$

where V_i is the value of the portfolio at time i.

7.2 Financial derivatives

7.2.1 Call and put options

Wikipedia:

> "A derivative is a financial contract which derives its value from the performance of another entity such as an asset, index, or interest rate, called the "underlying". Derivatives are one of the three main categories of financial instruments, the other two being equities (i.e., stocks) and debts (i.e., bonds and mortgages). Derivatives include a variety of financial contracts, including futures, forwards, swaps, options, and variations of these such as caps, floors, collars, and credit default swaps."

Definition 7.4. [Options]
(1) A *European call option* with strike price $K > 0$ and expiration time $t > 0$ gives the right (but not the obligation) to *buy* the stock at time t for K euros. The value at time t is $(S_t - K)^+ = \max\{S_t - K, 0\}$. The stock is the

underlying asset.
(2) A *European put option* with strike price $K > 0$ and expiration time $t > 0$ gives the right (but not the obligation) to *sell* the stock at time t for K euros. The value at time t is $(K - S_t)^+$.
(3) *American call options or put options* can be exercised at any time *before* the expiration time t.

In the remainder of this section we focus on European call options and ask what is a *fair price*: What price should be charged to the buyer of the option for being allowed to exercise the right?

7.2.2 Option pricing: one period

What is the price of a European call option at time 0 with expiration time 1?

Let us consider an example. Suppose that $S_0 = 4$, $u = \frac{1}{d} = 2$ and $r = \frac{1}{4}$. Hence

$$S_1(H) = uS_0 = 8, \qquad S_1(T) = dS_0 = 2. \tag{7.12}$$

Suppose that you want to determine the price of a European call option at strike price $K = 5$ and expiration time $t = 1$. This goes as follows. Note that the value of the option at time $t = 1$ is

$$(S_1(\omega) - K)^+ = \begin{cases} 3, & \omega = H, \\ 0, & \omega = T. \end{cases} \tag{7.13}$$

We are going to *replicate* the option by constructing a portfolio with the *same performance*. Suppose that your initial wealth is $X_0 = 1.2$ and you buy $\Delta_0 = \frac{1}{2}$ shares of stock at time 0, which costs

$$\Delta_0 S_0 = \frac{1}{2} \times 4 = 2. \tag{7.14}$$

In order to do so you have to borrow at the money market 0.8 euro, namely,

$$X_0 - \Delta_0 S_0 = 1.2 - 2 = -0.8. \tag{7.15}$$

At time $t = 1$, your cash position is

$$(X_0 - \Delta_0 S_0)(1 + r) = -1, \tag{7.16}$$

i.e., you have to pay back 1 euro. Your stock $\Delta_0 S_1$ is worth 4 euros if $\omega = H$ and 1 euro if $\omega = T$. Thus, the value of the portfolio at time $t = 1$ is

$$V_1(\omega) = (X_0 - \Delta_0 S_0)(1 + r) + \Delta_0 S_1(\omega) = \begin{cases} (-1) + 4 = 3, & \omega = H, \\ (-1) + 1 = 0, & \omega = T. \end{cases} \tag{7.17}$$

Note that the value of the portfolio at time $t = 1$ in (7.17) is precisely the value of the option at time $t = 1$ in (7.13), i.e.,

$$V_1(\omega) = (S_1(\omega) - K)^+. \tag{7.18}$$

The initial wealth $X_0 = 1.2$, needed to set up the replicating portfolio above, is actually the *no-arbitrage price of the option of time* 0. Why is that?

- If we can sell the option for more, e.g. 1.21 euro, then we can replicate the option for 1.20 and put the excess 0.01 into the money market with a guaranteed profit. At time $t = 1$ we can pay off the option, and we are guaranteed a profit of 0.0125 euro.

- Conversely, if we can buy the option for less than 1.20, e.g. 1.19 euro, then we can proceed as follows: we short sell $\Delta_0 = \frac{1}{2}$ shares of stock (giving us 2 euros), buy 1 option for 1.19, and put the difference $2-1.19 = 0.81$ into the money market. At time 1 we have to return the share, i.e., pay either 4 ($\omega = H$) or 1 ($\omega = T$). But the option will gives us either 3 euros ($\omega = H$) or 0 euros ($\omega = T$), while the money market gives us $0.81 \times (1+r) = 1.0125$ euro. Hence our position is covered (i.e., we can fulfil our obligations), and we are guaranteed a profit of 0.0125 euro.

More generally, given that the value of an option at time $t = 1$ is $C_1(\omega)$, we have to find X_0 and Δ_0 such that

$$\begin{aligned} V_1(\omega) &= (X_0 - \Delta_0 S_0)(1+r) + \Delta_0 S_1(\omega) \\ &= X_0(1+r) + (\Delta_0 S_1(\omega) - \Delta_0 S_0(1+r)) =^{\text{def}} C_1(\omega). \end{aligned} \tag{7.19}$$

Picking $\omega = H$ or $\omega = T$, we get

$$\begin{aligned} X_0 + \Delta_0\Big(\frac{1}{1+r} S_1(H) - S_0\Big) &= \frac{1}{1+r} C_1(H), \\ X_0 + \Delta_0\Big(\frac{1}{1+r} S_1(T) - S_0\Big) &= \frac{1}{1+r} C_1(T). \end{aligned} \tag{7.20}$$

Exercise 7.4. Show that the solution of (7.20) is given by

$$\Delta_0 = \frac{C_1(H) - C_1(T)}{S_1(H) - S_1(T)}, \quad X_0 = \frac{1}{1+r}[qC_1(H) + (1-q)C_1(T)], \tag{7.21}$$

where $q = \frac{1+r-d}{u-d} \in (0,1)$ because $d < 1+r < u$.

We conclude the following.

- The arbitrage-free price of the European call option expiring after period 1 is given by

$$C_0 = \frac{1}{1+r}[qC_1(H) + (1-q)C_1(T)], \quad q = \frac{1+r-d}{u-d}. \tag{7.22}$$

- Replication of the portfolio amounts to: sell one call option at price C_0, buy

$$\Delta_0 = \frac{C_1(H) - C_1(T)}{S_1(H) - S_1(T)} \tag{7.23}$$

  shares at price S_0.

7.2.3 Option pricing: multiple periods

We would like to repeat the above reasoning for N periods rather than one period. The answer, which we state without proof, is given in the following theorem.

Theorem 7.5 (Replication in the multi-period binomial model)**.** *Consider an N-period Binomial Asset Pricing Model with no arbitrage. Let*

$$q = \frac{1+r-d}{u-d}, \quad 1-q = \frac{u-1-r}{u-d}. \tag{7.24}$$

Let $C_N = C_N(\omega_1, \dots, \omega_N)$ be a random variable that describes a derivative security paying off at time N. For example, in the case of the European call option,

$$C_N = (S_N(\omega_1, \dots, \omega_N) - K)^+. \tag{7.25}$$

(I) Define, recursively backwards in time, a sequence of random variables $C_{N-1},\dots,C_0$ *by putting*

$$C_n(\omega_1,\dots,\omega_n) = \frac{1}{1+r}\Big(qC_{n+1}(\omega_1,\dots,\omega_n,H) + (1-q)C_{n+1}(\omega_1,\dots,\omega_n,T)\Big), \qquad n = N-1,\dots,0, \tag{7.26}$$

and note that C_n *depends on* $(\omega_1,\dots,\omega_n)$.
(II) Define

$$\Delta_n(\omega_1,\dots,\omega_n) = \frac{C_{n+1}(\omega_1,\dots,\omega_n,H) - C_{n+1}(\omega_1,\dots,\omega_n,T)}{S_{n+1}(\omega_1,\dots,\omega_n,H) - S_{n+1}(\omega_1,\dots,\omega_n,T)}. \tag{7.27}$$

(III) Set $X_0 = C_0$ *and define, recursively forwards in time, the portfolio values* $X_1,\dots,X_N$ *by putting*

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n), \tag{7.28}$$

i.e., at period n*, the portfolio consists of* $X_n - \Delta_n S_n$ *euros and of* Δ_n *shares at price* S_n *euros each.*

Then

$$X_n(\omega_1,\dots,\omega_n) = C_n(\omega_1,\dots,\omega_n) \qquad \forall\ (\omega_1,\dots,\omega_n) \in \{H,T\}^n,\ n = 1,\dots,N. \tag{7.29}$$

In particular, $X_N(\omega_1,\dots,\omega_N)$ *– the value of the portfolio after* N *periods – is equal to the value of the option* $C_N(\omega_1,\dots,\omega_N)$.

Theorem 7.5 not only gives the price, it also shows how to *replicate* the option with the help of the constructed dynamic portfolio. Therefore, similarly as in the case of the pricing formula for a single period, we conclude from Theorem 7.5 that $X_0 = C_0$ is the *fair price* for the European call option at expiration time $t = N$: any deviation in option price from C_0 will create an opportunity for arbitrage.

Exercise 7.5. Give a proof of Theorem 7.5.

7.3 Black-Scholes theory

7.3.1 Discrete Black-Scholes formula

To find C_0, we need to work out the backward recursion in Theorem 7.5. To that end we analyse the call option prices $C_n(\omega_1,\dots,\omega_n)$ given by (7.26): We have

$$\begin{aligned} C_0 &= \frac{1}{1+r}\Big(qC_1(H) + (1-q)C_1(T)\Big),\\ C_1(H) &= \frac{1}{1+r}\Big(qC_2(HH) + (1-q)C_2(HT)\Big),\\ C_1(T) &= \frac{1}{1+r}\Big(qC_2(TH) + (1-q)C_2(TT)\Big). \end{aligned} \tag{7.30}$$

Since $C_2 = (S_2 - K)^+$ and $S_2(HT) = S_2(TH) = udS_0$, we find that $C_2(HT) = C_2(TH)$, and hence

$$\begin{aligned} C_0 &= \frac{1}{(1+r)^2}\Big(q^2C_2(HH) + 2q(1-q)C_2(HT) + (1-q)^2C_2(TT)\Big)\\ &= \frac{1}{(1+r)^2}\Big(q^2(u^2S_0 - K)^+ + 2q(1-q)(udS_0 - K)^+ + (1-q)^2(d^2S_0 - K)^+\Big). \end{aligned} \tag{7.31}$$

More generally,

$$C_0 = \frac{1}{(1+r)^N}\sum_{k=0}^{N}\binom{N}{k}q^k(1-q)^{N-k}\Big(u^k d^{N-k}S_0 - K\Big)^+. \tag{7.32}$$

Exercise 7.6. Prove (7.32) with the help of induction, i.e., prove that for a system of N recurrence relations given by (7.26) (and (7.25)) C_0 is given by (7.32).

The formula in (7.32) is complete, but we can give a more convenient expression that allows for a proper interpretation as well. Let a be the smallest non-negative integer such that

$$u^a d^{N-a} S_0 > K. \tag{7.33}$$

Then

$$(u^k d^{N-k} S_0 - K)^+ = \begin{cases} 0, & k < a, \\ u^k d^{N-k} S_0 - K, & k \geq a. \end{cases} \tag{7.34}$$

Therefore (7.32) becomes

$$C_0 = \frac{1}{(1+r)^N} \sum_{k=a}^{N} \binom{N}{k} q^k (1-q)^{N-k} \Big(u^k d^{N-k} S_0 - K \Big). \tag{7.35}$$

Splitting this expression into two sums, we get

$$\begin{aligned} C_0 &= S_0 \sum_{k=a}^{N} \binom{N}{k} q^k (1-q)^{N-k} \Big[\frac{u^k d^{N-k}}{(1+r)^N} \Big] - \frac{K}{(1+r)^N} \sum_{k=a}^{N} \binom{N}{k} q^k (1-q)^{N-k} \\ &= S_0 \sum_{k=a}^{N} \binom{N}{k} \Big(\frac{qu}{1+r} \Big)^k \Big(\frac{(1-q)d}{1+r} \Big)^{N-k} - \frac{K}{(1+r)^N} \sum_{k=a}^{N} \binom{N}{k} q^k (1-q)^{N-k}. \end{aligned} \tag{7.36}$$

Abbreviate

$$\tilde{q} = q \frac{u}{1+r} \in (0,1). \tag{7.37}$$

Note that

$$\begin{aligned} 1 - \tilde{q} = 1 - \frac{qu}{1+r} &= \frac{1 + r - u\frac{1+r-d}{u-d}}{1+r} = \frac{(1+r)(u-d) - u(1+r-d)}{(u-d)(1+r)} \\ &= \frac{ud - (1+r)d}{(u-d)(1+r)} = \frac{u-1-r}{u-d} \frac{d}{1+r} = (1-q)\frac{d}{1+r}. \end{aligned} \tag{7.38}$$

Therefore (7.36) becomes

$$C_0 = S_0 \sum_{k=a}^{N} \binom{N}{k} \tilde{q}^k (1-\tilde{q})^{N-k} - \frac{K}{(1+r)^N} \sum_{k=a}^{N} \binom{N}{k} q^k (1-q)^{N-k}. \tag{7.39}$$

Equivalently,

$$C_0 = S_0 \mathbb{P}[Y \geq a] - \frac{K}{(1+r)^N} \mathbb{P}[Z \geq a], \tag{7.40}$$

where Y is $\text{Bin}(N, \tilde{q})$ and Z is $\text{Bin}(N, q)$.

The above computation in summary gives:

Theorem 7.6 (Binomial Option Pricing Formula)**.** *The no-arbitrage price C_0 equals*

$$C_0 = S_0 \Phi(a; N, \tilde{q}) - \frac{K}{(1+r)^N} \Phi(a; N, q),$$

where

$$\tilde{q} = q \frac{u}{1+r}, \quad q = \frac{1+r-d}{u-d},$$

$$\begin{aligned} a &= \min\{k \in \mathbb{N}_0 : \; u^k d^{N-k} S_0 \geq K\} \\ &= \min\Big\{ k \in \mathbb{N}_0 : \; k \geq \frac{\log K - \log(S d^N)}{\log u - \log d} \Big\}, \end{aligned}$$

and $\Phi(a; N, p) = \sum_{k \geq a} \binom{N}{k} p^k (1-p)^{N-k}$ *is the complementary binomial distribution function. If* $a > N$*, then* $C = 0$.

7.3.2 ⋆lack-Scholes Option Pricing Formula

In this section we move from discrete to continuous time, replacing Simple Random Walk by Brownian Motion as a model for stock prices. The rationale behind this is that the financial market moves very fast: every second thousands of transactions take places worldwide. It is therefore justified to think of time as running forward in very short units.

Figure 7.3: Fischer Black, Myron Scholes, Robert Merton.

Wikipedia:

> “The Black-Scholes or Black-Scholes-Merton model is a mathematical model of a financial market containing certain derivative investment instruments. From the model, one can deduce the Black-Scholes formula, which gives a theoretical estimate of the price of European-style options. The formula led to a boom in options trading and legitimised scientifically the activities of the Chicago Board Options Exchange and other options markets around the world. It is widely used, although often with adjustments and corrections, by options market participants. Many empirical tests have shown that the Black-Scholes price is "fairly close" to the observed prices ...
>
> The Black-Scholes model was first published by Fischer Black and Myron Scholes in their 1973 paper, "The Pricing of Options and Corporate Liabilities", published in the Journal of Political Economy. They derived a partial differential equation, now called the Black-Scholes equation, which estimates the price of the option over time. The key idea behind the model is to hedge the option by buying and selling the underlying asset in just the right way and, as a consequence, to eliminate risk. This type of hedging is called delta hedging and is the basis of more complicated hedging strategies such as those engaged in by investment banks and hedge funds.”

Fischer Black, Myron Scholes and Robert Merton (see Fig. 7.3) won the Nobel prize in economics for their invention.

We make the following assumptions on the assets:

(1) (“Riskless Rate”) The rate of return on the riskless asset is constant and is called the risk-free interest rate.

(2) ("Brownian Motion") The instantaneous log returns of the stock price form an infinitesimal random walk with drift, more precisely, a *geometric Brownian motion* (also known as exponential Brownian motion). We will assume that its drift μ and volatility σ are constant:

$$\log \frac{S_{t+\Delta t}}{S_t} = \mu \Delta t + \sigma \sqrt{\Delta t}\, \epsilon_t, \qquad 0 < \Delta t \ll 1. \tag{7.41}$$

Here, ϵ_t is the random component of the price evolution. To formalise the above model, we consider the corresponding Stochastic Differential Equation (SDE):

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{7.42}$$

Even though SDE's are not treated in this course, we mention that (7.42) has an explicit solution

$$S_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right], \tag{7.43}$$

where $(W_t)_{t\geq 0}$ is the standard Brownian motion. The process $(S_t)_{t\geq 0}$ given by (7.43) is called geometric Brownian motion.

(3) The stock does not pay a dividend.

Exercise 7.7. Consider (7.43). Show with the help of explicit computations that, for all $t \geq 0$,

$$\begin{aligned} \mathbb{E}(S_t) &= S_0 e^{\mu t}, \\ \mathrm{Var}(S_t) &= S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right). \end{aligned} \tag{7.44}$$

Hint: First prove that if Z has distribution $\mathcal{N}(0, t)$, then

$$E(e^{aZ}) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} e^{ax}\, e^{-x^2/2t}\, dx = e^{a^2 t/2}, \qquad a \in \mathbb{R},\ t > 0. \tag{7.45}$$

We make the following assumptions on the market:

(1) There is no arbitrage (i.e., there is no way to make a riskless profit).

(2) It is possible to borrow and lend any amount of cash, even fractional, at the riskless rate.

(3) It is possible to buy and sell any amount, even fractional, of the stock (which includes short selling).

(4) The above transactions do not incur any fees or costs (i.e., frictionless market).

Theorem 7.7. *The Black-Scholes formula for the arbitrage-free price C of a call option with strike price K and expiration time t is given by*

$$C = S\Phi(d_1) - Ke^{-rt}\Phi(d_2), \tag{7.46}$$

where

$$d_1 = \frac{\log(S/K) + \left(r + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t} = \frac{\log(S/K) + \left(r - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}, \tag{7.47}$$

and

$$\begin{aligned} S &= \text{ current stock price,} \\ K &= \text{ strike price,} \\ r &= \text{ (continuously compounded) risk-free interest rate,} \\ \sigma &= \text{ volatility of stock price} \\ &\quad \text{(standard deviation of short-term returns),} \\ t &= \text{ time remaining until expiration} \\ &\quad \text{(expressed as a percent of a year).} \end{aligned} \tag{7.48}$$

and $\Phi(d) = \int_d^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ *is the standard normal cumulative distribution function.*

Note that the formula in (7.46)–(7.47) does not depend on the value of the drift μ.

7.3.3 ⋆ Derivation

In this section we give the proof of Theorem 7.7. This is done by taking the formula in Theorem 7.5 and passing to the limit $N \to \infty$ after speeding up time by N. The proof below is straightforward but technical.

Suppose that the expiration time is t. We divide the time interval $[0,t]$ into $n \gg 1$ intervals of equal length $\Delta t = t/n$. Consider the Binomial Asset Pricing Model with time horizon n and parameters

$$u = u_n, \quad d = d_n, \quad \hat{r} = \hat{r}_n, \tag{7.49}$$

which we choose later.

(A) Matching interest rates:

In the continuous Black-Scholes model, 1 euro invested in a risk-free asset (bond) will result in e^{rt} euros. In the discrete binomial asset pricing model with one-period interest $\hat{r}$, it will result in $(1+\hat{r})^n$ euros. Hence

$$(1+\hat{r})^n = e^{rt} \quad \Rightarrow \quad 1+\hat{r} = e^{r\frac{t}{n}} = e^{r\Delta t} \quad \Rightarrow \quad \hat{r} = e^{r\Delta t} - 1 = r\Delta t + O((\Delta t)^2). \tag{7.50}$$

(B) Matching up and down factors:

In order to be consistent with (7.43), abbreviate $\tilde{\mu} = \mu - \frac{\sigma^2}{2}$ and put

$$\begin{aligned} d = d_n &= \exp\Big(\tilde{\mu}\Delta t - \sigma\sqrt{\Delta t}\Big) = \exp\Big(\tilde{\mu}\frac{t}{n} - \sigma\sqrt{\frac{t}{n}}\Big), \\ u = u_n &= \exp\Big(\tilde{\mu}\Delta t + \sigma\sqrt{\Delta t}\Big) = \exp\Big(\tilde{\mu}\frac{t}{n} + \sigma\sqrt{\frac{t}{n}}\Big). \end{aligned} \tag{7.51}$$

As we will see later, the value of $\tilde{\mu}$ does not affect the price of the option. Then

$$S_n = S_0 \exp\left(\tilde{\mu}t + \sigma\left[\sqrt{t}\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k\right]\right), \tag{7.52}$$

where $X_k = 1$ if $\omega_k = H$ (up movement) and $X_k = -1$ if $\omega_k = T$ (down movement). In this way, S_n matches the limiting price that follows the geometric Brownian motion

$$S_t = S_0 \exp(\tilde{\mu}t + \sigma W_t) \tag{7.53}$$

when $p = \frac{1}{2}$: $\mathbb{P}[\omega_k = H] = \mathbb{P}[\omega_k = T] = \frac{1}{2} = \mathbb{P}[X_k = 1] = \mathbb{P}[X_k = -1]$. The corresponding parameters q and $\tilde{q}$ are given by

$$q = \frac{1+\hat{r}-d}{u-d} = \frac{e^{r\Delta t} - e^{\tilde{\mu}\Delta t - \sigma\sqrt{\Delta t}}}{e^{\tilde{\mu}\Delta t + \sigma\sqrt{\Delta t}} - e^{\tilde{\mu}\Delta t - \sigma\sqrt{\Delta t}}}, \quad \tilde{q} = q\frac{u}{1+\hat{r}} = \frac{1 - e^{(\tilde{\mu}-r)\Delta t - \sigma\sqrt{\Delta t}}}{1 - e^{-2\sigma\sqrt{\Delta t}}}. \tag{7.54}$$

Exercise 7.8. ⋆ For a function $h\colon \mathbb{R} \to \mathbb{R}$ with $h(0) = 0$, the notation $h(x) = O(x^k)$ as $x \to 0$ for some $k \in \mathbb{N}_0$ means that there exist $\delta > 0$ and $0 < M < \infty$ such that

$$|h(x)| \le M|x|^k \qquad \forall\, x \in (-\delta, \delta). \tag{7.55}$$

This statement is equivalent to $\limsup_{x\to 0, x\neq 0} |h(x)/x^k| < \infty$. Similarly, $h(x) = i(x) + O(x^k)$ means that $h(x) - i(x) = O(x^k)$. It is easy to see that if $h(x) = O(x^k)$ and $i(x) = O(x^l)$ for some $k, l \in \mathbb{N}_0$ with $k \le l$, then

(i) $i(x) = O(x^k)$.

(ii) $ah(x) + bi(x) = O(x^k)$ for all $a, b \in \mathbb{R}$.

(iii) $h(x)i(x) = O(x^{k+l})$.

Show the following:

(a) If $h(x) = O(x^k)$ for some $k \in \mathbb{N}_0$, then $1/(1+h(x)) = O(1)$.

(b) If $h(x) = O(x^2)$ and $i(x) = O(x^2)$, then $(ax+b+h(x))/(c+i(x)) = (ax+b)/c + O(x^2)$ for all $a, b, c \in \mathbb{R}$ with $c \neq 0$.

(c) The exponential function satisfies $e^x = \sum_{n=0}^{N} \frac{x^n}{n!} + O(x^{N+1})$ for all $N \in \mathbb{N}_0$.

(d) For $\Delta t \downarrow 0$,

$$q = \frac{1}{2} + \frac{1}{2}\frac{r - \tilde{\mu} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t} + O(\Delta t), \tag{7.56}$$

$$\tilde{q} = \frac{1}{2} + \frac{1}{2}\frac{r - \tilde{\mu} + \frac{1}{2}\sigma^2}{\sigma}\sqrt{\Delta t} + O(\Delta t). \tag{7.57}$$

Hint: Use (7.54) for (7.56), and $\tilde{q} = (\frac{u}{1+\hat{r}})q$ for (7.57).

It remains to determine the minimal $a \in \mathbb{Z}_+$ such that

$$u^a d^{n-a} S_0 \geq K, \tag{7.58}$$

or, equivalently,

$$\exp\left(\tilde{\mu}t + 2a\sigma\sqrt{\frac{t}{n}} - n\sigma\sqrt{\frac{t}{n}}\right) S_0 \geq K. \tag{7.59}$$

Thus, a is the minimal integer such that

$$a \geq \frac{1}{2\sigma\sqrt{\frac{t}{n}}}\left(\log(K/S_0) + \sigma\sqrt{nt} - \tilde{\mu}t\right) = \frac{\log(K/S_0)}{2\sigma\sqrt{t}}\sqrt{n} + \frac{1}{2}n - \frac{\tilde{\mu}\sqrt{t}}{2\sigma}\sqrt{n}. \tag{7.60}$$

Abbreviate the right-hand side by a_n. Suppose that V is $\mathrm{Bin}(n, p)$. Then

$$\mathbb{P}[V_n \geq a_n] = \mathbb{P}\left[\frac{V_n - np}{\sqrt{n}} \geq \frac{a_n - np}{\sqrt{n}}\right]. \tag{7.61}$$

By the CLT we have

$$\frac{V_n - np}{\sqrt{n}} \Longrightarrow Z, \quad Z \sim N\big(0, p(1-p)\big), \qquad n \to \infty. \tag{7.62}$$

Hence if $\lim_{n\to\infty}(a_n - np)/\sqrt{n} = z$, then

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}[V_n \geq a_n] &= \lim_{n\to\infty}\mathbb{P}\left[\frac{V_n - np}{\sqrt{n}} \geq \frac{a_n - np}{\sqrt{n}}\right] \\ &= \mathbb{P}[Z \geq z] = \frac{1}{\sqrt{2\pi p(1-p)}}\int_z^{+\infty}\exp\left(-\frac{x^2}{2p(1-p)}\right)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\frac{z}{\sqrt{p(1-p)}}}^{+\infty}\exp\left(-\frac{x^2}{2}\right)dx = 1 - \Phi\Big(\frac{z}{\sqrt{p(1-p)}}\Big),\end{aligned} \tag{7.63}$$

where Φ is the cumulative distribution function of the standard normal random variable $\mathcal{N}(0,1)$.

With a little bit of extra work the above can be generalised to the case where p depends on n as well (like our probabilities q, $\tilde{q}$), as stated in the following lemma the proof of which is left to the reader.

Lemma 7.8. *Suppose that V_n is Bin(n, p_n) with* $\lim_{n\to\infty} p_n = p \in (0,1)$. *Suppose further that a_n is such that*

$$\lim_{n\to\infty} \frac{a_n - np_n}{\sqrt{n}} = z \in \mathbb{R}. \tag{7.64}$$

Then

$$\lim_{n\to\infty} \mathbb{P}[V_n \geq a_n] = 1 - \Phi\Big(\frac{z}{\sqrt{p(1-p)}}\Big). \tag{7.65}$$

By Exercise 7.8, we have

$$\begin{aligned} q = q_n &= \frac{1}{2} + \frac{1}{2}\frac{r - \tilde{\mu} - \frac{1}{2}\sigma^2}{\sigma}\sqrt{\frac{t}{n}} + O\Big(\frac{t}{n}\Big), \\ \tilde{q} = \tilde{q}_n &= \frac{1}{2} + \frac{1}{2}\frac{r - \tilde{\mu} + \frac{1}{2}\sigma^2}{\sigma}\sqrt{\frac{t}{n}} + O\Big(\frac{t}{n}\Big), \end{aligned} \tag{7.66}$$

and hence $\lim_{n\to\infty} q_n = \frac{1}{2}$.

Finally, let us compute the limits (recall (7.47) for the definition of d_1, d_2)

$$\begin{aligned} \frac{a_n - nq_n}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\left[\frac{\log(K/S_0)}{2\sigma\sqrt{t}}\sqrt{n} + \frac{1}{2}n - \frac{\tilde{\mu}\sqrt{t}}{2\sigma}\sqrt{n} - \frac{n}{2} - \frac{r - \tilde{\mu} - \frac{1}{2}\sigma^2}{2\sigma}\sqrt{nt} + O(t)\right] \\ &= \frac{\log(K/S_0)}{2\sigma\sqrt{t}} - \frac{r - \frac{1}{2}\sigma^2}{2\sigma}\sqrt{t} + \frac{1}{\sqrt{n}}O(t) \to -\frac{1}{2}d_2, \qquad n \to \infty. \end{aligned} \tag{7.67}$$

and

$$\begin{aligned} \frac{a_n - n\tilde{q}_n}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\left[\frac{\log(K/S_0)}{2\sigma\sqrt{t}}\sqrt{n} + \frac{1}{2}n - \frac{\tilde{\mu}\sqrt{t}}{2\sigma}\sqrt{n} - \frac{n}{2} - \frac{r - \tilde{\mu} + \frac{1}{2}\sigma^2}{2\sigma}\sqrt{nt} + O(t)\right] \\ &= \frac{\log(K/S_0)}{2\sigma\sqrt{t}} - \frac{r + \frac{1}{2}\sigma^2}{2\sigma}\sqrt{t} + \frac{1}{\sqrt{n}}O(t) \to -\frac{1}{2}d_1, \qquad n \to \infty. \end{aligned} \tag{7.68}$$

The no-arbitrage price for the n-period European call option equals

$$C^{(n)} = S_0\mathbb{P}[Y \geq a_n] - \frac{K}{(1+\hat{r})^n}\mathbb{P}[Z \geq a_n] \tag{7.69}$$

converges as $n \to \infty$ to

$$\begin{aligned} C &= S_0\left(1 - \Phi\left(\frac{-\frac{1}{2}d_2}{\sqrt{\frac{1}{4}}}\right)\right) - \frac{K}{e^{rt}}\left(1 - \Phi\left(\frac{-\frac{1}{2}d_1}{\sqrt{\frac{1}{4}}}\right)\right) \\ &= S_0\Big(1 - \Phi\big(-d_2\big)\Big) - Ke^{-rt}\Big(1 - \Phi\big(-d_1\big)\Big) \\ &= S_0\Phi(d_2) - Ke^{-rt}\Phi(d_1). \end{aligned} \tag{7.70}$$

This completes the proof of Theorem 7.7.