Part C: Interacting Particle Systems

Paul Chleboun

Last updated:

November 9, 2018
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Ising and Potts models</td>
<td>44</td>
</tr>
<tr>
<td>12 Gibbs states</td>
<td>45</td>
</tr>
<tr>
<td>13 Mixing and relaxation</td>
<td>46</td>
</tr>
</tbody>
</table>
PART C
INTERACTING PARTICLE SYSTEMS
16 lectures MT 2018

Aims
The aims is to introduce fundamental probabilistic and combinatorial tools, as well as key models, in the theory of discrete disordered systems. We will examine the large-scale behavior of systems containing many interacting components, subject to some random noise. Models of this type have a wealth of applications in statistical physics, biology and beyond, and we will see several key examples in the course. Many of the tools we will discuss are also of independent theoretical interest, and have far reaching applications. For example, we will study the amount of time it takes for these process to reach the stationary distribution (mixing time). This concept is also important in many statistical applications, such as studying the run time of MCMC methods.

Prerequisites:
Discrete and continuous time Markov process on countable state space, as covered for example in Part A A8 Probability and Part B SB3a Applied Probability.

Synopsis:
• Percolation and phase transitions.
• Uniform spanning trees, loop-erased random walks, and Wilson's algorithm.
• Random walks on graphs and electrical networks (discrete potential theory).
• Important models: Stochastic Ising model (Glauber dynamics), Random-cluster model, Contact process, Exclusion process, Hard-core model.
• Important tools: Monotonicity, coupling, duality, FKG inequality.
• Gibbs measures and a discussion of phase transitions in this context.
• Mixing times and the spectral gap.

Reading
Classes and problem sheets

There will be 16 hours of lectures; $L1-L16$. There will be 4 problem classes and assignment sheets. The 1st class will be shorter than usual covering material from the first week of lectures, it will run in Week 2. The 2nd and 3rd will be 'standard' and run in Weeks 4-5 and Weeks 6-8 respectively. Class 4 will take place in Week 1 HT. The final sheet will be longer, and include more stretching material, the sort of problems that would be nice to ask but no time during term. Full solutions will be provided to the students.
Chapter 1

Introduction and Notation

1.1 Introduction

Interacting particle systems (IPS) are Markov processes, in continuous or discrete time, which describe ‘particles’ moving in some underlying discrete space, subject to some random noise and interactions. Such models often arise naturally in theoretical physics (statistical mechanics), and there are now numerous applications of these systems across the natural and social sciences, including in; population genetics, epidemiology, computer science, electrical engineering, and economics. A few examples are given below and on the associated slides.

Interacting particle systems also provide a natural framework to study fundamental phenomena which occur in these applications, such as phase transitions, metastability and relaxation to equilibrium. One of the main goals in this field is to understand and predict emergent behaviour on macroscopic (large) scales, as a result of the microscopic dynamics and interactions of the individual components. In the past decades this field has grown in importance and established itself as one of the most active branches of probability theory.

The microscopic dynamics of interacting particle systems are often very simple to describe, however they are frequently very challenging to analyse, in particular on large or infinite scales. On the other-hand, they are often simple enough to be open to detailed mathematical (rigorous) analysis while still capturing the main features of interest in applications. Qualitative changes in large scale macroscopic behavior depending on the system parameters are often known as ‘collective phenomena’ or phase transitions, and are of particular interest. A couple of real world examples:

1. Figure 1 shows colour-coded data for car speed on the M25 as a function of space and time, as recorded by sensors under the road. The striped patterns of low-speed (red) correspond to stop-and-go waves during rush hour. Often there is no obvious external cause such as an accident, so the pattern is interpreted as an intrinsic collective phenomenon emerging from the interactions of cars on a busy road. A minimal mathematical description of this situation in terms of IPS would be to take a one-dimensional lattice (or a subset thereof), and at each site denote the presence or absence of a car with an occupation number 1 or 0 respectively. In terms of dynamics, we simplify the system so that we can determine if the observed phenomena of interest is a result of relatively simple interactions. We only want to model normal traffic on a single lane road without car crashes or overtaking. So cars are allowed to proceed one lattice site to the right, say, with a given rate 1, provided that the site in front is not occupied by another car. The rate may depend on the surrounding configuration of cars (e.g. number of empty sites ahead), and relatively simple choices depending only on three or four neighboring sites can already lead to interesting patterns and the emergence of stop-and-go waves. The defining features of this process in terms of IPS are that no particles are created or destroyed (conservation of the number of particles) and that there is at most one particle per site (exclusion rule). We will see the most basic example of a model in this class during the course, called the simple exclusion processes, which not only has man applications (including in biological transport systems) but has attracted a great deal of attention in the mathematical literature and is at the heart of many very important results.
2. Figure 2 shows segregation patterns of a microbial species (E. coli) when grown in a nutrient rich medium from a mixed initial population of circular shape. Each individual appears either red or green as a result of a neutral genetic marker that only affects the colour. A possible IPS model of this system might evolve on a two dimensional lattice (say $\mathbb{Z}^2$ for simplicity), where a state is described by assigning to each vertex one of three colours. The dynamics can be modeled by letting each individual split into two with a given rate, and then place the offspring on an empty (white) neighboring site. If there is no empty neighboring site the reproduction rate is zero (or equivalently the offspring is immediately killed). Therefore we have two equivalent species competing for the same resource (empty sites), and spatial segregation is a result of the fact that once the red particles died out in a certain region due to fluctuations, all the offspring is descending from green ancestors. Note that in contrast to the first example, the number of particles in this model is not conserved. The Eden model is a well known model of pure growth in this way. The simplest such process to model extinction or survival of a single species is called contact process.

3. Figure 3 shows a snapshot of a simplified model of magnetic materials, called the Ising model. This is one of the main prototypical models in statistical physics for exploring phase transitions and one of the most studied lattice spin models. Many important developments have been driven by studying this model (and slight generalisations), including the interplay between static (equilibrium properties) and dynamical features. In the Ising model each vertex of $\mathbb{Z}^2$ (or some subset) is assigned a value of either $+1$ (associated with spin-up) or $-1$ (associated with spin-down). The model is then defined by a probability measure on this space, which is parametrized by the system temperature and an external magnetic field. Natural Markovian dynamics can be associated with the model, an a particular special class called the Glauber dynamics have been extensively studied. The figure shows the system switching from one metastable state (mostly $-1$) to a stable one (mostly $+1$) by droplet growth. Metastable dynamics of stochastic systems is currently a rich and diverse research area in its own right.

4. Figure 4 shows a snapshot from a kinetically constrained model under Glauber type dynamics, relaxing to equilibrium. Kinetically constrained models where introduced in the physics literature as minimal models for the dynamics of glassy (or amorphous) materials - such as supercooled liquids. These models have recently attracted significant attention in the mathematics community. Developments here are occurring simultaneously in the more applied and theoretical communities and there is a great deal of interdisciplinary research in the area of glassy-dynamics.

In these notes we will cover the basic definitions, concepts, and tools, that are used in the study of interacting particle systems. Along the way we will introduce some classical examples of these systems systems. We will then focus mainly on the large scale behavior, such as phase transitions, and dynamical properties such as mixing and relaxation. Many of the tools we will discuss are also of independent theoretical interest, and have far reaching applications. For example, we will study the amount of time it takes for these process to reach the stationary distribution (mixing time). This concept is also important in many statistical applications, such as studying the run time of MCMC methods.

### 1.2 Notation

I will attempt to stick to the following notation and terminology throughout the notes. There maybe some need to stray from this now and again to improve clarity.
We will typically denote the state space (i.e., the set of all possible configurations of the system) by \( \Omega \), and configurations will typically be denoted by small Greek letters \( \eta, \sigma, \ldots \in \Omega \). For interacting particle systems \( \Omega \) will be of the form \( S^\Lambda \), where \( S \subset \mathbb{Z} \) is called the local state space, which will typically be finite (e.g., \( S = \{0, 1\} \) corresponding to the presence or absence of a particle at a site for exclusion type models and \( S = \{-1, +1\} \) for the Ising model). \( \Lambda \) will be a countable set which represents the lattice on which the 'particles' live. When \( S \) is finite \( \Omega \) is compact and metrizable with respect to the product topology. We will try to always denote vertices (or sites) in the lattice \( \Lambda \) by small Latin letters, \( x, y, z \in \Lambda \). For \( \eta \in \Omega \) and \( x \in \Lambda \) the local occupation number is denoted by \( \eta(x) \in S \). So a configuration \( \eta \in \Omega \) can be thought of as a vector \( (\eta(x))_{x \in \Lambda} \) indexed by the lattice \( \Lambda \), or as a function \( \eta : \Lambda \to S \). Clearly if \( S \) and \( \Lambda \) are both finite then \( \Omega \) is finite, if \( S \) is countable and \( \Lambda \) is finite then \( \Omega \) is countable, we will typically focus on these two cases. Often when discussing the 'large-scale' behavior we will be considering a sequence of increasingly large systems \( \Omega_n = S^{\Lambda_n} \), where \( |\Lambda_n| \to \infty \) as \( n \to \infty \). As soon as \( |S| > 1 \) and \( \Lambda \) is infinite the state space \( \Omega \) is uncountable. This is often the case in models in statistical mechanics which are defined on the entire lattice \( \mathbb{Z}^d \). For a measurable structure in this case we will take the Borel \( \sigma \)-algebra generated by open cylinders, denoted by \( \mathcal{B}(\Omega) \). We will denote the set of probability measures on \( \Omega \) by \( \mathcal{P}(\Omega) \). Probability measures on \( \Omega \) will typically be called \( \mu, \nu \) or \( \pi \in \mathcal{P}(\Omega) \).

When the state space is relatively low dimensional, e.g., \( \Omega \subset \mathbb{Z} \) we may use capital letters to denote the process e.g. \( (X_t)_{t \geq 0} \).

We will consider chains in continuous time,

\[
t \in \mathbb{R}^+ = [0, \infty)
\]

and also sometimes in discrete time

\[
n \in \mathbb{N}_0 = \{0, 1, \ldots\}.
\]

The letters \( n, m, k \) will typically denote integers, while we reserve \( t \) and \( s \) for real numbers (times).

We write \( (\eta_t)_{t \geq 0} \) for a continuous time process and \( (\eta_n)_{n \geq 0} \) for a discrete time process, where it is hopefully clear from context which one we are using. We will typically denote the probability measure associated with the process (on path space) by \( \mathbb{P} \). In continuous time the canonical path space is given by

\[
D[0, \infty) = \{ \eta : [0, \infty) \to \Omega \ \text{càdlàg} \},
\]

we will be more specific about the probability space etc. when we look at specific models. The associated expectation of a random variable \( f \) will be written as \( \mathbb{E}[f] \).

For a graph \( G = (V, E) \) we will write \( \sigma \sim \eta \iff \{\sigma, \eta\} \in E \).

### 1.3 Brief recap

This section gives a very brief recap of discrete and continuous time Markov processes on countable state spaces, and some important concepts such as reversibility. For more detail see for example Part A Probability and Part B Applied probability notes, or the book 'Markov Chains' by J.R. Norris.

Let \( \Omega \) be a countable set. Let \( P = (P(u, v))_{u, v \in \Omega} \) be a stochastic matrix on \( \Omega \), i.e.,

\[
P(\eta, \sigma) \geq 0, \ \forall \eta, \sigma \in \Omega \quad \text{and} \quad \sum_{\sigma \in \Omega} P(\eta, \sigma) = 1 \ \forall \eta \in \Omega.
\]
We say that the $\Omega$-valued random variable $(X_n)_{n \geq 0}$ is a Markov chain with transition matrix $P$ and initial distribution $\nu$ if its law, denoted by $\mathbb{P}_\nu$, is determined by the following: For all $n \geq 1$, $\xi, \zeta, \sigma_1, \ldots, \sigma_{n-1} \in \Omega$, and all events $H_{n-1} = \bigcap_{i=0}^{n-1} \{X_i = \sigma_i\}$ satisfying $\mathbb{P}_\nu(H_{n-1} \cap \{X_n = \xi\}) \neq 0$, we have

(i) $\mathbb{P}_\nu(X_0 = \xi) = \nu(\xi)$;

(ii) $\mathbb{P}_\nu(X_{n+1} = \xi \mid H_{n-1} \cap \{X_n = \xi\}) = P(\xi, \xi)$.

Equation (ii), often called the Markov property, means that the conditional probability of proceeding from state $\xi$ to state $\sigma$ is the same no matter what the sequence of states preceding the current state $\xi$, i.e. the past and future are conditionally independent given the present. If $\nu = \delta_\sigma$ then we write $\mathbb{P}_\sigma$.

**Lemma 1.1.** A discrete-time random process $(X_n)_{n \geq 0}$ is a Markov chain with initial distribution $\nu$ and transition metric $P$ if and only if for all $n \geq 0$ and $\sigma_0, \ldots, \sigma_n \in \Omega$

$$\mathbb{P}(X_0 = \sigma_0, X_1 = \sigma_1, \ldots, X_n = \sigma_n) = \nu(\sigma_0)P(\sigma_0, \sigma_1)P(\sigma_1, \sigma_2)\ldots P(\sigma_{n-1}, \sigma_n).$$

**Proof.** Exercise. \qed

We denote the distribution at time $n$ started from $\nu$ by $\nu_n$, i.e.

$$\nu_n(\sigma) : = \mathbb{P}_\nu(X_n = \sigma) = (\nu_{n-1}P)(\sigma) = \sum_{\xi \in \Omega} \nu_{n-1}(\xi) P(\xi, \sigma),$$

or $\nu_n = \nu_{n-1}P$, so $\nu_n = \nu_0P^n \forall n \geq 0$.

The transition matrix acts on functions (observables), which by convention we think of as column vectors, as

$$Pf(\sigma) = \sum_{\xi \in \Omega} P(\sigma, \xi) f(\xi) = \sum_{\xi \in \Omega} f(\xi) \mathbb{P}_\sigma(X_1 = \xi) = \mathbb{E}_\sigma[f(X_1)]$$

in general $P^n f(\sigma) = \mathbb{E}_\sigma[f(X_n)]$ for $n \geq 1$.

That is the $\sigma^{th}$ entry of $Pf$ tells us the expected value of the function $f$ after one time step, given the initial distribution was $\delta_\sigma$.

**Definition 1.2** (Stationary distribution). A distribution probability $\pi$ on $\Omega$ is called stationary for the chain if $\pi P = \pi$, equivalently

$$\pi(\sigma) = \sum_{\xi \in \Omega} \pi(\xi) P(\xi, \sigma) \quad \forall \sigma \in \Omega.$$ 

Clearly if $\pi$ is stationary and $\nu_0 = \pi$ then $\nu_n = \pi$ for all $n \geq 0$.

The Markov chain is said to be irreducible if it is possible to go from any state to any other using only transitions of positive probability.

**Definition 1.3** (Irreducible). A chain $P$ is irreducible if for any $\sigma, \xi \in \Omega$ there exists an $n \in \mathbb{N}$ such that $P^n(\sigma, \xi) > 0$.

Recall the definition of recurrence and positive recurrence.

**Theorem 1.4.** Let $P$ be irreducible Markov transition matrix. Then

(i) $P$ has a stationary distribution if and only if $P$ is positive recurrent.
BRIEF RECAP

(ii) In that case, the stationary distribution is unique and is given by \( \pi(\sigma) = \frac{1}{\mathbb{E}_{\tau_{\sigma}^+}} \) where \( \tau_{\sigma}^+ := \min\{n \geq 1 \mid X_n = \sigma\} \) is the first return time to state \( \sigma \).

Proof. Seen previously, see e.g. Prop 1.14 and Corollary 1.17 in [2].

We will call a chain ergodic if it has a unique stationary distribution.

**Definition 1.5** (Aperiodic). A discrete time Markov chain is called aperiodic if for all \( \sigma \in \Omega \), \( P^n(\sigma, \sigma) \) is eventually positive, i.e.

\[
\text{there exists } N_\sigma \in \mathbb{N} \text{ such that } P^n(\sigma, \sigma) > 0 \text{ for all } n \geq N_\sigma.
\]

**Theorem 1.6** (Convergence to equilibrium). An irreducible, aperiodic, Markov chain with finite state space is ergodic, i.e. it has a unique stationary distribution \( \pi \) and

\[
P^n(\sigma, \xi) = \mathbb{P}_\sigma(X_n = \xi) \to \pi(\xi) \text{ as } n \to \infty, \quad \forall \sigma, \xi \in \Omega.
\]

**Theorem 1.7** (Ergodic Theorem). Let \( f \) be a real valued function on \( \Omega \) and \( (X_n)_{n \geq 0} \) an irreducible, positive recurrent, Markov chain with stationary distribution \( \pi \), then for any initial distribution \( \nu \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} f(X_m) \to \mathbb{E}_\pi[f] \text{ a.s. as } n \to \infty.
\]

In words “time averages converge to space averages.”

Proof. See e.g. [2] Section 4.3

A Markov chains past and future are independent given the present. However, consider an ergodic chain (irreducible and aperiodic) starting from a point mass on a single state, convergence to equilibrium demonstrates an asymmetrical behavior, i.e. increasing entropy (or uncertainty). On the other hand, a Markov chain in equilibrium (started from a stationary distribution) run backward is again a Markov chain, although possibly not the same one.

**Theorem 1.8.** Suppose \( (X_n)_{n \geq 0} \) is a Markov chain with irreducible (positive recurrent) transition matrix \( P \) and initial distribution \( \pi \) the unique stationary distribution. Fix \( N > 0 \), then \( (\tilde{X}_n)_{n=0}^{N} \) given by \( \tilde{X}_n = X_{N-n} \) is a Markov chain with initial distribution \( \pi \) and transition matrix

\[
\tilde{P}(\sigma, \xi) = \frac{\pi(\xi)P(\xi, \sigma)}{\pi(\sigma)}.
\]

Proof. See [Thm 1.9.1 in Markov Chains, Norris].

**Definition 1.9.** A Markov chain is called reversible with respect to \( \pi \) if the forward and reversed chains have the same distribution, i.e. \( P = \tilde{P} \).

**Definition 1.10.** A transition matrix \( P \) satisfies the detailed balance condition with respect to measure \( \pi \in \mathcal{P}(\Omega) \) if

\[
\pi(\sigma)P(\sigma, \xi) = \pi(\xi)P(\xi, \sigma) \quad \forall \sigma, \xi \in \Omega.
\]

**Proposition 1.11.** If \( P \) satisfies the detailed balance condition with respect to measure \( \pi \), then \( \pi \) is a stationary distribution for the Markov chain generated by \( P \). A Markov chain is reversible with respect to \( \pi \) if and only if it’s transition matrix satisfies the detailed balance condition with respect to \( \pi \).

Proof. Exercises.
1.3. BRIEF RECAP

Remark 1.12. A Markov chain with transition matrix $P$ is reversible with respect to $\pi$ if and only if $P$ is self adjoint in $\ell^2(\pi)$, with inner product $\langle f, g \rangle_\pi = \sum_{\sigma \in \Omega} f(\sigma) g(\sigma) \pi(\sigma)$. That is, for all square integrable random variables $f, g \in \ell^2(\pi)$

$$\langle P f, g \rangle_\pi := \sum_{\sigma \in \Omega} \pi(\sigma) P f(\sigma) g(\sigma) = \sum_{\sigma \in \Omega} \pi(\sigma) g(\sigma) \sum_{\xi \in \Omega} P(\sigma, \xi) f(\xi)$$

$$= \sum_{\sigma, \xi \in \Omega} \pi(\xi) P(\xi, \sigma) g(\sigma) f(\xi) = \langle f, P g \rangle_\pi$$

You should revise/be familiar with the analogous results for continuous time chains on countable state space. In particular in terms of the ‘Q-matrix’ or (infinitesimal) generator of the process, and the correspondence with the jump-chain/holding time description (see Fig. 1.1).

We will often denote the generator (Q-matrix) of a continuous time process on countable state space as either $Q = \{q(\sigma, \eta)\}_{\sigma, \eta \in \Omega}$ or by $L$. Recall that if the state space is finite then the continuous time transition matrix (Markov semi-group) can be expressed as $P_t = e^{tL}$. These satisfy the backward and forward equations

$$P'_t = L P_t = P_t L \quad \text{with} \quad P_0 = I.$$  \hfill (1.3.1)
Chapter 2

Electrical networks

Any Markov process on a countable state space can be thought of as a random walk on a (directed) graph, possibly with self loops, where the vertices’ are the states of the process and edges represent possible transitions. It turns out that there is an intimate connection between reversible random walks on graphs and electrical networks. This connection gives the discrete analogue of the deep connection between continuous potential theory and Brownian motion.

The purpose of this chapter is to give an introduction to this discrete potential theory. In the next lecture we will see a nice application of this theory to transience and recurrence of random walks in $\mathbb{Z}^d$, summarised in the beautiful theorem of Pólya.

In countable or finite state spaces the electrical network analogy looks essentially the same in continuous and discrete time (as a consequence of the jump chain/holding time construction). We will write things predominantly in terms of a discrete time chain, but try to make it clear how things would work similarly in continuous time as we go. When the state space is uncountable there are some extra subtleties that we will not discuss here. A nice and much more detailed account of this subject is given by Doyle and Snell, Random Walks and Electric Networks (available on arXiv).

2.1 Set-up

To any reversible Markov chain on finite state space $\Omega$, with transition matrix $P$ and reversible distribution $\pi$, we can associate an (undirected) graph $G = (\Omega, E)$ and weights $c: E \to (0, \infty)$. We will drop the usual assumption that $G$ is irreflexive, that is $E$ may contain ‘loops’ i.e. edges of the form $\{\eta, \eta\}$ for $\eta \in \Omega$.

The edge set is given by $\{\eta, \sigma\} \in E$ if and only if $P(\eta, \sigma) > 0$ (or analogously in continuous time $q(\eta, \sigma) > 0$). Note that the definition of the edge set makes sense since $P(\eta, \sigma) > 0$ if and only if $P(\sigma, \eta) > 0$ by reversibility (the detailed balance condition). We put a weight function on the edges, $c: E \to (0, \infty)$, specified by $c(\{\eta, \sigma\}) = \pi(\eta) P(\eta, \sigma)$. This is symmetric with respect to switching $\eta$ and $\sigma$ again by detailed balance. We think of $c(\{\eta, \sigma\})$ as the conductance of the edge $\{\eta, \sigma\}$, or equivalently the edge has a resistor on it with resistance $r(\{\eta, \sigma\}) = 1/c(\{\eta, \sigma\})$. With a slight abuse of notation we will write $c(\sigma, \eta)$. It is straightforward to check that $\pi(\eta) = c(\eta) := \sum_{\sigma \sim \eta} c(\eta, \sigma)$. We will call a graph together with a weight function of this kind a network.

We could also start with a finite graph and associated conductances $c: E \to (0, \infty)$, and use these to construct a reversible Markov process via the relation $P(\eta, \sigma) = c(\eta, \sigma)/c(\eta)$. In this case it turns out that the chain is reversible with respect to the probability distribution given by $\pi(\eta) = c(\eta)/c_\Omega$ where $c_\Omega = \sum_{\eta \in \Omega} c(\eta)$.

For $A \subset \Omega$ define the hitting time and first return time as

$$\tau_A = \min\{n \geq 0 : X_n \in A\}$$
$$\tau_A^+ = \min\{n \geq 1 : X_n \in A\}.$$

In problems such as the gambler’s ruin we are often interested in the probability of reaching some point of the state space before another. Consider the probability that the chain started from $\eta \in \Omega$ hits a set $A$ before the set $B$ where $A \cap B = \emptyset$, and let

$$f(\eta) = \mathbb{P}_\eta(\tau_A < \tau_B).$$
Then \( f |_A \equiv 1, f |_B \equiv 0 \) and by the Markov property

\[
f(\eta) = \sum_{\sigma \sim \eta} P(\eta, \sigma)f(\sigma).
\]

In words this says that the value of the function at a point is equal to its local average, \( f(\eta) = \mathbb{E}_\eta[f(X_1)] \) (notice that this does not require the chain to be reversible). This property turns out to be very important and \( f \) is called harmonic on \((A \cup B)^c\) with respect to \( P \). These harmonic functions share many of the properties that you may be familiar with from the 'non discrete setting' (i.e. harmonic functions on \( \mathbb{R} \)), such as the mean value property above and the following theorems. We introduce some notation for the external boundary of a set \( A \subset \Omega \) in \( G \)

\[
\partial A = \{\sigma \in A^c : \sigma \sim \eta \text{ for some } \eta \in A\}.
\]

The following are stated for finite graphs but with week conditions can be easily extended to countable graphs.

**Theorem 2.1** (Maximum Principle). Let \( A \subset \Omega \) be a connected component of \( G \), if \( f : \Omega \to \mathbb{R} \) is harmonic on \( A \) and attains its supremum at \( \eta_0 \in A \), then \( f \) is constant on \( A \cup \partial A \).

**Proof.** Let \( K := \{\sigma \in \Omega : f(\sigma) = f(\eta_0)\} \). If \( \eta \in A \cap K \) and \( P(\eta, \sigma) > 0 \) then \( \sigma \in K \), since \( f \) is harmonic at \( \eta \) and takes its largest value at \( \eta \). We can proceed by iteration. \( \square \)

**Theorem 2.2** (Uniqueness Principle). Let \( G \) be connected, and \( A \) a proper subset of \( \Omega \). If \( f, g : \Omega \to \mathbb{R} \) are harmonic on \( A \) and \( f(\eta) = g(\eta) \) for all \( \eta \in A^c \) then \( f \equiv g \).

**Proof.** The proof follows from the Maximum Principle, exercise. \( \square \)

**Theorem 2.3** (Existence Principle). Let \( A \) be a proper subset of \( \Omega \), if \( f_0 : \Omega \setminus A \to \mathbb{R} \) then there exists an \( f : \Omega \to \mathbb{R} \) such that \( f |_{A^c} \equiv f_0 \) and \( f \) is harmonic on \( A \).

**Proof.** Check that \( f(\eta) = \mathbb{E}_\eta[f_0(X_{\tau_A})] \) gives such a function by using the usual one-step analysis. \( \square \)

### 2.2 Electrickery

We will now define potential differences (voltages) and current flows on the network described in the previous section, and give some probabilistic interpretations. Several nice results then follow from physical principles of electrical networks. We will denote the set of oriented edges \( \mathcal{E} = \{(\eta, \sigma) \in \Omega^2 : P(\eta, \sigma) > 0\} \).

We may think of real valued functions on \( \Omega \) as potentials. To any potential function \( \varphi : \Omega \to \mathbb{R} \) we may associate an induced current \( j : \mathcal{E} \to \mathbb{R} \) (and vice-versa) according to Ohm’s Law, which states that the current is the conductance times the gradient of the potential.

**Ohm’s Law:** If \( \eta \sim \sigma \) then the current \( j(\eta, \sigma) \) from \( \eta \) to \( \sigma \) is given by

\[
j(\eta, \sigma) = c(\eta, \sigma)(\varphi(\eta) - \varphi(\sigma)).
\]

Note that \( j \) is always anti-symmetric, i.e. \( j(\eta, \sigma) = -j(\sigma, \eta) \), since the conductances are symmetric. We will find the following definition very useful in this context.

**Definition 2.4.** Let \( A \) and \( B \) be disjoint subsets of \( \Omega \). We call an anti-symmetric function \( \theta : \mathcal{E} \to \mathbb{R} \) a flow from \( A \) to \( B \) if

- \( \text{div } \theta(\eta) := \sum_{\sigma \sim \eta} \theta(\eta, \sigma) = 0 \) for all \( \eta \notin A \cup B \) (“flow in is equal to flow-out”).
• \( \text{div } \theta(\sigma) \geq 0 \) for all \( \sigma \in A \).
• \( \text{div } \theta(\sigma) \leq 0 \) for all \( \sigma \in B \).

The strength of a flow \( |i| = \sum_{\sigma \in A} \text{div } \theta(\sigma) \), and if \( |\theta| = 1 \) we call \( \theta \) a unit flow.

**Exercises (uniqueness):** Check that if \( i \) and \( j \) are both flows from \( A = \{a\} \) to \( B = \{b\} \) with strength \( |i| = |j| \) then \( i \equiv j \).

If we attach a battery to the graph such that the vertices in set \( A \) are at a positive voltage, \( \phi \), and those in \( B \) at zero, then an equilibrium current will flow which satisfies Kirchhoff’s Node Law.

**Kirchhoff’s Current Law:** The equilibrium current, \( i_{AB}^\phi : E \rightarrow \mathbb{R} \), associated with a fixed voltage \( \phi \) across \( A \) and \( B \), is a flow from \( A \) to \( B \).

There are several interpretations of these laws. Firstly you could take Ohm’s law as a definition of the current given a potential. Then Kirchhoff’s law is a property of any current given by a potential which is fixed on \( A \) and \( B \) and harmonic on \((A \cup B)^c\). Alternatively you could take Kirchhoff’s law together with Ohm’s law as a definition of the equilibrium current and voltage. We described this more precisely next.

Suppose we attach a battery with voltage \( \phi \) on \( A \) and 0 on \( B \). Then by Kirchhoff’s current law this induces some current which is a flow called \( i_{AB}^\phi : E \rightarrow \mathbb{R} \). By Ohm’s law this current must be the gradient of some potential, we assume that this potential must be \( \phi \) on \( A \) and 0 on \( B \), then it turns out this uniquely determines the potential and the current flow. We denote by \( v_{AB}^\phi : \Omega \rightarrow \mathbb{R} \) the potential induced by this voltage \( \phi \), i.e., by its restriction to the sets \( A \) and \( B \), together with Kirchhoff’s current law and Ohm’s law. Ohm’s law together with Kirchhoff’s current law imply that the voltage must be a harmonic function on \((A \cup B)^c\), and then its restriction to \( A \) and \( B \) determine it uniquely. This immediately gives us a probabilistic interpretation of the voltage.

**Theorem 2.5.** The voltage (potential function) \( v_{AB}^\phi : \Omega \rightarrow \mathbb{R} \) induced by a voltage \( \phi \) applied to sets \( A \) and \( B \), that is \( v_{AB}^\phi | A = \phi, v_{AB}^\phi | B = 0 \) and \( v_{AB}^\phi \) satisfies Ohm’s law with respect to the current flow \( i_{AB}^\phi \), is harmonic on \((A \cup B)^c\). Furthermore \( v_{AB}^\phi(\eta) = \phi \mathbb{P}_\eta(\tau_A < \tau_B) \).

**Proof.** Let \( U = (A \cup B)^c \), and suppose \( v \) satisfies the conditions of the theorem, then by Kirchhoff’s current law and Ohm’s law

\[
\text{div } i(\eta) = \sum_{\sigma \sim \eta} i(\eta, \sigma) = \sum_{\sigma \sim \eta} c(\eta, \sigma)(v(\eta) - v(\sigma)) = 0 \quad \text{for } \eta \in U.
\]

Rearranging the final equality and using the fact that the conductances are symmetric we find

\[
v(\eta) = \frac{1}{c(\eta)} \sum_{\sigma \sim \eta} c(\eta, \sigma)v(\sigma)
\]

i.e. \( v \) is harmonic on \( U \). So, by the uniqueness theorem, \( v_{AB}^\phi \) is the unique harmonic function on \( U \) satisfying the boundary conditions \( v | A = \phi \) and \( v | B = 0 \). Since \( v'(\cdot) = v_{AB}^\phi(\cdot)/\phi \) is harmonic on \( U \) and satisfies \( v' | A = 1 \) and \( v' | B = 0 \) (simply check), then by uniqueness \( v'(\eta) = \mathbb{P}_\eta(\tau_A < \tau_B) \), i.e. \( v(\eta) = \phi \mathbb{P}_\eta(\tau_A < \tau_B) \).

There are also other interpretations of the voltage as well as probabilistic and combinatorial interpretations for the current flow. We will see some more of these in the next section. We now
2.3 Effective Resistance and Energy

The strength of the current, $i$, that flows when we apply a voltage $\phi$ on $A$ and 0 on $B$, as in Theorem 2.5, can be calculated as follows,

$$\text{div} i^\phi_{A,B}(a) = \sum_{\sigma \sim a} i_{A,B}^\phi(a, \sigma) = \sum_{\sigma \sim a} c(a, \sigma) \left[ \phi - \phi P_\sigma(\tau_A < \tau_B) \right]$$

$$= \phi \sum_{\sigma \sim a} c(a, \sigma) \left[ 1 - P_\sigma(\tau_A < \tau_B) \right] = \phi \sum_{\sigma \sim a} c(a, \sigma) P_\sigma(\tau_A \geq \tau_B)$$

$$= \phi \, \pi(a) \sum_{\sigma \sim a} P(a, \sigma) P_\sigma(\tau_A \geq \tau_B) = \phi \, \pi(a) P_a(\tau^+_A > \tau_B).$$

It follows that

$$|i^\phi_{A,B}| = \sum_{a \in A} \text{div} i^\phi_{A,B}(a) = \phi \sum_{a \in A} \pi(a) P_a(\tau^+_A > \tau_B). \tag{2.3.1}$$

Now in analogy with Ohm’s law we define the effective conductance as

$$C(A, B) := \sum_{a \in A} \pi(a) P_a(\tau^+_A > \tau_B),$$

which is independent of the applied voltage. The effective resistance $R(A, B) = 1/C(A, B)$. Notice that for a given voltage $\phi$ over the sets $A$ and $B$, the strength of the associated current flow is $\phi \, C(A, B)$, equivalently an applied voltage equal to $R(A, B)$ induces a unit current flow. This holds for the original network and the reduced network $G' = (\{A, B\}, \{\{A, B\}\})$ with $c(A, B) = C(A, B)$. The associated voltage that induces this unit flow is denoted $v_{A,B}(\cdot) := R(A, B) i_{A,B}^\phi(\cdot)$. We will call this (equilibrium) unit current flow $i_{A,B}$, furthermore it is simple to check that $i_{A,B}(\cdot) \equiv i_{A,B}^\phi(\cdot) R(A, B)/\phi$.

We now define the energy dissipated in a flow, and observe that the energy dissipated in the flow associated with potential $\phi$ over $A$ and $B$ is the same on $G$ as on a single resistor with resistance $R_{A,B}$ (further motivation for calling it effective resistance).

Definition 2.7. The energy of a flow $\theta$ on $G$ is given by

$$\mathcal{D}(\theta) = \sum_{e \in E} \theta(e)^2 r(e) = \frac{1}{2} \sum_{e \in E} \theta(e)^2 r(e).$$

Note the first sum is over unoriented edges (which makes sense because we square an antisymmetric) and the second is over all oriented edges in the graph. Consider the energy dissipated by a current $I_{A,B}$ flowing across a single edge with resistance $R(A, B)$ when a voltage
of \( \phi \) is applied, given by
\[
D(I_{A,B}) = \phi^2 C(A,B)^2 R(A,B) = \phi |I_{A,B}|.
\]
We observe that this is the same as the energy of the equilibrium flow \( i^\phi_{A,B} \) on \( G \) described above (suppressing the sub and superscripts for clarity),
\[
D(i^\phi_{A,B}) = \frac{1}{2} \sum_{(\eta,\sigma) \in E} i^\phi_{A,B}(\eta,\sigma)(v(\eta) - v(\sigma)) = \frac{1}{2} \left( \sum_{\eta \in \Omega} v(\eta) \text{div} i(\eta) + \sum_{\sigma \in \Omega} v(\sigma) \text{div} i(\sigma) \right)
\]
(since the potential (voltage) on \( B \) is zero) = \( \phi \sum_{\eta \in A} \text{div} i(\eta) = \phi |i| = \phi |I_{A,B}| \).

where in the final equality we have used the fact that the strength of the induced flow is equal to \( \phi C(A,B) \) on both networks.

We may give an equivalent definition of the effective resistance in terms of the energy.

**Theorem 2.8 (Thomson's principle).** Let \( A \) and \( B \) be disjoint subsets of a finite connected (weighted) graph \( G \), then
\[
R(A,B) = \frac{1}{C(A,B)} = \inf \{ D(\theta) : \theta \text{ is a unit flow from } A \text{ to } B \},
\]
and the unique minimum is attained by \( i_{A,B} \) (the unit strength equilibrium flow defined above).

**Proof.** It is clear from the calculation above that \( D(i_{A,B}) = R(A,B) \). Let \( j \) be a unit flow from \( A \) to \( B \) and define \( \delta = j - i_{A,B} \). It is straightforward to check that \( \delta \) is a strength zero flow from \( A \) to \( B \), i.e.
\[
\sum_{a \in A} \text{div} \delta(a) = \sum_{b \in B} \text{div} \delta(b) = \text{div} \delta(\eta) = 0 \quad \text{for each } \eta \notin A \cup B.
\]
Also,
\[
D(j) = D(i) + D(\delta) + \sum_{e \in E} \delta(e) i(e) r(e)
\]
\[
= D(i) + D(\delta) + \sum_{(\eta,\sigma) \in E} (v(\eta) - v(\sigma)) \delta(\eta,\sigma)
\]
\[
= D(i) + D(\delta).
\]
It follows that \( D(j) > D(i) \) as soon as \( \delta \neq 0 \). \( \square \)

We can also define the energy dissipated by a function \( f : \Omega \to \mathbb{R} \) acting as a potential (voltage). With a slight abuse of notation we use the same letter since it will be clear from context if the argument is a ‘voltage’ or a current flow.

**Definition 2.9.** The energy of a function \( f : \Omega \to \mathbb{R} \) on the weighted graph \( G \), also called the **Dirichlet form** of \( f \), is given by
\[
D(f) = \frac{1}{2} \sum_{(\eta,\sigma) \in E} c(\eta,\sigma)(f(\eta) - f(\sigma))^2.
\]

The Dirichlet form is often expressed in terms of the inner product on \( l^2(\pi) \), this will be explored a little on the first assignment sheet. Note that, from Ohm’s law, \( D(i^\phi_{A,B}) = D(v^\phi_{A,B}) \), so it makes sense to use related notation.

\[2.3. \text{EFFECTIVE RESISTANCE AND ENERGY}\]
2.4. SOME FURTHER PROBABILISTIC INTERPRETATIONS

**Theorem 2.10** (Dirichlet’s principle). Let $A$ and $B$ be disjoint subsets of a finite connect (weighted) graph $G$, then

$$C(A, B) = \inf \{D(f) : f : \Omega \rightarrow \mathbb{R}, f \upharpoonright_A \equiv 1 \text{ and } f \upharpoonright_B \equiv 0\},$$

and the minimum is attained at $v^{1}_{A,B}$.

Note that the effective conductance $C(A, B)$ is often also called the capacity. In our ‘dimensionless’ electrical network setting it turns out that these two physical concepts coincided.

**Proof.** The proof is essentially the same as for Thomson’s principle.

To see that $C(A, B) = D(v_{A,B})$ we appeal to Thomson’s principle. By Eq. 2.3.1 we have $|i^1_{A,B}| = C(A, B)$, i.e. $i^1_{A,B} = C(A, B)i_{A,B}$, so $D(v_{A,B}) = D(i^1_{A,B}) = C(A, B)^2/C(A, B)$ by Thomson’s principle.

Fix a function $f : \Omega \rightarrow \mathbb{R}$ such that $f \upharpoonright_A \equiv 1$ and $f \upharpoonright_B \equiv 0$, then define $h := f - v_{A,B}$. A simple calculation gives

$$D(f) = D(v_{A,B}) + D(h) + \sum_{(\eta, \sigma) \in E} c(\eta, \sigma)(v_{A,B}(\eta) - v_{A,B}(\sigma))(h(\eta) - h(\sigma))$$

$$= D(v_{A,B}) + D(h) + \sum_{(\eta, \sigma) \in E} (h(\eta) - h(\sigma))i^1_{A,B}(\eta, \sigma)$$

$$= D(v_{A,B}) + D(h),$$

where the last equality follows since $h \upharpoonright_A \equiv h \upharpoonright_B \equiv 0$ and $\text{div } i^1_{A,B}(\eta) = 0$ for $\eta \notin A \cup B$. It follows that $D(f) > D(v_{A,B})$ whenever $h \neq 0$.

Together Thomson’s principle and Dirichlet’s principle give two variational principles for the effective conductance, the former can be used to get lower bounds on the effective conductance by considering ‘test flows’ on $G$, and the latter can give upper bounds by consider ‘test functions’ on $\Omega$.

2.4 Some further probabilistic interpretations

The *Green’s function*, $\mathcal{G}_\tau(a, \eta)$, for a random walk started from $a$ and stopped at a stopping time $\tau$ is defined by the expected number of visits to $\eta$ before time $\tau$, i.e.

$$\mathcal{G}_\tau(a, \eta) := \mathbb{E}_a \left[ \sum_{n=0}^{\infty} 1(X_n = \eta, n < \tau) \right].$$

The following interpretation of the effective resistance is a simple consequence of the calculations in the previous section.

**Lemma 2.11.** Let $(X_n)$ be the random walk on weighted graph $G$, fix $B \subset \Omega$ and $a \notin B$, then

$$\mathcal{G}_{\tau_B}(a, a) = \pi(a)R(\{a\}, B),$$

and more generally (recall $\bar{v}_{(\{a\}, B)}$ is the voltage the induces a unit current flow from $\{a\}$ to $B$)

$$\mathcal{G}_{\tau_B}(a, \eta) = \pi(\eta)\bar{v}_{(\{a\}, B)}(\eta),$$

**Sketch proof.** The number of visits to $a$ before visiting $B$, started from $a$, has a geometric distribution with parameter $\mathbb{P}_a(\tau_B < \tau^+_a)$, by definition of the effective resistance this completes the proof by the first part.

For the second part we observe that it is sufficient to show that $\mathcal{G}_{\tau_B}(a, \eta)/\pi(\eta)$ is harmonic on $(\{a\} \cup B)^c$. This follows from one-step analysis together with the identity $\mathcal{G}_{\tau_B}(a, \eta)\pi(a) = \mathcal{G}_{\tau_B}(\eta, a)\pi(\eta)$ from problem Sheet 1.
We give another probabilistic interpretation of the current flow, which is particularly appealing if we take a naive view of electrical current as charged particles moving on the network. If a unit current flows from $a$ to $b$ then the current flow on all the edges $(\eta, \sigma)$ is equal to the net number of times that the process started from $a$ will cross $(\eta, \sigma)$ before it reaches $b$. As we have already observe the unit current is proportional to $i_{[a],[b]}^1$ and the constant of proportionality is the effective resistance.

**Lemma 2.12.** Let $(X_n)$ be a random walk on $G$ that is absorbed at $b$, and for $(\eta, \sigma) \in \mathcal{E}$ let $S_{\eta,\sigma}$ be the number of transitions from $\eta$ to $\sigma$ before $(X_n)$ is absorbed, then $\mathbb{E}_a[S_{\eta,\sigma} - S_{\sigma,\eta}] = i_{[a],[b]}(\eta, \sigma)$.

**Proof.** The proof follows by a simple calculation and using Lemma 2.11, and is left as an exercises. \qed
Chapter 3

Pólya’s Theorem

Up to this point we have focused on finite networks. In this Chapter we will see how the electrical network analogy can help us to establish if processes on countable graphs are transient or recurrent in term of limits of finite graphs.

3.1 Network reduction

There are several ways of simplifying electrical networks without changing the quantities of interest, which come in very handy when trying to do calculations. This is one reason that the electrical network analogy can be helpful. The simplest, and most familiar of these, will be the Series and Parallel laws. For the following fix a graph $\mathcal{G} = (\Omega, \mathcal{E})$ with conductances (resistances) $c : \mathcal{E} \to \mathbb{R}$ (respectively $r$), and $A, B \subset \Omega$ disjoint with potential difference $\phi$ applied across $A$ and $B$.

Lemma 3.1 (Series Law). Two resistors $r_1$ and $r_2$ in series are equivalent to a single resistor with resistance $r_1 + r_2$. That is if $\eta \in (A \cup B)^c$ is a vertex with degree 2 with neighbours $\sigma_1, \sigma_2$, and we replace the edges $\{\eta, \sigma_1\}, \{\eta, \sigma_2\}$ with a single edge $\{\sigma_1, \sigma_2\}$ with resistance $r(\eta, \sigma_1) + r(\eta, \sigma_2)$ then all the potentials and currents in $\mathcal{G} \setminus \{\eta\}$ are unchanged, and the current flowing from $\sigma_1$ to $\sigma_2$ equals $i(\sigma_1, \eta)$.

Note that, since this network reduction does not change the potential or current flows on any of the vertices or edges in the reduced graph, we must have that all the effective resistances calculated between sets that do not contain $\eta$ must also be the same (we define them in terms of the strength of the currents and potentials). In particular if we can reduce the entire network in this way to a single resistor between the sets we are interested in then the value of the resistor must be equal to the effective resistance.

Example 3.2. Consider a random walk on $\mathbb{Z}$ that moves right with probability $p$ and left with probability $q = 1 - p$. Now for the usual Gamblers ruin problem we would be interested in $\mathbb{P}_k(\tau_0 < \tau_N)$. From detailed balance we have $c(i, i+1) = (p/q)^i$ (e.g. take as reversible measure $\pi(i) = \frac{1}{q} \left(\frac{p}{q}\right)^i$ which is actually a probability measure on $\mathbb{N}_0$ if $p < q$, but we don’t mind if it is not normalised). We may now apply the series law ‘on either side’ of the initial state $k$. The resistance between 0 and $k$ is $R(\{0\}, \{k\}) = \sum_{i=0}^{k-1} (q/p)^i$ and the resistance between $k$ and $N$ is $R(\{k\}, \{0\}) = \sum_{i=k}^{N-1} (q/p)^i$. By our previous observations $v^1_{0,N}(k) = \mathbb{P}_k(\tau_0 < \tau_N)$. Since potentials and currents are unchanged by the application of the series law we know the strength of the flow $i^1_{0,N}$ is equal in both networks, and is $C(\{0\}, \{N\})$ by definition. By Ohm’s law on the reduced network $v^1_{0,N}(0) - v^1_{0,N}(k) = 1 - v^1_{0,N}(k) = C(\{0\}, \{N\})R(\{0\}, \{k\})$, therefore it follows that the probability of ruin is $[(p/q)^N - 1] / [(p/q)^N - 1]$.

Lemma 3.3 (Parallel Law). Two conductors $c_1$ and $c_2$ in parallel are equivalent to a single conductor with conductance $c_1 + c_2$. That is if $e_1$ and $e_2$ are both edges that join $\eta, \sigma \in \Omega$ they can be replaced with a single edges $\{\eta, \sigma\}$ with conductance $c(e_1) + c(e_2)$ and all the potentials and currents in $\mathcal{G} \setminus \{e_1, e_2\}$ are unchanged, and the current flowing from $\eta$ to $\sigma$ equals $i(e_1) + i(e_2)$. 

18
3.2. PÓLYA’S THEOREM AND PROOF

Proof. The proofs of the above lemmas just requires checking Kirchoff’s and Ohm’s laws with the respective current flows, and that this does not change the voltage function. Or, equivalently check that the new voltage function on the reduced network is still harmonic on the appropriate set.

Lemma 3.4 (Gluing). If two neighbouring vertices have the same voltage, then no current flows between them, and we can glue them in to a single vertex while keeping all other edges, without changing the potential or current flow.

Thomson’s principle immediately gives us another useful tool for bounding resistances.

Theorem 3.5 (Rayleigh monotonicity principle). The effective resistance $R(A, B)$ is a non-decreasing function of the edge resistances $(r(e))_{e \in \mathcal{E}}$.

Proof. Let $i_{A,B}$ and $i'_{A,B}$ be the two unit currents associated with a voltage across $A$ and $B$, with resistances $(r(e))_{e \in \mathcal{E}}$ and $(r'(e))_{e \in \mathcal{E}}$ respectively (we drop the bar). Suppose $r(e) \leq r'(e)$ for all $e \in \mathcal{E}$. By Thomson’s principle we have

$$R(A, B) = \frac{1}{2} \sum_{e \in \mathcal{E}} (i_{A,B}(e))^2 r(e) \leq \frac{1}{2} \sum_{e \in \mathcal{E}} (i'_{A,B}(e))^2 r'(e).$$

Now by assumption on the edge resistances and Thomson’s principle again

$$R(A, B) \leq \frac{1}{2} \sum_{e \in \mathcal{E}} (i'_{A,B}(e))^2 r(e) \leq \frac{1}{2} \sum_{e \in \mathcal{E}} (i'_{A,B}(e))^2 r'(e) = R'(A, B).$$

Corollary 3.6. Adding edges, that are not incident with $A$ can only increase the probability of escaping $A$.

3.2 Pólya’s Theorem and proof

We now apply the tools we have developed to the problem of identifying for which dimensions $d$ the simple random walk on $\mathbb{Z}^d$ is transient. Slightly more generally we consider an infinite connected graph $G = (\Omega, E)$ with finite vertex degree and conductances $(c(e))_{e \in \mathcal{E}}$, and a distinguished vertex $0 \in \Omega$ called the origin. Let $d(\cdot, \cdot)$ denote the graph distance, i.e. the length of the shortest path between two vertices, then let

$$\Omega_n = \{\eta \in \Omega : d(0, \eta) \leq n\} \quad \text{and} \quad \partial\Omega_n = \{\eta \in \Omega : d(0, \eta) = n\}$$

and $G_n = (\Omega_n, E_n)$ be the sub-graph of $G$ given by all the edges between vertices in $\Omega_n$. For each $n$ we construct a graph $G_n^* = (\Omega_n^*, E_n^*)$ by gluing all the vertices in $\Omega \setminus \Omega_n$ together, i.e. $\Omega_n^* = \Omega_n \cup \{\zeta_n\}$ and $E_n^* = E_n \cup \{\{\eta, \zeta_n\} : \eta \in \partial\Omega_n\}$. The weight function on $G_n^*$ is equal to $c$ on all edges present in $G$ and on the edges that connect with $\zeta_n$ we put $c(\eta, \zeta_n) = \sum_{\xi \in \partial\Omega_n} c(\eta, \xi)$. Let $R_{\text{eff}}(n) = R(\{0\}, \{\zeta_n\})$, it follows from the Rayleigh's monotonicity principle that the limit

$$R_{\text{eff}} = \lim_{n \to \infty} R_{\text{eff}}(n)$$

exists, and therefore

$$\mathbb{P}_0(\tau_0^+ = \infty) = \lim_{n \to \infty} \mathbb{P}_0(\tau_{\zeta_n} < \tau_0^+) = \lim_{n \to \infty} \frac{1}{R_{\text{eff}}(n)} = \frac{1}{R_{\text{eff}}}.$$

We define a flow from 0 to $\infty$ on $G$ as an anti-symmetric function $\theta$ satisfying $\text{div} \theta(\eta) = 0$ for all $\eta \neq 0$. 


Theorem 3.9 (Pólya’s Theorem). The simple random walk on $\mathbb{Z}^d$ is recurrent if $d = 1, 2$ and transient if $d \geq 3$.

Proof. It is clear by the previous corollary that we can restrict to the cases $d = 2$ and $d = 3$. Firstly, for $d = 2$ we aim to show that $R_{\text{eff}} = \infty$, and we therefore only need a lower bound on $R_{\text{eff}}(n)$ that is diverging with $n$. By Rayleigh’s monotonicity principle we reduce the effective resistance by adding edges (formally this corresponds to reducing the resistance between states from $\infty$). In this case we connect all the states in $\partial \Omega_n$ with edges that have zero resistance, or equivalently $\infty$ conductance. These vertices must be at the same voltage and all neighbours in
the new graph, so we may glue them. We are left with a one dimensional graph that we can reduce using the series/parallel laws, so

\[ R_{\text{eff}}(n) \geq \sum_{i=1}^{n-1} \frac{1}{4(2^i - 1)}, \]

and hence \( R_{\text{eff}}(n) \geq C \log n \to \infty \) as \( n \to \infty \). Alternatively we can construct Nash-Williams cut sets using the same sets (see Peres and Lyons *Probability on trees and networks*).

Now suppose \( d = 3 \), we wish to construct a non-zero flow from 0 to \( \infty \) such that \( D(j) < \infty \). There are many ways to do this of course, but it makes sense to make the most of the symmetry and take inspiration from the obvious flow field on \( \mathbb{R} \) given by a unit source at the origin. We construct such a flow as follows. Let \( u \sim \text{Uniform}(S^3) \) be a uniform random point on the unit sphere (with associate measure \( P \)), and let \( L_u \) be the unique straight half-line from 0 to \( \infty \) that passes through the point \( u \). Let \( \gamma^u \) be the (almost surely) unique path in \( G = (\mathbb{Z}^3, E = \{(x, y) : |x - y|_1 = 1\}) \) from 0 to \( \infty \) that is closest to \( L_u \) in the Hausdorff metric, thought of as a sequence of directed edges. Clearly \( \gamma^u \) stays within a distance of 4 from \( L \). Now define the flow by

\[ j(e) = \sum_{n \geq 0} \left[ P(\gamma^n = e) - P(\gamma^n = -e) \right]. \]

You should check that this defines a flow (in fact a unit flow) from 0 to \( \infty \), this trick of defining unit flows as the expectation of random unit flows is a more general one. We now show that \( j(e) \) has finite energy. By construction, for any given edge \( e \in E \) whose midpoint is a distance \( R \) from the origin, there is a constant \( A \) (independent of \( R \)) such that \( P(e \in \gamma^u) \leq A/R^2 \). Also, there is a constant \( B \) such that there are at most \( Bn^2 \) edges a distance between \( n \) and \( n + 1 \) from the origin. It follows that \( D(j) \leq \sum_{n=1}^{\infty} A^2n^{-4}Bn^2 \), which is finite, and the result follows from Proposition 3.7 2..
Chapter 4

Uniform Spanning Tree and Wilson’s algorithm

We have already seen the idea of uniform spanning trees (USTs) appear briefly in the previous chapter (see Theorem 2.6). It turns out that connection between USTs and random walks on graphs (and the electrical network analogy) goes much deeper. Also these objects turn out to have connections with our next topic - percolation.

In this chapter we will see our first correlation inequality, these turn out to be extremely useful, and we will revisit such concepts in more detail later. In large graphs there are typically a huge number of spanning trees, and it is not obvious how to generate one in an efficient algorithmic way. We will see one very effective way called Wilson’s algorithm which is efficient and theoretically very important, with links to loop erased random walks. At the end of the section we will touch on the idea of weak limits of finite-volume measures, a concept that will also be important later.

A spanning tree of an undirected graph will be composed of undirected edges, since we connect with the previous electrical network picture we will also consider directed graphs and directed spanning trees. It should be clear from context which case we are dealing with.

Definition 4.1. Let $G = (\Omega, E)$ be a finite connected graph. A spanning tree of $G$ is a sub-graph of $G$ that is connected, contains all vertices of $G$, and is a tree (has no cycles). Let $T$ denote the set of all spanning trees of $G$.

We call $T$ a uniform spanning tree, abbreviated UST, if it is distributed according to the uniform measure on $T$, i.e.

$$P(T = \omega) = \frac{1}{|T|}, \quad \omega \in T.$$  

We may think of $T$ as being a random subset of all the edges $E$ (satisfying the constraints to be a spanning tree), in this sense we can identify $T$ with a special subset of $\Omega_E = \{0, 1\}^E$, that is for $w \in T$ we may think of $\omega(e) = 1$ if the edge $e \in E$ is present in the tree and $\omega(e) = 0$ if the edge does not belong to the tree.

4.1 Wilson’s algorithm

There are many ways to generate a UST of $G$. The following algorithm highlights the close connection between USTs and random walks, and exploits some hidden independence in Markov chains. In fact we will present the algorithm in sufficient generality that it can not only sample UST be other weighted random trees.

We will consider the directed, and connect, graph $\mathcal{G}$ generated by a finite, irreducible, Markov process with transition matrix $P$. For $e = (\eta, \sigma) \in \mathcal{E}$ we call $e^- = \eta$ the tail of the edge, and $e^+ = \sigma$ the head.

Definition 4.2. We call a sub-graph of $\mathcal{G}$, with one vertex $r \in \Omega$ marked as a root, a (directed and rooted) spanning tree (sometimes called a spanning arborescence) if it includes every vertex, there is no cycle, and every vertex other than the root is the tail of exactly one edge.
The root \( r \) will be arbitrary but fixed from this point onwards. The edges in the directed spanning tree all point towards the root. It is clear that if we ignore the root and the edge directions then we obtain a spanning tree. Wilson’s algorithm will select spanning trees proportional to the weights

\[
\Psi(T) = \prod_{e \in T} P(e^-, e^+),
\]

by explicitly using the process with transition matrix \( P \). If the process is reversible, recall \( c(e) = \pi(e^-)P(e) \), so that

\[
\Psi(T) = \frac{\prod_{e \in T} c(e)}{\prod_{e \notin T} \pi(e)} \propto \prod_{e \in T} c(e),
\]

since the root is fixed. In particular for the simple random walk on \( G \) we have \( P(\eta, \sigma) = 1/\deg(\eta) \) if \( \eta \sim \sigma \), and all other entries are zero, and the reversible stationary measure satisfies \( \pi(\eta) \propto \deg(\eta) \), i.e. in this case we will select a UST (when we ignore the direction and root).

To describe Wilson’s algorithm we must first define the important concept of loop-erasure (meaning cycle erasure since we have already used loop in the context of the graph having self loops). For any path \( \Gamma = (\eta_0, \eta_1, \ldots, \eta_k) \) on the graph \( G \), meaning \( (\eta_i, \eta_{i+1}) \in E \) for \( 0 \leq i < k \), which can be thought of as a connected set of edges, we can construct a non-intersecting sub-path, we call \( \text{LE}(\Gamma) \), by removing cycles in the order they appear. Note that paths can be thought of as a sequence of edges and induce a sub-graph in the obvious way. More precisely let \( J = \min \{ j \geq 1 : \eta_j = \eta_i \text{ for some } i < j \} \) be the first time you re-visit a vertex in the path, and let \( I \) be the index of the point in the path it hits, i.e. \( I \) is the unique value of \( i \) satisfying \( i < J \) and \( w_i = w_j \). Let \( \Gamma' = (\eta_0, \ldots, \eta_I, \eta_{I+1}, \ldots, \eta_k) \) be the sub path attained after removing the first cycle. This procedure is iterated until we arrive at a path with no cycles, which is denoted \( \text{LE}(\Gamma) \) (Pictured in lectures).

**Wilson’s algorithm:** First let \( (\sigma_1, \sigma_2, \ldots, \sigma_n) \) be an arbitrary fixed ordering of \( \Omega \setminus \{r\} \), and let \( T_0 = \{r\} \).

1. If \( T_i \) spans \( G \) we are done and the algorithm stops. Otherwise pick the smallest element of \( \Omega \) that belongs to \( T_i \), call this \( \eta_i \), and run the process with transition matrix \( P \) on \( G \) started from \( \eta_i \) until it hits \( T_i \), call the path it followed \( \Gamma_i \).

2. Let \( T_{i+1} = T_i \cup \text{LE}(\Gamma_i) \) be the union of the two edge disjoint paths.

It is clear that each stage of the algorithm gives a sub-tree of \( G \), directed toward the root, also the final tree must be a spanning tree since it contains every element of \( \Omega \). Marvellously, the random tree grown in this way has the desired distribution.

**Theorem 4.3.** Wilson’s algorithm yields a random spanning tree rooted at \( r \) with distribution proportional to \( \Psi(\cdot) \).

Note that, in the reversible case, if we forget the root and the orientation then Wilson’s algorithm will construct a weighted spanning tree of \( G \) were each tree has probability proportional to \( \prod_{e \in E} c(e) \).

In order to prove this result we will construct a Markov process with transition matrix \( P \) in a special way and remove cycles in a way that is called ‘cycle popping.’ We therefore have some setting up to do, and first prove a deterministic result at the heart of the algorithm. We think of fixing the randomness of the process in advance, as follows. At each state we assign an infinite sequence of moves (stack of instructions) with the correct distribution. That is, for each \( \eta \in \Omega \setminus \{r\} \) let \( S_i^\eta \), \( i \geq 1 \), be independent random variables laws

\[
P(S_i^\eta = \sigma) = P(\eta, \sigma), \quad \text{for } \sigma \in \Omega.
\]
Since we only care about the process up to the first time it hits the root we put an empty stack on the root. Then the process which moves according to the next instruction in the stack each time it visits any state, and then discards that instruction, is clearly a Markov process with the given transition matrix (absorbed at the root).

The top of these stacks, or visible moves, always determines a directed graph via the edges \((\eta, S^n_{\eta})\), \(\eta \in \Omega\) and \(i \geq 1\), which we call the visible graph. If this visible graph contains no directed cycles then it is a spanning tree rooted at \(r\). Otherwise we will proceed by removing visible cycles in a certain order, called ‘cycle popping’. When a cycle is popped, the moves in that cycle get removed from the top of the stack. It turns out the actual order that visible cycles are popped does not matter.

**Lemma 4.4.** The order in which cycles are popped is irrelevant. Either every order props an infinite number of cycles, or every order pops the same finite set of cycles, thus leaving the same tree on the top of the stacks at the end.

**Proof.** We will track the position of the moves in the stack, we say that the edge \((\eta, S^n_{\eta})\) has colour \(i\). A coloured cycle is a visible cycle all of whose edges have an associated colour. The edges in a coloured cycle do not all have to be of the same colour. Although the same cycle could be popped several times, a coloured cycle can be popped at most once. We will show that if \(C\) is any coloured cycle that can be popped, i.e. there exists a sequence \(C_1, C_2, \ldots, C_n = C\) that may be popped in that order, but we instead chose to pop some other coloured cycle \(C' \neq C_1\) first, then either; (1) \(C = C'\) or else, (2) \(C\) can still be popped from the stack after \(C'\). Showing this completes the proof, since we either never stop popping or every cycle that can pop will be popped.

If all the vertices of \(C'\) are disjoint from those of \(C_1, \ldots, C_n\) then \(C\) can obviously still be popped. Otherwise, let \(k\) be the first cycle with at least one vertex in common with \(C'\). Fix \(\eta \in C' \cap C_k\), since all edges in \(C'\) have colour 1 by assumption, and \(\eta \not\in C_1, \ldots, C_{k-1}\), the edge with head \(\eta\) in \(C_k\) must also have colour 1. It follows that the edge with head \(\eta\) in \(C'\) and \(C_k\) must both point to the same state, lets call it \(\sigma\). We can now repeat the same argument for the state \(\sigma\), and it follows that \(C' = C_k\). Therefore \(C' = C\) or we can pop \(C\) in the order \(C', C_1, C_2, \ldots, C_{k-1}, C_{k+1}, \ldots, C_n\). \(\square\)

We are now in a position to prove Theorem 4.3.

**Proof of Theorem 4.3.** It is clear that Wilson’s argument stops with probability 1, it also pops visible cycles in some order, and so by the previous lemma the output is independent of the choice of order in the implementation.

We now show that the final distribution is the desired one. We think of the stacks (up to the point we stop popping) as defining a finite set of coloured cycles (loops) \(O = (C_1, C_2, \ldots, C_n)\) lying over a non coloured spanning tree \(T\), and let \(\Pi\) be the set of all such pairs we can construct. If \((O, T) \in \Pi\) then so is \((O, T')\) for any spanning tree \(T'\) since the final tree is dictated by the values on the bottom of the stacks after popping all the cycles, and the values of the preceding elements in the stack imparts no information on these. So \(\Pi = O \times T\) where \(O\) is the set of all sequences \(O\) that can occur, and \(T_r\) is the set of directed spanning trees rooted at \(r\). By construction the probability of seeing a pair \((O, T) \in \Pi\) is given by \(\prod_{C \in O} \prod_{e \in C} P(e^-e^+)\Psi(T)\). Since the probability factorises the random sequence of popped cycles and the final tree are independent and the marginal on the tree is the desired one. \(\square\)

The consequences of Wilson's algorithm are far reaching, we state some consequences of Theorem 4.3 here.

**Corollary 4.5.** Given two states \(\eta\) and \(\sigma\) in \(\Omega\) (finite), the distribution of the path in the weighted spanning tree from \(\eta\) to \(\sigma\) is equal to the distribution of the loop-erased path taken by the process from \(\eta\) to \(\sigma\).
The next corollary is somewhat less trivial, and although it can be proved in many ways it also follows from Theorem 4.3.

**Corollary 4.6.** The number of un-rooted trees with \( n \) vertices is \( n^{n-2} \).

One proof is to examine the USTs of the complete graph on \( \{1, 2, \ldots, n\} \), and in particular calculate the probability that a UST is given by a single path \( (1, 2, \ldots, n) \), which occurs with probability \( 1/n^{n-2} \), as required.

### 4.2 Electrical interpretations and negative association

In this section we will discuss some important theoretical consequence of the connection between random walks and spanning trees on finite networks, and in particular the connection again with electrical networks.

It turns out that knowing any given edge is present in the tree can never increase the probability that another edge is also in the tree. The most immediate way to express this property of negative association is as follows.

**Theorem 4.7.** For \( e_1, e_2 \in E \) such that \( e_1 \neq e_2 \),

\[
\Pr(e_2 \in T \mid e_1 \in T) \leq \Pr(e_2 \in T).
\] (4.2.1)

The proof makes use of the link between the current flow and the probability of the random walk crossing given edges. In particular the following results which you have already proved part of on the first example sheet.

**Lemma 4.8** (Kirchhoff’s Effective Resistance Formula). Let \( T \) be an un-rooted weighted spanning tree (with law \( \Pr \)) of a finite network \( G \) (associated with transition matrix \( P \), i.e. edge weights \( c(e) = \pi(e^-)P(e^-, e^+) \) and \( e \in E \), then

\[
\Pr(e \in T) = \Pr_{e^-}(1st \ hit \ e^+ \ via \ travelling \ along \ e) = i(e) = c(e)R(\{e^-\}, \{e^+\}),
\]

where \( i \) is the unit current flow from \( e^- \) to \( e^+ \).

**Proof.** Exercise. Hint use the current as edge crossing probabilities result from Sheet 1. Note that this result also follows immediately from Theorem 2.6.

**Proof of Theorem 4.7.** Denote the dependence of the random tree on the graph by \( T_G \). The contraction of the graph \( G \) along the edge \( e \) is defined by removing the edge \( e \) and identifying the end points, and is denoted by \( G/e \) (gluing the vertices connected via \( e \)’), and leaving the other edges and vertices the same. Note that even if \( G \) is a simple graph the contraction may be a multi-graph. There is a one-to-one correspondence between spanning trees of \( G \) containing the edge \( e \) and spanning trees of \( G/e \). Furthermore, the distribution of \( T_{G/e} \) given \( e \in T_G \) is the same as that of \( T_{G/e} \). It follows from the previous lemma that

\[
\Pr(f \in T \mid e \in T) = \Pr(f \in T_{G/e}) = R_{G/e}(\{f^-\}, \{f^+\}) \leq R_G(\{f^-\}, \{f^+\}) = \Pr(f \in T),
\]

where the key inequality follows from Rayleigh’s monotonicity principle.

The negative association in Theorem 4.7 was generalised by Feder and Mihail in 1992 to more general ‘increasing’ events and random variables, as follows. Recall we can think if subgraphs of \( G \) as elements of \( \{0, 1\}^E \), on which there is an obvious partial ordering, \( \omega \leq \omega' \) if \( \omega(e) \leq \omega'(e) \) for all \( e \in E \) (this type of partial ordering will be come up again later on our typical statespace for IPS \( S^A \)). An event \( A \subset \{0, 1\}^E \) is called increasing if for all \( \omega, \omega' \in \{0, 1\}^E \) with \( \omega \leq \omega' \), we have that \( \omega' \in A \) whenever \( \omega \in A \), i.e. adding any edge to any configuration (subgraph) in \( A \) results in another configuration (subgraph) in \( A \). Examples of increasing events include, the edge \( e \) is in the sub-graph, and there is a path connecting \( x \) to \( y \) in the sub-graph. A random variable \( X : \{0, 1\}^E \rightarrow \mathbb{R} \) is increasing if \( X(\omega) \leq X(\omega') \) whenever \( \omega \leq \omega' \).
Theorem 4.9. Let $G$ be a finite network and $A$ an increasing event that ignores the edge $e \in E$, then $\mathbb{P}(A \mid e \in T) \leq \mathbb{P}(A)$. Furthermore, if $X$ and $Y$ are increasing random variables then $E[XY] \leq E[X]E[Y]$.

For proofs see Chapter 3, Section 4.2 in Lyons and Peres Probability on Trees and Networks. Later we will see an extremely useful counterpart to this (with the inequality the other way) called ‘positive association’ or positive correlations, which will be called the FKG inequality.

4.3 Weak limits

It turns out that Wilson’s algorithm can also be run using recurrent, irreducible Markov processes on countable graphs, provided we start the loop-erased walks appropriately so that each vertex belongs to the final tree. Can this object on the countable graph be interpreted in terms of weighted spanning trees? Usually the weighted spanning tree on the $d$-dimensional cubic lattice is not defined directly on the infinite graph, but instead on finite subgraphs $G_n$ which increase to the cubic lattice as $n \to \infty$. Then the probability of observing an event that only depends on finitely many edges converges to the probability of observing this event on the infinite lattice as $G_n \nearrow (\mathbb{Z}^d, \mathcal{E})$. Also, it turns out that these local probabilities determine the distribution of the infinite tree, that is the infinite object locally looks like a weighted spanning tree. To make this discussion more precise we will examine weak limits of probability measures.

4.3.1 A Topological Aside

The state space for random trees as viewed as subsets of $\{0, 1\}^E$ is of the more general form that will be of interest later for IPS and was discussed in the intro, i.e. $S^\Lambda$ for some finite $S$ and countable $\Lambda$. So the discussion here will be relevant also later in the course. We endow this space with the product topology, i.e. the coarsest topology with respect to which all the canonical projections, $p_e : \{0, 1\}^E \to \{0, 1\}$ given by $p_e(\omega) = \omega(e)$, are continuous. Equivalently this is the topology is generated by countable unions and finite intersections of sets of the form

$$p_e^{-1}(U) = \{\omega \in \{0, 1\}^E : \omega(e) \in U\} \text{ for } U \subset \{0, 1\},$$

called open cylinders. Note that since we take the discrete topology on $\{0, 1\}$ (or more generally finite $S$) every subset if clopen. Finite intersections of open cylinders, of the form

$$\{\omega \in \{0, 1\}^E : \omega(e_1) \in U_1, \ldots, \omega(e_n) \in U_n\} \text{ for } U_i \subset \{0, 1\}, \quad 1 \leq i \leq n,$$

are called cylinder sets, in words these are exactly the events that depend on only finitely many coordinates. Since $\{0, 1\}$ is compact (more generally if $S$ is finite), then by Tychonoff’s Theorem $\{0, 1\}^E$ is compact. It is also metrisable by

$$d(\omega, \omega') = \sum_{i=1}^{\infty} 2^{-i}|\omega(e_i) - \omega'(e_i)|, \quad \omega, \omega' \in \{0, 1\}^E,$$

moreover this gives us a complete separable metric space.

For a measurable structure on $\{0, 1\}^E$ we take the product $\sigma$-algebra $\mathcal{F}$ of $\{0, 1\}^E$ generated by the cylinder sets. Equivalently this is just the Borel $\sigma$-algebra with respect to the topology discussed above.

We would like to discuss convergence of probability measures, as such we will take the topology of weak convergence on the set $\mathcal{M}_1(\{0, 1\}^E)$ of probability measures on our state space. We say $\mu_n$ converges weakly to $\mu$, written $\mu_n \rightharpoonup \mu$ as $n \to \infty$, if

$$\mu_n(f) \to \mu(f) \quad \text{as } n \to \infty,$$
for all bounded continuous functions $f : \{0,1\}^E \to \mathbb{R}$. Recall $\mu(f)$ denotes the expectation of the function $f$ with respect to $\mu$. Since $\Omega$ is compact $\mathcal{M}_1$ is also compact with respect to this topology, and every family of probability measures on $\{(0,1)^E, \mathcal{F}\}$ is relatively compact (this turns out to be very useful). Since all cylinder events are both open and closed, and they generate $\mathcal{F}$, we also have $\mu_n \to \mu$ if and only if $\lim_{n \to \infty} \mu_n(C) = \mu(C)$ for all cylinder sets $C$.

4.3.2 Back to the UST measure

We make a slight shift in notation in line with the topological discussion previously. Let $\mu_n$ be the distribution of the UST on the $d$-dimensional box $\Lambda_n = [-n,n]^d$, with edge set given by the usual $\ell_1$ neighbours (i.e. the sub-graph of the $d$-dimensional lattice). As above, we think of this as a measure on $\Omega_n = \{0,1\}^{E_n}$ where $E_n$ are the edges contained in $\Lambda_n$ (we should be careful with our change of notation since we previously called the vertex set of the graph $\Omega$).

**Theorem 4.10.** The weak limit $\mu = \lim_{n \to \infty} \mu_n$ exists, it is a translation-invariant and (shift)-ergodic probability measure. It is supported in the set of forests of $(\mathbb{Z}^d, E)$ (i.e. no cycles but possibly not connected) with no bounded component.

For $x \in \mathbb{Z}^d$ let $\theta_x$ be the shift operator, i.e. $\theta_x(y) = y + x$ for $y \in \mathbb{Z}^d$. This induces an obvious action on the edge set and $\Omega$, by $\theta(x) = (\theta_x(e^-), \theta_x(e^+))$ and $\theta(x) = (\omega(\theta_x(e))_{e \in E_d}$). In this case the translation by $x$ is a graph automorphism on the $d$-dimensional lattice.

An event $A \in \mathcal{F}$ is called shift-invariant if $A = \theta_x(A) := \{\theta_x(\omega) : \omega \in A\}$ for each $x \in \mathbb{Z}^d$. A probability measure $\nu$ is called ergodic (with respect to the shift) if every shift invariant event has probability 0 or 1. That is in words, the operator mixes things up with respect to $\nu$, that is no event with non-trivial measure is left fixed by shifts. A probability measure is said to be supported on an event $F \in \mathcal{F}$ if $\nu(F) = 1$.

**Proof of Theorem 4.10.** In the interest of time we skip this proof, but it may be found for example in Grimmett *Probability on Graphs*.

An important question now is; when is this spanning forest actually a tree? We will not prove the following result, see for example Pemantle (1991).

**Theorem 4.11.** The limit measure $\mu$ is supported on the set of spanning trees if and only if $d \leq 4$.

For proof of a similar result in $d \leq 2$ (in fact more generally when the random walk is recurrent), using the current flows see Lyons and Peres, *Probability on Trees and Networks*.

Another important question becomes, does the limit in Theorem 4.10 depend on what we did with the boundary of the set $\Lambda_n$. As it stands we did not insist on anything special happening on the boundary (a sort of ‘free’ boundary condition), and so $\mu_n$ is sometimes called the free uniform spanning tree measure. However we could have imposed other boundary conditions, one natural thing would be to glue all the points outside $\Lambda_n$ into a single vertex connected to the boundary of $\Lambda_n$ (as we did earlier). This gives rise to the so called wired uniform spanning tree measure. These turn out to give rise to the same weak limit on the $d$-dimensional integers lattice, and the key reason why the limits are the same is because these graphs are edge amenable, meaning the the ratio of the size of the edge set and the volume of large boxes goes to zero

$$\lim_{n \to \infty} \frac{|\partial E_{\Lambda_n}|}{|\Lambda_n|} = 0,$$

where $\Lambda_n \neq \mathbb{Z}^d$ and $\partial E_{\Lambda_n}$ is the set of edges that connect to $\Lambda_n$ from outside. Then we could ask, when is the limiting object the same thing. It turns out in this case that we do get the same thing, but we will return to this in more generality later when we look at Gibbs states.
Chapter 5

Percolation

Percolation theory was introduced over 50 years ago as a mathematical framework for studying random spatial processes such as the flow of liquids through disordered porous mediums. It gave rise to a huge amount of work in both applied and pure mathematics. These relatively simple models to define have given rise to a spectacular array of challenging mathematical problems, with many deep results (with surprising connections to other areas of mathematics) now proven, and many more still conjectured. One of the central features of these models, that we will be particularly interested in during this course, is the concept of a phase transition. This is a point in parameter space where the behaviour of the model changes in some abrupt way (we will be more precise later). More generally, these models give rise to very rich large scale random geometries.

5.1 Model definition

We will focus on independent bond percolation on $\mathbb{Z}^d$, as opposed to site percolation. We fix the underlying graph $G = (\mathbb{Z}^d, E)$, where $E = \{ \{x, y\} : |x - y|_1 = 1 \}$, and a parameter $p \in [0, 1]$. To each edge $e \in E$ we assign a state 0 or 1, corresponding to closed or open respectively. Our probability space will therefore be $\Omega = \{0, 1\}^E$ endowed with the Borel $\sigma$-algebra generated by finite dimensional cylinder sets that we discussed at the end of the last chapter. For a probability measure we take the product measure $P_p$, parametrised by $p$, under-which each edge is independently open with probability $p$. We think of an open edge as being open to the passage of some material, such as a liquid through a porous media.

It turns out that the following construction, that simultaneously couples bond percolation on $G$ for all parameter values $p \in [0, 1]$ on the same probability space, is often useful. The idea of coupling is an important one, the concept is deceptively simple looking and extremely general, but it is a very handy tool. By coupling, we simply mean that the random variables are defined on a single probability space, with a single measure, such that the marginal on each of the random variables is the desired one. In this way there is a huge amount of freedom in specification of the joint law. Suppose $(U_e : e \in E)$ is a family of independent uniform random variables on $[0, 1]$, then define $\omega_p(e) = \begin{cases} 1 & \text{if } U_e < p, \\ 0 & \text{if } U_e \geq p, \end{cases}$ and we call the associated probability measure $P$ (without the $p$ subscript). Then the marginal law on $\omega_p \in \Omega$ for a fixed $p \in [0, 1]$ is bond percolation on $G$ with parameter $p$, i.e. $\omega_p(e) : e \in E$ is independent in its components and $P(\omega_p(e) = 1) = P_p(\omega(e) = 1) = p$.

We are typically interested in the geometry of the open sub-graph, in particular the connected components, or clusters, of this sub-graph containing a given vertex $x$, called $C_x$. For $x, y \in V$ we write $x \leftrightarrow y$ if there exists a path of open edges connecting $x$ and $y$, and so $C_x = \{ y \in V : x \leftrightarrow y \}$.

We also write $x \leftrightarrow \infty$ if $|C_x| = \infty$. Since the graph $G$ is translation invariant (vertex transitive - i.e. for each pair $x, y \in V$ there exists a graph automorphism that takes $x$ to $y$), it is clear that $|C_x|$ has the same distribution as $|C_0|$, and so we will focus on the latter.
One of the principle concerns is whether or not there exists an infinite open cluster.

**Definition 5.1** (Percolation probability and the critical probability). The percolation probability as a function of the parameter $p$ is defined as

$$\theta(p) = \mathbb{P}_p(|C_0| = \infty).$$

The critical probability is defined by

$$p_c(d) = \sup\{p \in [0, 1] : \theta(p) = 0\}.$$

By the coupling given above we immediately get the following lemma.

**Lemma 5.2.** The percolation probability, $\theta(p)$, is an increasing function of $p$, and

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

### 5.2 The phase transition

In one dimension it is easy to check that $p_c = 1$, and there is no non-trivial phase transition. However it turns out that for all dimensions $d \geq 2$ there is a non-trivial phase transition.

**Theorem 5.3.** For all $d \geq 2$ we have $p_c \in (0, 1)$.

It turns out the for all $p > p_c$ the percolation probability $\theta(p)$ is a $C^\infty$ function of $p$ for $p \in (p_c, 1]$, by definition it is 0 for $p < p_c$, and by the previous coupling it must be increasing (this gives the typical picture for the function $\theta(p)$ in Fig. given in lecture). However, it is still a significant outstanding problem to show that there is continuity at $p_c$.

**Conjecture 1.** The percolation probability $\theta(p)$ is continuous at $p_c$ for each $d \geq 3$.

The result is known to hold for $d = 2$, and in high dimensions (currently for $d \geq 11$), but not in between - and in particular it is not known for $d = 3$.

**Proof of Theorem 5.3.** By considering the obvious embedding it is clear that $p_c(d + 1) \leq p_c(d)$. Therefore it is sufficient to show that $p_c(d) > 0$ for all $d \geq 2$ and $p_c(2) < 1$. We do the former by a path counting argument, and the latter using a technique which is often referred to as a Peierls contour argument, after Peierls who used a similar argument applied to the two dimensional Ising model.

We first show that $\theta(p) = 0$ for $p$ sufficiently close to zero, i.e. $p_c(d) > 0$ for $d \geq 2$. We do this by counting self avoiding paths. A self avoiding walk (SAW) is a path in the graph that visits each site at most once. Let $N_n$ be the the (random) number of SAWs of length $n$, starting from the origin, in the percolation sub-graph (using only open edges). Then

$$\theta(p) = \mathbb{P}_p(|C_0| = \infty) = \mathbb{P}(N_n \geq 1 \text{ for all } n \geq 1)$$

$$= \lim_{n \to \infty} \mathbb{P}_p(N_n \geq 1) \leq \lim_{n \to \infty} \mathbb{E}_p(N_n), \quad (5.2.1)$$

where the last inequality follows from Markov’s inequality. We upper bound the expectation by a crude upper bound on all possible SAWs in $\mathbb{Z}^d$ and using the independence structure. Let $\sigma_n$ denote the total number of SAWs, starting from the origin, in the $d$-dimensional lattice (deterministic graph). There are $2d$ choices for the first edge, and at most $2d - 1$ choices for each
subsequent edge, since they certainly can’t backtrack. Since each edge is open independently with probability \( p \) it follows that

\[
E_p(N_n) \leq \sigma_n p^n \leq 2d(d-1)^n - 1 p^n = \frac{2d}{2d-1} ((2d-1)p)^n .
\]

Combining with Inequality (5.2.1) we find that \( p_c > 1/(2d-1) \).

We now use planar duality and a Peierls argument in two dimensions to show that \( \theta(p) > 0 \) for \( p \) sufficiently large. We now define the dual graph (see figure in lectures). The (planer) dual of a planer graph \( G \) (i.e. a graph embedded in \( \mathbb{R}^2 \) without edges crossing) is described as follows. Each face of \( G \) (with respect to the edges of \( G \)), is associated with a vertex of the dual, and there is an edge between each of these vertices which have an edge in common (share a boundary). The dual of the \( \mathbb{Z}^2 \) lattice is isomorphic to the \( \mathbb{Z}^2 \) lattice, i.e. it is self dual, and in the obvious embedding the dual is just shifted by (1/2, 1/2). There is a 1-to-1 correspondence between edges of the dual and edges of the original (primal) graph since each edge of the dual crosses a unique edge of the primal graph. The percolation measure on the primal graph induces a percolation on the dual by calling each edge of the dual open if and only if it crosses an open edge.

Suppose that \( |C_0| < \infty \) then there exists a closed cycle in the dual percolation which encloses \( C_0 \). The converse also holds, if the origin lies inside a closed loop in the dual then \( |C_0| < \infty \). The proof of this fact is straightforward by tedious to write out, so we take it as given (see the figure in lectures). It was originally proved by Kesten in ’82. We now count again. Let \( M_n \) be the number of closed loops in the dual of length \( n \), containing the origin, then

\[
1 - \theta(p) = P_p(|C_0| < \infty) = P(\sum_n > 1 M_n > 1) \leq E_p(\sum_n > 1 M_n) = \sum_n > 1 E_p[M_n] \leq \sum_n > 1 \rho_n (1-p)^n,
\]

where \( \rho_n \) is the (deterministic) number of loops (closed paths) of length \( n \) containing the origin in the dual lattice. Every loop of length \( n \) which contains the origin must use an edge of the form (\( k + 1/2, 1/2 \)) for some \( 0 \leq k < n \) (with the obvious embedding in \( \mathbb{R}^2 \)), and so by applying previous bound on the number of SAWs we have \( \rho_n \leq n \sigma(n-1) < n4^n \). So by taking \( p \) sufficiently small we can make the sum on the right hand side of (5.2.2) smaller than 1, as required.

From the self duality in two dimensions you might guess that \( p_c(2) = 1/2 \). That is, closed edges in the dual form a percolation model where each edge is present independently with probability \( 1-p \), if the origin belongs to an infinite cluster in the primal graph then, due to the constrained geometry, it seems unlikely that there is simultaneously an infinite closed cluster in teh dual. On the other hand, for \( p < p_c \) we may picture many finite ‘islands’ of open edges in the primal graph, surrounded by a see of closed edges in the dual, and presumably this see contains some infinite cluster. Surprisingly this result was not made rigorous until Kesten in 1980, building on the work of Russo, Seymour and Welsh.

Lemma 5.4. Let \( A_{\infty} \) be the event that there exists an infinite cluster. Then the following dichotomy holds:

(a) If \( \theta(p) = 0 \) then \( P_p(A_{\infty}) = 0 \).

(b) If \( \theta(p) > 0 \) then \( P_p(A_{\infty}) = 1 \).

Proof. This result also follows from the fact that \( P_p \) turns out to be shift-ergodic, which we will see later. For now we can deduce the result immediately from Kolmogorov’s zero-one law.

(a) Suppose \( \theta(p) = 0 \), the result follows by a simple union bound and translation invariance,

\[
P_p(A_{\infty}) \leq \sum_{x \in \mathbb{Z}^d} P_p(|C_x| = \infty) = \sum_{x \in \mathbb{Z}^d} \theta(p) = 0 .
\]
(b) Recall Komogorov’s 0-1 law, if \((X_i)_{i \in \mathbb{N}}\) are i.i.d. r.v.s, \(\mathcal{F}_n = \sigma(X_k : k \geq n)\) and \(\mathcal{F}_\infty = \bigcap_{n \geq 0} \mathcal{F}_n\) is the asymptotic \(\sigma\)-algebra, then \(\mathcal{F}_\infty\) is trivial, i.e. for all \(A \in \mathcal{F}_\infty\) either \(\mathbb{P}(A) = 1\) or \(\mathbb{P}(A) = 0\). Now suppose \(\theta(p) > 0\). Order the edges of \(Z^d\) then \((w(e_i))_{i \in \mathbb{N}}\) are i.i.d., and \(A_\infty \in \mathcal{F}_\infty\) since changing finitely many edges does not change the occurrence of \(A_\infty\). Furthermore \(\mathbb{P}_p(A_\infty) \geq \theta(p) > 0\), and so \(\mathbb{P}_p(A_\infty) = 1\).  \(\square\)
Chapter 6

Percolation toolbox

In this chapter we present some of the fundamental tools that are often applied in the study of percolation, and more broadly in other areas of probability, for example in the study of random graphs and interacting particle systems.

Recall the natural partial order on $\Omega$ and the definition of increasing events and functions. Note that the events $\{|C_0| = \infty\}$ and $\{x \leftrightarrow y\}$ are increasing. We can take the consequences of the partial order on the state space yet another step, in that it also induces a partial order on probability measures on $\Omega$ (we do not always make explicit reference to the probability space but it is always the one we discussed in Section 4.3).

**Definition 6.1.** Given two probability measures $\mu_1, \mu_2 \in \mathcal{M}_1(\Omega)$ we say $\mu_2$ stochastically dominates $\mu_1$, written $\mu_1 \leq \mu_2$, if

$$
\mu_1(f) \leq \mu_2(f) \quad \text{for all increasing functions } f .
$$

(Equivalently we could formulate this in terms of increasing events).

The following results, which states that if $\mu_1 \leq \mu_2$ then the two measures can be coupled in a ‘pointwise monotone’ way, is often called Strassen’s theorem.

**Theorem 6.2** (Strassen’s Theorem). Let $\mu_1$ and $\mu_2$ be probability measures on $\Omega$ then a necessary and sufficient condition for $\mu_1 \leq \mu_2$ is that there exists a probability measure $\nu$ on $\Omega^2$ such that

$$
\nu\left(\{(\eta, \sigma) : \eta \leq \sigma\}\right) = 1 ,
$$

and whose marginals are $\mu_1$ and $\mu_2$.

**Proof.** Non examinable.

This should be compared with the increasing coupling we gave in the previous chapter for bond percolation. The coupling has the following immediate consequences (which we have already used in the context of $\theta(p)$).

**Proposition 6.3.** If $f$ is an increasing random variable then

$$
E_{p_1}(f) \leq E_{p_2}(f) \quad \text{whenever } p_1 \leq p_2 ,
$$

so long as the expectation exists. If $A$ is an increasing event then

$$
P_{p_1}(A) \leq P_{p_2}(A) \quad \text{whenever } p_1 \leq p_2 .
$$

**Proof.** Exercise.
6.1 Harris-FKG inequality

We have previously seen negative association for weighted spanning trees, we will now see an extremely useful converse to this of positive association. Often we which to calculate the probability of certain intersections, or conditional expectations of certain random variables. Unfortunately when the events (random variables) are not independent this is often difficult or impossible to do directly. However the following inequalities can give very useful bounds for a certain class of events (or random variables).

**Theorem 6.4** (Harris inequality). Let A and B be two increasing events then,

\[ \mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A) \mathbb{P}_p(B) \quad \text{i.e.} \quad \mathbb{P}_p(A \mid B) \geq \mathbb{P}(A). \]

More generally if f and g are increasing, bounded, functions then

\[ \mathbb{E}_p[fg] \geq \mathbb{E}_p[f] \mathbb{E}_p[g]. \]

**Remark 6.5.** Intuitively the meaning of the inequality is that conditioning on an increases the chance that edges are open, and therefore increases the probability of observing another increasing event.

**Example 6.6.** For bond percolation on any infinite graph G with countable edge set, the critical probability is independent of the vertex you pick. Let \( p^c_x = \sup\{p : \theta^x(p) = \mathbb{P}_p(\{C_x| = \infty\}\right) \), then \( p^c_x = p^c_y \) for all \( x, y \in V \), since

\[ \theta^x(p) \geq \mathbb{P}_p(\{x \leftrightarrow y\} \cap \{y \leftrightarrow \infty\}) \geq \mathbb{P}_p(x \leftrightarrow y) \theta_y(p), \]

so \( p^c_x \leq p^c_y \). By the same argument we can reverse the inequality.

**Proof of the Harris inequality.** The second inequality is more general than the first (take \( f = 1_A \) and \( g = 1_B \)), so we prove the second. Since \( E \) is countable we let \( E = \{e_i\}_{i \geq 1} \) and let \( \omega_i = \omega(e_i) \). Fix \( f \) and \( g \) bounded and increasing functions. We first show that it is sufficient to consider functions that depend on only a finite number of the edges. Let \( f_n := \mathbb{E}_p[f \mid \omega_1, \ldots, \omega_n] \) and \( g_n := \mathbb{E}_p[g \mid \omega_1, \ldots, \omega_n] \), which are functions only of the first \( n \) edge variables. If \( \mathcal{G}_n = \sigma(\omega_1, \ldots, \omega_n) \) then \( \{\mathcal{G}_n\}_{n \geq 1} \) is a filtration, and trivially we have

\[ \mathbb{E}_p[f_{n+1} \mid \mathcal{G}_n] = \mathbb{E}_p[\mathbb{E}_p[f \mid \mathcal{G}_{n+1}] \mid \mathcal{G}_n] = \mathbb{E}_p[f \mid \mathcal{G}_n] = f_n, \]

since \( \mathcal{G}_n \subset \mathcal{G}_{n+1} \), i.e. \( (f_n)_{n \geq 1} \) is a martingale. So by the martingale convergence theorem we have \( f_n \to f \) and \( g_n \to g \), \( \mathbb{P}_p \)-a.s. and in \( L^2(\mathbb{P}_p) \). We also have \( \mathbb{E}_p[f g] = \lim_{n \to \infty} \mathbb{E}_p[f_n g_n] \) since

\[ \mathbb{E}_p[f_n g_n - f g] \leq \mathbb{E}_p[|f_n g_n + g f_n - g f|] \]

\[ \leq (\mathbb{E}_p[|f_n - f|] \mathbb{E}_p[|g_n^2|]^{1/2} + (\mathbb{E}_p[|g_n - g|] \mathbb{E}_p[f^2])^{1/2} \to 0 \quad \text{as} \ n \to \infty, \]

where we use the triangle inequality in the first line and Cauchy-Schwarz inequality in the second. It therefore suffices to show that \( \mathbb{E}_p[f_n g_n] \geq \mathbb{E}_p[f_n] \mathbb{E}_p[g_n] \) for each \( n \geq 1 \). We do this by induction on \( n \).

For \( n = 1 \) we have \( f_1, g_1 : \{0, 1\} \to \mathbb{R} \), and \( f_1(\omega_1) - f_2(\omega_2) \geq 0 \) if \( \omega_1 \geq \omega_2 \) and vice versa, and the same for \( g_1 \), therefore

\[ (f_1(\omega_1) - f_2(\omega_2))(g_1(\omega_1) - g_2(\omega_2)) \geq 0 \quad \text{for all} \ \omega_1, \omega_2 \in \{0, 1\}. \]

It follows that

\[ 0 \leq \sum_{a, b \in \{0, 1\}} \mathbb{P}_p(\omega_1 = a) \mathbb{P}_p(\omega_2 = b)(f_1(a) - f_2(b))(g_1(a) - g_2(b)) = 2(\mathbb{E}_p[f_1 g_1] - \mathbb{E}_p[f_1] \mathbb{E}_p[g_1]), \]
6.1. **HARRIS-FKG INEQUALITY**

i.e. the required inequality holds.

Now assume that for an function depending on \( k < n \) edges the required inequality holds. The following gets notationally a little nasty. We firstly use the tower property, 

\[
\mathbb{E}_p[f_n g_n] = \mathbb{E}_p[\mathbb{E}_p[f_n g_n | \omega_1, \ldots, \omega_{n-1}]] \\
= \mathbb{E}_p[\mathbb{E}_{\omega_n}[f_n(\omega_1, \ldots, \omega_{n-1}, \cdot)] g_n(\omega_1, \ldots, \omega_{n-1}, \cdot)],
\]

where \( \mathbb{E}_{\omega_n}[f_n(\omega_1, \ldots, \omega_{n-1}, \cdot)] \) is understood to be the expected value with respect to the Bernoulli random variable on edge \( e_n \) with \( \omega_1, \ldots, \omega_{n-1} \) fixed, therefore \( f_n(\omega_1, \ldots, \omega_{n-1}, \cdot) \) is understood to be a function of a single edge configuration, and we may therefore apply the ‘\( k = 1 \) case’. It follows that

\[
\mathbb{E}_p[f_n g_n] \geq \mathbb{E}_p[\mathbb{E}_{\omega_n}[f_n(\omega_1, \ldots, \omega_{n-1}, \cdot)] \mathbb{E}_{\omega_n}[g_n(\omega_1, \ldots, \omega_{n-1}, \cdot)]] \\
= \mathbb{E}_p[\mathbb{E}_{\omega_n}[f_n | \omega_1, \ldots, \omega_{n-1}] \mathbb{E}_{\omega_n}[g_n | \omega_1, \ldots, \omega_{n-1}]] \\
\geq \mathbb{E}_p[f_n] \mathbb{E}_p[g_n],
\]

where the final inequality follows from the ‘\( k = n - 1 \) case’. This completes the proof. \( \square \)

The Harris inequality has been extended to measure more general than the product measure on \( \Omega \), by Fortuin, Kasteleyn and Ginibre (1971). The following result is often called the FKG inequality and is very useful in many interacting particle systems. We first give a precursor, from which the FKG inequality follows. For \( \omega_1, \omega_2 \in \Omega \) we define the (pointwise) maximum and minimum by

\[
(\omega_1 \vee \omega_2)(e) = \max\{\omega_1(e), \omega_2(e)\}, \\
(\omega_1 \wedge \omega_2)(e) = \min\{\omega_1(e), \omega_2(e)\},
\]

for \( e \in E \). A probability measure \( \mu \) on \( \Omega \) is called positive if it assigns a strictly positive probability measure to each point \( \omega \in \Omega \), i.e. \( \mu(\omega) > 0 \).

**Theorem 6.7** (Holley inequality). *Suppose \( \mu_1 \) and \( \mu_2 \) are two positive probability measures on \( \Omega \), if

\[
\mu_1(\omega_1 \wedge \omega_2) \mu_2(\omega_1 \vee \omega_2) \geq \mu_1(\omega_1) \mu_2(\omega_2), \quad \text{for all } \omega_1, \omega_2 \in \Omega,
\]

then \( \mu_1 \leq \mu_2 \).

**Proof.** The typical proof is via stochastic monotonicity, which we will come to later, and constructing suitable coupled processes and applying Strassen’s theorem. \( \square \)

The FKG inequality follows as a consequence of the Holley inequality.

**Theorem 6.8** (FKG inequality). *If \( \mu \) is a positive probability measure on \( \Omega \) such that

\[
\mu(\omega_1 \wedge \omega_2) \mu(\omega_1 \vee \omega_2) \geq \mu(\omega_1) \mu(\omega_2), \quad \text{for all } \omega_1, \omega_2 \in \Omega,
\]

then \( \mu \) is positively associated (has positive correlations) i.e.

\[
\mu(fg) \geq \mu(f) \mu(g) \quad \text{for all increasing } f, g : \Omega \to \mathbb{R}.
\]
6.2 BK inequality

In this section we will see a partial converse to the FKG inequality which holds in the special case of product measures on $\Omega$. The inequality, typically called the BK-inequality after Berg and Kesten, obviously can’t hold in general for events of the form $A \cap B$ for increasing $A$ and $B$, so we have to replace the intersection with something else (but closely related) called the ‘disjoint occurrence’ or ‘disjoint intersection’. In words, we will define $A \circ B$ to be the event that $A$ and $B$ occur on disjoint edge sets.

Suppose $A$ and $B$ are increasing events that depend on finitely many edges $\{e_i\}_{i=1}^n$, each $\omega = (\omega_1, \ldots, \omega_n)$ is specified uniquely by the set of edges on which it is non-zero, $K(\omega) = \{e_i : \omega(e_i) = 1\}$. We define $A \circ B$ to be the set of all $\omega$ such that there exists a subset $H$ of $K(\omega)$ with $\omega' \in A$ whenever $K(\omega') = H$, and $\omega'' \in B$ whenever $K(\omega'') = K(\omega) \setminus H$. In other-words, if

$$\omega_H(e) = \begin{cases} \omega(e) & \text{for } e \in H, \\ 0 & \text{for } e \not\in H, \end{cases}$$

then

$$A \circ B = \{\omega : \text{there exists } H(\omega) \subset E \text{ s.t. } \omega_H \in A, \ \omega_{E \setminus H} \in B\}.$$ 

In words, there are disjoint edge sets (possibly depending on $\omega$) such that the configuration being open on the first is sufficient to belong to $A$ and the configuration being open on the second is sufficient to belong to $B$. Note that $A \circ B$ is both commutative and associative.

Example 6.9. Suppose $G$ is a finite sub graph of the $d$-dimensional lattice, and let $A_G(x,y) = \{\text{there is an open path in } G \text{ connecting } x \text{ and } y\} = \{x \rightarrow_G y\}$, then $A_G(u,v) \circ A_G(x,y)$ is the event that there exists two edge-disjoint paths in $G$ connecting $x$ to $y$ and $u$ to $v$. It is then eminently plausible that

$$\mathbb{P}(A_G(u,v) \circ A_G(x,y) \mid A_G(u,v)) \leq \mathbb{P}(A_G(x,y)),$$

since the disjoint intersection can’t “re-use” the edges used for the event “$A_G(u,v)$.”

This conclusion is the assertion of the BK inequality.

Theorem 6.10 (BK inequality). If $A$ and $B$ are increasing events that depend on finitely many edges, then

$$\mathbb{P}_p(A \circ B) \leq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

Proof. Guided exercise on the final sheet. \hfill \Box

The restriction to events that depend on only finitely many edges is important, but it can be dropped in practice for some important applications (with a little extra work).

It was conjectured by Berg and Kesten that the inequality should hold in general without the restriction to increasing events, this was shown later in 1997 by Reimer. The following result actually contains the BK and Harris inequality but has not seen any application so far that was not one of these two special cases.

For $\omega \in \Omega$ and $F \subset E$ we define the cylinder event $C(\omega, F)$ generate by $\omega$ on $F$ as

$$C(\omega, F) = \{\omega' \in \Omega : \omega'(e) = \omega(e) \text{ for all } e \in F\}.$$ 

We define $A \sqcap B$ as the set of all $\omega$ such that there exists a $K(\omega) \subset E$ with $C(\omega, K) \subset A$ and $C(\omega, K^c) \subset B$. Alternatively, if we write $\omega|_I$ for the restriction of $\omega$ to $I \subset E$, then if $\omega \in A$ we say $I$ is a witness to the event $A$ for $\omega$ if $\omega' \in A$ whenever $\omega'|_I = \omega|_I$. Then in words $A \sqcap B$ is the set of configurations $\omega$ such that there exists two disjoint witnesses $I(\omega)$ and $J(\omega)$ of $A$ and $B$ for $\omega$. 

Remark 6.11. The disjoint occurrence satisfies:

- $A \square B \subset A \cap B$.
- If $A$ and $B$ are increasing then so is $A \square B$ and $A \circ B = A \cap B$.
- If $A$ is increasing and $B$ is decreasing then $A \square B = A \cap B$.

Theorem 6.12 (Reimer's inequality). If $A$ and $B$ are two events that depend on finitely many edges then

$$
\mathbb{P}_p(A \square B) \leq \mathbb{P}_p(A) \mathbb{P}_p(B).
$$

Corollary 6.13. Consider a finite set $S \subset \mathbb{Z}^d$ containing the origin and $x \notin S$, then,

$$
\mathbb{P}_p(o \leftrightarrow x) \leq \sum_{y \in \partial^o S} \mathbb{P}_p(0 \leftrightarrow y) \mathbb{P}_p(y \leftrightarrow x),
$$

where $\partial^o S = \{x \in S : \text{there exists a } y \in S^c \text{ s.t. } \{x, y\} \in E\}$ is the internal boundary of $S$.

Proof. For $n \geq 1$ let $\Lambda_n = [-n, n]^d$. Fix $n$ large enough such that $S \subset \Lambda_n$. If $\omega \notin \{0 \leftarrow \Lambda_n \rightarrow x\}$ then there is a SAW of open edges in $\omega$ from 0 to $x$ that must cross $S$, let $y$ be the first vertex in $\partial^o S$ that this path visits, then necessarily $\omega \in \{0 \leftrightarrow \delta_y \} \circ \{y \leftarrow \Lambda_n \rightarrow x\}$ since the set of edges in the path up to $y$ is witness to $\{0 \leftrightarrow \delta_y \}$ for $\omega$ and the remainder of the path is witness to $\{y \leftarrow \Lambda_n \rightarrow x\}$. It follows that $\{0 \leftarrow \Lambda_n \rightarrow x\} \subset \cup_{y \in \partial^o S} \{0 \leftrightarrow \delta_y \} \circ \{y \leftarrow \Lambda_n \rightarrow x\}$, and so by a union bound we have

$$
\mathbb{P}_p(o \leftarrow \Lambda_n \rightarrow x) \leq \sum_{y \in \partial^o S} \mathbb{P}_p(\{0 \leftrightarrow \delta_y \} \circ \{y \leftarrow \Lambda_n \rightarrow x\}).
$$

Then the BK inequality implies that

$$
\mathbb{P}_p(o \leftarrow \Lambda_n \rightarrow x) \leq \sum_{y \in \partial^o S} \mathbb{P}_p(0 \leftrightarrow \delta_y) \mathbb{P}_p(y \leftarrow \Lambda_n \rightarrow x),
$$

and the claim follows by letting $n$ tend to infinity. \hfill \Box

6.3 Margulis-Russo formula

We have already seen that $\mathbb{P}_p(A)$ is non decreasing in $p$ if $A$ is an increasing event. The final tool we will examine in this chapter will help to control the rate at which this probability changes as a function of $p$.

Consider our increasing coupling using the i.i.d. uniform r.v.s $\{U_e\}_{e \in E}$ (recall we denote the measure by $\mathbb{P}$). For an increasing event $A$ that depends on only finitely many edges, using this coupling, we have

$$
\mathbb{P}_{p+\delta}(A) - \mathbb{P}_p(A) = \mathbb{P}(\eta_{p+\delta} \in A, \eta_p \notin A).
$$

If $\eta_{p+\delta} \in A$ and $\eta_p \notin A$ we know there must exist at least one edge $e \in E$ with $p \leq U_e < p + \delta$, denote the set of all such edges by $E_{p, \delta}$. Since the edges have independent weights it is clear that $\mathbb{P}(|E_{p, \delta}| \geq 2) = o(\delta)$, i.e. with high probability there is a single edge with $U_e \in [p, p + \delta)$. If $e$ is the unique edge satisfying $U_e \in [p, p + \delta)$ then this edge must be ‘essential’ for the event $A$ in the sense that $\eta_p \notin A$ but $\eta_p^e \in A$, where in general for $\omega \in \Omega$ and $e \in E$ we define

$$
\omega^e(f) = \begin{cases} 
\omega(f) & \text{if } f \neq e, \\
1 - \omega(f) & \text{if } f = e.
\end{cases}
$$
to be the configuration ‘flipped’ at edge $e$. Note that this ‘essentialness’ of this particular edge could depend on $\eta_p \in A$ but not the specific value of $\eta_p(e)$. It follows that this specific edge contributes to (6.3.1) an amount $\mathbb{P}(U_e \in [p, p+\delta]) \mathbb{P}_P(e \text{ is ‘essential’} \text{ for } A) = \delta \mathbb{P}_P(e \text{ is ‘essential’} \text{ for } A)$. Now summing over all such contributions and taking $\delta \searrow 0$ we have

$$\frac{d}{dp} \mathbb{P}_P(A) = \sum_{e \in E} \mathbb{P}_P(e \text{ is ‘essential’} \text{ for } A).$$

It turns out that this heuristic is essentially valid. Now we make the argument precise.

Fix $A$ an increasing even depending only finitely many edges and $\omega \in \Omega$, we call the edge $e \in E$ pivotal for the pair $(A, \omega)$ if $1_A(\omega) \neq 1_A(\omega^e)$, and define the event

$$\text{piv}_e(A) = \{ e \text{ is pivotal for } A \} = \{ \omega \in \Omega : e \text{ is pivotal for the pair } (A, \omega) \}.$$

**Example 6.14.** If $A = \{0 \leftrightarrow \partial^0 \Lambda_n \}$ then

$$\text{piv}_e(A) = \{ \omega : e \text{ is open then } 0 \leftrightarrow \partial^0 \Lambda_n \text{ and if } e \text{ is closed then } 0 \not\leftrightarrow \partial^0 \Lambda_n \}$$

which is all the $\omega \in \Omega$ such that $e^- \not\leftrightarrow 0, e^+ \leftrightarrow \partial^0 \Lambda_n$, and there exists a closed loop in the dual surrounding $0$ (except for the edge crossing $e$).

**Theorem 6.15** (Russo’s formula). If $A$ is an increasing event, that depends on only finitely many edges, then

$$\frac{d}{dp} \mathbb{P}_P(A) = \sum_{e \in E} \mathbb{P}_P(\text{piv}_e(A)) = \mathbb{E}_P[N(A)],$$

where $N(A)(\omega) = | \{ e \in E : e \text{ is pivotal for } (A, \omega) \} |$.

Again, it is important that $A$ depends on only finitely many edges, consider for example $\theta(p)$ - we expect that this is not differentiable at the critical point.

**Proof.** We start by showing that if $A$ depends on finitely many edges then $f(p) = \mathbb{P}_P(A)$ is differentiable. In fact (with slight abuse of notation since really we have average with respect to the variables on $E \setminus \{e_1, \ldots, e_n\}$),

$$f(p) = \mathbb{E}_P[1_A] = \sum_{\omega_1, \ldots, \omega_n} 1_A(\omega) p^{\omega}(1-p)^{n-\omega},$$

where $\omega = \sum_{i=1}^n \omega_i$. So in particular $f(p)$ is a polynomial in $p$. We can simply calculate the derivative and collect terms to find (check),

$$f'(p) = \frac{1}{p(1-p)} \sum_{i=1}^n \sum_{\omega_1, \ldots, \omega_n} (\omega_i - p) 1_A(\omega) p^{\omega}(1-p)^{n-\omega}$$

$$= \frac{1}{p(1-p)} \sum_{i=1}^n \mathbb{E}_P[1_A(\omega)(\omega_i - p)] = \frac{1}{p(1-p)} \sum_{i=1}^n \mathbb{C}_P[1_A(\omega), \omega_i]. \tag{6.3.2}$$

This is already in a form which may come in handy later, but to express it in the form stated we make a couple more observations. Firstly, if $\omega \in A \cap \text{piv}_e(A)^c$ then $\omega^e \in A$ by definition of $\text{piv}_e(A)$ and $\omega^e \in \text{piv}_e(A)^c$ since $\text{piv}_e(A)$ does not depend on $\omega(e)$, therefore

$$\mathbb{E}_P[1_A(\omega)(\omega_i - p) 1_{\text{piv}_e(A)^c}(\omega)] = \mathbb{E}_P[(\omega_i - p)] \mathbb{E}_P[1_A \cap \text{piv}_e(A)^c] = 0.$$ 

It follows by the total law that

$$\mathbb{E}_P[1_A(\omega)(\omega_i - p)] = \mathbb{E}_P[(\omega_i - p)] 1_A(\omega) 1_{\text{piv}_e(A)}(\omega),$$

and since $A$ is increasing, if $\omega \in \text{piv}_e(A)$, then $1_A(\omega) = 0$ if $\omega_i = 0$ and 1 if $\omega_i = 1$. Therefore

$$\mathbb{E}_P[1_A(\omega)(\omega_i - p)] = (1-p) \mathbb{P}_P(\text{piv}_e(A) \cap \{w_i = 1\}) = (1-p) P(\text{piv}_e(A)),$$

which completes the proof. \qed
Chapter 7

Further percolation

We start this section with some important results on ergodicity (with respect to shifts) and some important results about the infinite cluster. Recall the discussion in Section 4.3 where we discussed the details of the measurable structure of $\Omega$, and we defined shift-ergodic.

**Lemma 7.1.** The measure $P_p$ is invariant under translations and is shift-ergodic.

**Proof.** (will appear) \qed

**Corollary 7.2.** If $p > p_c$ then there is an infinite cluster $P_p$-a.s.

Note we already saw one proof using Kolmogorov’s zero-one law earlier, but it is also an immediate consequence of the previous lemma. It turns out that the infinite cluster (if it exists) is also unique.

**Theorem 7.3.** If $p > p_c$ then there is a unique infinite cluster $P_p$-a.s.

This result was first proved by Aizenman, Kesten and Newman in 1987, and there have been various alternative proof since. I will add references.

### 7.1 Exponential decay in the subcritical regime

It turns out that (strictly) inside the subcritical region, not only is there almost surely no infinite cluster, but the probability that the cluster containing the origin is large decays exponentially quickly in the diameter. More precisely:

**Theorem 7.4** (Exponential decay in the diameter). Fix $d \geq 2$, for every $p \in (0, p_c)$ there exists $c_p > 0$ such that for all $n \geq 1$ we have $P_p(0 \leftrightarrow \partial^o \Lambda_n) \leq e^{-c_p n}$. Furthermore, there exists a $c > 0$ such that for $p \in (0, p_c)$, $\theta(p) \geq c(p - p_c)$.

This theorem was first proved by Aizenman and Barksy in 1987, but we now present are more recent proof by Duminil-Copin and Tassion from 2016. Clearly $P_p(0 \leftrightarrow \partial^o \Lambda_n)$ is at least $p^n = e^{-\left(\log 1/p\right) n}$ since this is the probability that the path from the origin straight up to the boundary is open.

**Proof.** Let $\theta_n(p) = P_p(0 \leftrightarrow \partial^o \Lambda_n)$, and denote the edge boundary of $S \subset \mathbb{Z}^d$ by $\Delta S = \{e \in E : e^- \in S, e^+ \notin S\}$, and finally define

$$\varphi_p(S) = p \sum_{e \in \Delta S} P_p(0 \leftrightarrow e^-),$$

which is the expected number of open edges in the boundary of the connected component of $S$ containing $0$. Clearly if $0 \notin S$ then $\varphi_p(S) = 0$ The proof now relies on the following claims (we postpone their proof to the end):

**Claim 7.5.** If there exists a finite set $S$ containing $0$ such that $\varphi_p(S) < 1$, then there exists $c_p > 0$ such that for all $n \geq 1$ we have $\theta_n(p) \leq e^{-c_p n}$.
For \( n \geq 1 \) define the random subset of \( \Lambda_n \),
\[
\mathcal{S} = \{ x \in \Lambda_n : x \not\leftrightarrow \partial^o \Lambda_n \}.
\]
The boundary of \( \mathcal{S} \) is the ‘out-most blocking surface’ from the boundary of \( \Lambda_n \), you can ‘discover’ \( \mathcal{S} \) by exploring all the sites you can reach following open paths started from the boundary, this will leave ‘unexplored’ islands of sites which can’t be reached from the boundary (belonging to \( \mathcal{S} \)), in a ‘sea’ of the connected component of the boundary. Thus the event \( \{ \mathcal{S} = S \} \) is measurable with respect to all the edges which have at one end outside of \( S \).

**Claim 7.6.** For every \( p \in (0, 1) \) we have
\[
\theta_n'(p) = \frac{1}{p(1 - p)} \mathbb{E}_p[\varphi_p(\mathcal{S})].
\]
(7.1.1)

**Remark 7.7.** Since the edge boundary of \( \mathcal{S} \) must be closed (or else the vertex attached by an open edge would also be in \( \mathcal{S} \)), we can interpret \( \varphi_p(S) \) (up to constant factors) as the expected number of closed pivotal points of \( \{ 0 \leftrightarrow \partial^o \Lambda_n \} \) conditioned on the connected component of the boundary being \( S^c \).

We now complete the proof in light of these claims. This is achieved through a new definition of the critical point that turns out to be equivalent; let
\[
\tilde{p}_c := \sup\{ p \in [0, 1] : \exists \text{ a finite set } S \text{ containing } 0 \text{ such that } \varphi(S) < 1 \}.
\]
By Claim 7.5, for \( p < \tilde{p}_c \) there exists a \( c_p > 0 \) such that \( \theta_n(p) \leq e^{-c_p n} \) for every \( n \geq 1 \), in particular \( \theta(p) = 0 \). Therefore \( p \leq \tilde{p}_c \), so \( \tilde{p}_c \leq p_c \). We now show \( \tilde{p}_c \geq p_c \), i.e. \( \theta(p) > 0 \) for all \( p > \tilde{p}_c \). Fix \( p > p_c \), so by definition \( \varphi(S) \geq 1 \) for all \( S \) containing the origin. Claim 7.6 then gives
\[
\theta'_n(p) = \frac{1}{p(1 - p)} \mathbb{E}_p[\varphi_p(\mathcal{S})] \geq \frac{1}{p(1 - p)} \mathbb{P}_p(0 \in \mathcal{S}) = \frac{1}{p(1 - p)} (1 - \theta_n).
\]
Some basic re-arranging then gives
\[
\left( \log \left( \frac{1}{1 - \theta_n} \right) \right)' \geq \left( \log \left( \frac{p}{1 - p} \right) \right),'
\]
which by integrating between \( \tilde{p}_c \) and \( p_c \) implies that for all \( n \geq 1 \), \( \theta_n(p) \geq \frac{p - \tilde{p}_c}{p(1 - p)} \). Letting \( n \) tend to infinity we have a positive lower bound on \( \theta(p) \) since \( p > \tilde{p}_c \). This completes the proof that \( p_c = p_c \) together with Claim 7.5 this completes the proof of the first part of the theorem. Also, substituting \( \tilde{p}_c = p_c \) into the previous bound on \( \theta(p) \) gives \( \theta(p) \geq c(p - p_c) \) as required (the so-called mean field bound).

**Proof of Claim 7.5.** A sketch could be helpful here. Fix \( S \) such that \( \varphi_p(S) < 1 \), since \( S \) is finite there exists an \( L > 0 \) such that \( S \subset \Lambda_{L-1} \). Now fix \( k \in \mathbb{N} \) and consider the ‘big box’ \( \Lambda_{kL} \), then by the same argument as in the proof of Corollary 6.13, using the BK inequality,
\[
\theta_{kL}(p) \leq p \sum_{e \in \Delta S} \mathbb{P}_p(0 \stackrel{\mathcal{S}}\leftrightarrow e^-) \mathbb{P}_p(e^+ \leftrightarrow \partial^o \Lambda_{kL}) \leq \varphi_p(S) \theta_{(k-1)L}(p),
\]
where in the second inequality we used translation invariance and the fact that \( e^+ \) must be a distance at least \((k - 1)L\) of \( \partial^o \Lambda_{kL} \). [The proof of the above inequality is left as an exercise on Sheet 3]. The claim follows by iteration on \( k \), which give \( \theta_{kL}(p) \leq \varphi_p(S)^k \), i.e. exponential decay of \( \theta_n(p) \) in \( n \).
Proof of Claim 7.6. Here we use Russo’s formula. Let \( A = \{0 \leftrightarrow \partial^0 \Lambda_n\} \), then note that
\[
(1 - p)\mathbb{P}_p(e \text{ pivotal for } A) = \mathbb{P}_p(\{e \text{ pivotal for } A, \omega(e) = 0\}) = \mathbb{P}_p(\{e \text{ pivotal for } A, 0 \not\leftrightarrow \partial^0 \Lambda_n\}),
\]
since (as discussed previously) \( \{e \text{ pivotal for } A\} \) does not depend on the configuration of edge \( e \). The by Russo’s formula,
\[
\theta'_n(p) = \frac{1}{1 - p} \sum_{e \in E_n} \mathbb{P}_p(\{e \text{ pivotal for } A, 0 \not\leftrightarrow \partial^0 \Lambda_n\})
\]
\[
= \frac{1}{1 - p} \sum_{e \in E_n} \mathbb{P}_p(0 \leftrightarrow e^-, e^+ \leftrightarrow \partial^0 \Lambda_n, 0 \not\leftrightarrow \partial^0 \Lambda_n)
\]
where \( E_n \) are the edges between vertices both in \( \Lambda_n \). Now we partition on \( \{S = S\} \), and observe that on the event \( \{S = S\} \) the event on the right hand side above can be interpreted neatly,
\[
\{0 \leftrightarrow e^-, e^+ \leftrightarrow \partial^0 \Lambda_n, 0 \not\leftrightarrow \partial^0 \Lambda_n, S = S\} = \{0 \leftrightarrow S e^-, e^+ \not\in S, 0 \in S, S = S\}.
\]
it follows immediately that we have
\[
\theta'_n(p) = \frac{1}{p(1 - p)} \sum_{S \subseteq \Lambda_n : 0 \in S} \sum_{e \in \Delta S} p \mathbb{P}_p(0 \leftrightarrow S e^-, S = S)
\]
\[
= \frac{1}{p(1 - p)} \sum_{S \subseteq \Lambda_n : 0 \in S} \left( \sum_{e \in \Delta S} p \mathbb{P}_p(0 \leftrightarrow S e^-) \right) \mathbb{P}_p(S = S)
\]
\[
= \frac{1}{p(1 - p)} \mathbb{E}_p[\varphi_p(S)],
\]
where in the second equality we used that \( \{0 \leftrightarrow S e^-\} \) depends only on edges strictly inside \( S \), and \( \{S = S\} \) can be determined from all the other edges (exploring from the boundary in using open paths).

\[\square\]

7.2 Kesten’s theorem

7.3 The Russo-Seymour-Welsh theory

We did not have time to cover this in lectures, but we might explore some of these concepts on the longer final example sheet.
Chapter 8

Exclusion, contact and voter models
Chapter 9

Graphical construction and couplings
Chapter 10

Duality
Chapter 11

Ising and Potts models
Chapter 12

Gibbs states
Chapter 13

Mixing and relaxation