PCMI
Undergraduate Summer School

Lecture notes

(Preliminary Version)

Marek Biskup

Park City, Utah; July 20, 2007

Contents

Chapter 1

Prerequisites

1.1 Basic probability concepts

The purpose of this introductory section is to introduce the basic concepts that will be needed throughout the course. We will for the most part stay away from stating theorems, lemmas, etc.; these will be the subject of the next section.

Probability originated in games of chance. A prime example associated with probability is the experiment of *tossing a coin*. This is a procedure intended to produce a random choice out of two answers: Heads or Tails. Generally, the outcomes of such a random experiment are collected in a set Ω called the *sample space*. For the coin toss, we simply have $\Omega = \{H, T\}$.

Before probability theory was axiomatized and embedded into the rest of mathematics, the word probability referred to the *frequencies* with which the various outcomes — i.e., values from Ω — were seen in repeated random experiments. For instance, it is a known fact that tossing a coin many times generally results in about the same fraction of heads and tails. If ω denotes the result of the coin toss, we extrapolate from this

$$\mathbb{P}(\omega = \text{ heads}) = \mathbb{P}(\omega = \text{ tails}) = \frac{1}{2}. \tag{1.1}$$

We read: the probability to get heads is $1/2$, and similarly for the tails. For more general discrete sets Ω — finite or countably infinite — we similarly need to extract the frequences with which each element z of Ω occurs. The value thus obtained is then proclaimed to be the probability to get z, namely, $\mathbb{P}(\omega = z)$.

Once we know the frequencies for all elements of Ω, we can start asking more general questions, e.g., for a set $A \in \Omega$, how likely it is that the outcome will fall into A. An example of this is *rolling a die*. Here $\Omega = \{1, \dots, 6\}$ with $\mathbb{P}(\omega = j) = 1/6$ for each $j = 1, \dots, 6$. However, if we take $A = \{5, 6\}$ we may ask what is the probability that $\omega \in A$? It turns out

$$\mathbb{P}(\omega \in A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \tag{1.2}$$

where we used our common sense to conclude that the frequency with which we can see 5 or 6 is simply the sum of the frequencies to see 5 and to see 6.

The previous observation quickly results in the conclusion that it is actually better to define probability as a *function of subsets of* Ω which we call *events*. We will thus talk of *probability of a set* A, writing $\mathbb{P}(A)$, or *probability that* A *occurs*. From its nature $\mathbb{P}(A)$ has to be a number between 0 and 1 and such that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$. We will also require that if A and B are *disjoint* events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{1.3}$$

which reflect the fact that the relative frequency of results from $A \cup B$ is the sum of relative frequencies of a result in A and the relative frequency of a result in B. Since *something* has to come up as a result of the experiment, we have

$$\sum_{\omega \in \Omega} \mathbb{P}\big(\{\omega\}\big) = 1 \tag{1.4}$$

whenever Ω is a countable set to make the sum meaningful.

Unfortunately, there are random experiments that take uncountably many values — even though that is an abstraction in the sense that no physical devise measures with absolute precision and so there is always a rounding error. For this reason, we set things up even more axiomatically:

Definition 1.1 [Probability and probability space] *Let Ω be a set and let $\mathscr{F}$ be a collection of subsets of Ω that*

(1) contains $\emptyset$ and Ω, i.e., $\emptyset, \Omega \in \mathscr{F}$.

(2) is closed under complements, i.e., $A \in \mathscr{F}$ implies $A^{\mathrm{c}} \in \mathscr{F}$.

(3) is closed under countable unions, i.e., if $A_1, A_2, \cdots \in \mathscr{F}$ then $\bigcup_{n\geq 1} A_n \in \mathscr{F}$.

A function $\mathbb{P}\colon \mathscr{F} \to [0,1]$ is then called probability *if it has the properties of*

(A1) countable additivity: If $A_1, A_2, \cdots \in \mathscr{F}$ are disjoint then

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \mathbb{P}(A_n). \tag{1.5}$$

(A2) normalization: $\mathbb{P}(\Omega) = 1$.

The triplet of objects $(\Omega, \mathscr{F}, \mathbb{P})$ is referred to as the probability space.

The conditions (1-3) are very natural as they basically ensure that the set $\mathscr{F}$ is closed under all basic set-theoretical operations.

Exercise 1.2 Show that $\mathscr{F}$ is closed also under countable intersections,

$$A_1, A_2, \cdots \in \mathscr{F} \quad \Rightarrow \quad \bigcap_{n\geq 1} A_n \in \mathscr{F} \tag{1.6}$$

and set theoretical differences

$$A, B \in \mathscr{F} \quad \Rightarrow \quad B \setminus A, A \setminus B \in \mathscr{F}. \tag{1.7}$$

Also the axioms (A1-A2) are fairly natural. Indeed, (A1) is a generalization of the statement that the probability of a set of individual outcomes is simply the sum of individual probabilities, while (A2) ensures that "something must occur," i.e., the frequencies of all outcomes add up to one.

Before we pass to some examples, let us record some basic facts about probability spaces. The proofs of these are omitted for brevity. Nevertheless, the reader may find them to be very interesting exercises.

Lemma 1.3 *Suppose $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space. Then $\mathbb{P}$ has the properties of*

(1) Monotonicity: If $A \subset B$ are events then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(2) Countable subadditivity: If $A_1, A_2, \ldots$ are events, then

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) \leq \sum_{n\geq 1} \mathbb{P}(A_n). \tag{1.8}$$

(3) Inclusion-exclusion: If $A_1, \ldots, A_n$ are events, then

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{\ell=1}^{n} (-1)^{\ell-1} \sum_{1\leq i_1<\cdots<i_\ell\leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_\ell}) \tag{1.9}$$

Lemma 1.4 [Continuity] *The probability $\mathbb{P}$ is continuous with respect to monotone sequences of events. Explicitly,*

(1) if $A_1 \supset A_2 \supset \ldots$, then

$$\mathbb{P}\Big(\bigcap_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{1.10}$$

(2) if $A_1 \subset A_2 \subset \ldots$, then

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{1.11}$$

We proceed by listing a few examples of the above setting:

Example 1.5 *Tossing a fair coin n-times*: The result can be viewed as a random *sequence* of elements from $\{0,1\}$. In particular, the sample space is $\Omega_n = \{0,1\}^n$. Now if $\omega = (\omega_i)_{i=1}^n \in \Omega_n$ is one such sequence, the probability that it occurs is

$$\mathbb{P}\big(\{\omega\}\big) = \frac{1}{2^n}. \tag{1.12}$$

To see that this is reasonable we appeal to our intuition that — for the coin to be fair — each sequence should have the same probability. Since there are 2^n distinct sequences, each should have probability 2^{-n}.

Example 1.6 *Rolling a die*: A die has six sides and if it is fair—i.e., no "sticky" sides—all have equal chance to come up as a result of rolling it. The sample space is $\Omega = \{1, 2, \ldots, 6\}$ and the probability is given by $\mathbb{P}(\{\omega\}) = 1/6$ for all $\omega \in \Omega$.

Example 1.7 *Rolling a die n-times*: We proceed as for coin tosses. The sample space is $\Omega_n = \{1, 2, \dots, 6\}^n$ and each sequence $\omega \in \Omega_n$ of n numbers from $\{1, 2, \dots, 6\}$ has probability

$$\mathbb{P}\big(\{\omega\}\big) = \frac{1}{|\Omega_n|} = \frac{1}{6^n}. \tag{1.13}$$

Example 1.8 *Recording only the 6's*: We add a twist to the previous example. Again we will roll a die 6 times, but now we only want to record whether we got a "6" or not. The result will be a sequence of 1's and 0's and so our sample space is $\Omega'_n = \{0, 1\}^n$, that is, the same as for the coin tosses! However, the probability is very different. Indeed, if $\omega \in \Omega'_n$ is one such sequence, its probability will be

$$\mathbb{P}\big(\{\omega\}\big) = \prod_{k=1}^{n} \Big(\frac{1}{6}\Big)^{\omega_k} \Big(\frac{5}{6}\Big)^{1-\omega_k}, \tag{1.14}$$

to be compared with $1/2^n$ the corresponding probability for coin tosses.

To justify the previous formula, we go back to the full representation of the experiment on the space $\Omega_n = \{1, 2, \dots, 6\}^n$ and, for each $k = 1, \dots, n$, introduce a map $X_k \colon \Omega_n \to \{0, 1\}$ that assigns $\omega \in \Omega_n$ value one if $\omega_k = 6$ and zero otherwise. Now let $\omega \in \Omega'_n$ be a sequence of zeros and ones. Then

$$\mathbb{P}\big(X_k(\omega) = \omega_k,\, k = 1, \dots, n\big) = \sum_{\substack{\omega \in \Omega_n \\ \eta \to \omega}} \mathbb{P}\big(\{\eta\}\big) = \frac{1}{|\Omega_n|} 5^{\#\{k\colon\, \omega_k = 1\}}. \tag{1.15}$$

Here $\eta \to \omega$ means that η_k equals six if and only if $\omega_k = 1$, for all $k = 1, \dots, n$, and the last expression is the consequence of the fact that there are five different choices for η_k whenever $\omega_k = 0$. As is easy to check, this is exactly (1.14).

The quantity X_k we defined is our first example of an important concept:

Definition 1.9 [Random variable] *Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ a (real-valued)* random variable *X is a function $X \colon \Omega \to \mathbb{R}$ for which*

$$\{\omega \in \Omega \colon X(\omega) \in I\} \in \mathscr{F} \tag{1.16}$$

for any interval $I \subset \mathbb{R}$. A random variable is vector valued if it takes values in $\mathbb{R}^d$ (in this case we substitute intervals by balls in (1.16)*).*

The condition (1.16) is entirely technical and can be omitted except for its role in the following definition:

Definition 1.10 [Distribution function] *Consider a random variable X taking values in $\mathbb{R}$. Then the function*

$$F(x) = \mathbb{P}(\{\omega \colon X(\omega) \le x\}) \tag{1.17}$$

is called the distribution function*.*

Clearly, we need (1.16) for at least intervals of the form $I = (-\infty, x]$ in order to be able to define the distribution function. Generally, we refer to expressions of the form $\mathbb{P}(X \in A)$ as the *distribution* of X.

Problem 1.11 Show that F takes values in $[0,1]$, is non-decreasing and right continuous with $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$. Moreover,

$$\mathbb{P}(X = z) = F(z) - \lim_{a\uparrow z} F(a). \tag{1.18}$$

In particular, if F is continuous then X takes no value with positive probability.

Definition 1.12 [Discrete vs continuous] *Let X be a random variable with distribution function F. We say that X has* discrete distribution *if F is piecewise constant and* continuous distribution *if there exists a function $f\colon \mathbb{R} \to [0,\infty)$, called the* probability density*, such that for every $a \in \mathbb{R}$,*

$$F(a) = \int_{-\infty}^{a} f(x)\mathrm{d}x \tag{1.19}$$

Going back to the random variables $X_k(\omega)$ in Example 1.8, a moment's thought reveals that (1.15) can be generalized as follows: Let $A_1, \dots, A_n$ be subsets of $\{0,1\}$. Then the probability that $X_k \in A_k$ for all k factors into the product of probabilities. This natural property is referred to as *independence*. Here is a formal definition:

Definition 1.13 [Independence] *A collection $X_1, X_2, \dots, X_n$ of $\mathbb{R}^d$-valued random variables is said to be* independent *if for any sequence of balls $A_1, \dots, A_n$ in $\mathbb{R}^d$,*

$$\mathbb{P}\Big(\bigcap_{k=1}^{n}\{X_k \in A_k\}\Big) = \prod_{k=1}^{n}\mathbb{P}(X_k \in A_k). \tag{1.20}$$

The reason for this name is seen from the following observation:

Problem 1.14 Suppose $X_1, X_2, \dots, X_n$ can only take values in a countable set R. Show that knowing the value of one random variable, say X_1, does not influence the distribution of the others. Explicitly, for any $a_1, \dots, a_n \in R$,

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) = \mathbb{P}(X_1 = a_1)\mathbb{P}(X_2 = a_2, \dots, X_n = a_n) \tag{1.21}$$

Let us go back to Example 1.8. The right-hand side in (1.14) is not the most general imaginable on Ω_n. An important extension is as follows:

Example 1.15 *Bernoulli random variables*: Consider the sample space $\Omega_n = \{0,1\}^n$ and let $p \in [0,1]$. For each sequence $\omega \in \Omega_n$, define

$$\mathbb{P}\big(\{\omega\}\big) = \prod_{k=1}^{n} p^{\omega_k}(1-p)^{1-\omega_k} \tag{1.22}$$

This is the *Bernoulli distribution*. Note that the outcomes of the experiment with this distribution are manifestly independent in the above sense.

Consider a collection of Bernoulli (i.e., independent zero-one valued) random variables X_k where $k = 1, 2, \dots$. Define

$$S_n = X_1 + \cdots + X_n, \tag{1.23}$$

that is, the sum of the first n elements of the random sequence (X_k). As is easy to check, S_n can take any integer values between 0 and n. Suppose we wish to calculate the *distribution* of S_n, i.e., the collection of numbers $\mathbb{P}(S_n = k)$ for all $k = 0, 1, \ldots, n$. The answer is as follows:

Lemma 1.16 [Binomial distribution] *Let (X_j) be Bernoulli with parameter $p \in [0, 1]$. For each $k = 0, 1, \ldots, n$, we have*

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.24}$$

A random variable with this distribution is called Binomial *with parameters n and p.*

Proof. Recall that X_j are, technically, functions $\Omega_n \to \{0, 1\}$ such that $X_j(\omega) = \omega_j$. Now to ensure that $S_n = k$, exactly k of the X_j have to evaluate to one and $n - k$ to zero. This means that the *event* $\{S_n = k\}$ is the collection of $\omega \in \Omega_n$ with exactly k ones and $n - k$ zeros. Each such sequence has probability $p^k(1-p)^{n-k}$ and there are $\binom{n}{k}$ such sequences. This proves the desired claim. □

A binomial random variable is an example of a discrete random variable. Another such example is the *Poisson* random variable X that takes values in $\{0\} \cup \mathbb{N}$ with probabilities

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda}, \qquad n \geq 0. \tag{1.25}$$

Here $\lambda > 0$ is a parameter. The last two examples are actually closely related:

Problem 1.17 *Poisson convergence*: Consider a binomial random variable $Z_{n,p}$ with parameters n and p. Set $p = \lambda/n$ with $\lambda > 0$. Show that then for each $m \geq 0$,

$$\mathbb{P}(Z_{n,\lambda/n} = m) \underset{n\to\infty}{\longrightarrow} \frac{\lambda^m}{m!} \mathrm{e}^{-\lambda} \tag{1.26}$$

In technical terms, the random variables $Z_{n,\lambda/n}$ converge *in distribution* to a Poisson random variable with parameter λ.

A yet another example of discrete random variables is the geometric random variable that can be arrived at as follows:

Problem 1.18 Suppose $X_1, X_2, \ldots$ be independent 0-1-valued random variables with $\mathbb{P}(X_k = 1) = p$ and $\mathbb{P}(X_k = 0) = 1 - p$. We may think of these as results of tossing an unfair coin. Define T to be the time when the first 1 appeared,

$$T = \inf\{n \geq 1 \colon X_n = 1\}. \tag{1.27}$$

Show that T takes all values in $\mathbb{N}$ with

$$\mathbb{P}(T = n) = (1-p)^{n-1} p, \qquad n \geq 1. \tag{1.28}$$

We call this distribution *geometric*.

Examples of random variables with continuous distribution are easier to state because one only has to give the corresponding probability density function $f(x)$. Note that in these cases

$$\mathbb{P}(X \in A) = \int_A f(x)\mathrm{d}x \tag{1.29}$$

Example 1.19 *Normal*: A normal random variable — denoted by $\mathcal{N}(\mu, \sigma^2)$ — is characterized by two parameters, the *mean* μ and *variance* σ^2. The probability density is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2} \tag{1.30}$$

The special case of $\mathcal{N}(0,1)$, where the probability density takes the form

$$f(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2} \tag{1.31}$$

is referred to as *standard normal*.

Exercise 1.20 Show that if X has distribution of $\mathcal{N}(\mu, \sigma^2)$, then

$$Y = \frac{X-\mu}{\sigma} \tag{1.32}$$

has the distribution of $\mathcal{N}(0,1)$.

Example 1.21 *Exponential*: This random variable takes values in $[0, \infty)$; the probability density is

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1.33}$$

Example 1.22 *Cauchy*: A Cauchy random variable takes all values in $\mathbb{R}$ and it has the probability density

$$f(x) = \frac{1}{\pi}\,\frac{1}{1+x^2} \tag{1.34}$$

Having given a sufficient number of representative examples, let us address the last item on the "basic probability list," namely, computing expectations.

Definition 1.23 [Expectation] *Let X be a random variable taking values in $\mathbb{R}$. Then the* expectation, $\mathbb{E}X$, *is defined by*

$$\mathbb{E}X = \sum_a a\,\mathbb{P}(X = a) \tag{1.35}$$

if X has discrete distribution, and by

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \tag{1.36}$$

if X has continuous distribution with probability density f. We say that $\mathbb{E}X$ does not exist if either the sum or the integral is not well defined (e.g., is of type $\infty - \infty$, etc).

The motivation for introducing the notion of expectation is as follows. We already agreed that $\mathbb{P}(X = a)$ corresponds to the relative frequency of seeing a in many samples of random variable X, i.e., in n trials we should see order $n\mathbb{P}(X = a)$ occurrences of a. If we are interested in the the total sum of all values, then the sum should be order n times $\sum_a a\mathbb{P}(X = a) = \mathbb{E}X$. This is expressed more precisely by the *Laws of Large Numbers*.

Problem 1.24 Let X be a *positive* random variable with probability density $f(x)$ and distribution function $F(x)$. Show that then

$$\mathbb{E}X = \int_0^\infty \big(1 - F(x)\big)\mathrm{d}x. \tag{1.37}$$

Since a function of a random variable is a random variable, this defines expectation of $g(X)$ for any function $g\colon \mathbb{R} \to \mathbb{R}$. However, technically, this expectation is expressed using the probability density of $g(X)$. Nevertheless, it is not hard to re-express it back using the probability density of X:

Lemma 1.25 [Change of variables] *Let $g\colon \mathbb{R} \to \mathbb{R}$ and let X be a random variable with probability density $f(x)$. Suppose that the expectation $\mathbb{E}g(X)$ exists. Then*

$$\mathbb{E}g(X) = \int_{-\infty}^\infty g(x)f(x)\mathrm{d}x. \tag{1.38}$$

We omit the proof for brevity. The idea of the proof is easier to see in the analogous statement for discrete random variables:

Exercise 1.26 Let X have a discrete distribution and let us define $\mathbb{E}g(X)$ using the probability mass function of $g(x)$. Show that

$$\mathbb{E}g(X) = \sum_a g(a)\,\mathbb{P}(X = a). \tag{1.39}$$

Finally, we introduce another common concept:

Definition 1.27 [Moments and variance] *Given a random variable X taking values in $\mathbb{R}$, its k-*th moment *is simply the expectation $\mathbb{E}(|X|^k)$. The* variance of X, *denoted by* $\mathrm{Var}(X)$, *is given by*

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}X)^2\big). \tag{1.40}$$

Exercise 1.28 Show that the variance of X can also be computed by

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \tag{1.41}$$

provided the second moment of X exists.

Finally we also note the relation between expectation and independence:

Lemma 1.29 *Suppose $X_1, X_2, \ldots, X_n$ are independent and $g_1, \ldots, g_n\colon \mathbb{R} \to \mathbb{R}$ are functions. Then*

$$\mathbb{E}\Big(\prod_{k=1}^n g_k(X_k)\Big) = \prod_{k=1}^n \mathbb{E}\big(g_k(X_k)\big) \tag{1.42}$$

Chapter 2

Random walks

Random walks are one of the basic objects studied in probability theory. The motivation comes from observations of various random motions in physical and biological sciences. The most well-known example is the erratic motion of pollen grains immersed in a fluid — observed by botanist Robert Brown in 1827 — caused, as we now know, by collisions with rapid molecules. The latter example serves just as well for the introduction of Brownian motion. As will be discussed in a parallel course, Brownian motion is a continuous analogue of random walk and, not surprisingly, there is a deep connection between both subjects.

2.1 Random walks and limit laws

The definition of a random walk uses the concept of independent random variables whose technical aspects are reviewed in Chapter 1. For now let us just think of independent random variables as outcomes of a sequence of random experiments where the result of one experiment is not at all influenced by the outcomes of the other experiments.

Definition 2.1 [Random walk] *Suppose that $X_1, X_2, \dots$ is a sequence of $\mathbb{R}^d$-valued independent and identically distributed random variables. A* random walk *started at $z \in \mathbb{R}^d$ is the sequence $(S_n)_{n\geq 0}$ where $S_0 = z$ and*

$$S_n = S_{n-1} + X_n, \qquad n \geq 1. \tag{2.1}$$

The quantities (X_n) are referred to as steps *of the random walk.*

Our interpretation of the above formula is as follows: The variable S_n marks the position of the walk at time n. At each time the walk chooses a step at random — with the same step distribution at each time — and adds the result to its current position. The above can also be written as

$$S_n = z + X_1 + \cdots + X_n \tag{2.2}$$

for each $n \geq 1$. Note that while the steps $X_1, X_2, \dots$ are independent as random variables, the actual positions of the walk $S_0, S_1, \dots$ are not.

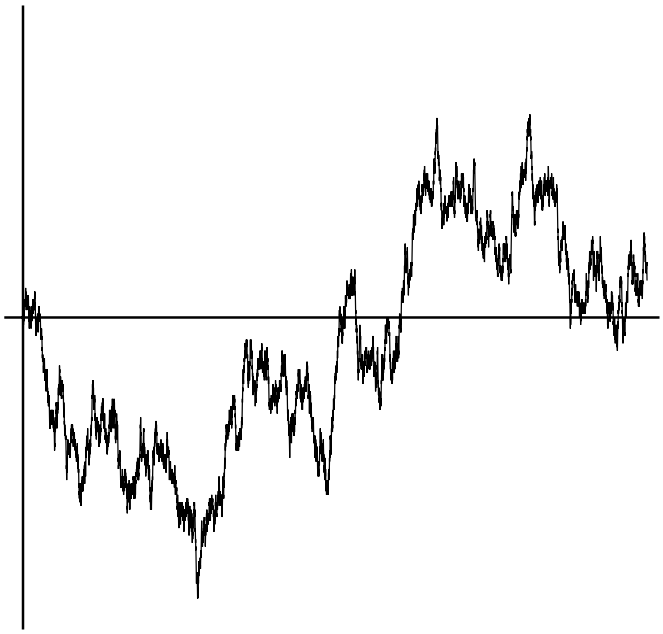

Figure 2.1: A path of length 10^4 of the simple random walk on $\mathbb{Z}$ drawn by interpolating linearly between the points with coordinates (n, S_n), $n = 0, \ldots, 10^4$.

Exercise 2.2 Let $(S_n)_{n\geq 0}$ be a random walk. Show that $S_{2n} - S_n$ and S_n are independent and have the same distribution.

Here are some representative examples of random walks:

Example 2.3 *Simple random walk (SRW) on* $\mathbb{Z}$: This is the simplest of all random walks — hence the name. Here X_1 takes values in $\{+1, -1\}$ and the walk S_n started from 0 is thus confined to the set of all integers $\mathbb{Z}$. Often enough, X_1 takes both values with equal probabilities, i.e.,

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2 \tag{2.3}$$

The walk then jumps left or right equally likely at each time. This case is more correctly referred to as the "simple *symmetric* random walk," but the adjective "symmetric" is almost invariably dropped. In the other cases, i.e., when

$$\mathbb{P}(X_1 = 1) = p \quad \text{and} \quad \mathbb{P}(X_1 = -1) = 1 - p \tag{2.4}$$

with $p \neq 1/2$, the walk is referred to as *biased*. The bias is to the right when $p > 1/2$ and to the left when $p < 1/2$.

Example 2.4 *Simple random walk on* $\mathbb{Z}^d$: This is a d-dimensional version of the first example. Here X_1 takes values in $\{\pm\hat{e}_1, \ldots, \pm\hat{e}_d\}$ where $\hat{e}_k$ is the "coordinate vector" $(0, \ldots, 0, 1, 0, \ldots, 0)$ in $\mathbb{R}^d$ with the "1" appearing in the k-th position. This random walk is confined to the set of points in $\mathbb{R}^d$ with integer coordinates,

$$\mathbb{Z}^d = \big\{(n_1, \ldots, n_d) \colon n_1, \ldots, n_d \in \mathbb{Z}\big\}. \tag{2.5}$$

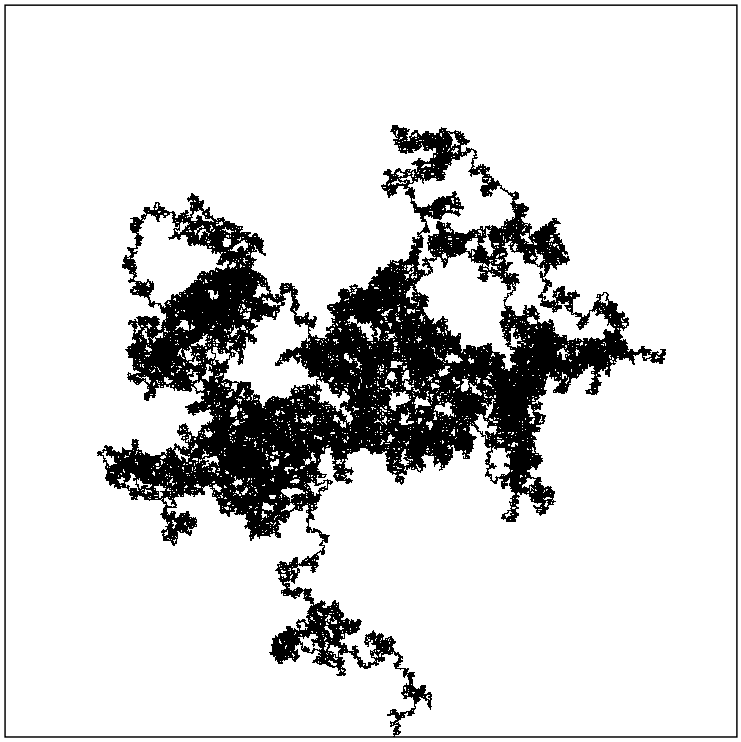

Figure 2.2: The set of vertices visited by a two-dimensional simple random walk before it exited a box of side 10^3. The walk was started at the center of the box and it took 682613 steps to reach the boundary.

The easiest example to visualize is the case of $d = 2$ where the set $\mathbb{Z}^2$ are the vertices of a square grid. Thinking of $\mathbb{Z}^2$ as a graph, the links between the neighboring vertices represent the allowed transitions of the walk. A majority of appearances of this random walk is in the symmetric case; i.e., when X_1 takes any of the $2d$ allowed values with equal probabilities.

Example 2.5 *"As the knight jumps" random walk on* $\mathbb{Z}^2$: This random walk takes steps allowed to the knight in the game of chess; i.e., there are 8 allowed jumps

$$2\hat{e}_1 + \hat{e}_2, \quad \hat{e}_1 + 2\hat{e}_2, \quad -2\hat{e}_1 - \hat{e}_2, \quad -\hat{e}_1 - 2\hat{e}_2, \tag{2.6}$$

$$2\hat{e}_1 - \hat{e}_2, \quad \hat{e}_1 - 2\hat{e}_2, \quad -2\hat{e}_1 + \hat{e}_2, \quad -\hat{e}_1 + 2\hat{e}_2. \tag{2.7}$$

Some experience with chess reveals that the random walk can reach every vertex of $\mathbb{Z}^2$ in a finite number of steps. This fails to be true if we further reduce the steps only to those in the top line; the random walk is then restricted to the fraction of $1/3$ of all vertices in $\mathbb{Z}^2$; see Fig. 2.3.

Example 2.6 *Gaussian random walk*: This random walk has steps that can take any value in $\mathbb{R}$. The probability distribution of X_1 is normal (or Gaussian) with mean

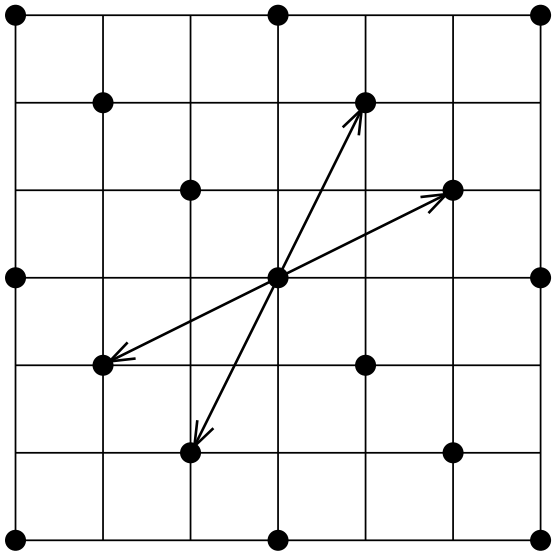

Figure 2.3: The set of allowed steps (arrows) and reachable vertices (dots) for the random walk discussed in Example 2.5.

zero and variance 1, i.e., $X_1 = \mathcal{N}(0,1)$ or, explicitly,

$$\mathbb{P}(X_1 \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-x^2/2} \, \mathrm{d}x \tag{2.8}$$

A distinguished feature of this walk that the distribution of S_n is also normal with mean zero but variance $\sqrt{n}$. A typical displacement of this random walk after n steps is thus "order-$\sqrt{n}$" — a scale that, as we will see in Theorem 2.11, is quite typical for random walks with zero mean.

Example 2.7 *Heavy tailed random walk*: To provide contrast to the previous example, we can also take a random walk on $\mathbb{R}$ with a step distribution that is symmetric but has "heavy tails." (We discuss these briefly in Chapter 1.) For instance, take X_1 continuous with probability density

$$f(x) = \begin{cases} \frac{\alpha}{2} \, \frac{1}{|x|^{\alpha+1}}, & \text{if } |x| \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

where α is a parameter with $0 < \alpha < 2$. As is seen by comparing Fig. 2.1 and Fig. 2.4, a distinction between this random walk and the SRW is clear at first sight.

We finish our introduction to random walks by adapting standard limit theorems for sequences of i.i.d. random variables to the quantity $S_n = X_1 + \cdots + X_n$. Note the requirement of a particular moment condition in each theorem.

We begin by extracting the leading order (linear) scaling of S_n:

Theorem 2.8 [Strong Law of Large Numbers] *Suppose that $E|X_1| < \infty$. Then, with probability one,*

$$\lim_{n\to\infty} \frac{S_n}{n} \text{ exists and equals } \mathbb{E}X_1 \tag{2.10}$$

The expectation $\mathbb{E}X_1$ thus defines the *asymptotic velocity* of the walk. In particular, if $\mathbb{E}X_1 \neq 0$ then the walks moves away from the starting point at linear speed while for $\mathbb{E}X_1 = 0$ the speed is zero.

Exercise 2.9 Show that if $\mathbb{E}X_1 \neq 0$, the probability that the random walk with steps $X_1, X_2, \ldots$ visits the starting point infinitely often is zero.

Problem 2.10 An example of a heavy tailed random walk is the *Cauchy random walk* where X_1 has Cauchy distribution characterized by the probability density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}. \tag{2.11}$$

This example is analogous — or technically, in same "basin of attraction" — as the $\alpha = 1$ random walk discussed in Example 2.7. Show that if $X_1, \ldots, X_n$ are independent Cauchy, then so is S_n/n for each n. In particular, the conclusion of Theorem 2.8 fails in this case.

Next we will describe the fluctutations of the position S_n around its mean:

Theorem 2.11 [Central Limit Theorem] *Consider a one-dimensional random walk with $\mathbb{E}(X_1^2) < \infty$. Then, as $n \to \infty$,*

$$\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}} \tag{2.12}$$

has asymptotically normal distribution with mean zero and variance $\sigma^2 = \mathrm{Var}(X_1)$.

The crux of this result is that, for the walks with $\mathbb{E}X_1 = 0$, the distribution of the endpoint is asymptotically very close to that of the Gaussian random walk with a properly adjusted variance. This is a manifestation of a much more general *invariance principle* that deals with the distribution of the entire path of the random walk. The limiting object there is Brownian motion.

Problem 2.12 Consider the Gaussian random walk of length n. Show that the largest step is of size order $\sqrt{\log n}$ and that the difference between the first and second largest positive step tends to zero as $n \to \infty$.

Exercise 2.13 Suppose $1 < \alpha < 2$ and consider the first n steps of the heavy tailed random walk from Example 2.7. Show that the probability that the largest step is twice as large than any other step is bounded away from zero uniformly in n.

Problem 2.14 Suppose now that $0 < \alpha < 1$. Show that that with probability that is uniformly positive in n, the largest step of a heavy tailed random walk of length n is larger than the sum of the remaining steps. See Fig. 2.4.

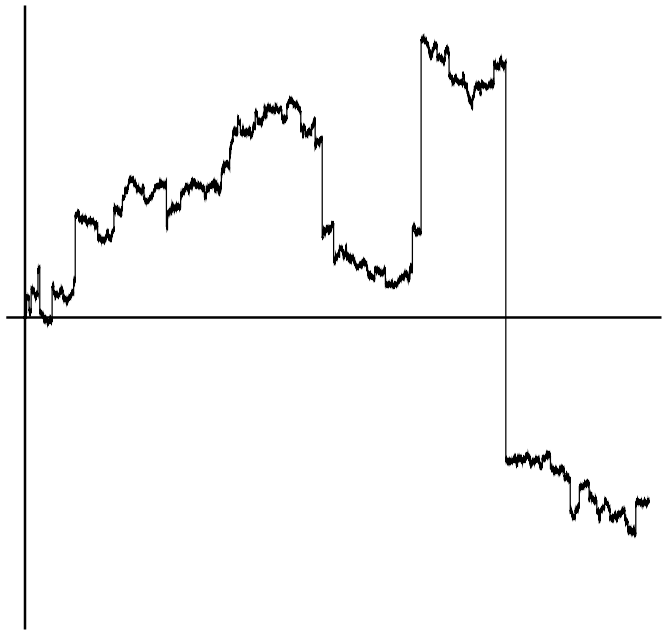

Figure 2.4: A plot of 25000 steps of the heavy tailed random walk from Example 2.7 with $\alpha = 1.2$. The defining feature of heavy tailed random walks is the presence of "macroscopic" jumps, i.e., those comparable with the typical distance of the walk from the starting point at the time of their occurrence. In particular, the Central Limit Theorem does not apply due to the lack of the second moment of X_1.

2.2 Transition in $d = 2$: Recurrence vs transience

In the previous section we introduced random walks in quite some generality. However, to make our discussion easier, we will henceforth assume that

$$\text{all random walks have step distribution concentrated on } \mathbb{Z}^d$$

Our next business is to try to address two basic questions:

(1) Under what conditions does a random walk come infinitely often back to its starting position?

(2) When do the paths of two independent copies of the same random walk intersect infinitely often?

The interest in these is bolstered by the fact that the answer depends sensitively on the dimension. Explicitly, for rather generic step distributions, the character of the answer changes as dimension goes from 2 to 3 for the first question and from 4 to 5 for the second question.

Throughout this section we will focus on the first question.

Definition 2.15 [Recurrence & transience] *We say that a random walk is* recurrent *if it visits its starting position infinitely often with probability one and* transient *if it visits its starting position finitely often with probability one.*

Our analysis begins by showing that every random walk is either recurrent or transient; no intermediate scenarios take place. Let N be the number of visits of (S_n) to its starting point S_0,

$$N = \sum_{n\geq 0} 1_{\{S_n = S_0\}}. \tag{2.13}$$

Recurrence then means $\mathbb{P}(N = \infty) = 1$ while transience means $\mathbb{P}(N < \infty) = 1$ and so absence of intermediate scenarios is equivalent to showing $P(N < \infty) \in \{0, 1\}$. Let τ denote the first time the walk is back to the starting point:

$$\tau = \inf\big\{n \geq 1 \colon X_1 + \cdots + X_n = 0\big\} \tag{2.14}$$

If no such visit exists, then $\tau = \infty$. Note that

$$\mathbb{P}(N = 1) = \mathbb{P}(\tau = \infty). \tag{2.15}$$

Lemma 2.16 [Either recurrent or transient] *For each $n \geq 1$,*

$$\mathbb{P}(N = n) = \mathbb{P}(\tau = \infty)\mathbb{P}(\tau < \infty)^{n-1}. \tag{2.16}$$

Then either $\mathbb{P}(\tau = \infty) = 0$ which implies $\mathbb{P}(N < \infty) = 0$, or $\mathbb{P}(\tau = \infty) > 0$ which implies $\mathbb{P}(N < \infty) = 1$. In particular, every random walk is either recurrent or transient.

Proof. We first prove the identity

$$\mathbb{P}(N = n+1) = \mathbb{P}(N = n)\mathbb{P}(\tau < \infty), \qquad n \geq 1. \tag{2.17}$$

Consider the first visit back to the origin and suppose it occurred at time $\tau = k$. Then $N = n + 1$ implies that the walk $S'_m = S_{k+m} - S_k$ — namely, the part of the walk (S_n) after time k — makes n visits back to its starting point, $S'_0 = 0$. But the walk S'_m is independent of the event $\{\tau = k\}$ because $\tau = k$ is determined by $X_1, \ldots, X_k$ while S'_m is a function of only $X_{k+1}, X_{k+2}, \ldots$. This implies

$$\begin{aligned}\mathbb{P}(N = n+1 \;\&\; \tau = k) &= \mathbb{P}\Big(\sum_{m\geq 0} 1_{\{S'_m = 0\}} = n \;\&\; \tau = k\Big) \\ &= \mathbb{P}\Big(\sum_{m\geq 0} 1_{\{S'_m = 0\}} = n\Big)\,\mathbb{P}(\tau = k) = \mathbb{P}(N = n)\mathbb{P}(\tau = k)\end{aligned} \tag{2.18}$$

where we used that the walk S'_m has the same distribution as S_m. Summing

$$\mathbb{P}(N = n+1 \;\&\; \tau = k) = \mathbb{P}(N = n)\mathbb{P}(\tau = k) \tag{2.19}$$

over k in the range $1 \leq k < \infty$ gives (2.17).

To get (2.16), plug (2.15) in (2.17) and solve recursively for $\mathbb{P}(N = n)$. For the rest of the claim, we note that if $\mathbb{P}(\tau = \infty) = 0$, then $\mathbb{P}(N = n) = 0$ for all $n < \infty$

implying $\mathbb{P}(N < \infty) = 0$. If, on the other hand, $\mathbb{P}(\tau = \infty) > 0$ then $\mathbb{P}(\tau < \infty) < 1$ and, by (2.17), the probabilities $\mathbb{P}(N = n)$ form a geometric sequence. Summing over all n in the range $1 \leq n < \infty$ gives

$$\mathbb{P}(N < \infty) = \frac{\mathbb{P}(\tau = \infty)}{1 - \mathbb{P}(\tau < \infty)} = 1 \tag{2.20}$$

as desired. □

Problem 2.17 Suppose (S_n) is a random walk and let x be such that $\mathbb{P}(S_n = x) > 0$ for some $n \geq 0$. Prove that with probability one (S_n) visits x only finitely often if (S_n) is transient and infinitely often if (S_n) is recurrent.

The main technical point of the previous derivations is that transience can be characterized in terms of finiteness of $\mathbb{E}N$:

Lemma 2.18 *A random walk is transient if $\mathbb{E}N < \infty$ and recurrent if $\mathbb{E}N = \infty$.*

Proof. If $\mathbb{E}N < \infty$ then $\mathbb{P}(N < \infty) = 1$ and the walk is transient. However, the other implication is more subtle. Assume $\mathbb{P}(N < \infty) = 1$ and note that then also $\mathbb{P}(\tau = \infty) > 0$. Then sequence $\mathbb{P}(N = n)$ thus decays exponentially and so

$$\begin{aligned} \mathbb{E}N &= \sum_{n=1}^{\infty} n\mathbb{P}(N = n) = \mathbb{P}(\tau = \infty) \sum_{n=1}^{\infty} n\mathbb{P}(\tau < \infty)^{n-1} \\ &= \frac{\mathbb{P}(\tau = \infty)}{[1 - \mathbb{P}(\tau < \infty)]^2} = \frac{1}{\mathbb{P}(\tau = \infty)} \end{aligned} \tag{2.21}$$

Hence $\mathbb{E}N < \infty$ as we intended to show. □

Exercise 2.19 As noted in the proof, the fact that $\mathbb{E}N < \infty$ implies $\mathbb{P}(N < \infty) = 1$ is special for the context under consideration. To see this is not true in general, find an example of an integer valued random variable $Z \geq 0$ such that $\mathbb{P}(Z < \infty) = 1$ but $\mathbb{E}Z = \infty$.

Exercise 2.20 Show that the probability $\mathbb{P}(S_n = 0)$ for the simple symmetric random walk in $d = 1$ decays like $n^{-1/2}$. Conclude that the walk is recurrent.

A practical advantage of the characterization using the finiteness of $\mathbb{E}N$ is that the expectation can be explicitly computed:

Lemma 2.21 [Expectation formula] *Consider a random walk on $\mathbb{Z}^d$ with steps denoted by $X_1, X_2, \ldots$ and let*

$$\varphi(k) = \mathbb{E}(\mathrm{e}^{\mathrm{i}k\cdot X_1}) := \mathbb{E}\cos(k \cdot X_1) + \mathrm{i}\,\mathbb{E}\sin(k \cdot X_1). \tag{2.22}$$

Then

$$\mathbb{E}N = \lim_{t\uparrow 1} \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d} \frac{1}{1 - t\varphi(k)} \tag{2.23}$$

Proof. The proof is based on the formula

$$1_{\{S_n=0\}} = \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d}\, \mathrm{e}^{\mathrm{i}k\cdot S_n}, \tag{2.24}$$

which is a consequence of d-fold application of the Fourier identity

$$\int_{[-\pi,\pi]} \frac{\mathrm{d}\theta}{2\pi}\, \mathrm{e}^{\mathrm{i}n\theta} = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \in \mathbb{Z} \setminus \{0\}. \end{cases} \tag{2.25}$$

(Here is where we used the fact that the walk is confined to *integer* lattice.)
Taking expectation in (2.24), we thus get

$$\mathbb{P}(S_n = 0) = \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d}\, \mathbb{E}(\mathrm{e}^{\mathrm{i}k\cdot S_n}). \tag{2.26}$$

Since S_n is the sum $X_1 + \cdots + X_n$, we have $\mathrm{e}^{\mathrm{i}S_n} = \mathrm{e}^{\mathrm{i}k\cdot X_1} \cdots \mathrm{e}^{\mathrm{i}k\cdot X_n}$. Moreover, as $X_1, \ldots, X_n$ are independent, then so $\mathrm{e}^{\mathrm{i}k\cdot X_1}, \cdots, \mathrm{e}^{\mathrm{i}k\cdot X_n}$. Using that expectation of a product of independent random variables is a product of expectations,

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}k\cdot S_n}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}k\cdot X_1}) \cdots \mathbb{E}(\mathrm{e}^{\mathrm{i}k\cdot X_n}) = \varphi(k)^n \tag{2.27}$$

Next multiply (2.26) by t^n for some $t \in [0, 1)$ and sum on $n \geq 0$. This gives

$$\begin{aligned} \sum_{n=0}^{\infty} t^n \mathbb{P}(S_n = 0) &= \sum_{n\geq 0} t^n \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d} \varphi(k)^n \\ &= \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d} \sum_{n\geq 0} \big[t\varphi(k)\big]^n \\ &= \int_{[-\pi,\pi]^d} \frac{\mathrm{d}k}{(2\pi)^d} \frac{1}{1 - t\varphi(k)} \end{aligned} \tag{2.28}$$

where we used that $|t\varphi(k)| \leq t < 1$ to see that the sum and integral can be interchanged in the second line. Taking the limit $t \uparrow 1$ makes the left-hand side tend to $\sum_{n\geq 0} \mathbb{P}(S_n = 0) = \mathbb{E}N$. □

These observation allow us to characterize when the simple random walk is recurrent and when it is transient:

Theorem 2.22 [Recurrence/transience of SRW] *The simple symmetric random walk on $\mathbb{Z}^d$ is recurrent in dimensions $d = 1, 2$ and transient in dimensions $d \geq 3$.*

Proof. To apply the previous lemma, we need to calculate φ for the SRW. Using that the walk makes steps only in (positive or negative) coordinate directions, we get

$$\begin{aligned} \varphi(k) &= \frac{1}{2d}\mathrm{e}^{\mathrm{i}k_1} + \frac{1}{2d}\mathrm{e}^{-\mathrm{i}k_1} + \cdots + \frac{1}{2d}\mathrm{e}^{\mathrm{i}k_d} + \frac{1}{2d}\mathrm{e}^{-\mathrm{i}k_d} \\ &= \frac{1}{d}\cos(k_1) + \cdots + \frac{1}{d}\cos(k_d). \end{aligned} \tag{2.29}$$

This shows that $\varphi(k) = 1$ on $[-\pi, \pi]^d$ if an only if $k_1 = \cdots = k_d = 0$ and so $k = 0$ is the only point that could make the integral diverge in the limit as $t \uparrow 1$. To find out what happens precisely, we will need to control the behavior of the function $1 - t\varphi(k)$ around $k = 0$ for t close to one. First we note that

$$1 - \cos(x) = 2\sin^2(x/2) \quad \text{and} \quad \frac{2x}{\pi} \leq \sin(x) \leq x \tag{2.30}$$

yield

$$2\frac{k_i^2}{\pi^2} \leq 1 - \cos(k_i) \leq \frac{k_i^2}{2} \tag{2.31}$$

Plugging this in the definition of $\varphi(k)$ shows that

$$1 - t + 2t\frac{|k|^2}{\pi^2 d} \leq 1 - t\varphi(k) \leq 1 - t + \frac{|k|^2}{2d}. \tag{2.32}$$

Taking the limit we find that the function $k \mapsto 1 - t\varphi(k)$ is uniformly integrable around $k = 0$ if and only if the function $k \mapsto |k|^2$ is integrable, i.e.,

$$\mathbb{E}N < \infty \quad \text{if and only if} \quad \int_{|k|<1} \frac{\mathrm{d}k}{|k|^2} < \infty \tag{2.33}$$

The integral is finite if and only if $d \geq 3$. □

A famous quote sums up the previous theorem as follows: "A drunken man will always find his way home but a drunken bird may get lost forever." This, of course, assumes that the spontaneous motion of intoxicated biological material is described by a random walk with similar properties as the SRW.

Exercise 2.23 Show that a biased simple random walk on $\mathbb{Z}$ — i.e., the walk on $\mathbb{Z}$ with $\mathbb{P}(X_1 = +1) = p = 1 - \mathbb{P}(X_1 = -1)$ — is transient for all $p \neq 1/2$.

Problem 2.24 Use the above techniques to show that the random walk described in Example 2.5 is recurrent in dimensions $d = 2$ and transient otherwise — the generalization to $d \geq 3$ is straightforward: The walk jumps by two units in one lattice direction and one in some other. How about the walk in Fig. 2.3? (Here both jumps are either both positive or both negative.)

Problem 2.25 Consider a random walk on $\mathbb{Z}$ with step distribution

$$\mathbb{P}(X_1 = n) = \frac{1}{2}\Big(\frac{1}{|n|^\alpha} - \frac{1}{(|n|+1)^\alpha}\Big), \qquad n \neq 0. \tag{2.34}$$

Characterize the values of $\alpha > 0$ for which the walk is recurrent.

2.3 Transition in $d = 4$ & Loop-erased random walk

In this section will be devoted to the second question from Section 2.2 which concerns the *non-intersection of the paths* of independent copies of the same random

walk. Consider two independent copies (S_n) and $(\tilde{S}_n)$ of the same random walk. We are interested in the cardinality of the set

$$\mathfrak{I}(S, \tilde{S}) := \{S_n \colon n \geq 0\} \cap \{\tilde{S}_n \colon n \geq 0\}. \tag{2.35}$$

First we note that in some cases the question can be answered directly:

Exercise 2.26 Use Problem 2.17 to show that paths of two independent copies of a (non-constant) recurrent random walk meet at infinitely many distinct points.

This allows us to focus, as we will do from now on, on transient random walks only. Some of these can be still handled by geometric arguments:

Problem 2.27 Show that the paths of two independent copies of a simple random walk on $\mathbb{Z}$, biased or symmetric, intersect infinitely often with probability one.

To address the general case, instead of $|\mathfrak{I}(S, \tilde{S})|$ we will work with the number

$$N^{(2)} = \sum_{m,n \geq 0} 1_{\{S_m = S_n\}} \tag{2.36}$$

that counts the number of pairs of times when the walks collided. To see this comes at no loss, we note that

$$N^{(2)} < \infty \quad \Rightarrow \quad \bigl|\mathfrak{I}(S, \tilde{S})\bigr| < \infty \tag{2.37}$$

To get the opposite implication, we note:

Lemma 2.28 *Suppose the random walks S and $\tilde{S}$ are transient. Then*

$$\mathbb{P}(N^{(2)} = \infty) = 1 \quad \textit{if and only if} \quad \mathbb{P}\bigl(|\mathfrak{I}(S, \tilde{S})| = \infty\bigr) = 1 \tag{2.38}$$

Proof. Let n_x be the number of visits of (S_n) to x,

$$n_x = \sum_{n \geq 0} 1_{\{S_n = x\}} \tag{2.39}$$

and let $\tilde{n}_x$ be the corresponding quantity for $\tilde{S}_n$. By the assumption of transience, $n_x < \infty$ and $\tilde{n}_x < \infty$ for every x with probability one. Next we note

$$N^{(2)} = \sum_{m,n \geq 0} \sum_{x \in \mathfrak{I}(S,\tilde{S})} 1_{\{S_n = x\}} 1_{\{\tilde{S}_m = x\}} = \sum_{x \in \mathfrak{I}(S,\tilde{S})} n_x \tilde{n}_x \tag{2.40}$$

If $|\mathfrak{I}(S, \tilde{S})| < \infty$, then the sum would be finite implying $N^{(2)} < \infty$. Thus, if $\mathbb{P}(N^{(2)} = \infty) = 1$ then we must have $|\mathfrak{I}(S, \tilde{S})| = \infty$ with probability one. □

We now proceed to characterize the transient random walks which $N^{(2)}$ is finite with probability one. The analysis is analogous to the question of recurrence vs transience but some steps are more tedious and so will be a bit sketchy at times.

Using arguments that we omit for brevity, one can again show that $\mathbb{P}(N^{(2)} < \infty)$ takes only values zero and one and

$$\mathbb{P}(N^{(2)} < \infty) = 1 \quad \text{if and only if} \quad \mathbb{E}N^{(2)} < \infty. \tag{2.41}$$

Next we prove:

Lemma 2.29 *Consider a random walk on $\mathbb{Z}^d$ with steps $X_1, X_2, \ldots$ and let, as before, $\varphi(k) = \mathbb{E}(e^{ik\cdot X_1})$. Then*

$$\mathbb{E}N^{(2)} = \lim_{t\uparrow 1} \int_{[-\pi,\pi]^d} \frac{dk}{(2\pi)^d} \frac{1}{|1 - t\varphi(k)|^2} \tag{2.42}$$

Proof. A variant of the formula (2.24) gives

$$1_{\{S_n=\tilde{S}_m\}} = \int_{[-\pi,\pi]^d} \frac{dk}{(2\pi)^d} e^{ik\cdot(S_n-\tilde{S}_m)}. \tag{2.43}$$

Applying (2.27) we have

$$\mathbb{E}(e^{ik\cdot(S_n-\tilde{S}_m)}) = \varphi(k)^n \overline{\varphi(k)}^m \tag{2.44}$$

and so taking expectations on both sides of (2.43) leads to

$$\mathbb{P}(S_n = \tilde{S}_m) = \int_{[-\pi,\pi]^d} \frac{dk}{(2\pi)^d} \varphi(k)^n \overline{\varphi(k)}^m \tag{2.45}$$

Multiplying by t^{n+m} and summing over $n, m \geq 0$ we get

$$\sum_{m,n\geq 0} t^{m+n} \mathbb{P}(S_n = \tilde{S}_m) = \int_{[-\pi,\pi]^d} \frac{dk}{(2\pi)^d} \frac{1}{|1 - t\varphi(k)|^2}. \tag{2.46}$$

As before, from here the claim follows by taking the limit $t \uparrow 1$. □

The principal conclusion is now as follows:

Theorem 2.30 *The paths of two independent simple symmetric random walks on $\mathbb{Z}^d$ intersect infinitely often with probability one in dimensions $d \leq 4$ and only finitely often with probability one in dimensions $d \geq 5$.*

Proof. Using the same estimates as in Theorem 2.22, we get

$$\mathbb{E}N^{(2)} < \infty \quad \text{if and only if} \quad \int_{|k|<1} \frac{dk}{|k|^4} < \infty. \tag{2.47}$$

The integral is finite if and only if $d \geq 5$. □

Problem 2.31 For what values of $\alpha > 0$ do the paths of independent copies of the walk described in Problem 2.25 intersect infinitely often.

There is a heuristic explanation of the above phenomena: The fact that the random walk is recurrent in $d = 2$, but just barely, means that the path of the walk is *two dimensional*. (This is actually a theorem if we interpret the dimension in the sense of Hausdorff dimension.) Now it is a fact from geometry two generic two-dimensional subspaces of $\mathbb{R}^d$ do not intersects in dimension $d \geq 5$ and they do in dimensions $d \leq 4$. Hence we *should* expect that Theorem 2.30 is true, except perhaps for the subtle boundary case $d = 4$.

Problem 2.32 To verify the above heuristics, let us investigate the intersections of m paths of SRW. Explicitly, let $S^{(1)}, \dots, S^{(m)}$ be m independent SRW and define

$$N^{(m)} = \sum_{\ell_1,\dots,\ell_m \geq 0} 1_{\{S_{\ell_1} = \cdots = S_{\ell_m}\}} \tag{2.48}$$

Find in what dimensions we have $\mathbb{E}N^{(m)} < \infty$.

The non-intersection property of the simple random walk above 4 dimensions plays a crucial role in the understanding of a walk derived from the SRW called the *loop-erased random walk* (LERW). Informally, the LERW is extracted from a finite path of the SRW by sequentially erasing all cycles on the path of the SRW in the order they were created. The main point of doing this is that the resulting LERW has *self-avoiding paths* — i.e., paths that visit each point at most once.

Definition 2.33 [Loop erasure of SRW path] *Let $S_0, S_1, \dots, S_n$ be a finite path of the SRW. Define the sequence of times $T_0 < T_1 < \dots$ by setting $T_0 = 0$ and*

$$T_{k+1} = 1 + \sup\{m \colon T_k \leq m \leq n \ \&\ S_m = S_{T_k}\} \tag{2.49}$$

The loop erasure *of (S_k) is then the sequence (Z_m) where*

$$Z_m = S_{T_m}, \qquad T_m \leq n. \tag{2.50}$$

The subject of the LERW goes way beyond the level and scope of these notes. (Indeed, it has only been proved recently that, in all dimensions, the LERW has a well defined scaling limit which is understood in $d = 2$ — see Fig. 2.5 — and $d \geq 4$, but not in $d = 3$.) However, the analysis of the path-avoiding property of the SRW allows us to catch at least a glimpse of what is going on in dimensions $d \geq 5$.

The key notion to be studied in high dimension is that of a *cut point*. The cleanest way to define this notion is for the *two sided* random walk which is a sequence of random variables $(S_n)_{n\in\mathbb{Z}}$ indexed by (both positive and negative) integers, where S_n is defined by

$$S_n = \begin{cases} X_1 + \cdots + X_n, & n \geq 1, \\ 0, & n = 0, \\ X_{n+1} + \cdots + X_0, & n < 0 \end{cases}$$

where (X_n) is a doubly infinite sequence of independent copies of X_1.

Definition 2.34 [Cut point] *Consider a two sided random walk on $\mathbb{Z}^d$. Then $x \in \mathbb{Z}^d$ is said to be a* cut point *if there exists k such that $S_k = x$ and the paths of one sided walks*

$$S'_n = S_{n+k} \quad \text{and} \quad S''_n = S_{k-n}, \qquad n \geq 0, \tag{2.51}$$

intersect only at time $n = 0$ — the starting point. The time k is then referred to as the cut time *of the random walk (S_n).*

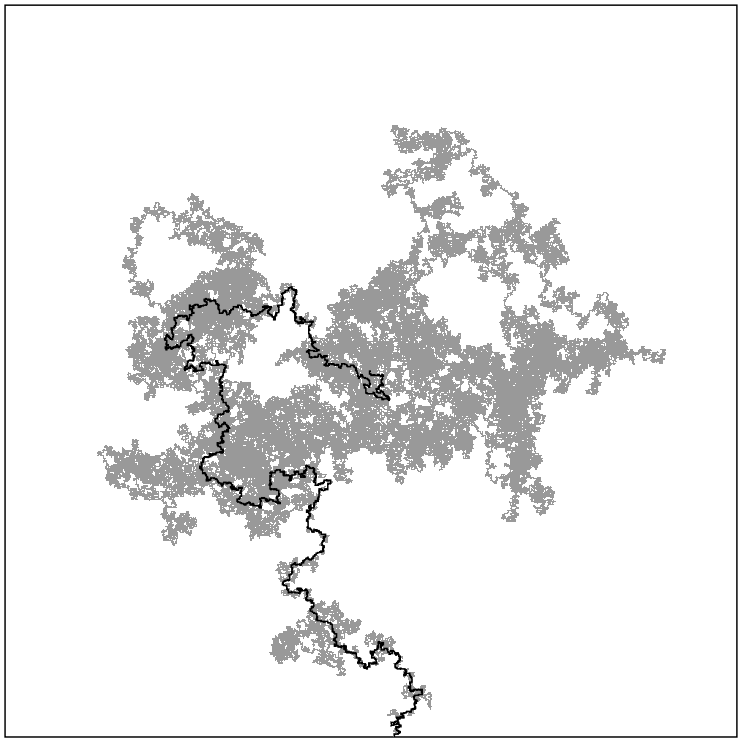

Figure 2.5: A path of the loop erased random walk obtained by loop-erasure from the SRW from Fig. 2.2. The trace of the SRW is depicted in light gray. While the SRW needed 682613 steps to exit the box, its loop erasure took only 3765 steps.

Lemma 2.35 *Consider the two sided random walk (S_n) and let (Z_n) be the loop erasure of the $n \geq 0$ portion of the path. Then the sequence (Z_n) visits all cutpoints (of the two-sided path) on the $n \geq 0$ portion of the path (S_n) in chronological order.*

Proof. The loop erasure removes only vertices on the path that are inside cycles. Cut points are never part of a cycle and so they will never be loop-erased. □

The fact that the SRW and the LERW agree on all cutpoints has profound consequences provided we can control the frequency of occurrence of cutpoints. We state a very weak claim to this extent:

Lemma 2.36 *Let R_n be the number of cut times — in the sense of Definition 2.34 — in the set of times $\{1, \ldots, n\}$. Then*

$$\mathbb{E}R_n = n\mathbb{P}(N^{(2)} = 1). \tag{2.52}$$

In particular, for each $\epsilon \in [0, 1)$,

$$\mathbb{P}(R_n \geq \epsilon n) \geq \frac{\mathbb{P}(N^{(2)} = 1) - \epsilon}{1 - \epsilon}. \tag{2.53}$$

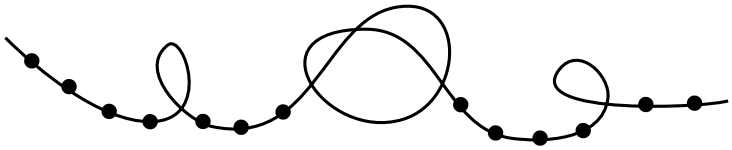

Figure 2.6: A schematic picture of the path of a two sided random walk which, in high dimension, we may think of as a chain of little tangles or knots separated by cutpoints (marked by the bullets).

Proof. We have

$$R_n = \sum_{k=1}^{n} 1_{\{k \text{ is a cut time}\}} \tag{2.54}$$

Taking expectation we get

$$\mathbb{E}R_n = \sum_{k=1}^{n} \mathbb{P}(k \text{ is a cut time}). \tag{2.55}$$

But the path of the two-sided random walk looks the same from every time and so $\mathbb{P}(k \text{ is a cut time})$ equals the probability that 0 is a cut time. That probability in turn equals $\mathbb{P}(N^{(2)} = 1)$. This proves (2.52). To get also (2.53) we note

$$\mathbb{E}R_n \le \epsilon n\big[1 - \mathbb{P}(R_n \ge \epsilon n)\big] + n\mathbb{P}(R_n \ge \epsilon n). \tag{2.56}$$

Then (2.53) follows from (2.52) and some simple algebra. □

Of course having a positive density of points where the SRW and the LERW agree is not sufficient to push the path correspondence through. However, if we can show that the "tangles" between the cutpoints have negligible diameter and that none of them consumes a macroscopic amount of time, then on a large scale the paths of the LERW and the SRW will be hardly distinguishable.

2.4 Harmonic analysis and electric networks

Random walks have a surprising connection to *electric* or, more specifically, *resistor networks*. This connection provides very efficient means to estimate various hitting probabilities and other important characteristics of random walks. The underlying common ground is the subject of *harmonic analysis*.

We begin by a definition of a resistor network:

Definition 2.37 [Resistor network] *A* resistor network *is an unoriented (locally finite) graph $G = (V, E)$ endowed with a collection $(c_{xy})_{(x,y)\in E}$ of positive and finite numbers — called* conductances *— that obey the symmetry*

$$c_{xy} = c_{yx}, \qquad (x, y) \in E, \tag{2.57}$$

and local boundedness

$$\sum_{y \in V} c_{xy} < \infty, \qquad x \in V, \tag{2.58}$$

conditions. The reciprocal values, $r_e = 1/c_e$ are referred to as resistances.

The above definition builds on applications of electric networks in engineering where one often considers circuits with *nodes* and *links*. The links transport *electric current* between the nodes and the resistance of the link characterizes energy dissipation — generally due to heat production — of the link. The nodes, on the other hand, are characterized by the value of the *electric potential*. The currents and voltages are in one-to-one correspondence via Ohm's Law, which we will regard as a definition of the currents:

Definition 2.38 [Ohm's Law] *Suppose $G = (V, E)$ is an resistor network with conductances (c_{xy}). Let $u\colon V \to \mathbb{R}$ be an electric potential at the nodes. Then the* electric current *i_{xy} from x to y is given by*

$$i_{xy} = c_{xy}\big[u(y) - u(x)\big]. \tag{2.59}$$

For ease of exposition, we also introduce the notation $i(x)$ for the total current,

$$i(x) := \sum_{y \in V} i_{xy} \tag{2.60}$$

out of vertex x. There are two basic engineering questions that one may ask about resistor networks:

(1) Suppose the values of the potential u are fixed on a set $A \subset V$. Find the potential at the remaining nodes.

(2) Suppose that we are given the total current $i(x)$ out of the vertices in $A \subset V$. Find the potential at the nodes of V that is consistent with these currents.

The context underlying these questions is sketched in Figs. 2.7 and 2.8.

Of course, the above questions would not come even close to having a unique solution without imposing an additional physical principle:

Definition 2.39 [Kirchhoff's Law of Currents] *We say that a collection of currents (i_{xy}) obeys Kirchhoff's law of currents in the set $W \subset V$ if the total current out of any vertex in W is conserved, i.e.,*

$$i(x) = 0, \qquad x \in W. \tag{2.61}$$

The imposition of Kirchhoff's law has the following simple consequence:

Lemma 2.40 *For a function $f\colon V \to \mathbb{R}$, define $(\mathcal{L}f)\colon V \to \mathbb{R}$ by*

$$(\mathcal{L}f)(x) = \sum_{y \in V} c_{xy}\big[f(y) - f(x)\big] \tag{2.62}$$

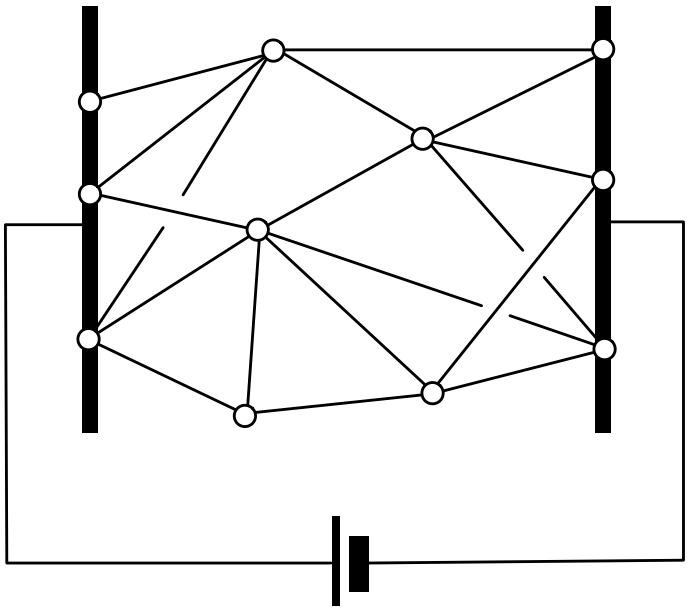

Figure 2.7: A circuit demonstrating the setting in the first electrostatic problem mentioned above. Here vertices on the extreme left and right are placed on conducting plates that, with the help of a battery, keep them at a constant electrostatic potential. The problem is to determine the potential at the "internal" vertices.

Let $W \subset V$ and suppose u is an electric potential for which the currents defined by Ohm's law satisfy Kirchhoff's law of currents in W. Then

$$(\mathcal{L}u)(x) = 0, \qquad x \in W. \tag{2.63}$$

Proof. Using Ohm's Law, the formula for the current out of x becomes

$$i(x) = \sum_{y \in V} i_{xy} = \sum_{y \in V} c_{xy}\big[u(y) - u(x)\big] = (\mathcal{L}u)(x) \tag{2.64}$$

Thus if $i(x) = 0$, then $\mathcal{L}u$ vanishes at x. □

Definition 2.41 [Harmonic function] *We say that $f\colon V \to \mathbb{R}$ is* harmonic in W with respect to $\mathcal{L}$ *if $(\mathcal{L}f)(x) = 0$ for each x in a subset $W \subset V$.*

The object $\mathcal{L}$ is a *linerar operator* in the sense that it operates on a function to get another function and the operation is linear. In Markov chain theory which we will touch upon briefly in the next section, $\mathcal{L}$ is referred to as *generator*.

Note that while the definition of harmonicity of f speaks only about the vertices in W, vertices outside W may get involved due to the non-local nature of $\mathcal{L}$. Harmonic functions are special in that they satisfy the Maximum Principle. Given a set $W \subset V$, we use

$$\partial W = \big\{y \in V \setminus W\colon \exists x \in W \text{ such that } (x, y) \in E\big\} \tag{2.65}$$

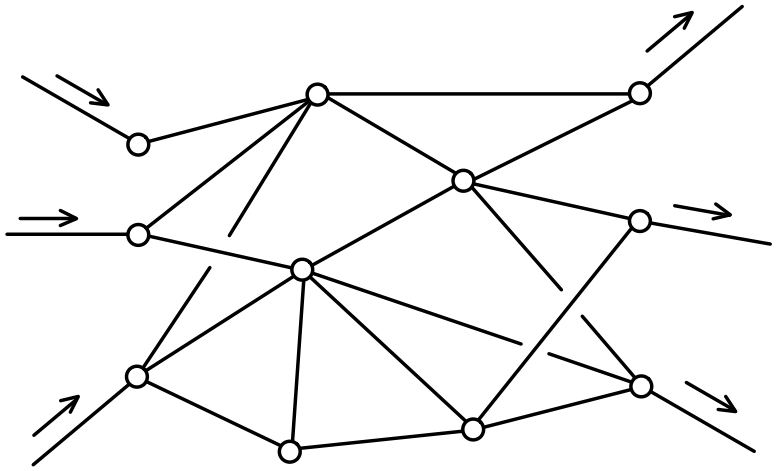

Figure 2.8: A circuit demonstrating the setting in the second electrostatic problem above. The topology of the circuit is as in Fig. 2.7, but now the vertices on the sides have a prescribed current flowing in/out of them. The problem is again to determine the electrostatic potential consistent with these currents.

to denote its outer boundary. Then we have:

Theorem 2.42 [Maximum Principle] *Let $W \subsetneq V$ be finite and connected and suppose $f\colon V \to \mathbb{R}$ is harmonic on W with respect to $\mathcal{L}$. Then*

$$\inf_{y\in\partial W} f(y) \leq \min_{x\in W} f(y) \leq \max_{x\in W} f(x) \leq \sup_{y\in\partial W} f(y), \tag{2.66}$$

In fact, f cannot have a strict local maximum on W and if has a local maximum on W then it is constant on $W \cup \partial W$.

Proof. Let $x \in W$ be a local maximum of f on $W \cup \partial W$. We claim that then

$$f(y) = f(x) \text{ for all } y \text{ with } (x,y) \in E \tag{2.67}$$

Indeed, if $f(y) \leq f(x)$ for all neighbors of x with at least one inequality strict, then

$$\sum_{y\in V} c_{xy} f(x) > \sum_{y\in V} c_{xy} f(y). \tag{2.68}$$

But that is impossible because the difference of the left and right-hand side equals $(\mathcal{L}f)(x)$ which is zero because $x \in W$ and because f is harmonic at x.

Now suppose that the right-inequality on (2.66) does not hold. Then the maximum of f over $W \cup \partial W$ occurs on W. We claim that then f is constant on $W \cup \partial W$. Indeed, if $x \in W \cup \partial W$ were a vertex where f is not equal its maximum but that has a neighbor where it is, then we would run into a contradiction with the first part of the proof by which f must be constant on the neighborhood of any local maxima. Hence, f is constant on $W \cup \partial W$. But then the inequality on the right of

(2.66) *does* hold and so we have a contradiction anyway. The inequality on the left is equivalent to that on the right by passing to $-f$. □

Corollary 2.43 [Rayleigh's Principle] *Let $W \subset V$ be finite and let $u_0 \colon V \setminus W \to \mathbb{R}$ a function that vanishes outside $W \cup \partial W$. Then there exists a unique $u \colon V \to \mathbb{R}$ which is harmonic on W with respect to $\mathcal{L}$ and satisfies*

$$u(x) = u_0(x), \qquad x \in V \setminus W. \tag{2.69}$$

Moreover, this function is the unique minimizer of the Dirichlet energy functional,

$$\mathcal{E}(u) = \frac{1}{2} \sum_{\substack{x,y \in V \\ (x,y) \in E}} c_{xy} \big[u(y) - u(x)\big]^2 \tag{2.70}$$

subject to the condition (2.69).

Proof. First we will establish uniqueness. Suppose u and $\tilde{u}$ are two distinct functions which are harmonic on W with respect to $\mathcal{L}$ and both of which obey (2.69). Then $v = u - \tilde{u}$ is harmonic on W and vanishes on $V \setminus W$. But the Maximum Principle implies

$$0 = \inf_{y \in \partial W} v(y) \leq \min_{x \in W} v(x) \leq \max_{x \in W} v(x) \leq \sup_{y \in \partial W} v(y) = 0 \tag{2.71}$$

and so $v \equiv 0$. It follows that $u \equiv \tilde{u}$.

To see that the desired harmonic function in W exists for each "boundary condition" u_0, we will use the characterization by means of the minimum of $\mathcal{E}(u)$. The function $u \mapsto \mathcal{E}(u)$, regarded as a function of variables $\{u(x) \colon x \in W\}$ — with u_0 substituted for u outside W — is continuous and bounded from below, and so it achieves its minimum. The minimizer is characterized by the vanishing of partial derivatives,

$$\frac{\partial}{\partial u(x)} \mathcal{E}(u) = -2 \sum_{y \in V} c_{xy} \big[u(y) - u(x)\big] = -2(\mathcal{L}u)(x) \tag{2.72}$$

for all $x \in W$. This means that it is harmonic on W. □

The word *energy* appears because $\mathcal{E}(u)$ represents the total electrostatic energy of the current flowing through the network; another name for $\mathcal{E}(u)$ is *Dirichlet form*. An important consequence of the characterization in terms of a minimizer of the Dirichlet energy is the monotonicity in c_{xy}. Indeed, we note:

Lemma 2.44 [Monotonicity] *Let $W \subset V$ be finite or infinite. The (value of the) minimum of $u \mapsto \mathcal{E}(u)$ subject to $u = u_0$ on $V \setminus W$ is non-decreasing in all variables c_{xy}.*

Proof. This is an immediate consequence of the fact that the minimum of a family of non-decreasing functions is non-decreasing. □

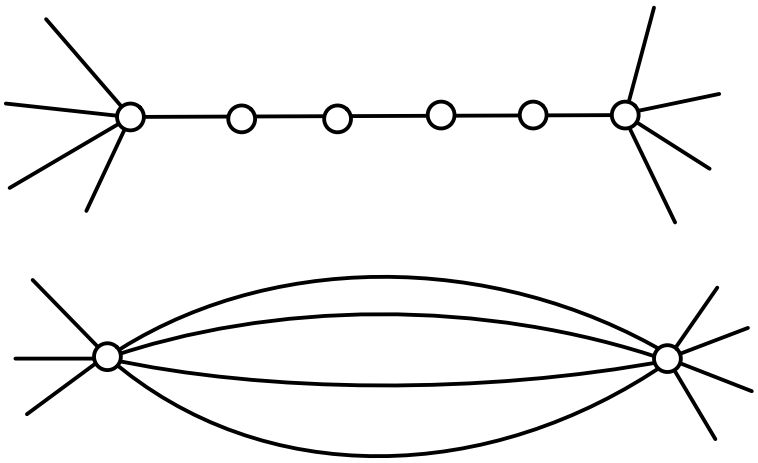

Figure 2.9: The setting for the application of the serial law (top) and parallel law (bottom). In the top picture the sequence of nodes is replaced by a single link whose resistance is the sum of the individual resistances. In the bottom picture, the cluster of parallel links can be replaced by a single link whose conductatance is the sum of the individual conductances.

Let us go back to the two questions we posed above and work them out a little more quantitatively. Suppose A and B are disjoint sets in V and suppose that A is kept at potential $u = 0$ and B at a constant potential $u = U > 0$. A current I will then flow from A to B. Thinking of the whole network as just one resistor, the natural question is what is its *effective resistance* $R_{\text{eff}} = U/I$. A formal definition of this quantity is as follows:

Definition 2.45 [Effective resistance] *Let $A, B \subset V$ be disjoint. The* effective resistance $R_{\text{eff}}(A, B)$ *is a number in* $[0, \infty]$ *defined by*

$$R_{\text{eff}}(A, B)^{-1} = \inf\bigl\{\mathcal{E}(u) \colon 0 \le u \le 1,\ u \equiv 0 \text{ on } A,\ u \equiv 1 \text{ on } B\bigr\}. \tag{2.73}$$

Problem 2.46 Show that adding or removing the condition $0 \le u \le 1$ does not change the value of the infimum.

A consequence of Lemma 2.44 is that the effective resistance $R_{\text{eff}}(A, B)$ increases when the individual resistances r_{xy} are increased.

Exercise 2.47 Show that adding an extra link of positive conductance to the graph G *decreases* the effective resistance $R_{\text{eff}}(A, B)$ between any two disjoint sets A and B.

Problem 2.48 Abusing the notation slightly, we write $R_{\text{eff}}(x, y)$ for $R_{\text{eff}}(\{x\}, \{y\})$. Show that $(x, y) \mapsto R_{\text{eff}}(x, y)$ defines a *metric distance* on G, e.g., a non-negative function which is symmetric,

$$R_{\text{eff}}(x, y) = R_{\text{eff}}(y, x), \qquad x, y \in V, \tag{2.74}$$

and obeys the triangle inequality,

$$R_{\text{eff}}(x, y) \le R_{\text{eff}}(x, z) + R_{\text{eff}}(y, z), \qquad x, y, z \in V. \tag{2.75}$$

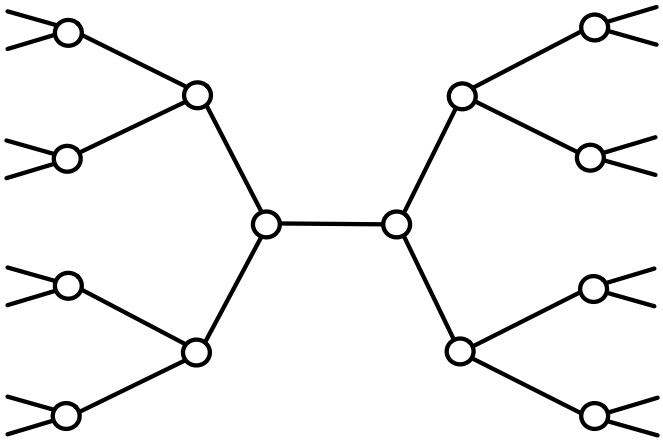

Figure 2.10: A regular ternary tree.

Exercise 2.49 Find the effective resistance $R_{\text{eff}}(x, y)$ between any pair of vertices of the ring $\{0, \dots, N-1\}$ where $N-1$ is considered a neighbor of 0 and all edges have a unit conductance.

Problem 2.50 Consider a two-dimensional torus $\mathbb{T}_N$ of $N \times N$ vertices. Explicitly, $\mathbb{T}_N$ is a graph with vertex set $V = \{0, \dots, N-1\} \times \{0, \dots, N-1\}$ and edges between any pair $(x_1, y_1) \in V$ and $(x_2, y_2) \in V$ with

$$\big(|x_1 - x_2| \bmod N\big) + \big(|y_1 - y_2| \bmod N\big) = 1. \tag{2.76}$$

All edges have a unit conductance. Find the effective resistance $R_{\text{eff}}(x, y)$ between a neighboring pair (x, y) of vertices of $\mathbb{T}_N$. *Hint*: Use discrete Fourier transform.

As it turns out, the most important instance of effective resistance $R_{\text{eff}}(x, y)$ is when one of the points is "at infinity." The precise definition is as follows:

Definition 2.51 [Resistance to infinity] *Consider an infinite resistor network and let B_R be a sequence of balls of radius R centered at a designated point 0. The* resistance $R_{\text{eff}}(x, \infty)$ from x to ∞ *is then defined by the monotone limit*

$$R_{\text{eff}}(x, \infty) = \lim_{R \to \infty} R_{\text{eff}}\big(\{x\}, B_R^{\text{c}}\big). \tag{2.77}$$

Exercise 2.52 Show that the value of $R_{\text{eff}}(x, \infty)$ does not depend on the choice of the designated point 0.

Exercise 2.53 Show that $R_{\text{eff}}(x, \infty) = \infty$ for a network given as an infinite chain of vertices, i.e., $G = \mathbb{Z}$ with the usual nearest neighbor structure.

Apart from monotonicity, the resistor networks have the convenient property that certain parts of the network can be modified without changing effective resistances between sets non-intersecting the modified part. The most well known examples of these are the *parallel* and *serial* laws.

For the sake of stating these laws without annoying provisos, we will temporarily assume that vertices of G may have multiple edges between them. (Each such edge then has its own private conductance.) In graph theory, this means that we allow G to be an unoriented *multigraph*. The parallel and serial laws tell us how to reinterpret such networks back in terms of graphs.

Lemma 2.54 [Serial Law] *Suppose a resistor network contains a sequence of incident edges $e_1, \dots, e_\ell$ of the form $e_j = (x_{j-1}, x_j)$ such that the vertices x_j, $j = 1, \dots, \ell - 1$, are all of degree 2. Then the effective resistance $R_{\text{eff}}(A, B)$ between any sets A, B not containing $x_1, \dots, x_{\ell-1}$ does not change if we replace the edges $e_1, \dots, e_\ell$ by a single edge e with resistance*

$$r_e = r_{e_1} + \cdots + r_{e_\ell} \tag{2.78}$$

Lemma 2.55 [Parallel Law] *Suppose two vertices x, y have multiple edges $e_1, \dots, e_n$ between them. Then the effective resistance $R_{\text{eff}}(A, B)$ between any sets A, B does not change if we replace these by a single edge e with conductance*

$$c_e = c_{e_1} + \cdots + c_{e_n} \tag{2.79}$$

Problem 2.56 Prove the parallel and serial laws.

Problem 2.57 Consider the tree as in Fig. 2.10. Use the aforementioned facts about effective resistance to show that, for any vertex v, we have $R_{\text{eff}}(v, \infty) < \infty$. Compare with Exercise 2.53. *Hint*: First prove the result for the rooted tree.

2.5 Random walks on resistor networks

To demonstrate the utility of resistor networks for the study of random walks, we will now define a *random walk on a resistor network*. Strictly speaking, this will *not* be a random walk in the sense of Definition 2.1 because resistor networks generally do not have any underlying (let alone Euclidean) geometry. However, the definition will fall into the class of *Markov chains* that are natural generalizations of random walks to non-geometric setting.

Definition 2.58 [Random walk on resistor network] *Suppose we have a resistor network — i.e., a connected graph $G = (V, E)$ and a collection of conductances c_e, $e \in E$. A* random walk *on this network is a collection of random variables $Z_0, Z_1, \dots$ such that for all $n \geq 1$ and all $z_1, \dots, z_n \in V$,*

$$\mathbb{P}(Z_1 = z_1, \dots, Z_n = z_n) = \mathsf{P}(z, z_1)\mathsf{P}(z_1, z_2) \cdots \mathsf{P}(z_{n-1}, z_n) \tag{2.80}$$

where $\mathsf{P} \colon V \times V \to [0, 1]$ is a symmetric matrix defined by

$$\mathsf{P}(x, y) = \frac{c_{xy}}{\pi(x)} \quad \text{with} \quad \pi(x) = \sum_{y \in V} c_{xy}. \tag{2.81}$$

We say that the random walk starts at z if

$$\mathbb{P}(Z_0 = z) = 1. \tag{2.82}$$

To mark the initial condition explicitly, we will write $\mathbb{P}^z$ for the distribution of the walks subject to the initial condition (2.82).

Example 2.59 Any *symmetric* random walk on $\mathbb{Z}^d$ is a random walk on the resistor network with nodes $\mathbb{Z}^d$ and an edge between any pair of vertices that can be reached in one step of the random walk. Indeed, if $X_1, X_2, \dots$ denote the steps of the random walk (S_n) with $S_0 = z$, then

$$\mathbb{P}^z(S_1 = z_1, \dots, S_n = z_n) = \mathbb{P}(X_1 = z_1 - z) \cdots \mathbb{P}(X_n = z_n - z_{n-1}). \tag{2.83}$$

To see that this is of the form (2.80–2.81), we define the conductance c_{xy} by

$$c_{xy} = \mathbb{P}(X_1 = y - x) \tag{2.84}$$

and note that symmetry of the step distribution implies $c_{xy} = c_{yx}$ while the normalization gives $\pi(x) = 1$.

The symmetry assumption is crucial for having $\mathsf{P}(x, y)$ of the form (2.81). If one is content with just the *Markov property* (2.80), then any random walk on $\mathbb{Z}^d$ will do. The simplest example of a symmetric random walk is the simple random walk, which just chooses a neighbor at random and passes to it. This "dynamical rule" generalizes to arbitrary graphs:

Example 2.60 *Random walk on a graph*: Consider a locally finite unoriented graph G and let $d(x)$ denote the degree of vertex x. Define

$$c_{xy} = 1, \qquad (x, y) \in E. \tag{2.85}$$

This defines a resistor network; the random walk on this network is often referred to as *random walk on G* because the probability to jump from x to neighbor y is

$$\mathsf{P}(x, y) = \frac{1}{d(x)}, \qquad (x, y) \in E, \tag{2.86}$$

which corresponds to choosing a neighbor at random. In this case $\pi(x) = d(x)$.

We proceed by a characterization of the electrostatic problems for resistor networks by means of the random walk $Z_0, Z_1, \dots$ on the network:

Lemma 2.61 [Dirichlet problem in finite domain] *Let $W \subset V$ be a finite subset of the resistor network and let*

$$\tau_{W^c} = \inf\{n \geq 0 \colon Z_n \in W^c\} \tag{2.87}$$

be the first time the walk (Z_n) visits W^c. Then

$$\mathbb{P}^z(\tau_{W^c} < \infty) = 1, \qquad z \in V. \tag{2.88}$$

In addition, if $u_0 \colon W^c \to \mathbb{R}$ is a bounded function then

$$u(x) = \mathbb{E}^x\big(u_0(Z_{\tau_{W^c}})\big), \qquad x \in V, \tag{2.89}$$

defines the unique function $u \colon V \to \mathbb{R}$ that is harmonic on W with respect to $\mathcal{L}$ defined in (2.62) *and that coincides with u_0 on W^c.*

Problem 2.62 Prove (2.88) by showing that $\mathbb{P}^z(\tau_{W^c} > n) \le \mathrm{e}^{-\delta n}$ for some $\delta > 0$. *Hint*: Show that $\epsilon := \min_{z\in W} \mathbb{P}^z(\tau_{W^c} \le k) > 0$ when k is the diameter of W. Then iterate along multiples of k to prove that $\mathbb{P}^z(\tau_{W^c} > nk) \le (1-\epsilon)^n$.

Proof of Lemma 2.61. Let u be given by (2.89). We clearly have $\mathbb{P}^z(\tau_{W^c} = 0) = 1$ when $z \in W^c$ and so $u = u_0$ outside W. Since there is only one harmonic function for each boundary condition u_0, we just need to show that u defined above is harmonic on W with respect to $\mathcal{L}$. This is quite easy: If the walk started at $x \in W$, then $\tau_{W^c} \ge 1$. Fixing explicitly the value of Z_1 yields

$$u(x) = \sum_{y\in V} \mathbb{E}^x\big(u_0(Z_{\tau_{W^c}})1_{\{Z_1=y\}}\big). \tag{2.90}$$

Since the probability of each path factors into a product (2.80), we have

$$\mathbb{E}^x\big(u_0(Z_{\tau_{W^c}})1_{\{Z_1=y\}}\big) = \mathsf{P}(x,y)\mathbb{E}^y\big(u_0(Z_{\tau_{W^c}})\big) = \mathsf{P}(x,y)u(y) \tag{2.91}$$

and so

$$u(x) = \sum_{y\in V} \mathsf{P}(x,y)u(y). \tag{2.92}$$

In explicit terms,

$$\pi(x)u(x) = \sum_{y\in V} c_{xy}u(y). \tag{2.93}$$

But $\pi(x)$ is the sum of c_{xy} over all y and so we can write the difference of the right and left-hand side as $(\mathcal{L}u)(x) = 0$. □

The probabilistic interpretation of the solution allows us to rewrite the formula for effective resistance as follows:

Lemma 2.63 *Let $W \subset V$ be a finite set and let $x \in W$. Let T_x denote the first return time of the walk (Z_n) to x,*

$$T_x = \inf\{n \ge 1 \colon Z_n = x\}. \tag{2.94}$$

Then

$$R_{\mathrm{eff}}\big(\{x\}, W^c\big)^{-1} = \pi(x)\,\mathbb{P}^x\big(T_x \ge \tau_{W^c}\big). \tag{2.95}$$

Proof. Consider the function

$$u(z) = \begin{cases} 1, & z = x, \\ \mathbb{P}^z\big(T_x < \tau_{W^c}\big), & x \in W \setminus \{x\}, \\ 0, & x \in W^c. \end{cases} \tag{2.96}$$

Then u is a solution to the Dirichlet problem in $W \setminus \{x\}$ with boundary condition $u = 1$ on $\{x\}$ and 0 on W^c. In particular,

$$R_{\mathrm{eff}}\big(\{x\}, W^c\big)^{-1} = \mathcal{E}(u). \tag{2.97}$$

We thus have to show that $\mathcal{E}(u)$ equals the RHS of (2.95). For that we insert

$$\big[u(y)-u(z)\big]^2 = u(y)\big[u(y)-u(z)\big] + u(z)\big[u(z)-u(y)\big] \tag{2.98}$$

into the definition of $\mathcal{E}(u)$ to get

$$\mathcal{E}(u) = \sum_{z\in V} u(z) \sum_{y\in V} c_{yz}\big[u(z)-u(y)\big] = -\sum_{z\in V} u(z)(\mathcal{L}u)(z), \tag{2.99}$$

where we used that $c_{yz} = c_{zy}$ to write the contribution of each term on the right of (2.98) using the same expression. But $\mathcal{L}u(z) = 0$ for $z \in W \setminus \{x\}$ and $u(z) = 0$ for $z \in W^{\mathrm{c}}$. At $z = x$ we have $u(z) = 1$ and

$$\begin{aligned} -(\mathcal{L}u)(x) &= \sum_{y\in V} c_{xy}\big[1-\mathbb{P}^y(T_x < \tau_{W^{\mathrm{c}}})\big] \\ &= \pi(x) - \pi(x)\sum_{y\in V} P(x,y)\mathbb{P}^y(T_x < \tau_{W^{\mathrm{c}}}) \\ &= \pi(x) - \pi(x)\mathbb{P}^x(T_x < \tau_{W^{\mathrm{c}}}) = \pi(x)\mathbb{P}^x(T_x \geq \tau_{W^{\mathrm{c}}}) \end{aligned} \tag{2.100}$$

Plugging this in (2.99) yields $\mathcal{E}(u) = \pi(x)\mathbb{P}^x(T_x \geq \tau_{W^{\mathrm{c}}})$. □

Theorem 2.64 [Effective resistance and recurrence vs transience] *Recall the notation* (2.94) *for* T_x. *Then*

$$R_{\mathrm{eff}}(x,\infty) = \infty \quad \Leftrightarrow \quad \mathbb{P}^x(T_x < \infty) = 1 \tag{2.101}$$

and

$$R_{\mathrm{eff}}(x,\infty) < \infty \quad \Leftrightarrow \quad \mathbb{P}^x(T_x = \infty) > 0. \tag{2.102}$$

Proof. It clearly suffices to prove only (2.102). By Lemma 2.63 and Exercise 2.52, for the sequence of balls B_R of radius R centered at any designated point,

$$\pi(x)\lim_{R\to\infty} \mathbb{P}^x(T_x > \tau_{B_R^{\mathrm{c}}}) = R_{\mathrm{eff}}(x,\infty)^{-1}. \tag{2.103}$$

But $\tau_{B_R^{\mathrm{c}}} \geq R$ and so

$$\lim_{R\to\infty} \mathbb{P}^x(T_x > \tau_{B_R^{\mathrm{c}}}) \leq \lim_{R\to\infty} \mathbb{P}^x(T_x \geq R) = \mathbb{P}^x(T_x = \infty) \tag{2.104}$$

On the other hand,

$$\mathbb{P}^x(T_x > \tau_{B_R^{\mathrm{c}}}) \geq \mathbb{P}^x(T_x = \infty) \tag{2.105}$$

and so

$$\pi(x)\mathbb{P}^x(T_x = \infty) = R_{\mathrm{eff}}(x,\infty)^{-1}. \tag{2.106}$$

As $\pi(x) > 0$, the claim now directly follows. □

Exercise 2.65 Use Exercise 2.53 and Problem 2.57 to show that the random walk on $\mathbb{Z}$ is recurrent and that on a regular ternary tree is transient.

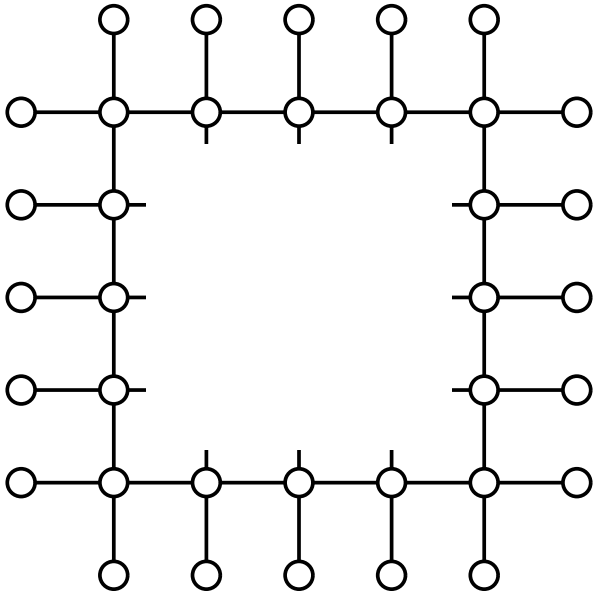

Figure 2.11: The setting for the proof of the Nash-Williams estimate. Only the edges between $B_n(0)$ and its complement are fully drawn. By setting the potential on the boundary vertices of $B_n(0)$ to a constant, these vertices are effectively fused into one. The resistivity between the fused vertex and $B_n(0)^c$ is $|\partial B_n|^{-1}$.

Corollary 2.66 [Monotonicity of recurrence/transience] *Suppose that the random walk (Z_n) on a resistor network with conductances (c_{xy}) is recurrent. Then so is the random walk on the network with conductances $(\tilde{c}_{xy})$ where $\tilde{c}_{xy} \leq c_{xy}$ for all $(x,y) \in E$.*

Similarly, if the random walk is transient for conductances (c_{xy}) then it is transient also for any conductances $(\tilde{c}_{xy})$ such that $\tilde{c}_{xy} \geq c_{xy}$ for all $(x,y) \in E$.

Proof. This follows because the effective resistance to infinity, $R_{\text{eff}}(x,\infty)$, is a decreasing function of the conductances. □

Exercise 2.67 Show that there exists a critical dimension $d_c \in \mathbb{N} \cup \{\infty\}$ such that the d-dimensional SRW is recurrent in dimensions $d \leq d_c$ and transient in $d > d_c$.

Problem 2.68 Show that a removal of a single edge from $\mathbb{Z}^d$ does not change it recurrence/transience properties.

To demonstrate the usefulness of the above technology let us provide a completely different proof of the fact that the SRW in $d = 1,2$ is recurrent. The key step is provided by the following lemma:

Lemma 2.69 [Nash-Williams estimate] *For electric network $G = (V,E)$ with conductances $c_{xy} = 1$, abbreviate $R_n = R_{\text{eff}}(0, B_n(0)^{-1})$. Then*

$$R_{n+1} \geq R_n + \frac{1}{|\partial B_n|} \tag{2.107}$$

where

$$|\partial B_n| = \#\big\{(x,y) \in E\colon x \in B_n(0),\, y \in B_n(0)^{-1}\big\}. \tag{2.108}$$

Proof. By (2.73), we get an upper bound on R_{n+1}^{-1} by setting u to a constant on the outer vertices of $B_n(0)$. This effectively means we fuse all vertices on the outer boundary of B_n into one, see Fig. 2.11. As there are $|\partial B_n|$ edges between $B_n(0)$ and its complement, the parallel law sets the resistance between the fused vertex and $B_n(0)^{-1}$ to $1/|\partial B_n|$. The bound (2.107) then follows by the serial law. □

Corollary 2.70 *The SRW is recurrent in dimensions $d = 1, 2$.*

Proof. The SRW can be thought of as a random walk on the electrical network with graph structure $\mathbb{Z}^d$ and a unit conductance on each edge. In the notation of Lemma 2.69 we have $|\partial B_n| \le 2dn^{d-1}$ and so

$$R_n \ge \sum_{k=1}^{n-1} \frac{1}{|\partial B_k|} \ge \frac{1}{2d} \sum_{k=1}^{n-1} k^{1-d}. \tag{2.109}$$

The right hand side grows proportional to n in $d = 1$ and $\log n$ in $d = 2$ and so $R_{\text{eff}}(0,\infty) = \lim_{n\to\infty} R_n = \infty$ in these dimensions. By Theorem 2.64, the SRW is recurrent in $d = 1, 2$. □

Chapter 3

Branching processes

In this chapter we begin studying another classical process considered in probability theory: *branching*. The motivation comes from attempts to understand the dynamics of genealogical trees — as was the case for Galton and Watson who invented branching processes — but the real interest comes from applications that reach beyond this limited, and to most people little appealing, context.

3.1 Galton-Watson branching process

A branching process is uniquely determined by its *offspring distribution* which is a sequence $\mathfrak{p}_n$ of non-negative numbers such that

$$\sum_{n\geq 0} \mathfrak{p}_n = 1. \tag{3.1}$$

An informal definition of a branching process is as follows: At each time, the process has a certain number of living individuals. To get the next generation, each of the living individuals *splits* into a random number of offspring, and then *dies*. The number of offspring is sampled from the offspring distribution independently of all other neighbors. A formal definition is as follows:

Definition 3.1 [Galton-Watson branching process] *Let $(\xi_{n,m})_{m,n\geq 1}$ be i.i.d. integer-valued random variables with $\mathbb{P}(\xi_{n,m} = k) = \mathfrak{p}_k$. Define the sequence of random variables (X_n) by setting*

$$X_0 = 1 \tag{3.2}$$

and solving recursively

$$X_{n+1} = \begin{cases} \xi_{n+1,1} + \cdots + \xi_{n+1,X_n}, & \text{if } X_n > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

A Galton-Watson branching process *with offspring distribution $(\mathfrak{p}_n)$ is a sequence of random variables that has the same law as (X_n).*

The second line in the formula for X_{n+1} shows that once the sequence (X_n) hits zero — i.e., once the family has died out — it will be zero forever. We refer to this situation *extinction*; the opposite case is referred to as *survival*. The first goal is to characterize offspring distributions for which extinction occurs with probability one — or, complementarily, survival occurs with a positive probability.

Consider the moment-generating function

$$\phi_n(s) = \mathbb{E}(\mathrm{e}^{-sX_n}), \qquad s \geq 0. \tag{3.4}$$

The reason for looking at this function is that if $\mathbb{P}(X_n \geq 1)$ is bounded uniformly away from zero, say, by $\epsilon > 0$, then we have $\phi(s) \leq 1 - \epsilon \mathrm{e}^{-s}$. On the other hand, if $\mathbb{P}(X_n \geq 1)$ tends to zero, we clearly get $\phi(s) \to 1$ for all $s \geq 0$.

Lemma 3.2 *For each $s \geq 0$, let*

$$\lambda(s) := -\log \mathbb{E}(\mathrm{e}^{-s\xi_{1,1}}) = -\log\Big(\sum_{n\geq 0} \mathrm{e}^{-sn}\mathfrak{p}_n\Big). \tag{3.5}$$

Then for all $n \geq 0$,

$$\phi_{n+1}(s) = \phi_n\big(\lambda(s)\big). \tag{3.6}$$

In particular,

$$\phi_n(s) = \exp\big\{-\lambda_n(s)\big\} \tag{3.7}$$

where λ_n denotes the n-fold iteration $\lambda \circ \cdots \circ \lambda$ of function λ.

Proof. First we derive (3.6). By the definition of X_{n+1},

$$\phi_{n+1}(s) = \mathbb{E}\big(\mathrm{e}^{-s(\xi_{n+1,1}+\cdots+\xi_{n+1,X_n})}\big) \tag{3.8}$$

where we interpret the sum in the exponent as zero when $X_n = 0$. Conditioning on X_n and using the independence of $\xi_{n+1,j}$'s of themselves as well as X_n, we get

$$\begin{aligned}
\mathbb{E}\big(\mathrm{e}^{-s(\xi_{n+1,1}+\cdots+\xi_{n+1,X_n})}\big) &= \sum_{k\geq 0} \mathbb{E}\big(\mathrm{e}^{-s(\xi_{n+1,1}+\cdots+\xi_{n+1,k})}\mathbf{1}_{\{X_n=k\}}\big) \\
&= \sum_{k\geq 0} \mathbb{E}\big(\mathrm{e}^{-s(\xi_{n+1,1}+\cdots+\xi_{n+1,k})}\big)\mathbb{P}(X_n = k) \\
&= \sum_{k\geq 0} \big(\mathbb{E}(\mathrm{e}^{-s\xi_{n+1,1}})\big)^k \mathbb{P}(X_n = k) \\
&= \sum_{k\geq 0} \mathrm{e}^{-\lambda(s)k}\,\mathbb{P}(X_n = k) = \mathbb{E}\big(\mathrm{e}^{-\lambda(s)X_n}\big)
\end{aligned} \tag{3.9}$$

The right-hand side is the moment-generating function of X_n at the point $\lambda(s)$. This is the content of (3.6). Note that $\lambda(s) \geq 0$ once $s \geq 0$ and so there is no problem with using $\lambda(s)$ as an argument of ϕ_n.

To derive the explicit formula for ϕ_n we first solve recursively for ϕ_n to get

$$\phi_n(s) = \phi_0\big(\lambda \circ \cdots \circ \lambda(s)\big) \tag{3.10}$$

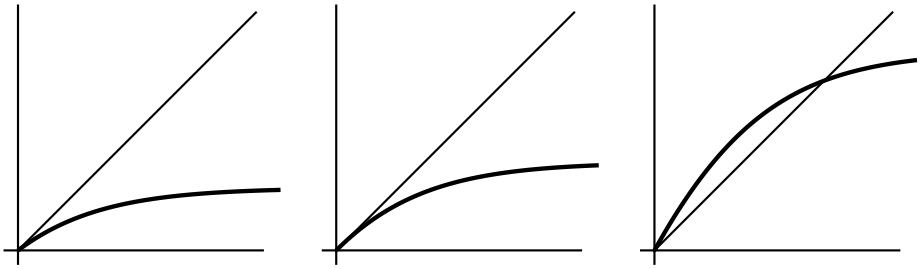

Figure 3.1: The plot of $s \mapsto \lambda(s)$ for offspring distribution with $\mathfrak{p}_0 = 1 - p$ and $\mathfrak{p}_3 = p$ for p taking values 0.25, 0.33 and 0.6, respectively. These values represent the three generic regimes distinguished — going left to right — by whether $\lambda'(0^+)$ is less than one, equal to one, or larger than one. As λ is strictly concave, non-negative with $\lambda(0) = 0$, only in the last case λ has a strictly positive fixed point. We refer to the three situations as subcritical, critical and supercritical.

(Again, we are using that λ maps $(0,\infty)$ into $(0,\infty)$ and so the iterated map is well defined.) From here (3.7) follows by noting that $\phi_0(s) = \mathrm{e}^{-s}$ due to (3.2). □

In light of (3.7), the question whether $\phi_n(s) \to 1$ or not now boils to the question whether $\lambda_n(s) \to 0$ or not. To find for the right criterion, we will need to characterize the analytic properties of λ:

Lemma 3.3 *Suppose that $\mathfrak{p}_n < 1$ for all n. Then λ is non-decreasing and continuous on $[0,\infty)$ and strictly concave and differentiable on $(0,\infty)$. In addition,*

$$\lim_{s\downarrow 0} \lambda'(s) = \sum_{n\geq 0} n\mathfrak{p}_n \tag{3.11}$$

and

$$\lim_{s\to\infty} \lambda'(s) = \inf\{n\colon \mathfrak{p}_n > 0\}. \tag{3.12}$$

Proof. Since $\mathfrak{p}_n \leq 1$, the series $\sum_{n\geq 0} \mathrm{e}^{-sn}\mathfrak{p}_n$ is absolutely summable locally uniformly on $s > 0$ and so it can be differentiated term-by-term. In particular, the first derivative $\lambda(s)$ is the expectation

$$\lambda'(s) = \sum_{n\geq 0} n\mathfrak{p}_n\, \mathrm{e}^{-ns+\lambda(s)} \tag{3.13}$$

for the probability mass function $n \mapsto \mathfrak{p}_n \mathrm{e}^{-sn}\mathrm{e}^{\lambda(s)}$ on $\mathbb{N}\cup\{0\}$, while the second derivative $\lambda''(s)$ is the negative of the corresponding variance. Under the condition $\mathfrak{p}_n < 1$ for all n, the variance is non-zero and so $\lambda''(s) < 0$ for all $s > 0$. This establishes differentiability and strict concavity $(0,\infty)$; continuity at $s = 0$ is directly checked. The limit of the derivatives (3.11) exists by concavity and equals the corresponding limit of (3.13). To prove (3.12), let k be the infimum on the right hand side. Then $\mathrm{e}^{-\lambda(s)} \sim \mathfrak{p}_k \mathrm{e}^{-sk}$ and so all terms but the k-th in (3.13) disappear in the limit $s \to \infty$. The k-th term converges to k and so (3.12) holds. □

Exercise 3.4 Compute $\lambda''(s)$ explicitly and show that it can be written as the negative of a variance. Use this to show that $\lambda''(s) < 0$ for all $s > 0$ once at least two of the $\mathfrak{p}_n$'s are non-zero.

The statements in the lemma imply that the function looks as in Fig. 3.1.

Exercise 3.5 Consider branching process with offspring distribution determined by $p_0 = p$ and $p_2 = 1 - p$ with $0 < p < 1$. Sketch the graph of $\lambda(s)$ and characterize the regime when λ has a non-zero fixed point.

We can now use these properties to control the iterations of λ:

Theorem 3.6 [Survival/extinction criteria] *Suppose* $0 < \mathfrak{p}_0 < 1$ *and let*

$$\mu = \sum_{n \geq 0} n\mathfrak{p}_n \in (0, \infty] \tag{3.14}$$

denote the mean offspring. Then we have:

(1) If $\mu \leq 1$ *then the branching process dies out,*

$$\mathbb{P}(X_n \geq 1) \underset{n\to\infty}{\longrightarrow} 0. \tag{3.15}$$

(2) If $\mu > 1$ *then the process dies out with probability* $\mathrm{e}^{-s_\star}$*,*

$$\mathbb{P}(X_n = 0) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-s_\star}, \tag{3.16}$$

where $s_\star$ *is the unique positive solution to* $\lambda(s) = s$*, and it survives forever and, in fact, the population size goes to infinity with complementary probability,*

$$\mathbb{P}(X_n \geq M) \underset{n\to\infty}{\longrightarrow} 1 - \mathrm{e}^{-s_\star} \tag{3.17}$$

for every $M \geq 1$*.*

Proof. Suppose first that $\mu \leq 1$. The strict convexity and the fact that $\lambda'(s) \to \mu$ as $s \downarrow 0$ ensure that $\lambda(s) < s$ for all $s > 0$. This means $\lambda_{n+1}(s) = \lambda(\lambda_n(s)) < \lambda_n(s)$, i.e., the sequence $n \mapsto \lambda_n(s)$ is strictly decreasing. Thus for each $s \geq 0$ the limit

$$r(s) = \lim_{n\to\infty} \lambda_n(s) \tag{3.18}$$

exists. But $\lambda_{n+1}(s) = \lambda(\lambda_n(s))$ and the continuity of λ imply

$$\lambda(r(s)) = r(s) \tag{3.19}$$

As $\lambda(s) < s$ for $s > 0$, the only point that satisfies this equality on $[0, \infty)$ is $r(s) = 0$. We conclude that $\lambda_n(s) \to 0$ yielding

$$\lim_{n\to\infty} \phi_n(s) = 1, \qquad s \geq 0. \tag{3.20}$$

From

$$\phi_n(s) \leq 1 - (1 - \mathrm{e}^{-s})\mathbb{P}(X_n \geq 1) \tag{3.21}$$

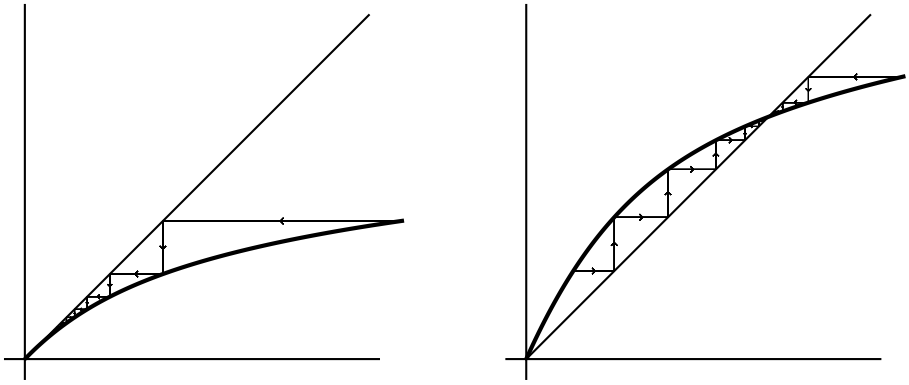

Figure 3.2: A graphical construction underlying the iterations $\lambda_n = \lambda \circ \cdots \circ \lambda$ in the proof of Theorem 3.6. For the subcritical process (left) $\lambda_n(s)$ converges to zero for any s. For the supercritical process (right), λ_n increases for $s < s_\star$ and decreases for $s < s_\star$ where $s_\star$ is the positive fixed point of λ.

it follows that

$$\mathbb{P}(X_n \geq 1) \underset{n\to\infty}{\longrightarrow} 0 \tag{3.22}$$

as we desired to prove.

Next let us assume $\mu > 1$. Then we have $\lambda(s) > s$ for s sufficiently small. On the other hand, due to $\mathfrak{p}_0 > 0$ and (3.12), we have $\lambda(s) < s$ once s is large. Thus there exists a non-zero solution to $\lambda(s) = s$. Strict convexity of λ implies that this solution is actually unique because if $s_\star$ is the least such positive solution then $\lambda'(s_\star) \leq 1$ and strict concavity tell us that $\lambda(s) < s$ for $0 < s < s_\star$ and $\lambda(s) > s$ for $s > s_\star$.

We claim that

$$\lambda_n(s) \underset{n\to\infty}{\longrightarrow} s_\star, \qquad s > 0. \tag{3.23}$$

This is proved in the same way as for $\mu \leq 1$ but we have to deal with $s > s_\star$ and $s < s_\star$ separately.

For $s > s_\star$ we have $s_\star < \lambda(s) < s$ and so the sequence $\lambda_n(s)$ is *decreasing*. The limit is again a fixed point of $\lambda(s) = s$ and since $s_\star$ is the only one available, we have $\lambda_n(s) \to s_\star$.

For $0 < s < s_\star$ we instead have $s < \lambda(s) < s_\star$, where we used that $\lambda(s)$ is increasing to get the second inequality. It follows that $\lambda_n(s)$ is *increasing* in this regime. The limit is a fixed point of λ and so $\lambda_n(s) \to s_\star$ in this case as well.

Having established (3.23), we now note that this implies

$$\phi_n(s) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-s_\star} \tag{3.24}$$

Since this holds for s arbitrary large, and since

$$\mathbb{P}(X_n = 0) \leq \phi_n(s) = \mathbb{P}(X_n = 0) + \mathrm{e}^{-s} \tag{3.25}$$

we must have

$$\mathbb{P}(X_n = 0) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-s_\star}. \tag{3.26}$$

Now fix $M \geq 1$ and suppose that $\epsilon := \liminf_{n\to\infty} \mathbb{P}(X_n = M) > 0$. Then

$$\phi_n(s) \geq \mathbb{P}(X_n = 0) + \mathrm{e}^{-sM}\mathbb{P}(X_n = M) \tag{3.27}$$

would imply $\liminf_{n\to\infty} \phi_n(s) \geq \mathrm{e}^{-s_\star} + \epsilon\mathrm{e}^{-sM} > \mathrm{e}^{-s_\star}$, a contradiction. Hence, $\mathbb{P}(X_n = M)$ tends to zero for any finite M, proving (3.17). □

Exercise 3.7 Suppose $\mathfrak{p}_0 = 0$ and $\mathfrak{p}_1 < 1$. Show that $X_n \to \infty$ with probability one.

Problem 3.8 Suppose $\mu < 1$. Show that $\mathbb{P}(X_n \geq 1)$ decays exponentially with n.

3.2 Critical process & duality

The analysis in the previous section revealed the following picture: Branching processes undergo an abrupt change of behavior when μ increases through one. This is a manifestation of what physicists call a *phase transition*. The goal of this section is to investigate the situation at the *critical point*, i.e., for generic branching processes with mean-offspring $\mu = 1$. Here is the desired result:

Theorem 3.9 *Suppose $\mu := \mathbb{E}\xi = 1$ and $\sigma^2 := \mathrm{Var}(\xi) \in (0, \infty)$. Then*

$$\mathbb{P}(X_n \geq 1) = \frac{2}{\sigma^2}\,\frac{1}{n}\,\big(1 + o(1)\big), \qquad n \to \infty. \tag{3.28}$$

Moreover,

$$\mathbb{P}\Big(\frac{X_n}{n} \geq z \,\Big|\, X_n \geq 1\Big) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-z(2/\sigma^2)}, \tag{3.29}$$

i.e., conditional on survival the law of X_n/n converges to that of an exponential random variable with parameter $2/\sigma^2$.

The key to the proof are two lemmas:

Lemma 3.10 *For any $s > 0$,*

$$\frac{1 - \phi_n(s)}{1 - \mathrm{e}^{-s}} = \sum_{k\geq 1} \mathrm{e}^{-s(k-1)}\mathbb{P}(X_n \geq k). \tag{3.30}$$

Proof. Abbreviate $a_k = \mathbb{P}(X_n \geq k)$. Then

$$\phi_n(s) = \sum_{k\geq 0} \mathrm{e}^{-ks}\,(a_k - a_{k+1}). \tag{3.31}$$

Rearranging the sum, we get

$$\phi_n(s) = a_0 + \sum_{k\geq 1} (\mathrm{e}^{-ks} - \mathrm{e}^{-(k-1)s})a_k. \tag{3.32}$$

Since $a_0 = 1$, subtracting both sides from one gives us

$$1 - \phi_n(s) = (1 - \mathrm{e}^{-s}) \sum_{k \geq 1} \mathrm{e}^{-s(k-1)} a_k \tag{3.33}$$

This is exactly (3.30). □

Lemma 3.11 *For any fixed $s > 0$,*

$$\frac{1}{\lambda_n(s)} = \frac{\sigma^2}{2}\, n \,\big(1 + o(1)\big), \qquad n \to \infty. \tag{3.34}$$

Moreover, if $\theta > 0$ is fixed, then also

$$\frac{1}{\lambda_n(\theta/n)} = \Big(\frac{1}{\theta} + \frac{\sigma^2}{2}\Big)\, n \,\big(1 + o(1)\big), \qquad n \to \infty. \tag{3.35}$$

Proof. First, let us get the intuitive idea for the appearance of $1/n$ scaling for λ_n. We know that $\lambda_n(s) \to 0$ for any $s \geq 0$. Since $\lambda_{n+1}(s) = \lambda(\lambda_n(s))$, we can thus expand λ about $s = 0$ to get a simpler recurrence relation. A computation shows

$$\lambda(s) = \mu\, s - \frac{\sigma^2}{2} s^2 + o(s^2) \tag{3.36}$$

and since $\mu = 1$, we thus have

$$\lambda_{n+1}(s) = \lambda_n(s) - \frac{\sigma^2}{2} \lambda_n(s)^2 \big(1 + o(1)\big). \tag{3.37}$$

This is solved to the leading order by setting $\lambda_n(s) = c/n$ which gets us $c = 2/\sigma^2$. We actually do not have to work much harder in order to get the above calculation under control. First, the existence of the first two derivatives of λ tells us that, for each $\epsilon > 0$, we can find $s_0(\epsilon) > 0$ such that

$$\frac{1}{s} + (1 - \epsilon)\frac{\sigma^2}{2} \leq \frac{1}{\lambda(s)} \leq \frac{1}{s} + (1 + \epsilon)\frac{\sigma^2}{2}, \qquad 0 < s < s_0. \tag{3.38}$$

Thus, we have

$$\frac{1}{\lambda_n(s)} + (1 - \epsilon)\frac{\sigma^2}{2} \leq \frac{1}{\lambda_{n+1}(s)} \leq \frac{1}{\lambda_n(s)} + (1 + \epsilon)\frac{\sigma^2}{2} \tag{3.39}$$

whenever $\lambda_n(s) < s_0$. Define $n_0 = n_0(s)$ be the first n for which this is true for iterations started from s, i.e.,

$$n_0 = \sup\big\{n \geq 0 \colon \lambda_n(s) \geq s_0(\epsilon)\big\}. \tag{3.40}$$

Since $s_0(\epsilon) \leq \lambda_{n_0}(s) \leq \lambda(s_0(\epsilon))$, summing the inequalites from n_0 on we get

$$\frac{1}{\lambda(s_0(\epsilon))} + (1 - \epsilon)\frac{\sigma^2}{2}(n - n_0) \leq \frac{1}{\lambda_n(s)} \leq \frac{1}{s_0(\epsilon)} + (1 + \epsilon)\frac{\sigma^2}{2}(n - n_0) \tag{3.41}$$

From here the limit (3.34) follows by taking $n \to \infty$ followed by $\epsilon \downarrow 0$.
To prove also (3.35), we now assume that ${}^{\theta}/_{n} < s_0(\epsilon)$. Then (3.39) tells us

$$\frac{n}{\theta} + (1-\epsilon)\frac{\sigma^2}{2}k \leq \frac{1}{\lambda_k({}^{\theta}/_{n})} \leq \frac{n}{\theta} + (1+\epsilon)\frac{\sigma^2}{2}k \tag{3.42}$$

Setting $k = n$, we get (3.35). □

Proof of Theorem 3.9. First we prove (3.28). Fix $s > 0$. By (3.34) we then have

$$1 - \phi_n(s) = \frac{2/\sigma^2}{n}\big(1 + o(1)\big), \qquad n \to \infty. \tag{3.43}$$

The bounds $0 \leq \mathbb{P}(X_n \geq k) \leq \mathbb{P}(X_n \geq 1)$ and Lemma 3.10 imply

$$\mathbb{P}(X_n \geq 1) \leq \frac{1 - \phi_n(s)}{1 - \mathrm{e}^{-s}} \leq \frac{\mathbb{P}(X_n \geq 1)}{1 - \mathrm{e}^{-s}}. \tag{3.44}$$

From here (3.28) follows by taking $n \to \infty$ and $s \to \infty$.
Next we plug $s = {}^{\theta}/_{n}$ into the left-hand side of (3.30) and apply (3.35) to get

$$\frac{1 - \phi_n({}^{\theta}/_{n})}{1 - \mathrm{e}^{-\theta/n}} = \frac{{}^{1}/_{\theta}}{{}^{1}/_{\theta} + {}^{\sigma^2}/_{2}} + o(1), \qquad n \to \infty. \tag{3.45}$$

The identity

$$\mathbb{P}(X_n \geq k) = \mathbb{P}(X_n \geq 1)\,\mathbb{P}(X_n \geq k | X_n \geq 1) \tag{3.46}$$

turns (3.30) into

$$\frac{1 - \phi_n({}^{\theta}/_{n})}{1 - \mathrm{e}^{-\theta/n}} = n\mathbb{P}(X_n \geq 1)\mathrm{e}^{\theta/n}\,\frac{1}{n}\sum_{k\geq 1}\mathrm{e}^{-\theta k/n}\mathbb{P}(X_n \geq k | X_n \geq 1) \tag{3.47}$$

and applying (3.28) and (3.45) then gets us

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k\geq 1}\mathrm{e}^{-\theta k/n}\mathbb{P}(X_n \geq k | X_n \geq 1) = \frac{1}{\theta + {}^{2}/_{\sigma^2}} \tag{3.48}$$

Thus, assuming that

$$\mathbb{P}(X_n \geq zn | X_n \geq 1) \underset{n\to\infty}{\longrightarrow} G(z) \tag{3.49}$$

for some non-increasing function $G\colon (0,\infty) \to (0,1)$, we can interpret the sum on the left of (3.48) as the Riemann sum of an (improper) integral to get

$$\int_0^\infty \mathrm{d}z\,\mathrm{e}^{-\theta z}\,G(z) = \frac{1}{\theta + {}^{2}/_{\sigma^2}} \tag{3.50}$$

By the properties of the Laplace transform (whose discussion we omit) this can only be true if

$$G(z) = \mathrm{e}^{-z(2/\sigma^2)}. \tag{3.51}$$

But this shows that the limit (3.49) must exist because from any subsequence we can always extract a limit by using Cantor's diagonal argument and the fact that G is decreasing (again we omit details here). □

From Theorem 3.9 we learn that, conditional on survival up to time n, the number of surviving individuals is of order n. Next we will look at what happens when we condition on extinction. Of course, this is going to have a noticeable effect only on the supercritical processes.

Theorem 3.12 [Duality] *Consider a Galton-Watson branching process with supercritical offspring distribution $(\mathfrak{p}_n)$. Assume $0 < \mathfrak{p}_0 < 1$, let $s_\star$ be the unique positive solution to $\lambda(s) = s$, and define*

$$\mathfrak{q}_n = \mathfrak{p}_n \mathrm{e}^{-s_\star n + \lambda(s_\star)}, \qquad n \geq 0. \tag{3.52}$$

Then $\sum_{n\geq 0} \mathfrak{q}_n = 1$ and

$$\sum_{n\geq 0} n\mathfrak{q}_n = \lambda'(s_\star) < 1, \tag{3.53}$$

i.e., $(\mathfrak{q}_n)$ is an offspring distribution of a subcritical branching process. Furthermore, if T_p denotes the branching tree for $(\mathfrak{p}_n)$ and T_q is the corresponding object for $(\mathfrak{q}_n)$, then for any finite tree T,

$$\mathbb{P}\big(T_p = T \big| \text{ extinction}\big) = \mathbb{P}(T_q = T) \tag{3.54}$$

i.e., the law of T_p conditioned on extinction is that of the dual process T_q.

Proof. Consider a finite tree T and let V be the set of vertices and E the set of edges. Let $n(v)$ denote the number of children of vertex $v \in T$. The probability that T occurs is then

$$\mathbb{P}(T_p = T) = \prod_{v\in T} \mathfrak{p}_{n(v)} \tag{3.55}$$

Conditioning on extinction multiplies this by $\mathrm{e}^{s_\star}$ which we can write using

$$|V| = |E| + 1 \tag{3.56}$$

and the fact that $\lambda(s_\star) = s_\star$ as

$$\mathrm{e}^{s_\star} = \prod_{e\in E} \mathrm{e}^{-s_\star} \prod_{v\in V} \mathrm{e}^{\lambda(s_\star)} = \prod_{v\in V} \mathrm{e}^{-s_\star n(v) + \lambda(s_\star)} \tag{3.57}$$

This shows

$$\mathbb{P}\big(T_p = T \big| \text{ extinction}\big) = \prod_{v\in T} \mathfrak{p}_{n(v)} \mathrm{e}^{-s_\star n(v) + \lambda(s_\star)} \tag{3.58}$$

which is exactly (3.54). □

An interesting special case of duality is the case when a process is *self-dual*. This loosely defined term refers to the situation when the dual process has a distribution "of the same kind" as the original process. Here are some examples:

Example 3.13 *Binomial distribution*: Let $\theta \in [0,1]$ and consider the offspring distribution $(\mathfrak{p}_n)$ which is Binomial(N,θ), i.e.,

$$\mathfrak{p}_n = \binom{N}{n} \theta^n (1-\theta)^{N-n}, \qquad 0 \leq n \leq N. \tag{3.59}$$

Then the dual process is also binomial, with parameters N and $\theta^\star$ where

$$\frac{\theta^\star}{1-\theta^\star} = \frac{\theta}{1-\theta}\mathrm{e}^{-s_\star} \tag{3.60}$$

where $s_\star$ is determined from $(1-\theta+\theta\mathrm{e}^{-s_\star})^N = \mathrm{e}^{-s_\star}$.

Example 3.14 *Poisson distribution*: A limit case of the above is the Poisson offspring distribution,

$$\mathfrak{p}_n = \frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}, \qquad n \geq 0, \tag{3.61}$$

with $\lambda > 1$. Here the dual $(\mathfrak{q}_n)$ is also Poisson but with parameter $\lambda^\star$ which is defined as the unique number less than one with

$$\lambda\mathrm{e}^{-\lambda} = \lambda^\star\mathrm{e}^{-\lambda^\star}. \tag{3.62}$$

The self-dual point is $\lambda = 1$.

Exercise 3.15 Verify the claims in the previous examples.

Problem 3.16 Characterize the distribution of the size of the first generation, X_1, conditioned on survival forever, i.e., the event $\bigcap_{n\geq 1}\{X_n \geq 1\}$.

Problem 3.17 Apply the above analysis to prove that if $\mu > 1$, then for each $\tilde{\mu} < \mu$, $\mathbb{P}(X_n \geq \tilde{\mu}^n | X_n \geq 1) \to 1$ as $n \to \infty$.

3.3 Tree percolation

One of the special, but extremely useful, applications of branching processes is to the problem of *percolation*. We will consider percolation on two graphs for which the connection with branching is quite apparent, a regular directed tree $\mathbb{T}_n$ and the complete graph K_n on n-vertices. Here we will address the former of the two.

A regular directed tree $\mathbb{T}_b$ is a connected infinite graph without cycles where each vertex except one — called the *root* and denoted by $\varnothing$ — has exactly $b+1$ neighbors; the root has only b neighbors. One of the $b+1$ neighbors of vertex $v \neq \varnothing$ — namely, the one on the unique path from v to the root — will be referred to as the *parent*, the other b vertices are the *children*. The root still has b children but no parent.

We will now define a version of percolation called bond percolation. The definition can be made on an arbitrary graph:

Definition 3.18 [Bond percolation] *Consider a finite or infinite graph G with vertex set V and edge set E. Fix $p \in [0,1]$ and consider a collection of Bernoulli random variables ω_e indexed by edges $e \in E$ such that*

$$\mathbb{P}(\omega_e = 1) = p = 1 - \mathbb{P}(\omega_e = 0). \tag{3.63}$$

The phrase bond percolation *then generally refers to the (random) subgraph $G_\omega = (V, E_\omega)$ of G with vertex set V and edge set*

$$E_\omega = \{e \in E \colon \omega_e = 1\}. \tag{3.64}$$

Performing bond percolation on the tree $\mathbb{T}_b$ results in a *forest* — i.e., a collection of trees. The pertinent question in the context of infinite graphs is for what values of p, or whether at all, the graph G_ω contains an infinite connected component. We will narrow this question to whether the root is in an infinite connected component or not. To this end, let $C_\omega(v)$ denote the set of vertices in the connected component of G_ω containing v. Denote

$$\theta(p) = \mathbb{P}_p\big(|C_\omega(\varnothing)| = \infty\big) \tag{3.65}$$

where the index p denotes we are considering percolation with parameter p. The principal observations concerning tree percolation are summarized as follows:

Theorem 3.19 [Bond percolation on $\mathbb{T}_b$] *Consider bond percolation on $\mathbb{T}_b$ with $b \geq 2$ and parameter p. Define $p_c = 1/b$. Then*

$$\theta(p) \begin{cases} = 0, & \text{for } p \leq p_c, \\ > 0, & \text{for } p > p_c. \end{cases} \tag{3.66}$$

Moreover, $p \mapsto \theta(p)$ is continuous on $[0,1]$ and strictly increasing on $[p_c, 1]$ and it vanishes linearly at $p = p_c$,

$$\theta(p) \sim (p - p_c), \qquad p \downarrow p_c. \tag{3.67}$$

Furthermore, the mean-component size, $\chi(p) = \mathbb{E}_p(|C_\omega(\varnothing)|)$ diverges as $p \uparrow p_c$,

$$\chi(p) = \frac{p_c}{p_c - p}, \qquad p < p_c, \tag{3.68}$$

and, at $p = p_c$, the component size distribution has a power-law tail,

$$\mathbb{P}_{p_c}\big(|C_\omega(\varnothing)| \geq n\big) \sim \frac{1}{\sqrt{n}} \tag{3.69}$$

To keep the subject of branching processes in the back of our mind, we begin by noting a connection between percolation and branching:

Lemma 3.20 [Connection to branching] *Let X_n be the number of vertices in $C_\omega(\varnothing)$ that have distance n to the root. Then $(X_n)_{n\geq 0}$ has the law of a Galton-Watson branching process with binomial offspring distribution $(\mathfrak{p}_n)$,*

$$\mathfrak{p}_n = \begin{cases} \binom{b}{n} p^n (1-p)^{b-n}, & \text{if } 0 \leq n \leq b, \\ 0, & \text{otherwise.} \end{cases} \tag{3.70}$$

Proof. Let V_n be the vertices of $\mathbb{T}_b$ that have distance n to the root. Suppose that X_n is known and let $v_1, \ldots, v_{X_n}$ be the vertices in $C_\omega(\varnothing) \cap V_n$. Define

$$\xi_{n+1,j} = \#\big\{u \in V_{n+1} \colon (v_j, u) \in E_\omega\big\}. \tag{3.71}$$

Then, clearly,

$$X_{n+1} = \xi_{n+1,1} + \cdots + \xi_{n+1,X_n} \tag{3.72}$$

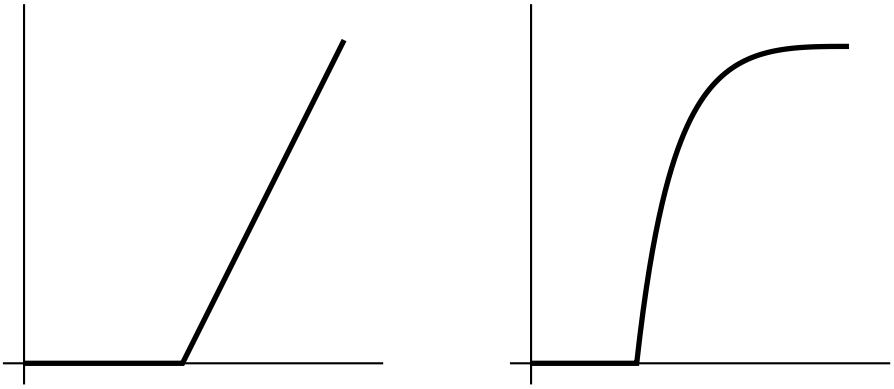

Figure 3.3: The graph of $p \mapsto \theta(p)$ for the tree with $b = 2$ and $b = 3$.

Since each $\xi_{n,j}$ is determined by different ω_e's which are in turn independent of X_n, we may think of $\xi_{n+1,1}, \ldots, \xi_{n+1,X_n}$ as the first X_n terms in the sequence of i.i.d. random variables. The law of these ξ's is binomial with parameters b and p and so, inspecting Definition 3.1, the sequence (X_n) is the Galton-Watson branching process with offspring distribution as stated above. □

Corollary 3.21 *The probability $\theta(p)$ to have an infinite component at the root is the maximal positive solution in $[0,1]$ to the equation*

$$\theta = 1 - \bigl(1 - p\theta\bigr)^b \tag{3.73}$$

The generating function of the component size distribution, $\Psi_p(x) = \mathbb{E}_p(x^{|C_\omega(\varnothing)|})$, is for $x > 0$ and $0 < p < 1$ the unique solution to

$$\Psi = x\bigl(1 - p + p\Psi\bigr)^b \tag{3.74}$$

that lies strictly between zero and one.

Proof. Recall the connection with branching. The event that $C_\infty(\varnothing) = \infty$ clearly coincides with the event that the branching process (X_n) lives forever. As the probability to die out equals $q = \mathrm{e}^{-s_\star}$, where the latter quantity is the smallest solution to the equation

$$q = \sum_{n \geq 0} \mathfrak{p}_n q^n = (pq + 1 - p)^b, \tag{3.75}$$

we easily check that $\theta = 1 - q$ solves (3.73).

To derive the formula for the component moment generating function, we note that if $v_1, \ldots, v_{X_1}$ are the neighbors of the root in $C_\omega(\varnothing)$, then

$$C_\omega(\varnothing) = \{\varnothing\} \cup \tilde{C}_\omega(v_1) \cup \cdots \cup \tilde{C}_\omega(v_{X_1}), \tag{3.76}$$

where $\tilde{C}_\omega(v)$ denote the connected component of v in the subtree of $\mathbb{T}_b$ rooted at v. The union on the right-hand side is disjoint and the components are actually

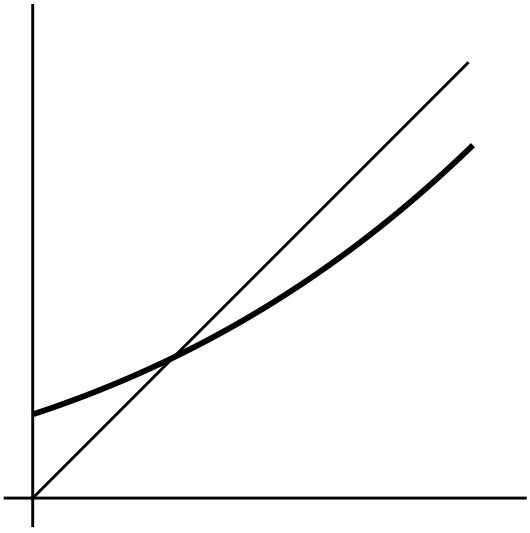

Figure 3.4: The graph of $\Psi \mapsto x(1-p+p\Psi)^b$ for the tree with $b=4$ at $x=0.8$ and $p=0.3$. The function is convex and has a unique fixed point with $\Psi \in (0,1)$.

independent copies of $C_\omega(\varnothing)$. Thus we have

$$\Psi_p(x) = x\,\mathbb{E}_p\big(x^{|C_\omega(v_1)|}\cdots x^{|C_\omega(v_{X_1})|}\big) = x\sum_{n=0}^{b}\mathfrak{p}_n\,\Psi_p(x)^n. \tag{3.77}$$

The equation now (3.74) follows from (3.70). The right-hand side of (3.74) is convex and it is strictly positive at $\Psi=0$ and less than one at $\Psi=1$. There is thus a unique intersection in $(0,1)$ which then has to be the value of $\Psi_p(x)$. □

Exercise 3.22 Set $b=2$ and/or $b=3$ and solve (3.73) explicitly. See Fig. 3.3.

Problem 3.23 Use the argument in the previous proof to show that $\mathbb{P}_p(|C_\omega(\varnothing)|\geq n)$ decays exponentially with n for all $p<p_c$.

Exercise 3.24 Suppose $p>p_c$. Use the duality for branching processes (Theorem 3.12) to show that the tail of the probability distribution of the *finite* components, $\mathbb{P}_p(n\leq|C_\omega(\varnothing)|<\infty)$, decays exponentially with n.

To get the asymptotic for the component size distribution at p_c, we also note:

Lemma 3.25 *For every $x<1$,*

$$\frac{1-\Psi_p(x)}{1-x} = \sum_{n\geq 1} x^{n-1}\,\mathbb{P}_p\big(|C_\omega(\varnothing)|\geq n\big) \tag{3.78}$$

Proof. The proof is essentially identical to that of Lemma 3.10. Let us abbreviate

$a_n = \mathbb{P}_p(|C_\omega(\varnothing)| \geq n)$. Then

$$\Psi_p(x) = \sum_{n\geq 1} x^n \, \mathbb{P}_p\big(|C_\omega(\varnothing)| = n\big) = \sum_{n\geq 1} x^n \, (a_n - a_{n+1}). \tag{3.79}$$

Rearranging the latter sum, we get

$$\Psi_p(x) = a_1 + \sum_{n\geq 1} (x^n - x^{n-1}) a_n. \tag{3.80}$$

The first term on the right-hand side is $a_1 = 1$ while the second term can be written as $(x-1)\sum_{n\geq 1} x^{n-1} a_n$. This and a bit of algebra yield (3.78). □

Proof of Theorem 3.19. The claims concerning $\theta(p)$ are derived by analyzing (3.73). Let us temporarily denote $\Phi(\theta) = 1 - (1 - p\theta)^b$. Then Φ is concave on $[0,1]$ and

$$\Phi(\theta) = pb\theta - \binom{b}{2} p^2\theta^2 + o(\theta^2) \tag{3.81}$$

This shows that for $pb \leq 1$ the only fixed point of Φ is $\theta = 0$, while for $pb > 1$ there are two solutions: one at $\theta = 0$ and one at $\theta = \theta(p) > 0$. This proves (3.66).

To get the critical scaling (3.67), we note that (3.81) can be rewritten as

$$\Phi(\theta) = \theta + b\theta\Big[(p - p_c) - \frac{b-1}{2} p^2\big(1 + o(1)\big)\theta\Big] \tag{3.82}$$

The positive solution $\theta(p)$ of $\Phi(\theta) = \theta$ will thus satisfy

$$\theta(p) = \frac{2b^2}{b-1}\big(1 + o(1)\big)\,(p - p_c), \qquad p \downarrow p_c. \tag{3.83}$$

To get the asymptotic for $\chi(p)$, we note

$$\chi(p) = \Psi_p'(1) \tag{3.84}$$

so we just need to find the good asymptotic of Ψ_p near $x = 1$. This end we denote $g(x) = 1 - \Psi_p(x)$ and note that (3.74) becomes

$$g(x) = 1 - x\big(1 - pg(x)\big)^b \tag{3.85}$$

Differentiating at $x = 1$ we get

$$g'(1) = -1 + bpg'(1) \tag{3.86}$$

and so $\Psi_p'(1) = (1 - bp)^{-1} = p_c/(p_c - p)$.

Finally, to get the component-size distribution at p_c, we will restrict ourselves only to $b = 2$. In that case the equation for $g(x)$ can be solved and we get

$$g(x) = \frac{2}{x}\Big(\sqrt{1-x} - (1-x)\Big) \tag{3.87}$$

where the sign of the square root was chosen to make g positive on $(0,1)$ as it should be. By (3.78) we in turn have

$$x\,\frac{g(x)}{1-x} = \sum_{n\geq 1} x^n\,\mathbb{P}_p\big(|C_\omega(\varnothing)|\geq n\big) \tag{3.88}$$

and so to find the probabilities $\mathbb{P}_p(|C_\omega(\varnothing)|\geq n)$ we just need to expand the left-hand side into a Taylor series about $x=0$ and use that the coefficients of this expansion are uniquely determined. This is said nearly as easily as it is done:

$$x\,\frac{g(x)}{1-x} = 2\Big(\frac{1}{\sqrt{1-x}}-1\Big) = 2\sum_{n\geq 1}\binom{2n}{n}\Big(\frac{x}{4}\Big)^n \tag{3.89}$$

and so, by identifying the coefficients,

$$\mathbb{P}_{p_c}\big(|C_\omega(\varnothing)|\geq n\big) = 2\binom{2n}{n}4^{-n}. \tag{3.90}$$

Stirling's formula, $n! = (n/\mathrm{e})^n\sqrt{2\pi n}(1+o(1))$, shows that

$$2\binom{2n}{n}4^{-n} = \frac{2+o(1)}{\sqrt{\pi n}}, \qquad n\to\infty, \tag{3.91}$$

and so we have $\mathbb{P}_p(|C_\omega(\varnothing)|\geq n)\sim 1/\sqrt{n}$ as claimed. □

Problem 3.26 Show that for any $b\geq 2$ the limit

$$\lim_{x\uparrow 1}\frac{1-\Psi_{p_c}(x)}{\sqrt{1-x}} \tag{3.92}$$

exists and is positive and finite. Compute its value.

Some of our computations above have perhaps been somewhat unnecessarily formal. For instance, (3.73) can be derived as follows:

Let $1-\theta$ be the probability that the root is in a finite component. For that to be true, every occupied edge from the root must end up at a vertex whose component in the forward direction is also finite. The probability that the vertex v_1 is like this is $1-p\theta$, and similarly for all b neighbors of $\varnothing$. As these events are independent for distinct neighbors, this yields

$$1-\theta = (1-p\theta)^b \tag{3.93}$$

which is (3.73). Similarly, we can also compute $\chi(p)$ directly: By (3.76),

$$|C_\omega(\varnothing)| = 1+\omega_{\varnothing,v_1}|\tilde{C}_\omega(v_1)|+\cdots+\omega_{\varnothing,v_b}|\tilde{C}_\omega(v_b)| \tag{3.94}$$

Using that $\omega_{\varnothing,v}$ is independent of $|\tilde{C}_\omega(v)|$, taking expectations we get

$$\chi(p) = 1+bp\chi(p) \tag{3.95}$$

whereby $\chi(p)=\frac{p_c}{p_c-p}$.

The recursive nature of the tree, and of the ensuing calculations, allows us to look at some more complicated variants of the percolation process:

Problem 3.27 *k-core percolation*: Consider the problem of so called k-core percolation on the tree $\mathbb{T}_b$. Here we take $k \geq 3$ and take a sample of $C(\varnothing)$. Then we start applying the following pruning procedure: If a vertex has less than k "children" to which it is connected by an occupied edge, we remove it from the component along with the subtree rooted in it. Applying this over and over, this gives us a *decreasing* sequence $C_n(\varnothing)$ of subtrees of $C(\varnothing)$. Let $\vartheta(p)$ denote the probability that $C_n(\varnothing)$ is infinite for all n. Show that $\vartheta(p)$ is the largest positive solution to

$$\vartheta = \pi_k(\vartheta p) \quad \text{where} \quad \pi_k(\lambda) = \sum_{\ell=k}^{b} \binom{b}{\ell} \lambda^\ell (1-\lambda)^{b-\lambda}, \tag{3.96}$$

i.e., $\pi_k(\lambda)$ is the probability that Binomial(b, λ) is at least k. Explain why $\vartheta(p)$ for 1-core percolation equals to $\theta(p)$ for the ordinary percolation.

3.4 Erdős-Rényi random graph

Having understood percolation on the regular rooted tree, we can now move on to a slightly complicated setting which is percolation on the complete graph K_n. This problem has emerged independent of the development in percolation as a model of a random graph and is named after its inventors Erdős and Rényi.

The complete graph K_n has vertices $\{1, \ldots, n\}$ and an (unoriented) edge between every pair of distinct vertices. Given $p \in [0, 1]$ we toss a biased coin for each edge and if it comes out heads — which happens with probability p — we keep the edge and if we get tails then we discard it. We call the resulting random graph $\mathscr{G}(n, p)$. The principal question of interest is the distribution of the largest connected component; particularly, when it is of order n.

Our main observation is that a percolation transition still occurs in the setting, even though the formulation is somewhat less clean due to the necessity to take $n \to \infty$:

Theorem 3.28 *For any $\alpha \geq 0$ and $\epsilon > 0$, let $\theta_{\epsilon,n}(\alpha)$ denote the probability that vertex "1" is in a component $\mathscr{G}(n, \alpha/n)$ of size at least ϵn. Then*

$$\theta(\alpha) = \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \theta_{\epsilon,n}(\alpha) \tag{3.97}$$

exists and equals the largest positive solution to

$$\theta = 1 - \mathrm{e}^{-\alpha\theta}. \tag{3.98}$$

In particular,

$$\theta(\alpha) \begin{cases} = 0, & \alpha \leq 1, \\ > 0, & \alpha > 1. \end{cases} \tag{3.99}$$

Exercise 3.29 Show that $\theta(\alpha)$ is the probability that the branching process with Poisson offspring distribution $\mathfrak{p}_n = \frac{\alpha^n}{n!}\mathrm{e}^{-\alpha}$ survives forever.

The key idea of our proof is to explore the component of vertex "1" using a *search algorithm*. The algorithm keeps vertices in three classes: *explored* and *active* — called

jointly "discovered" — and *undisovered*. Initially, we mark vertex "1" as active and the others as undiscovered. Then we repeat the following:

(1) Pick the active vertex v that has the least index.

(2) Find all undiscovered neighbors of v.

(3) Change the status of v to explored and its undiscovered neighbors to active.

The algorithm stops when we run out of active vertices. It is easy to check that this happens when we have explored the entire connected component of vertex "1".

Let A_k denote the number of active vertices at the k-th stage of the algorithm; the initial marking of "1" is represented by $A_0 = 1$. If $v \in A_k$ is the vertex with the least index, we use L_k to denote the number of as of yet undiscovered neighbors of v. We have

$$A_{k+1} = A_k + L_k - 1. \tag{3.100}$$

Note that $n - k - A_k$ is the number of undiscovered vertices after the k-th run of the above procedure.

Lemma 3.30 *Conditional on A_k, we have*

$$L_k = \text{Binom}(n - k - A_k, \alpha/n). \tag{3.101}$$

Proof. The newly discovered vertices are chosen with probability α/n from the set of $n - k - A_k$ undiscovered vertices. □

We will use the above algorithm to design a *coupling* with the corresponding search algorithm for bond percolation on a regular rooted tree — or, alternatively, with a branching process with a binomial offspring distribution. Recall that $\mathbb{T}_b$ denotes the rooted tree with forward degree b and K_n the complete graph on n vertices. To make all percolation processes notationally distinct, we will from now on write $\mathbb{P}_{\mathbb{T}_n}$ for the law of percolation on $\mathbb{T}_n$ and $\mathbb{P}_{K_n}$ for the corresponding law on K_n. In *all* cases below the probability that an edge is occupied is α/n.

Lemma 3.31 [Coupling with tree percolation] *For $m \le n$ and $r \le n - m$,*

$$\mathbb{P}_{\mathbb{T}_m}\big(|C(\varnothing)| \ge r\big) \le \mathbb{P}_{K_n}\big(|C(1)| \ge r\big) \le \mathbb{P}_{\mathbb{T}_n}\big(|C(\varnothing)| \ge r\big) \tag{3.102}$$

Proof. The proof is based on the observation that, as long as less than $n - m$ vertices have been discovered, we can couple the variables L_k for $\mathbb{T}_m$, K_n and $\mathbb{T}_n$ so that

$$L_k^{(\mathbb{T}_m)} \le L_k^{(K_n)} \le L_k^{(\mathbb{T}_n)} \tag{3.103}$$

Indeed, conditioning on the number of discovered vertices, we have

$$L_k^{(\mathbb{T}_b)} = \text{Binom}(b, \alpha/n) \tag{3.104}$$

Now think of the binomial random variable $\text{Binom}(b, \alpha/n)$ as the sum of the first b terms in a sequence of Bernoulli random variables that are 1 with probability α/n

and zero otherwise. To get $L_k^{(\mathbb{T}_m)}$ we then add only the first m, to sample $L_k^{(K_n)}$ we add the first $n-k-A_k$, and to get $L_k^{(\mathbb{T}_n)}$ we add the first n of these variables. Under the condition

$$m \le n-k-A_k \le n \tag{3.105}$$

we will then have (3.103). The upper bound in this condition is trivial and the lower bound will hold as long as $k + A_k \le n - m$.

This argument shows that, if the connected component $C(1)$ of vertex "1" in K_n is of size r, then so is the component $C(\varnothing)$ on $\mathbb{T}_n$, i.e., the right inequality in (3.102) holds. Similarly, thinking of adding the discovered vertices to the tree one by one, before the component $C(1)$ of K_n reaches the size $r \le n-m$, the component $C(\varnothing)$ of $\mathbb{T}_m$ will not be larger than r. This implies

$$\mathbb{P}_{K_n}\big(|C(1)| < r\big) \le \mathbb{P}_{\mathbb{T}_m}\big(|C(\varnothing)| < r\big), \qquad r \le n-m, \tag{3.106}$$

whose complement then yields the left inequality in (3.102). □

The following observation will be helpful in the proof:

Lemma 3.32 [Continuity in offspring distribution] *Let $\mathfrak{p}^{(m)} = (\mathfrak{p}_m^{(n)})$ be a family of offspring distributions and let $\mathfrak{p}$ be an offspring distribution such that $0 < \mathfrak{p}_0 < 1$. Suppose that, for each $m \ge 0$,*

$$\mathfrak{p}_m^{(n)} \underset{n\to\infty}{\longrightarrow} \mathfrak{p}_m. \tag{3.107}$$

Then for each $r \ge 1$,

$$\mathbb{P}_{\mathfrak{p}^{(n)}}\Big(r \le \sum_{\ell\ge 0} X_\ell < \infty\Big) \underset{n\to\infty}{\longrightarrow} \mathbb{P}_{\mathfrak{p}}\Big(r \le \sum_{\ell\ge 0} X_\ell < \infty\Big). \tag{3.108}$$

where $\mathbb{P}_{\mathfrak{p}^{(n)}}$ and $\mathbb{P}_{\mathfrak{p}}$ denote the law of the branching process with offspring distributions $\mathfrak{p}^{(n)}$ and $\mathfrak{p}$, respectively.

Problem 3.33 Define

$$\lambda^{(n)}(s) = -\log \sum_{m\ge 0} \mathfrak{p}_m^{(n)} \mathrm{e}^{-sm} \tag{3.109}$$

and show that $\lambda^{(n)}(s) \to \lambda(s)$, where $\lambda(s)$ is defined using $\mathfrak{p} = (\mathfrak{p}_m)$. Then use this to prove the lemma.

Now we are ready to prove our result for the Erdős-Rényi random graph:

Proof of Theorem 3.28. We will prove upper and lower bounds on the quantity

$$\theta_{\epsilon,n} := \mathbb{P}_{K_n}(|C(1)| \ge \epsilon n). \tag{3.110}$$

By Lemma 3.31, it follows that

$$\theta_{\epsilon,n} \ge \mathbb{P}_{\mathbb{T}_{(1-\epsilon)n}}\big(|C(\varnothing)| \ge \epsilon n\big) \ge \mathbb{P}_{\mathbb{T}_{(1-\epsilon)n}}\big(|C(\varnothing)| = \infty\big) \tag{3.111}$$

Equation (3.73) in Corollary 3.21 shows that the right-hand side is the largest positive solution to the equation

$$\theta = 1 - \big(1 - \tfrac{\alpha}{n}\theta\big)^{(1-\epsilon)n} \tag{3.112}$$

As

$$\big(1-\tfrac{\alpha}{n}\theta\big)^{(1-\epsilon)n} \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{(1-\epsilon)\alpha\theta} \tag{3.113}$$

this equation and, by convexity of the right-hand side, also its maximal positive solution converge to that of (3.98) in the limits $n \to \infty$ followed by $\epsilon \downarrow 0$. This proves (3.97) with the limits replaced by *limes inferior*.

It remains to prove the corresponding upper bound. Lemma 3.31 gives us

$$\theta_{\epsilon,n} \leq \mathbb{P}_{\mathbb{T}_n}\big(|C(\varnothing)| \geq \epsilon n\big). \tag{3.114}$$

In order to take the limits, we rewrite the right-hand side as follows:

$$\theta_{\epsilon,n} \leq \mathbb{P}_{\mathbb{T}_n}\big(|C(\varnothing)| = \infty\big) + \mathbb{P}_{\mathbb{T}_n}\big(n\epsilon \leq |C(\varnothing)| < \infty\big) \tag{3.115}$$

The first term on the right-hand side then converges to the solution to (3.73) and so it suffices to show that the second term vanishes after we take the required limits. To this end we fix $r \geq 1$ and write

$$\mathbb{P}_{\mathbb{T}_n}\big(n\epsilon \leq |C(\varnothing)| < \infty\big) \leq \mathbb{P}_{\mathbb{T}_n}\big(r \leq |C(\varnothing)| < \infty\big) \tag{3.116}$$

The only n-dependence on the right hand side is then through the offspring distribution. As

$$\binom{n}{m}\Big(\frac{\alpha}{n}\Big)^m\Big(1-\frac{\alpha}{n}\Big)^{n-m} \underset{n\to\infty}{\longrightarrow} \frac{\alpha^m}{m!}\mathrm{e}^{-\alpha}, \qquad m \geq 0, \tag{3.117}$$

the offspring distribution converges to Poisson, $\mathfrak{p}_n(\alpha) = \frac{\alpha^n}{n!}\mathrm{e}^{-\alpha}$, and so, by Lemma 3.32, we have

$$\mathbb{P}_{\mathbb{T}_n}\big(r \leq |C(\varnothing)| < \infty\big) \underset{n\to\infty}{\longrightarrow} \mathbb{P}_{\mathfrak{p}(\alpha)}\Big(r \leq \sum_{\ell\geq 0} X_\ell < \infty\Big). \tag{3.118}$$

The rest of the analysis has to be done depending on the value of α. Indeed, if $\alpha \leq 1$ — which corresponds to the (sub)critical regime — then the branching process dies out with probability one and so the right-hand side tends to zero as $r \to \infty$. For $\alpha > 1$, we can use duality (Theorem 3.12) to convert this to the similar question for $\alpha < 1$ and so the convergence to zero holds in this case as well. We conclude that, after the required limits $\theta_{\epsilon,n}$ is bounded by the maximal positive solution of (3.98). This proves (3.97) with *limes superior* and thus finishes the proof. □

The fact that all vertices of K_n look "the same" suggest that $\theta(\alpha)$ actually represents the fraction of vertices in components of macroscopic size. This is indeed the case, but we will not try to prove it here. In fact, more is known:

Theorem 3.34 *Given a realization of $\mathscr{G}(n, \alpha/n)$, let $\mathcal{C}_1, \mathcal{C}_2, \dots$ be the list of all connected components ranked decreasingly by their size. Then we have:*

(1) If $\alpha < 1$, then $|\mathcal{C}_1| = O(\log n)$, i.e., all components are at most logarithmic.

(2) If $\alpha > 1$, then $|\mathcal{C}_1| = \theta(\alpha)n + o(n)$ and $|\mathcal{C}_2| = O(\log n)$, i.e, there is a unique *giant component and all other components are of at most logarithmic size.*

(3) If $\alpha = 1$, then $|\mathcal{C}_1|, |\mathcal{C}_2|, \dots$ are all of order $n^{2/3}$ with a nontrivial limit distribution of $n^{-2/3}|\mathcal{C}_1|, n^{-2/3}|\mathcal{C}_2|, \dots$.

The case (3) is the one most interesting because corresponds to the critical behavior we saw at $p = p_c$ on the regular tree. The regime actually extends over an entire *critical window*, i.e., for values

$$p = \frac{1}{n} + \frac{\lambda}{n^{4/3}} \tag{3.119}$$

where λ is any fixed real number. One of the key tools to analyze this regime is the tree-search algorithm that we used at the beginning of this section.

Chapter 4

Percolation

Percolation was mentioned earlier in these notes in the context of infinite trees and complete graphs. Here we will consider percolation on the hypercubic lattice $\mathbb{Z}^d$, or more generally on graphs with underlying spatial geometry. The main goal is to prove a uniqueness of the infinite connected component by the famous Burton-Keane uniqueness argument.

4.1 Percolation transition

The goal of this section is to establish the existence of a unique *percolation threshold*. The argument is based on monotonicity in p. Intuitively it is clear that increasing p will result into a larger number of occupied edges which in turn means that the graph is more likely to contain large connected component. To make this more precise, we will couple percolation for all p's on one probability space and exhibit the monotonicity explicitly.

Lemma 4.1 [Coupling of all p's] *Consider a graph $G = (V, E)$ and consider a family of i.i.d. uniform random variables U_e indexed by the edges $e \in E$. For each $p \in [0, 1]$ define*

$$\omega_e^{(p)} = \begin{cases} 1, & \text{if } U_e \leq p, \\ 0, & \text{if } U_e > p. \end{cases} \tag{4.1}$$

Then $(\omega_e^{(p)})$ are Bernoulli with parameter p and the graph with vertices V and edges $E^{(p)} = \{e \in E \colon \omega_e^{(p)} = 1\}$ has the law of bond percolation on G with parameter p.

Proof. The random variables $\omega_e^{(p)}$ are independent because the U_e's are independent. As

$$\mathbb{P}(\omega_e^{(p)} = 1) = \mathbb{P}(U_e \leq p) = p \tag{4.2}$$

the distribution of $(\omega_e^{(p)}$ is Bernoulli with parameter p. □

The above allows us to establish the monotonicity of the quantity

$$\theta(p) = \mathbb{P}_p\big(|C_\omega(\varnothing)| = \infty\big). \tag{4.3}$$

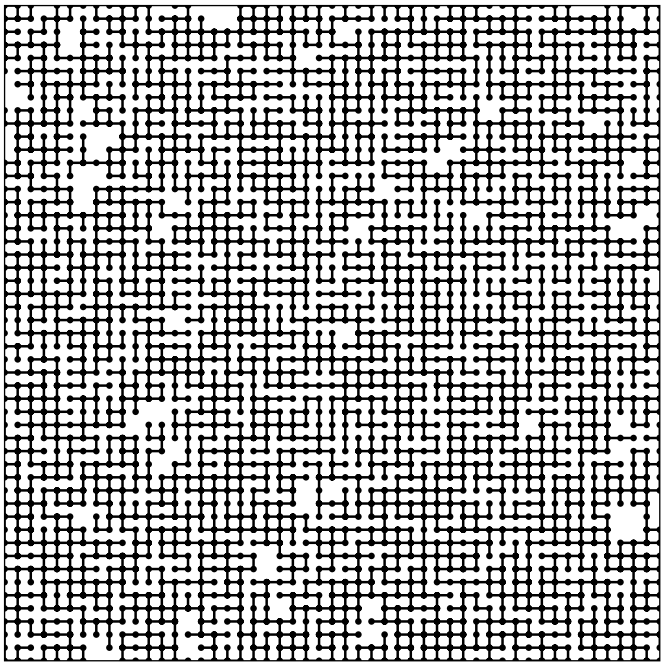

Figure 4.1: A sample from bond percolation on $\mathbb{Z}^2$ with $p = 0.65$. Only the sites that have a connection via occupied edges to the boundary of the 50×50 box are depicted; the others are suppressed.

Corollary 4.2 [Uniqueness of percolation threshold] *The function $p \mapsto \theta(p)$ is non-decreasing. In particular, there exists a unique $p_c \in [0,1]$ such that*

$$\theta(p) \begin{cases} = 0, & \text{for } p < p_c, \\ > 0, & \text{for } p > p_c. \end{cases} \tag{4.4}$$

Note that, unlike for the regular tree $\mathbb{T}_b$, we are not making any claim about the value of $\theta(p)$ at $p = p_c$. It is expected that $\theta(p_c) = 0$ in all $d \geq 1$ but proofs exist only for $d \leq 2$ and $d \geq 19$. This is one of the most annoying basic open problem in percolation theory.

Exercise 4.3 Show that $p_c = 1$ in $d = 1$.

Theorem 4.4 [Non-triviality of percolation threshold] *We have*

$$0 < p_c < 1, \qquad d \geq 2 \tag{4.5}$$

Proof of $p_c > 0$. Suppose $d \geq 2$. We claim that then

$$p_c \geq \frac{1}{2d-1}. \tag{4.6}$$

Our goal is to prove that there are no infinite components once p is sufficiently small. To this end we note that if $C(\varnothing)$ is infinite, then there exists an infinite

occupied edge-self-avoiding paths from $\varnothing$ "to infinity." Let A_n be the event that there exists such a path of length at least n and let $c_n(d)$ be the number of such paths of length n starting from the origin. Then

$$\mathbb{P}_p(A_n) \leq c_n(d)p^n \tag{4.7}$$

In order to bound the right-hand side, we need to estimate $c_n(d)$. We can do this as follows: The path has $2d$ ways to leave the origin. In every next step, there are at most $2d-1$ choices to go because the path cannot immediately "backtrack" to the previous vertex. This implies

$$c_n(d) \leq 2d(2d-1)^{n-1} \tag{4.8}$$

Thus, if $(2d-1)p < 1$ then $\mathbb{P}_p(A_n)$ decays exponentially with n and so

$$\mathbb{P}_p(|\mathsf{C}_\omega(\varnothing)| = \infty) \leq \lim_{n\to\infty} \mathbb{P}_p(A_n) = 0. \tag{4.9}$$

This proves (4.6). □

For the upper bound on p_c, it suffices to focus only on $d = 2$. The key observation is that, if p is very close to one, then the vacant edges are so unlikely that they cannot block the occupied edges from forming an infinite self-avoiding path containing the origin. The only way vacant blocks can block such a path in $d = 2$ is by forming a *circuit* around the origin which is entirely composed of vacant edges. This observation is the core of the so called *Peierls argument* named after R. Peierls who used it to argue for the existence of a phase transition in the Ising model.

In order to define the abovementioned circuit, we will consider a graph

$$(\mathbb{Z}^2)^* = (1/2, 1/2) + \mathbb{Z}^2, \tag{4.10}$$

i.e., the grid $\mathbb{Z}^2$ shifted by half lattice spacing in each direction. This graph has the property that each of its edges crosses exactly one edge in $\mathbb{Z}^2$ and *vice versa*. For this reason we will refer to $(\mathbb{Z}^2)^*$ as the *dual graph* and to the edge of $(\mathbb{Z}^2)^*$ crossing edge e of $\mathbb{Z}^2$ as the *dual edge*.

Proof of $p_c < 1$. We will prove that in all $d \geq 2$,

$$p_c \leq \frac{2}{3} \tag{4.11}$$

It suffices to show this in $d = 2$ because if $p > 2/3$ implies that $\mathbb{Z}^2$ contains an infinite connected component, the natural embedding $\mathbb{Z}^2 \subset \mathbb{Z}^d$ guarantees that the same will be true for all $d \geq 2$.

Let Γ be a closed edge self-avoiding path on $(\mathbb{Z}^2)^\star$ and let B_Γ be the event that all edges of $\mathbb{Z}^2$ crossing an edge in this path are vacant. It is easy to check that

$$\{\forall v \in \mathbb{Z}^d \colon |C_\omega(v)| < \infty\} \subset \bigcap_{n\geq 1} \bigcup_{\substack{\Gamma\colon \mathrm{Int}\Gamma \ni 0 \\ |\Gamma| \geq n}} B_\Gamma. \tag{4.12}$$

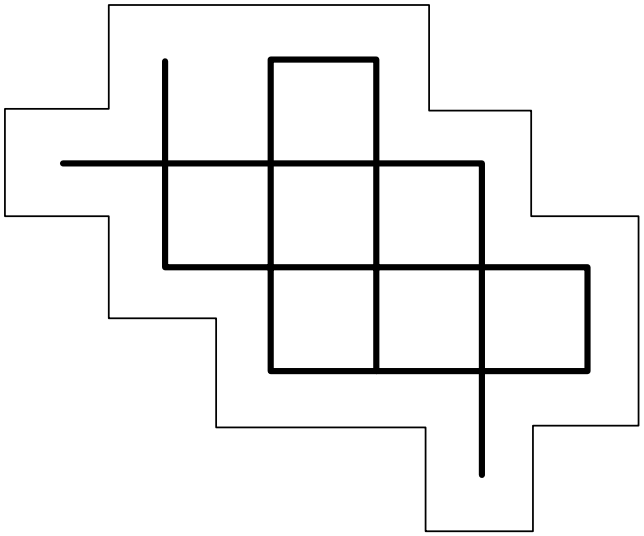

Figure 4.2: A connected component of occupied edges and a contour on the dual lattice bisecting (some of) the vacant edges adjacent to this component.

where IntΓ denotes the set of vertices of $\mathbb{Z}^2$ surrounded by Γ and $|\Gamma|$ denotes the number of edges in Γ. Now

$$\mathbb{P}_p(B_\Gamma) \leq (1-p)^{|\Gamma|} \tag{4.13}$$

and so we just need to estimate the number of possible paths of length n.

As the paths are self avoiding, the number of paths of length $|\Gamma| = n$ passing through a point $x \in (\mathbb{Z}^2)^*$ is bounded by c_n. However, any circuit surrounding the origin has to cross the positive part of the "x-axis" in $(\mathbb{Z}^2)^*$ within distance n from the origin. Hence the number of x's we need to consider is less than n. By truncating the intersection in (4.12) to $n \geq N$, we thus get

$$\mathbb{P}_p\big(\forall v\colon |C_\omega(v)| < \infty\big) \leq \sum_{n\geq N} nc_n(1-p)^n \tag{4.14}$$

for all $N \geq 1$. Once $(1-p)3 < 1$, the series on the right-hand side is absolutely summable and the $N \to \infty$ limit is thus zero. Therefore, the complementary event, $\{\exists v\colon |C_\omega(\varnothing)| = \infty\}$, has probability one. □

Historically, another critical value of p was introduced to characterize the values of p for which the mean-cluster size is finite. Let, as before,

$$\chi(p) = \mathbb{E}_p(|C_\omega(\varnothing)|). \tag{4.15}$$

The coupling of all p's implies that $p \mapsto \chi(p)$ is increasing and so it makes sense to define the critical value

$$\pi_c = \sup\big\{p \geq 0\colon \chi(p) < \infty\big\}. \tag{4.16}$$

The following, extremely non-trivial result has been established independently by Aizenman-Barsky and Menshikov:

Theorem 4.5 [Absence of intermediate phase] *For all $d \geq 2$, $p_c = \pi_c$.*

This very simple but deep theorem — whose validity we already checked for regular tree — has profound consequences. One of them is the following observation due to Hammersely. Consider the two-point connectivity function,

$$\tau_p(x,y) = \mathbb{P}_p\big(x \in C_\omega(y)\big) \tag{4.17}$$

Despite the apparent lack of symmetry on the right-hand side, we clearly have $\tau_p(x,y) = \tau_p(y,x)$ because the event boils down to having an occupied path between x and y.

Theorem 4.6 *Suppose $\chi(p) < \infty$. Then there exists $c > 0$ such that*

$$\tau_p(x,y) \leq \mathrm{e}^{-c|x-y|}. \tag{4.18}$$

Combined with the previous result, this says that the two-point connectivities decay exponential throughout the *entire* subcritical regime. Or, in other words, the result rules out the existence of so called *intermediate phase* which would be characterized by absence of percolation and yet non-exponential decay of connectivities. We will prove this theorem once we have introduced the BK-inequality.

Exercise 4.7 Prove Theorem 4.5 in $d = 1$.

The following is implied by the above for the subcritical regime:

Lemma 4.8 *Let $p < p_c$. There exists a constant $c_1 \in (0,\infty)$ and a random variable $L_0 = L_0(\omega) < \infty$ such that for each $L \geq L_0(\omega)$, the largest connected component intersecting the cube $[-L,L]^d \cap \mathbb{Z}^d$ is of diameter at most $c_1 \log L$.*

Proof. Let $R = R_L(\omega)$ denote the diameter of the largest connected component intersecting the cube $B_L = [-L,L]^d \cap \mathbb{Z}^d$. Then

$$\{R_L \geq n\} \subset \bigcup_{x \in B_L} \bigcup_{\substack{y \in \mathbb{Z}^d \\ |y-z| \geq n}} \big\{y \in C_\omega(x)\big\} \tag{4.19}$$

This implies

$$\mathbb{P}_p(R_L \geq n) \leq \sum_{x \in B_L} \sum_{\substack{y \in \mathbb{Z}^d \\ |y-z| \geq n}} \tau_p(x,y) \tag{4.20}$$

From Theorem 4.6 we know that $\tau_p(x,y) \leq \mathrm{e}^{-c|x-y|}$ and a calculation shows

$$\sum_{\substack{y \in \mathbb{Z}^d \\ |y-z| \geq n}} \mathrm{e}^{-c|x-y|} \leq c_2 \mathrm{e}^{-cn} \tag{4.21}$$

for some constant $c_2 < 0$. Using this in (4.20) gives

$$\mathbb{P}_p(R_L \geq n) \leq c_2(2L+1)^d \mathrm{e}^{-cn} \tag{4.22}$$

Setting, e.g., $c_1 = (d+2)/c$ and plugging $n = c_1 \log L$, the right-hand side is summable on L and so the event $R_L \geq c_1 \log L$ will occur only for finitely many L. Writing $L_0 - 1$ to denote the largest such L, the claim is proved. □

4.2 Uniqueness of infinite component

Having established non-triviality of the percolation transition on $\mathbb{Z}^d$ with $d \geq 2$, our next task is count the number of infinite connected component. We will prove that, if it exists, such a component is unique with probability one. To see this is a non-trivial fact, note the following:

Exercise 4.9 Show that on the regular rooted tree $\mathbb{T}_b$ with $b \geq 2$ there are infinitely many infinite components with probability one for every p with $p_c < p < 1$.

We will invoke some useful "abstract nonsense" theorems from probability theory. The first one goes back to Kolmogorov:

Definition 4.10 [Tail Event] *Let (η_j) be random variables indexed by $j \in \mathbb{N}$. We say that an event A is a* tail event *if the following holds: For any $\eta \in A$ and any $k \geq 1$, any η' such that $\eta'_j = \eta_j$ for $j \geq k$ obeys $\eta' \in A$.*

In other words, A is a tail event if changing any finite random of the random variables η_j won't affect the containment in A.

Theorem 4.11 [Kolmogorov's Zero-One Law] *Suppose (η_j) be i.i.d. Then every tail event has probability either zero or one.*

This law may be deemed responsible for many limit theorems in probability theory. E.g., it implies that the event $\lim_{n\to\infty}(X_1 + \cdots + X_n)/n$ exists and equals $\mathbb{E}X_1$ is tail event and so, for i.i.d. X_j's, it occurs with probability zero or one.

Exercise 4.12 Verify this statement.

We will use this to conclude an "abstract non-sense" result in percolation theory. Given a configuration (ω_e) of bond percolation, let $N = N(\omega)$ denote the number of infinite connected components in the graph generated by ω. This quantity can take any value in $\{0, 1, \dots\} \cup \{\infty\}$.

Corollary 4.13 *Consider bond percolation on $\mathbb{Z}^d$ with parameter $p \in [0,1]$. Then*

$$\{N = 0\}, \quad \{1 \leq N < \infty\} \quad \textit{and} \quad \{N = \infty\} \tag{4.23}$$

are tail events and so exactly one of them occurs with probability one and the others with probability zero.

Proof. Tail events are not affected by a change of any finite number of edges. It is easy to check that if there is no infinite connected component, then a change of any finite number of edges will not introduce one. Similarly, if there are infinitely many infinite components, changing any finite number of edges is not going to destroy more than a finite number of them. So $\{N = 0\}$ and $\{N = \infty\}$ are tail events. The event $\{1 \le N < \infty\}$, being the complement of $\{N = 0\} \cup \{N = \infty\}$, is then a tail event as well. □

Note that, for different p, a different event may be the one with probability one. E.g., for $p < p_c$ we definitely have $\mathbb{P}_p(N = 0) = 1$ while for $p > p_c$ we have $\mathbb{P}_p(N > 0) = 1$. The latter follows because as $\mathbb{P}(N > 0)$ has a positive probability, it must already have probability one.

Our other technical "abstract nonsense" tool will be the Ergodic Theorem. Consider a collection of random variables (η_x) indexed by vertices in $\mathbb{Z}^d$. We define the *translation by* z to be the function T_z that acts on the η's as

$$(T_z\eta)_x = \eta_{x+z}, \qquad x \in \mathbb{Z}^d. \tag{4.24}$$

In other words, T_z shifts η in such a way that the coordinate η_z lands at the origin. If $f(\eta)$ is a function of these random variables, then $f \circ T_z$ denotes the function such that $(f \circ T_z)(\eta) = f(T_z(\eta))$.

Definition 4.14 [Shift invariant events] *We say that an event A is* shift invariant *if for each $\eta \in A$,*

$$T_z(\eta) \in A \quad \textit{holds for all} \quad z \in \mathbb{Z}^d \tag{4.25}$$

Theorem 4.15 [Spatial Ergodic Theorem for Bernoulli] *Let (η_x) be i.i.d. random variables indexed by $x \in \mathbb{Z}^d$ and let $f = f(\eta)$ be a function thereof. Suppose that $\mathbb{E}|f(\eta)| < \infty$. Then, with probability one,*

$$\lim_{n\to\infty} \frac{1}{(2n+1)^d} \sum_{x:\ |x|\le n} f \circ T_x = \mathbb{E}f(\eta). \tag{4.26}$$

Moreover, every shift invariant event has probability zero or one.

This theorem has a number of proofs all of which go somewhat beyond the scope of these notes. However, one of the possible lines of attack gets actually very close:

Problem 4.16 Suppose $d = 1$ for simplicity and let $f = f(\eta)$ be a function that depends only on finitely many coordinates. Use the Strong Law of Large Numbers to show that (4.26) holds. *Note*: From here Theorem 4.15 follows by approximating every integrable function by functions of this type.

Let us see how bond percolation on $\mathbb{Z}^d$ fits into the above framework. For each vertex $z \in \mathbb{Z}^d$, we have a collection of d independent random variables $\omega_{e_1(z)}, \dots, \omega_{e_d(z)}$ where $e_1(z), \dots, e_d(z)$ denote the edges with one endpoint at z that are oriented in the positive coordinate direction. We then set

$$\eta_z = (\omega_{e_1(z)}, \dots, \omega_{e_d(z)}). \tag{4.27}$$

As each ω_e appears in exactly one η_z, these variables are i.i.d. and they encode the entire percolation configuration. Going back to the question about the number of infinite connected components, we can thus draw the following immediate consequence of Theorem 4.15:

Corollary 4.17 *Consider bond percolation on $\mathbb{Z}^d$ with parameter $p \in [0,1]$. Then there exists $k \in \{0,1,\dots\} \cup \{\infty\}$ such that $\mathbb{P}_p(N = k) = 1$.*

Proof. The event $\{N = k\}$ is translation invariant because shifting the ω's will not affect the component structure of the resulting random graph. Hence $\mathbb{P}_p(N = k)$ is either zero or one. But the union of these events over $k \in \{0\} \cup \mathbb{N} \cup \{\infty\}$ is everything and so one of these events actually must occur with probability one. □

Again, which k is the "lucky one" depends on the value of p. Next we observe:

Lemma 4.18 *Suppose that $0 < p < 1$. If $\mathbb{P}_p(N = k) > 0$ for some $k \geq 2$, then we also have $\mathbb{P}_p(N \leq k-1) > 0$.*

Proof. Without loss of generality assume $k < \infty$. If the full graph contains k clusters with positive probability, then there exists n such that the box $\Lambda_n = [-n,n]^d \cap \mathbb{Z}^d$ intersects at least two of them with positive probability. But every configuration in the box Λ_n has positive probability and, in particular, that where all edges with both endpoints in Λ_n are occupied does as well. In this configuration, all infinite components intersecting Λ_n became connected and, as there were at least two such components, in the new configuration we have one less than originally. It follows that $\mathbb{P}_p(N \leq k-1) > 0$ as claimed. □

Corollary 4.19 [Zero, One or Infinity] *Consider bond percolation on $\mathbb{Z}^d$ with parameter p. Then exactly one of the events $\{N = 0\}$, $\{N = 1\}$ or $\{N = \infty\}$ has probability one and others have probability zero.*

Proof. The previous lemma showed that if $\mathbb{P}(N = k)$ cannot have probability one for $2 \leq k < \infty$. So, by Corollary 4.17, it must have probability zero. □

Finally, we nail things down even further:

Theorem 4.20 [Burton-Keane's Uniqueness Theorem] *Consider bond percolation on $\mathbb{Z}^d$ with parameter p. Then $\mathbb{P}_p(N = \infty) = 0$.*

The proof is based on the following definition:

Definition 4.21 [Encounter point] *We say that $z \in \mathbb{Z}^d$ is an* encounter point *in configuration ω if $C_\omega(z)$ is infinite, z has exactly three adjacent occupied edges and, disconnecting these edges, $C_\omega(z)$ splits into three infinite connected components.*

Proof of Theorem 4.20. The first step of the proof is to realize that, if $\mathbb{P}_p(N = \infty) = 1$, then the origin has a positive probability to be an encounter point. Indeed, for a sufficiently large box Λ_n we will find at least three distinct infinite clusters that intersect Λ_n with a positive probability. We can connect these clusters by paths to

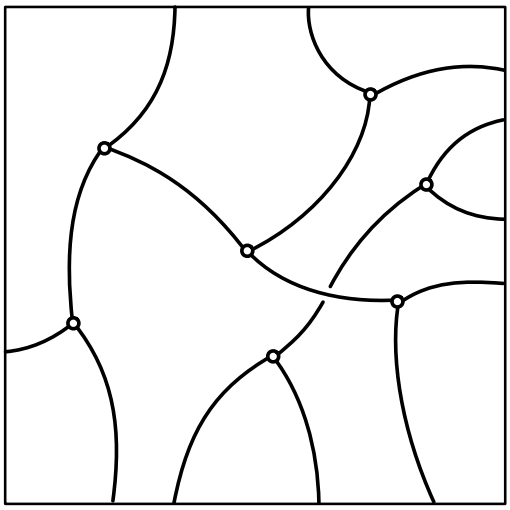

Figure 4.3: An illustration demonstrating the fundamental part of Burton-Keane's argument. There are 7 encounter points in two connected components which lead to at least 7 distinct exit point on the boundary. The number of encounter points in the box is thus bounded by the size of the boundary.

the origin and set all other edges in Λ_n to vacant. This will make the origin an encounter point.

Let $E(v)$ be the event that v is an encounter point. The Ergodic Theorem now implies that if

$$q := \mathbb{P}_p(E(v)) > 0 \tag{4.28}$$

then the box Λ_n will, with probability tending to one, contain at least $q/2|\Lambda_n|$ encounter points.

We will show that this leads to a contradiction: Label encounter points in Λ_n in some fashion, say, $v_1, \ldots, v_m$, and for each v_i define a distinct vertex w_i, called an *exit point*, on the boundary of Λ_n as follows: For v_1 pick a vertex in one of the components meeting at v_1. For v_2 do the same noting that even if v_2 lies on a path between v_1 and w_1, the fact that v_2 is encounter implies the existence of another vertex on the boundary that is not on this path. This applies for each v_i because even if it lies on a path between some v_j, $j < i$, and w_k, $k < i$, there will always be another path from v_i to the boundary that has not yet been used.

We conclude that, given we have $m \geq (q/2)|\Lambda_n|$ encounter points inside Λ_n, we can find as many distinct exit points on $\partial\Lambda_n$. That would imply

$$\frac{|\partial\Lambda_n|}{|\Lambda_n|} \geq \frac{q}{2} \tag{4.29}$$

which is a contradiction for n sufficiently large as $|\partial\Lambda_n|$ grows like n^{d-1} while $|\Lambda_n|$ grows like n^d. Hence we could not have $\mathbb{P}_p(N = \infty) > 0$ to begin with and so, by Corollary 4.19 either $\mathbb{P}_p(N = 0) = 1$ or $\mathbb{P}_p(N = 1) = 1$. □

Exercise 4.22 Derive a uniform lower bound on the probability that, under the condition (4.28), the box Λ_n will contain at least $q/2|\Lambda_n|$ encounter points. Use that the expected number of encounter points in this box is $q|\Lambda_n|$.