Interacting Stochastic Systems

Benjamin Graham
Fitzwilliam College and Statistical Laboratory
University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

August 2007
Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

I would like to thank my PhD supervisor, Geoffrey Grimmett. Chapters 2 and 3 were done in collaboration with him. We have agreed that 65% of the work is mine. The presentation is entirely my own. The work has also been submitted to journals as joint publications [31, 32]. I am the sole author of Chapter 4.

I would like to acknowledge the financial support I have received from the Engineering and Physical Sciences Research Council under a Doctoral Training Award to the University of Cambridge.
Abstract

Statistical mechanics is the branch of physics that seeks to explain the properties of matter that emerge from microscopic scale interactions. Mathematicians are attempting to acquire a rigorous understanding of the models and calculations arising from physicists’ work. The field of interacting particle systems has proved a fruitful area for mathematical research. Two of the most fundamental models are the Ising model and bond percolation.

The Ising model was suggested by Lenz in 1920. It is a probabilistic model for ferromagnetism. Magnetization can then be explored as correlation between spin random variables on a graph. Bond percolation was introduced by Simon Broadbent and John Hammersley in 1957. It is a model for long range order. Edges of a lattice graph are declared open, independently, with some probability \( p \), and clusters of open edges are studied. Both these models can be understood as aspects of the random-cluster model. In this thesis we study various aspects of mathematical statistical mechanics.

In Chapter 2 we create a diluted version of the random-cluster model. This allows the coupling of the Ising model to the random-cluster model to be extended to include the Blume–Capel model. Crucially, it retains some of the key properties of its parent model. This enables much of the random-cluster technology to be carried forward.

The key issue for bond percolation concerns the fraction of open edges required in order to have long range connectivity. Harry Kesten proved that this fraction is precisely one half for the square planar lattice. Recent development in the theory of influence and sharp thresholds allowed Béla Bollobás and Oliver Riordan to simplify parts of
his proof. In Chapter 3 we extend an influence result to apply to monotonic measures. This allows sharp thresholds to be shown for certain families of stochastically increasing monotonic distributions, including random-cluster measures.

In Chapter 4 we study time to convergence for a mean-field zero-range process. The problem was motivated by the canonical ensemble model of energy levels used in the derivation of Maxwell–Boltzmann distributions. By considering the entropy of the system, we show that the empirical distribution rapidly converges—in the sense of Kullback-Leibler divergence—to a geometric distribution. The proof utilizes arguments of spectral gap and log Sobolev type.
## Contents

1 **Introduction**  
1.1 Phase transition and statistical mechanics .......................... 1  
1.2 Bond percolation .................................................. 2  
1.3 Gibbs distributions ............................................. 3  
1.4 The Ising and Potts models ...................................... 6  
1.5 The random-cluster measure ..................................... 8  
1.6 The FKG inequality and Holley’s theorem .......................... 10  
1.7 Influence and sharp threshold .................................... 12  
1.8 Microcanonical ensemble Markov chains .......................... 14  

2 **Blume–Capel model**  
2.1 The Blume–Capel model ............................................ 16  
2.2 Notation .............................................................. 18  
2.3 FKG and Holley’s theorem ........................................ 19  
2.4 The diluted-random-cluster representation of the BCP model ... 20  
2.5 The vertex DRC marginal measure ................................ 24  
2.6 Edge comparisons .................................................. 29  
2.7 Infinite-volume measures ......................................... 31  
2.8 Uniqueness by convexity of pressure ............................. 37  
2.9 Phases and comparison results ................................... 40  

3 **Influence and monotonic measures**  
3.1 Influence beyond KKL ............................................. 46  
3.2 Symmetric events .................................................. 48  
3.3 Sharp thresholds for product measures ........................... 49
List of Figures

1.1 Approximate phase diagram for H$_2$O ........................................... 2
1.2 Percolation with $p = 0.48$, 0.50 and 0.52 on the $20 \times 20$ torus and
the $200 \times 200$ torus. The largest open cluster is black, the other
open edges are grey. .......................................................... 4
2.1 Approximate Blume–Capel phase diagram .......................... 11
2.2 Approximate Blume–Capel phase diagram .......................... 12
2.3 Blume–Capel comparison results .......................... 13
3.1 Finite graph and dual .......................................................... 51
3.2 Square planar lattice and isomorphic dual .......................... 51
4.1 A sample path: the empirical distribution for $N = 10^7$, $R = 20$ at
intervals of $40N$ steps .......................................................... 70
4.2 Hydrodynamic limit $X^f_R$ tends to $1/2$ .......................... 73
4.3 Graph of $\delta(x, y)$ as a function of $y$ .......................... 74
Chapter 1

Introduction

1.1 Phase transition and statistical mechanics

Statistical mechanics seeks to explain the properties of matter that emerge from microscopic scale interactions. It has for a long time been a source of interesting mathematical problems. The field goes back to the work of Boltzmann, Gibbs and Maxwell in the late 19th and early 20th century. Most models are not exactly solvable. Simple local interactions produce complex long range behaviour.

In 1895 Pierre Curie discovered that ferromagnetic substances exhibit a phase transition with temperature—in the case of iron at 1043 K. The phenomenon had been observed 60 years earlier by Claude Pouillet; however, he had lacked the means to measure temperature as precisely. In the classic physics experiment, an iron bar is placed with a strong magnetic field parallel to its axis. The bar becomes magnetized. When the magnetic field is reduced to zero, one of two things happens. If the temperature is above the Curie point, the bar retains no magnetism. When the temperature is below the Curie point, it does retain a magnetized state. This is a physical example of a threshold phenomenon.

Statistical physicists have devised stochastic models for magnetic spin systems in thermal equilibrium. They allow the study of the phase transition phenomenon. The most famous model for ferromagnetism is the Ising model. It exhibits phase transitions in two and higher dimensions. Another type of phase transition is
1.2 Bond percolation

It would be almost impossible to begin a thesis on statistical mechanics without first mentioning percolation. Bond percolation was introduced by Simon Broadbent and John Hammersley in 1957 [12]. It is one of the simplest spatial models with a phase transition phenomenon, yet still some basic question have proved difficult to answer. Bond percolation is defined on a lattice. Rather than defining exactly what we mean by a lattice we will just give the standard example. Let $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ be the set of integers and let $\mathbb{Z}^d$, $d \geq 2$, be the set of $d$-tuples of integers,

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

Figure 1.1: Approximate phase diagram for H$_2$O

that exhibited by H$_2$O. The phase diagram has a tri-critical point where three phases meet—the triple point of water, 273 K, 612 Pa. See Figure 1.1.
Define the $L_1$ distance $l : \mathbb{Z}^d \times \mathbb{Z}^d \to [0, \infty)$ by

$$l(x, y) = \sum_{i=1}^{d} |x_i - y_i|, \quad x, y \in \mathbb{Z}^d.$$ 

Take $E^d$ to be the pairs of lattice points at distance one, so $l$ is the normal graph-theoretic distance.

$$E^d = \{ (x, y) : x, y \in \mathbb{Z}^d, l(x, y) = 1 \}.$$ 

This is the set of ‘nearest-neighbours’ of $\mathbb{Z}^d$. The graph $L^d = (\mathbb{Z}^d, E^d)$ is called the $d$-dimensional hypercubic lattice.

Let $\Omega = \{0, 1\}^{\mathbb{Z}^d}$. For $\omega \in \Omega$, we say that an edge is $\omega$-open (or just open) if $\omega(e) = 1$, and $\omega$-closed otherwise. Let $(\Omega, \mathcal{F}, \mu_p)$ be the product probability space such that the edge states are independent and each edge is open with probability $p$. For $x, y \in \mathbb{Z}^d$, write $\{x \leftrightarrow y\}$ for the event that there is a chain of open edges connecting $x$ to $y$. Write $\{x \leftrightarrow \infty\}$ for the event that $x$ is the starting point of an arbitrarily long chain of open edges. The open clusters are the maximal sets of connected vertices. The model is said to percolate, or to be supercritical, if $\mu_p(0 \leftrightarrow \infty) > 0$. By Kolmogorov’s zero-one law, the system has an infinite cluster if and only if $\mu_p(0 \leftrightarrow \infty) > 0$. It is clear that $\mu_p(0 \leftrightarrow \infty)$ is non-decreasing in $p$. The critical probability is defined

$$p_c = \inf \{ p : \mu_p(0 \leftrightarrow \infty) > 0 \}.$$ 

The exact critical point for the hypercubic lattice is only known for $d = 2$; this case is particularly amenable to analysis due to the ‘self-duality’ of the planar lattice.

### 1.3 Gibbs distributions

Suppose a system can be described by a configuration $\sigma \in \Sigma$ with $\Sigma$ finite. Let $\mathcal{H}$ be the energy function, or Hamiltonian, $\mathcal{H} : \Sigma \to \mathbb{R}$. Let $T$ be the temperature of the system and take $\beta = (k_B T)^{-1}$ to be the inverse-temperature; the constant $k_B$ is Boltzmann’s constant. Define the partition function,

$$Z(\beta) = \sum_{\sigma} e^{-\beta \mathcal{H}(\sigma)}.$$
Figure 1.2: Percolation with $p = 0.48$, 0.50 and 0.52 on the $20 \times 20$ torus and the $200 \times 200$ torus. The largest open cluster is black, the other open edges are grey.
Then the Gibbs distribution $G_\beta$ is defined by

$$G_\beta(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z(\beta)}.$$ 

We are abusing notation by using $G_\beta$ to mean both a probability measure and the corresponding mass function. The motivation for the measure is as follows. Assume now for simplicity that $H$ is an integer valued function. As $\Sigma$ is finite it does not have interesting statistical properties by itself. Consider a canonical ensemble of $N$ copies of the system, with states $\sigma^{(1)}, \ldots, \sigma^{(N)}$. The total energy of the ensemble is

$$E_{\text{tot}} = \sum_{i=1}^{N} H(\sigma^{(i)}).$$

Imagine that the copies have been weakly interacting in some way that conserves the total energy, and have reached an equilibrium state. Let

$$\Omega(E_{\text{tot}}) = \sum_{\sigma^{(1)}} \cdots \sum_{\sigma^{(N)}} \delta_{E_{\text{tot}}, E_{\text{tot}}},$$

counting the number of configurations of the canonical ensemble with total energy $E_{\text{tot}}$. The basic assumption of statistical mechanics is that all compatible configurations are equally likely once the system has reached equilibrium. Such systems are said to have Maxwell–Boltzmann statistics. This gives

$$\mathbb{P}(\sigma^{(1)}, \ldots, \sigma^{(N)} \mid E_{\text{tot}}) = \frac{\delta_{E_{\text{tot}}, E_{\text{tot}}}}{\Omega(E_{\text{tot}})},$$

Take the limit $N \to \infty$ with $E = E_{\text{tot}}/N$ a fixed integer. We want to show that the marginal distribution of one system converges to a Gibbs distribution. If $p(\cdot) = \lim_{N \to \infty} \mathbb{P}(\sigma^{(1)} = \cdot \mid E_{\text{tot}} = N E)$, then the energy of $p$ is $\sum_{\sigma} p(\sigma) H(\sigma) = E$. Define entropy,

$$S(p) = -\sum_{\sigma} p(\sigma) \log p(\sigma).$$

Let $\delta : (0, \infty) \times [0, \infty) \to \mathbb{R}$ be given by

$$\delta(x, y) = y \log y - x \log x - (y - x) \left[ \frac{d}{dt} t \log t \bigg|_{t=x} \right] \geq 0.$$
Take inverse temperature $\beta$ such that $G_\beta$ has energy $\bar{E}$,

$$\bar{E} = -\frac{\partial}{\partial \beta} \log Z(\beta).$$

For all distributions $p$ with energy $\bar{E}$,

$$S(G_\beta) - S(p) = \sum_\sigma \delta(G_\beta(\sigma), p(\sigma)) \geq 0.$$

Therefore $G_\beta$ is the maximum entropy distribution. This implies,

$$G_\beta(\sigma) = \lim_{N \to \infty} P(\sigma^{(1)} = \sigma | E_{\text{tot}} = NE).$$

### 1.4 The Ising and Potts models

The Ising model was suggested by Lenz in 1920, and first analysed by his student Ising in 1925 [39]. It is a probabilistic model for ferromagnetism. Magnetization can be explored as correlation between simple random variables on a graph. Ising analysed the model on the set of integers with nearest neighbour interactions. At finite temperature, the Ising model is simply an ergodic two state Markov Chain—there is no long range order. Certain associated calculations may be done exactly on the $d = 2$ hypercubic lattice with uniform nearest neighbour interactions and no external magnetic field [45, 47]. There are many similar models for magnetism [2].

Let $G = (V, E)$ be a graph. A spin model consists of a set of spins $S$ that each vertex can take, and a Hamiltonian function $\mathcal{H} : \Sigma \to \mathbb{R}$ from the set of configurations $\Sigma = S^V$ to energy values. Hamiltonians are often written in an ill defined way, as sums that do not converge on unbounded graphs. However, if two configurations $\sigma_1, \sigma_2 \in \Sigma$ only differ at a finite number of vertices, it is standard to interpret the sums in such a way that $\mathcal{H}(\sigma_1) - \mathcal{H}(\sigma_2)$ is well-defined.

Let $k_B$ be Boltzmann’s constant and $T \in (0, \infty)$ be the temperature. Define inverse temperature $\beta = (k_B T)^{-1}$. The Ising model has spin states $S = \{+1, -1\}$. Let $J \in \mathbb{R}$ be the interaction strength and let $h \in \mathbb{R}$ be the strength of the external magnetic field. The Hamiltonian $\mathcal{H} : \Sigma \to \mathbb{R}$ is given by

$$\mathcal{H}(\sigma) = -J \sum_{(x,y) \in E} \sigma(x)\sigma(y) - h \sum_{x \in V} \sigma(x), \quad \sigma \in \Sigma = S^V.$$
1.4 The Ising and Potts models

The first sum is over the set of unordered edges and the second is over the set of vertices. If \( J > 0 \) the model is called ferromagnetic; the spins are positively correlated. This is easily seen using the random-cluster representation that will be introduced in the next section. An important special case is \( h = 0 \), the Ising model with no external magnetic field.

If \( G \) is finite we can immediately define the Ising measure \( \pi = \pi_{G, \beta J, \beta h} \) by the Gibbs formalism,

\[
\pi(\sigma) = \frac{1}{Z^I} \exp(-\beta H(\sigma)), \quad \sigma \in \Sigma,
\]

with

\[
Z^I = Z^I_{\beta J, \beta h}(G) = \sum_{\sigma \in \Sigma} \exp(-\beta H(\sigma))
\]

the necessary normalizing constant. \( Z^I \) is called the Ising partition function.

If \( G \) is infinite, constructing the Ising measure is not quite so straightforward. Define \( \Lambda = (V, E) \) a region of \( \mathbb{G} \). Let \( V \subset \mathbb{V} \) be a finite set of vertices. Let \( \partial V \) be the external vertex boundary of \( V \), the vertices in \( \mathbb{V} \setminus V \) that are adjacent to a member of \( V \). Let \( E \) be the subset of edges \( \mathbb{E} \) with both end vertices in \( V \cup \partial V \). Note \( (V \cup \partial V, E) \) is a subgraph of \( \mathbb{G} \). Let \( \xi \in \Sigma \), call \( \xi \) the boundary conditions. Let \( \Sigma^\xi_\Lambda \) be the set \( \{ \sigma \in \Sigma : \sigma(x) = \xi(x) \text{ for } x \in \mathbb{V} \setminus V \} \). The Ising measure on \( \Lambda \) with boundary conditions \( \xi \) is defined by,

\[
\pi^\xi_{\Lambda, \beta J, \beta h} (\sigma) = \frac{1}{Z^I} \exp \left\{ \beta J \sum_{(x,y) \in E} \sigma(x)\sigma(y) + \beta h \sum_{x \in V} \sigma(x) \right\}, \quad \sigma \in \Sigma^\xi_\Lambda.
\]

Again \( Z^I = Z^I_{\beta J, \beta h}(\xi, \Lambda) \) is the required normalizing constant. There are two standard ways to define the Ising model on an infinite graph \( \mathbb{G} \). They are both measures on \( (\Sigma, \mathcal{F}) \) where \( \mathcal{F} \) is the \( \sigma \)-algebra generated by the product topology on \( \Sigma = S^\mathbb{V} \).

The first is as the infinite-volume limit of finite Ising measures. Let \( (\Lambda_n) \) be an increasing sequence of regions that tend towards \( \mathbb{G} \) as \( n \to \infty \). \( \mathcal{W}_{\beta J, \beta h} \) is defined as the set of accumulation points of the sequences of measures \( \pi^\xi_{\Lambda_n, \beta J, \beta h} \). There is no loss of generality in using \( \xi \) instead of a series of boundary conditions \( (\xi_n) \), see [33]. Write \( \text{co} \mathcal{W}_{\beta J, \beta h} \) for the convex closure of \( \mathcal{W}_{\beta J, \beta h} \). If, for some \( \xi \in \Sigma \) and all increasing sequences \( (\Lambda_n) \) that tend to \( \mathbb{G} \) there is a single accumulation point, we
1.5 The random-cluster measure

can define that point to be \(\pi_{\beta J,\beta h}^\xi\). We can also define the infinite-volume limit with free boundary conditions. Let \((\Lambda_n)\) be an increasing sequence of subgraphs that tend to \(G\), and consider the accumulation points of the measures \(\pi_{\Lambda_n,\beta J,\beta h}\).

The second is the Dobrushin-Lanford-Ruelle (DLR) formalism. Let \(T_\Lambda\) be the \(\sigma\)-algebra generated by the states of vertices outside \(\Lambda\). A DLR measure is a measure \(\pi\) such that for \(\pi\)-almost every \(\xi \in \Sigma\), for all regions \(\Lambda\),

\[
\pi(A \mid T_\Lambda)(\xi) = \pi_{\Lambda,\beta J,\beta h}^\xi (A), \quad \forall A \in \mathcal{F}.
\]

The Ising model can be seen as the ‘\(q = 2\)’ case of the Potts model \([49]\). For \(q \in \{1, 2, \ldots\}\) the \(q\)-state Potts model has spin states \(S = \{1, 2, \ldots, q\}\). Let \(J \in \mathbb{R}\) be the interaction strength. The Hamiltonian is given by

\[
\mathcal{H}(\sigma) = -J \sum_{(x,y) \in E} \delta_{\sigma(x),\sigma(y)}, \quad \sigma \in \Sigma = S^V.
\]

This defines measures for inverse temperature \(\beta\) on graph \(G\), in the same ways as the Ising model. The Potts model may be viewed as a generalization of the \(h = 0\) Ising model. The \(-1, +1\) Ising states are represented in the \(q = 2\) Potts model by states 1, 2.

1.5 The random-cluster measure

A key step in the understanding of the Ising and Potts models was the development of the random-cluster model by Fortuin and Kasteleyn \([25]\). It provides a connection between bond percolation and the magnetic spin models. For a recent account of the random-cluster measure see \([30, 34]\). It is an edge model like bond percolation. Edges are open or closed, maximal set of vertices connected by open edge paths are called clusters.

Let \(G = (V, E)\) be a finite graph. Let \(\Omega = \{0, 1\}^E\). For \(\omega \in \Omega\), let \(k(\omega)\) be the number of \(\omega\)-open clusters. The random-cluster measure with parameters \(p \in (0, 1), q \in (0, \infty)\) is given by,

\[
\mu(\omega) = \frac{1}{Z_{\text{RC}}^p(G)} q^{k(\omega)} \prod_{e \in E} \left( \frac{p}{1-p} \right)^{\omega(e)}
\]
1.5 The random-cluster measure

with normalizing constant, the random-cluster partition function,

\[ Z_{p,q}^{RC}(G) = \sum_{\omega \in \{0,1\}^E} q^{k(\omega)} \prod_{e \in E} \left( \frac{p}{1-p} \right)^{\omega(e)}. \]  

(1.1)

If \( q = 1 \) this is just a product measure with density \( p \). If \( q > 1 \) then the measure favours having more clusters and so the edge density is driven down. The reverse holds for \( q < 1 \). This is a generalisation of percolation with a dependence structure. As described in the next section, measure \( \mu \) has the property of monotonicity if \( q \geq 1 \).

The density of an edge \( e \), conditional on the states of the other edges, takes a particularly simple form. Let \( J_e \) be the event edge \( e \) is open and let \( T_e \) be the \( \sigma \)-algebra generated by \( \{ J_f : f \neq e \} \). The value of \( \mu(J_e \mid T_e) \) depends on whether or not the edges \( E \setminus \{ e \} \) contain an open path between the end vertices of \( e \),

\[ \mu(J_e \mid T_e) = \begin{cases} p & \text{if the end vertices of } e \text{ are joined,} \\ \frac{p}{p+q(1-p)} & \text{otherwise.} \end{cases} \]  

(1.2)

The random-cluster measure can also be defined with boundary conditions on a region \( \Lambda = (V, E) \) of an infinite graph \( G \). The number of clusters \( k(\omega) \) is then taken to be the number of open edge clusters of \( G \) that intersect \( V \cup \partial V \).

The random-cluster model is naturally coupled to the Potts model. Let \( G = (V, E) \) be a finite graph with \( q \)-state Potts measure \( \phi \) at inverse temperature \( \beta \), and \( (p,q) \) random-cluster measure \( \mu \). If \( p = 1 - e^{-\beta J} \) there is a measure \( \pi \) on \( \{1, \ldots, q\}^V \times \{0,1\}^E \) with the following properties.

(i) The marginal measure on \( \{1, \ldots, q\}^V \) is \( \phi \).

(ii) The marginal measure on \( \{0,1\}^E \) is \( \mu \).

(iii) Given the edge configuration, spins are constant on connected clusters. Between the different clusters, the spins are independent. The common spin of each cluster is uniformly distributed.

(iv) Given the vertex spins, an edge \( \langle x, y \rangle \) is open with probability \( p \) if the spins at \( x \) and \( y \) agree, and closed otherwise.
We get the following formula relating long range connectivity in the random-cluster model and long range correlation in the Potts models. The two-point correlation function for the Potts model is,

$$\tau(x, y) = \pi(\sigma_x = \sigma_y) - q^{-1} = (1 - q^{-1}) \phi(x \leftrightarrow y).$$

In Chapter 2 we consider a variant of the Potts model, the Blume–Capel model. We analyse the Blume–Capel model by extending the coupling between the Potts model and the random-cluster model.

### 1.6 The FKG inequality and Holley’s theorem

Let $\Omega$ be a partially ordered set. $\Omega$ is called a **lattice** if for all $\omega_1, \omega_2 \in \Omega$, there exists a least upper bound $\omega_1 \vee \omega_2 \in \Omega$, and a greatest lower bound $\omega_1 \wedge \omega_2 \in \Omega$. A lattice is **distributive** if for all $\omega_1, \omega_2, \omega_3 \in \Omega$,

$$\omega_1 \wedge (\omega_2 \vee \omega_3) = (\omega_1 \wedge \omega_2) \vee (\omega_1 \wedge \omega_3),$$

$$\omega_1 \vee (\omega_2 \wedge \omega_3) = (\omega_1 \vee \omega_2) \wedge (\omega_1 \vee \omega_3).$$

Let $S \subset \mathbb{R}$ and let $\Omega = S^I$ for a set $I$. Then $\Omega$ is a distributive lattice with respect to the coordinate-wise partial order: $\omega_1 \leq \omega_2$ if $\forall i \in I, \omega_1(i) \leq \omega_2(i)$. For $\omega_1, \omega_2 \in \Omega$,

$$(\omega_1 \vee \omega_2)(i) = \max\{\omega_1(i), \omega_2(i)\},$$

$$(\omega_1 \wedge \omega_2)(i) = \min\{\omega_1(i), \omega_2(i)\}.$$  

A set $A \subset \Omega$ is called **increasing** if it is closed upwards with respect to the partial order,

$$\forall \omega_1, \omega_2 \in \Omega : \omega_1 \leq \omega_2 \text{ and } \omega_1 \in A \rightarrow \omega_2 \in A.$$  

Let $\mathcal{F}$ be the set of all subsets of $\Omega$. Again we abuse notation by using $\mu$ to mean both a measure on $(\Omega, \mathcal{F})$, and the corresponding mass function from $\Omega$ to $[0, 1]$. A probability measure $\mu$ is positively-associated if for all increasing $A, B \subset \Omega$,

$$\mu(A \cap B) \geq \mu(A)\mu(B).$$
An early result in percolation theory is that product measures on sets \( \{0, 1\}^I \) are positively-associated \[37\]. Motivated by the Potts measure, Fortuin, Kasteleyn and Ginibre showed the following \[26\].

**Theorem 1.3.** Let \( \mu \) be a probability measure on a finite distributive lattice \( \Omega \). If for all \( \omega_1, \omega_2 \in \Omega \),
\[
\mu(\omega_1 \lor \omega_2) \mu(\omega_1 \land \omega_2) \geq \mu(\omega_1) \mu(\omega_2)
\] (1.4)
then \( \mu \) is positively-associated.

Inequality (1.4) is known as the FKG lattice condition. Holley proved a related result \[38\]. We actually state a slight generalisation, see for example \[50\]. Given measures \( \mu_1, \mu_2 \) on \( \Omega \) say that \( \mu_1 \) is stochastically dominated by \( \mu_2 \), written \( \mu_1 \preceq \mu_2 \), if for all increasing \( A \subset \Omega \), \( \mu_1(A) \leq \mu_2(A) \).

**Theorem 1.5.** Let \( \Omega \) be a finite distributive lattice. Let \( \mu_1, \mu_2 \) be two measures on \( (\Omega, F) \) such that for all \( \omega_1, \omega_2 \in \Omega \),
\[
\mu_1(\omega_1 \land \omega_2) \mu_2(\omega_1 \lor \omega_2) \geq \mu_1(\omega_1) \mu_2(\omega_2).
\]
Then \( \mu_1 \preceq \mu_2 \).

We will also define strong positive-association and monotonicity. We call \( \mu \) positive if for all \( \omega \in \Omega \), \( \mu(\omega) > 0 \). Let \( \mu \) be a positive measure on a finite distributive lattice \( \Omega = S^I \). Let \( J \subset I \). Define \( \Omega_J = S^J \) and for \( \xi \in \Omega \) let
\[
\Omega^I_J = \{ \omega \in \Omega : \omega(i) = \xi(i) \text{ for } i \in I \setminus J \}.
\]
Let \( \mu^I_J \) be the measure \( \mu \), conditioned on the coordinates in \( I \setminus J \) taking the corresponding values from \( \xi \). Let \( X_i \) be the random variable \( X_i = \omega(i) \). For \( \eta \in \Omega_J \),
\[
\mu^I_J(\eta) = \mu(X_j = \eta(j) \mid X_i = \xi(i) \text{ for } i \in I \setminus J).
\]
We say that \( \mu \) is strongly positively-associated if for all \( J \subset I \) and for all \( \xi \in \Omega \), \( \mu^I_J(\cdot) \) is positively associated. We say that \( \mu \) is monotonic if for \( J \subset I \) and for all increasing \( A \subset \Omega \), \( \mu^I_J(A) \) is increasing in \( \xi \).

If \( \Omega = \{0, 1\}^I \) finite and \( \mu \) is a positive measure on \( (\Omega, F) \) then these results take a particularly simple form \[34\].
1.7 Influence and sharp threshold

Theorem 1.6. The following are equivalent:

(i) $\mu$ satisfies the FKG lattice condition \[1.4\].
(ii) $\mu$ is strongly positively-associated.
(iii) $\mu$ is monotonic.

1.7 Influence and sharp threshold

Another type of random graph is due to Erdős and Rényi \[21\]. The Erdős–Rényi graph $G(n, p)$ is a bond percolation model on the complete graph. Starting with the graph of $n$ vertices and all possible $\binom{n}{2}$ edges, edges are open with probability $p$ and otherwise closed. An alternative model is $G(n, M)$. All $M$-edge subgraphs of the complete graph are assigned probability $1/\binom{n}{M}$. Many properties of Erdős–Rényi graphs have sharp thresholds. It is natural to consider the limit $n \to \infty$ with $p = c/n$ for constant $c$. If $c < 1$ then the largest connected cluster has size $\Theta(\log n)$. If $c > 1$ then there is a unique giant cluster of size $\Theta(n)$, and the second largest cluster has size $\Theta(\log n)$.

A concept that has proved to be closely related to sharp-thresholds is that of influence. Ben-Or and Linial \[3\] considered influence in the context of computer science. They were studying coin tossing algorithms involving fair coins. How important is any given coin? For $N \in \mathbb{N}$ let $\Omega = \{0, 1\}^N$, the $N$-dimensional Hamming space. Recall that $\Omega$ has the coordinate-wise partial order. Let $\mu$ be the uniform measure on $\Omega$, so the coordinates are independent Bernoulli $1/2$ random variables. We will use the following notation to manipulate single coordinates of vectors in $\Omega$.

$$
\omega^j(i) = \begin{cases} 
1 & i = j \\
\omega(i) & i \neq j 
\end{cases}
$$

$$
\omega_j(i) = \begin{cases} 
0 & i = j \\
\omega(i) & i \neq j 
\end{cases}
$$

For a Boolean function $f : \Omega \to \{0, 1\}$ define the influence of the $j$-th coordinate,

$$
I_f(j) = \mu(\{\omega \in \Omega : f(\omega_j) \neq f(\omega^j)\}).
$$

Equivalently, for $A \subset \Omega$, let $f$ be the indicator function of $A$, so we have influence $I_A(j) = I_f(j)$. Clearly if $\mu(f)$ is very close to either 0 or 1, then the influence of
1.7 Influence and sharp threshold

every coordinate can be small. To avoid this, Ben-Or and Linial asked how small the maximum influence could be if $\mu(f) = 1/2$.

In this context, to study functions that minimize influence it is sufficient to consider only increasing functions. Let $f$ be a general Boolean function. Then by Proposition 2.1 of [3], there is an increasing Boolean function $g$ such that

$$\mu(f) = \mu(g), \quad \forall j, I_f(j) \geq I_g(j).$$

This can be shown using a technique from extremal set theory. The values of $f$ on pairs $(\omega_i, \omega_i')$ where $f(\omega_i) > f(\omega_i')$ can be swapped, making the function ‘more’ monotonic without increasing the coordinate-wise influences.

Let $f : \Omega \rightarrow \{0, 1\}$ be a Boolean function such that $\mu(f) = 1/2$. Consider for example $f(\omega) = \omega(i)$, the dictatorship of coordinate $i$. Then $I_f(j) = 1$ if $i = j$ and $I_f(j) = 0$ otherwise. An example with much smaller maximum influence is the ‘majority’ function. Suppose for convenience that $N$ is odd. Set $f(\omega) = 1$ if more than half of the $\omega(i)$ are 1. The influence of each coordinate is $\Theta(N^{-1/2})$, the probability exactly half of the other $N - 1$ coordinates take value 1.

They found they could do better, giving the ‘tribe’ function with maximum influence $\Theta(N^{-1} \log N)$. The $N$ coordinates are split into approximately $N/\log_2 N$ tribes of size $\log_2 N - \log_2 \log_2 N + O(1)$. Take $f(\omega) = 1$ when at least one tribe is unanimously 1. The influence of a coordinate is the probability that all the other members of the tribe take value 1, probability $\Theta(N^{-1} \log N)$, and that none of the other tribes is unanimously 1, probability $\Theta(1)$. They conjectured that for all $f$ with $\mu(f) = 1/2$, for some $j$, $I_f(j) = \Omega(N^{-1} \log N)$. This was proved by a set of authors known as KKL [41]. Their proof uses the observation that the influences for increasing function $f$, equation (1.7), are Fourier coefficients of the function on the group $\mathbb{Z}_2^n$. They actually prove a more general result, without the restriction $\mu(f) = 1/2$. One can take $C = 1/5$ below.

**Theorem 1.8 (KKL).** There is an absolute constant $C > 0$ such that for all Boolean functions $f$,

$$\exists j, I_f(j) \geq C \min\{\mu(f), 1 - \mu(f)\} \frac{\log N}{N}.$$
1.8 Microcanonical ensemble Markov chains

This definition of influence can be extended to more general product measure, see [27]. Let \( \Omega = [0, 1]^N \) and let \( \lambda \) be the uniform measure on \( \Omega \). For \( x = (x_1, \ldots, x_N) \) and \( j \in \{1, \ldots, N\} \), define the set

\[
S_j(x) = \{ y \in \Omega : y(i) = x(i) \text{ for } i \neq j \}.
\]

This is the subset of \( \Omega \) generated by freeing the \( j \)-th coordinate. Given a random variable \( f : [0, 1]^N \to \{0, 1\} \), define the influence of coordinate \( j \) on \( f \),

\[
I_f(j) = \lambda(\{x : f \text{ not constant on } S_j(x)\}).
\]

The result of KKL has been extended to this setting by a superset of authors known as BKKKL. We use this result to study influence and sharp thresholds for monotonic measures in Chapter 3.

1.8 Microcanonical ensemble Markov chains

We have introduced the Gibbs formalism and shown how Gibbs measures correspond to maximum entropy distributions. This has aroused some interest in Markov chain models for microcanonical ensembles, such as the Ehrenfest urn model [20, 44]. In Boltzmann’s \( H \) theory, entropy always increases. Markov chain models often have a seemingly paradoxical property. It seems natural that they should be time reversible. However, then by ergodicity any function of the system that goes up must also come down.

In the Ehrenfest urn model there is an urn of \( N \) coloured balls, each either red or green. A ball is picked at random and replaced with a ball of the other colour. To ensure aperiodicity, we will consider a lazy version: each step do nothing with probability one half. The equilibrium distribution is uniform on the set \( \{\text{red, green}\}^N \), the system has Maxwell–Boltzmann statistics [40]. If the system is started from a highly ordered state, say all green balls, after time \( O(N \log N) \) it is in a disordered state, with an approximately equal number of red and green balls [19, 48].

In Chapter 4, we study a different Markov chain model. We start with \( N \) boxes each containing \( R \) indistinguishable balls. The process evolves as follows.
Pick two boxes uniformly at random. If the first box has any balls in, move one ball to the second box. It is a discrete time ‘mean-field’ version of the zero-range process \[53\]. A zero-range process is defined on a directed graph. For simplicity I will only describe a finite one-dimensional version. Consider \(N\) boxes arranged in a circle with edges oriented clockwise. Distributed among the boxes are balls. Balls move along the directed edges. The rate at which balls leave a box may vary by location and box occupancy. The problem is hard to analyse; the model exhibits a phenomenon known as condensation where a large fraction of all the balls end up in one box \[22\].

We consider a zero-range process on the complete graph; there are directed edges in both directions between each pair of boxes. The process can be interpreted as a model for a microcanonical ensemble. Think of the boxes as particles and the balls as quanta of energy. The equilibrium distribution has Maxwell–Boltzmann statistics.
Chapter 2

Blume–Capel model

2.1 The Blume–Capel model

We have introduced the Ising model for ferromagnetism. Each vertex takes spin +1 or −1. The Blume–Capel model is a generalization of the Ising model, formulated by Blume in 1966 [7]. The set of spin states is $S = \{+1, 0, -1\}$. The Blume–Capel model is defined on a graph $G = (V, E)$ by the Hamiltonian,

$$
\mathcal{H}(\sigma) = -J \sum_{\langle x, y \rangle \in E} \sigma(x)\sigma(y) + D \sum_{x \in V} \sigma(x)^2 - h \sum_{x \in V} \sigma(x), \quad \sigma \in \Sigma = S^V.
$$

This can be considered a dilution of the Ising model in the following sense. Let $G$ be a finite graph. For a configuration $\sigma \in \Sigma$, define the undeleted graph to be $G_\sigma = (V_\sigma, E_\sigma)$, $V_\sigma = \{x : \sigma(x) \neq 0\}$ and $E_\sigma = \{\langle x, y \rangle \in E : x, y \in V_\sigma\}$. Condition the Blume–Capel measure on the event that the undeleted graph takes a particular value; the resulting measure is the Ising measure on the undeleted graph.

A first order phase transition is one where a property of the model, such as the density of +1 spins, is discontinuous as a function of the parameters. A second order phase transition is one without this phenomenon. Capel studied the ferromagnetic model ($J > 0$) by mean field approximations [14, 15, 16]. He predicted that in the case $h = 0$, on a lattice with degree $\delta$, there is a first order phase transition when \( \frac{1}{3} J \delta \log 4 < D < \frac{1}{2} J \delta \). He predicted a second order phase transition when $D < \frac{1}{7} J \delta \log 4$ and no long range order for any temperature when
2.1 The Blume–Capel model

\[ D > \frac{1}{2} J \delta. \] It is believed that even on two dimensional lattices there is a first-order phase transition. It is predicted that there is a tri-critical point at which the phase transition goes from first-order to second-order.

Some of the results we will show for the Blume–Capel model are in fact true in a more general setting. The generalization of the Ising model from two states to the \( q \)-state Potts model motivates an expansion of the Blume–Capel model with \( h = 0 \). The Blume–Capel zero state is kept, but the ±1 states are replaced by states 1, 2; and more generally 1, \ldots, \( q \).

Define the Blume–Capel–Potts (BCP) Hamiltonian as follows. Let \( q \) be a positive integer, let \( S = \{0, 1, \ldots, q\} \) and let \( \Sigma = S^V \). Again, define the deleted graph by \( V_\sigma = \{x : \sigma(x) \neq 0\} \) and \( E_\sigma = \{(x, y) \in E : x, y \in V_\sigma\} \). Let

\[ H(\sigma) = J|E_\sigma| - 2J \sum_{(x, y) \in E_\sigma} \delta_{\sigma(x), \sigma(y)} + D|V_\sigma|, \quad \sigma \in \Sigma. \]

Let \( K = \beta J \), \( \Delta = \beta D \). Let \( \Lambda = (V, E) \) be a region of \( \mathbb{G} \) and let \( \xi \in \Sigma \). Define \( \Sigma_\Lambda^\xi = \{\sigma \in \Sigma : \sigma(x) = \xi(x) \text{ for } x \notin V\} \). Let \( V_\sigma = V_\sigma \cap V \) and \( E_\sigma = E_\sigma \cap E \).

Define the BCP measure on region \( \Lambda \) by

\[ \pi_{\xi, K, \Delta, q}(\sigma) = \frac{1}{Z_{BCP}} \exp \left[-K|E_\sigma| + 2K \sum_{(x, y) \in E_\sigma} \delta_{\sigma(x), \sigma(y)} - \Delta|V_\sigma|\right], \quad \sigma \in \Sigma_\Lambda^\xi. \]

It is easy to see that the \( q = 2 \) case is equivalent to the Blume–Capel measure.

In Chapter 1 we described a coupling of the Potts spin model and the random-cluster bond model. In Section 2.4, we demonstrate a random-cluster representation for the Blume–Capel–Potts model—we call it the diluted-random-cluster model. Paths of open edges in the random-cluster model correspond to order amongst the spins of the Potts model. Likewise, paths of open edges in the diluted-random-cluster model correspond to order amongst the non-zero spins of the BCP model. We say that it is a diluted version of the random-cluster model. Some vertices, and the adjacent edges, are removed—they correspond to zero spins in the BCP model.

In Section 2.5 we show that the distribution of vertices deleted in the diluted-random-cluster model is monotonic. We also give various comparison results in Sections 2.5 and 2.6, indicating how the diluted-random-cluster measure varies
with the parameters of the model and with the boundary conditions. The comparison results relating to the boundary conditions allow us to define infinite-volume limit measures in Section 2.7 extending the coupling to the hypercubic lattice. In Section 2.8 we use the convexity of the BCP partition function to prove a uniqueness result. If $q = 2$, the diluted-random-cluster measure on the hypercubic lattice is independent of the boundary conditions except on a null subset of the parameter space. In Section 2.9 we use the coupling and comparison results to draw conclusions about the phase diagram of the $q = 2$ Blume–Capel–Potts model on the $d = 2$ hypercubic lattice.

2.2 Notation

Let $d \geq 2$. Let $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. For $x, y \in \mathbb{Z}^d$ define distance

$$l(x, y) = \sum_{i=1}^{d} |x_i - y_i|.$$ 

Then $L^d = (\mathbb{Z}^d, \mathbb{E}^d)$, the $d$-dimensional hypercubic lattice, is the nearest-neighbour graph on $\mathbb{Z}^d$,

$$\mathbb{E}^d = \{(x, y) : x, y \in \mathbb{Z}^d, l(x, y) = 1\}.$$ 

It is regular, every vertex has degree $\delta = 2d$. $L^d$ is simple, it has no loop edges or parallel edges. If $V \subset \mathbb{Z}^d$ then the boundary $\partial V$ is set of vertices in $\mathbb{Z}^d \setminus V$ adjacent to a member of $V$. The corresponding region $\Lambda = (V, E)$ of $L^d$ contains the edges with both end vertices in $V \cup \partial V$. Write $\Lambda_n \rightarrow L^d$ if $(\Lambda_n)$ is a sequence of regions with vertices tending to $\mathbb{Z}^d$.

If $\omega \in \Omega = \{0, 1\}^I$ for some set $I$, define configurations $\omega^x, \omega_x$ for $x \in I$ by

$$\omega^x(y) = \begin{cases} 1 & x = y, \\ \omega(y) & \text{otherwise,} \end{cases} \quad \omega_x(y) = \begin{cases} 0 & x = y, \\ \omega(y) & \text{otherwise.} \end{cases}$$

For convenience when manipulating multiple coordinates, we will write $\omega^{x,y}$ meaning $(\omega^x)^y$. We use the Kronecker-delta symbol,

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$
2.3 FKG and Holley’s theorem

In Chapter 1 we stated the FKG theorem and Holley’s theorem \[26, 38\]. They are useful tools in the study of the random-cluster model. In this section we will give two helpful additional results \[34, 35\]. Let \( I \) be a finite set and let \( \Omega = \{0, 1\}^I \). Let \((\Omega, \mathcal{F})\) be the \(\sigma\)-algebra with \( \mathcal{F} \) the set of all subset of \( \Omega \). We call probability measure \( \mu \) positive if \( \mu(\omega) > 0 \) for all \( \omega \in \Omega \).

**Theorem 2.1.** Let \( \mu \) be a positive probability measure on \((\Omega, \mathcal{F})\) such that for all \( i, j \in I \), for all \( \omega \in \Omega \) with \( \omega(i) = \omega(j) = 0 \),

\[
\mu(\omega)\mu(\omega^{i,j}) \geq \mu(\omega^i)\mu(\omega^j).
\]

Then \( \mu \) satisfies the FKG lattice condition (1.4). By the FKG theorem it is monotonic and strongly positively-associated.

**Theorem 2.2.** Let \( \mu_1, \mu_2 \) be positive probability measures on \((\Omega, \mathcal{F})\) such that \( \mu_1 \) is strongly positively-associated. Suppose that for all \( \omega \in \Omega \) with \( \omega(i) = 0 \),

\[
\mu_1(\omega)\mu_2(\omega^i) \geq \mu_1(\omega^i)\mu_2(\omega),
\]

and for all \( \omega \in \Omega \) with \( \omega(i) = \omega(j) = 0 \),

\[
\mu_1(\omega)\mu_2(\omega^{i,j}) \geq \mu_1(\omega^i)\mu_2(\omega^j).
\]

Then

\[
\forall \omega, \omega' \in \Omega : \mu_1(\omega \land \omega')\mu_2(\omega \lor \omega') \geq \mu_1(\omega)\mu_2(\omega'),
\]

so by Holley’s theorem \( \mu_1 \leq_{st} \mu_2 \).

Let \( G = (V, E) \) be a finite graph, let \( p \in (0, 1) \) and \( q \geq 1 \). The \((p, q)\) random-cluster measure on \( G \) satisfies the FKG lattice condition and hence is monotonic. The more edges are open, the more closed edges can be opened without reducing the number of clusters.
2.4 The diluted-random-cluster representation of the BCP model

In the same sense that the BCP model is a dilution of the Potts model, the measure we now define is a diluted version of the random-cluster measure. The parameters of the model are \( a \in [0, 1], p \in [0, 1) \) and \( q \in (0, \infty) \). Let \( G = (V, E) \) be a graph. Let \( \Psi = \{0, 1\}^V \) be the set of vertex configurations and let \( \Omega = \{0, 1\}^E \) be the set of edge configurations. For \( \psi \in \Psi \), call vertex \( x \) open if \( \psi(x) = 1 \) and closed otherwise. Define the undeleted graph \( G_\psi = (V_\psi, E_\psi) \) by \( V_\psi = \{x \in V : \psi(x) = 1\} \), \( E_\psi = \{\langle x, y \rangle \in E : x, y \in V_\psi\} \).

Call the edges not in \( E_\psi \) deleted. Let \( \omega \in \Omega \). Call edge \( e \) open if \( \omega(e) = 1 \) and closed otherwise. If all deleted edges are closed, say that \( \psi \) and \( \omega \) are compatible. Define \( \Theta = \{(\psi, \omega) : \psi, \omega \text{ are compatible}\} \). Let \( \Lambda = (V, E) \) be a region, and let \( \lambda = (\kappa, \rho) \in \Theta \). Let \( \Theta_\lambda^\Lambda = \{\theta \in \Theta : \theta \text{ off } \Lambda\} \), the set of configurations that agree with boundary conditions \( \lambda \) on the vertices \( V \setminus V \) and edges \( E \setminus E \). Let \( V_\psi = V_\psi \cap V \) and \( E_\psi = E_\psi \cap E \).

Two open vertices are connected, written \( x \leftrightarrow y \), if they are joined by a path of open edges. A cluster of the graph is a maximal set of connected open vertices. Let \( k(\theta) \) denote the number of clusters of the graph that include a vertex in \( V \cup \partial V \); do not count isolated closed vertices. For now assume \( a \in (0, 1) \), \( p \in [0, 1) \).

For notational convenience, set \( r = \sqrt{1 - p} \). Define the \((a, p, q)\) diluted-random-cluster (DRC) measure with boundary conditions \( \lambda \) on \( \Lambda \). For \( \theta = (\psi, \omega) \in \Theta_\lambda^\Lambda \),

\[
\phi_{\Lambda, a, p, q}^\lambda(\theta) = \frac{1}{Z_{DRC}^\theta} d^{E_\psi} q^{k(\theta)} \prod_{x \in V} \left( \frac{a}{1 - a} \right)^{\psi(x)} \prod_{e \in E_\psi} \left( \frac{p}{1 - p} \right)^{\omega(e)}
\]

and \( \phi_{\Lambda, a, p, q}^\lambda(\theta) = 0 \) otherwise. Here \( Z_{DRC}^\theta = Z_{DRC}^{\theta, \Lambda}(\lambda, \Lambda) \) is a normalizing constant. The above formula may be interpreted with \( a = 0 \) and \( a = 1 \). If \( a = 0 \), set all vertices and edges closed. If \( a = 1 \), require all vertices be open, with edges are distributed according to the normal \((p, q)\) random-cluster measure.

The motivation for this definition is provided by the following coupling. Let \( q \) be a positive integer and let \( s \in \{0, 1, \ldots, q\} \). We will abuse notation by using
2.4 The diluted-random-cluster representation of the BCP model

$s$ to represent a BCP spin configuration: the member of $\Sigma$ such that every vertex has spin integer $s$. If $s = 0$ let $b = 0$; if $s \in \{1, \ldots, q\}$ let $b = 1$. Again we will abuse notation, this time by taking $b$ to represent a DRC configuration: the element of $\Theta$ that assigns each vertex and edge state $b$. Let $\mathcal{S}$ be the set of triples of spin, vertex and edge configurations, $(\sigma, \psi, \omega) \in \Sigma \times \Psi \times \Omega$, such that

(i) $\sigma \in \Sigma^\Lambda$,  
(ii) $(\psi, \omega) \in \Theta^\Lambda$,  
(iii) for all $x \in \mathcal{V}$, $\sigma(x) = 0$ if and only if $\psi(x) = 0$, and  
(iv) for all $e = \langle x, y \rangle \in \mathcal{E}$, $\sigma(x) \neq \sigma(y)$ implies $\omega(e) = 0$.

So $\mathcal{S}$ is the set of configuration that have $E_\sigma = E_\psi$ and such that the $\sigma$-spins on $(\psi, \omega)$-clusters are constant. Define measure $\mu$ with support $\mathcal{S}$ and parameters,

$$p = 1 - e^{-2K}, \quad \frac{a}{1 - a} = e^{-\Delta}$$

by

$$\mu(\sigma, \psi, \omega) = \begin{cases} Z^{-1} p^{|E_\psi|} \prod_{x \in V} \left( \frac{a}{1 - a} \right)^{\psi(x)} \prod_{e \in E_\psi} \left( \frac{p}{1 - p} \right)^{\omega(e)} & (\sigma, \psi, \omega) \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

**Theorem 2.4.** Measure $\mu$ couples measures $\pi = \pi^{s}_{\Lambda, K, \Delta, q}$ and $\phi = \phi^{b}_{\Lambda, a, p, q}$ where $a, p, K, \Delta$ satisfy (2.3).

**Proof.** Let $\sigma \in \Sigma$. This uniquely determines vertex configuration $\psi$ and the set of edges that may be open, $A_\sigma = \{ \langle x, y \rangle \in E_\sigma : \sigma(x) = \sigma(y) \}$. Then

$$\pi(\sigma) = \frac{1}{Z_{\text{BCP}}} \exp \left[ -K |E_\sigma| + 2K \sum_{\langle x, y \rangle \in E_\sigma} \delta_{\sigma(x), \sigma(y)} - \Delta |V_\sigma| \right]$$

$$= \frac{1}{Z_{\text{BCP}}} e^{-K |E_\psi|} e^{-\Delta |V_\psi|} e^{2K |A_\sigma|}$$

$$= \frac{1}{Z_{\text{BCP}}} \prod_{x \in V} \left( \frac{a}{1 - a} \right)^{\psi(x)} \prod_{e \in A_\sigma} \left[ 1 + \left( \frac{p}{1 - p} \right)^{\omega(e)} \right]$$

$$= \frac{1}{Z_{\text{BCP}}} \sum_{\omega(\sigma, \psi, \omega) \in S} \prod_{x \in V} \left( \frac{a}{1 - a} \right)^{\psi(x)} \prod_{e \in E_\psi} \left( \frac{p}{1 - p} \right)^{\omega(e)}$$

$$= \frac{Z}{Z_{\text{BCP}}} \sum_{\omega(\sigma, \psi, \omega) \in S} \mu(\sigma, \psi, \omega).$$
2.4 The diluted-random-cluster representation of the BCP model

By summing over $\sigma$, $Z = Z^{BCP}$ and $\mu$ has first marginal $\pi$.

Now fix $\theta = (\psi, \omega) \in \Theta_A^b$. There are $q^{k(\theta)-b}$ BCP configurations $\sigma \in \Sigma_A^s$ such that $(\sigma, \psi, \omega) \in S$; there are $q$ possible spins on each of the $k(\theta) - b$ open clusters that do not meet the boundary.

$$\phi(\psi, \omega) = \frac{1}{Z^{DRC}} \frac{Z}{r_{E_{\psi}}} q^{k(\theta)} \prod_{x \in V} \left( \frac{a}{1-a} \right)^{\psi(x)} \prod_{e \in E_{\psi}} \left( \frac{p}{1-p} \right)^{\omega(e)}$$

$$= \frac{Z}{Z^{DRC}} q^b \sum_{\sigma} \mu(\sigma, \psi, \omega).$$

So $Z^{DRC} = q^b Z$ and the second marginal is $\phi$. \qed

Note that $Z^{DRC} = q^b Z^{BCP}$. This is used to give a uniqueness result in Section 2.8. Conditional on the BCP configuration, the coupled diluted-random-cluster edge states are

(i) zero on the deleted edges,

(ii) zero on the undeleted edges $(x, y)$ where $\sigma(x) \neq \sigma(y)$, and

(iii) independent Bernoulli $p$ random variables on the other edges.

Conditional on the diluted-random-cluster configuration, the coupled BCP spin states are

(i) constant and non-zero on clusters,

(ii) zero on deleted vertices,

(iii) equal to the boundary condition $s$ if the cluster meets the boundary, and

(iv) independent between the other clusters, with the spins uniformly distributed on $S = \{1, \ldots, q\}$.

Suppose $\psi(x) = \psi(y) = 1$. If $x \leftrightarrow y$, $\sigma(x) = \sigma(y)$. If $x \not\leftrightarrow y$ then $\sigma(x)$ and $\sigma(y)$ are independent and they agree with probability $1/q$. It follows that for $q > 1$,

$$\pi(\sigma(x) = \sigma(y) \neq 0) - q^{-1} \pi(\sigma(x)\sigma(y) \neq 0) = (1 - q^{-1}) \phi(x \leftrightarrow y). \quad (2.5)$$

This will provide a precise connection between the phase diagrams of the DRC and BCP models. Existence of large open edge clusters in the DRC model implies
2.4 The diluted-random-cluster representation of the BCP model

that non-zero spins in the BCP model are correlated in a non-trivial way over large distances. Also, the distribution of the set of 0-spins in both models is the same. The DRC model has an infinite vertex cluster of zero states if and only if the same is true for the corresponding BCP model.

Unlike the random-cluster measure, the DRC measure is not monotonic. Let $G$ be the graph with two vertices and one edge, $V = \{x, y\}$ and $E = \{e = \langle x, y \rangle\}$. Let $0 < a, p < 1$ and $q \in (0, \infty)$. Set $r = \sqrt{1 - p} < 1$. Let $\phi$ be the $(a, p, q)$ DRC measure for $G$. Then

$$
\phi(\psi(y) = 1 \mid \psi(x) = 0, \omega(e) = 0) = \frac{qa}{qa + 1 - a} \geq \phi(\psi(y) = 1 \mid \psi(x) = 1, \omega(e) = 0) = \frac{qar}{qar + 1 - a}.
$$

Hence $\phi$ is not monotonic with respect to $\{0,1\}^V \times \{0,1\}^E$; the probability that $y$ is open is not increasing with the other states. However, we will show in Section 2.5 that $\phi$ does have the weaker property of monotonicity in the vertex states.

There are some special cases of the diluted-random-cluster measure that are well understood. We will use them later for comparison purposes. If $p = 0$, the edge configuration is zero. The vertex states are independent Bernoulli random variables with density $qa/(1 - a + qa)$ This corresponds to the $K = 0$ Blume–Capel–Potts measure with independent spins.

On graphs with regular degree, the $q = 1$ BCP and DRC measures may be viewed as the Ising model. For the BCP model, let $\eta(x) = 2\sigma(x) - 1$. This takes us from BCP configuration $\sigma \in \{0, 1\}^V$ to Ising configuration $\eta \in \{+1, -1\}^V$.

$$
\pi_{K,\Delta}^\xi(\sigma) \propto \exp \left[ -K|E_\sigma| + 2K \sum_{\langle x, y \rangle \in E_\sigma} \delta_{\sigma(x), \sigma(y)} - \Delta|V_\sigma| \right]
\propto \exp \left[ -\Delta \sum_{x \in V} \sigma(x) + K \sum_{\langle x, y \rangle \in E} \sigma(x)\sigma(y) \right]
\propto \exp \left[ \sum_{x \in V} \eta(x) \frac{K\delta - 2\Delta}{4} + \frac{1}{4} K \sum_{\langle x, y \rangle \in E} \eta(x)\eta(y) \right].
$$

By the coupling theorem, the same holds for the vertex states of the corresponding DRC measure.
2.5 The vertex DRC marginal measure

Let $\phi = \phi^\lambda_{A,a,p,q}$ be a diluted-random-cluster measure on a region $\Lambda = (V,E)$ in $\mathbb{G}$ with boundary conditions $\lambda = (\kappa, \rho)$. Define vertex DRC measure $\Phi = \Phi^\lambda_{A,a,p,q}$ as the marginal measure of $\phi$ on the vertex states $\Psi = \{0,1\}^V$,

$$\Phi(\psi) = \sum_{(\psi,\omega) \in \Theta^\lambda_{\Lambda}} \phi(\psi,\omega).$$

Let $a \in (0,1)$, $p \in [0,1)$ and $q \in [1,2]$. We exclude the cases $a = 0$ and $a = 1$; the corresponding vertex measures are trivial. Assume $\mathbb{G}$ is simple with maximum vertex degree $\delta$. Each vertex, conditional on the states of the other vertices, is open with probability bounded away from 0 and 1; $\Phi$ has what is known as the ‘finite energy property’ [13].

**Theorem 2.6.** Let $J_x$ be the event that $x \in V$ is open, and let $T_x$ be the $\sigma$-field generated by the states of vertices $V \setminus \{x\}$,

$$\frac{qa}{1 - a + qa} \leq \Phi(J_x | T_x) \leq \frac{qa}{r^\delta(1 - a) + qa}.$$

**Theorem 2.7.** $\Phi$ is monotonic and strongly positively-associated.

The following theorems will be used to show the existence of infinite-volume limit DRC measures.

**Theorem 2.8.** Let $\lambda_1, \lambda_2 \in \Theta$. If $\lambda_1 \leq \lambda_2$ then $\Phi^{\lambda_1}_{A,a,p,q} \leq_{st} \Phi^{\lambda_2}_{A,a,p,q}$.

**Theorem 2.9.** Let $\Lambda_1, \Lambda_2$ be two regions in $\mathbb{G}$ such that $\Lambda_1$ is contained in $\Lambda_2$,

$$\Phi^0_{\Lambda_1,a,p,q} \leq_{st} \Phi^0_{\Lambda_2,a,p,q}, \quad \Phi^1_{\Lambda_1,a,p,q} \geq_{st} \Phi^1_{\Lambda_2,a,p,q}.$$

We can identify a number of situations in which stochastic inequalities hold between versions of the vertex DRC measure with different parameter values.

**Theorem 2.10.** Let $a_i \in (0,1), p_i \in [0,1)$ and $q_i \in [1,2]$ for $i = 1,2$. Let $\Phi_i = \Phi^\lambda_{A,a_i,p_i,q_i}$ for $i = 1,2$. Let $\delta$ be the maximum degree of $\mathbb{G}$. Then $\Phi_1 \leq_{st} \Phi_2$ if any of the following hold.

(a) That $a_1 \leq a_2$, $p_1 \leq p_2$ and $q_1 = q_2$. 

24
2.5 The vertex DRC marginal measure

(b) That

\[ q_2 \left( \frac{a_2}{1-a_2} \right) \geq q_1 \left( \frac{a_1}{1-a_1} \right) (1-p_1)^{-\delta/2}. \]

(c) That \( p_1 \leq p_2, q_1 \geq q_2 \) and

\[ q_2 \left( \frac{a_2}{1-a_2} \right) (1-p_2)^{\delta/2} \geq q_1 \left( \frac{a_1}{1-a_1} \right) (1-p_1)^{\delta/2}. \]

(d) That \( q_1 \leq q_2, \frac{p_1}{q_1(1-p_1)} \leq \frac{p_2}{q_2(1-p_2)} \) and

\[ q_2 \left( \frac{a_2}{1-a_2} \right) (1-p_2)^{\delta/2} \geq q_1 \left( \frac{a_1}{1-a_1} \right) (1-p_1)^{\delta/2}. \]

Let \( \Lambda(\lambda, \psi) \) be the undeleted graph \( G_\psi \) with boundary conditions \( \lambda \) on \( \Lambda \); identify the vertices in the boundary \( \partial V \) joined by open edges of \( \lambda \) outside \( \Lambda \). Then with reference to (1.1) we can write,

\[ \Phi(\psi) = \frac{1}{Z_{\text{RC}}^{p,q}(\Lambda(\lambda, \psi))} \left( \frac{|E|}{\mu} \right)^{|V|} Z_{\text{RC}}^{p,q}(\Lambda(\lambda, \psi)). \]  

(2.11)

**Lemma 2.12.** Let \( \Phi_i, i = 1, 2, \) be the vertex DRC measures on \( \Lambda \) with boundary conditions \( \lambda \in \Theta \) and parameters \( a_i \in (0, 1), p_i \in [0, 1), q_1 \in [1, \infty) \) and \( q_2 \in [1, 2] \). Suppose that \( \psi \in \{0, 1\}^V \), \( x \in V \) with \( \psi(x) = 0 \). Let \( b \) be the number of edges incident to \( x \) in \( \Lambda(\lambda, \psi) \) and let \( I_x \) be the event that all the edges incident to \( x \) are closed. Let \( \mu_i^G \) be the \((p_i, q_i)\) random-cluster measure on graph \( G \).

(i) If

\[ q_2 \left( \frac{a_2}{1-a_2} \right) \left( \frac{1-p_2}{\mu_2^{\Lambda(\lambda, \psi^x)}(I_x)} \right)^{b/2} \geq q_1 \left( \frac{a_1}{1-a_1} \right) \left( \frac{1-p_1}{\mu_1^{\Lambda(\lambda, \psi^x)}(I_x)} \right)^{b/2} \]

then

\[ \Phi_1(\psi) \Phi_2(\psi^x) \geq \Phi_1(\psi^x) \Phi_2(\psi). \]  

(2.13)

(ii) If \( y \in V \setminus \{x\} \) with \( \psi(y) = 0 \), let \( f \in \{0, 1\} \) be the number of edges between \( x \) and \( y \). Then

\[ q_2 \left( \frac{a_2}{1-a_2} \right) \left( \frac{1-p_2}{\mu_2^{\Lambda(\lambda, \psi^{x,y})}(I_x)} \right)^{(b+f)/2} \geq q_1 \left( \frac{a_1}{1-a_1} \right) \left( \frac{1-p_1}{\mu_1^{\Lambda(\lambda, \psi^{x,y})}(I_x)} \right)^{(b+f)/2} \]

implies

\[ \Phi_1(\psi) \Phi_2(\psi^{x,y}) \geq \Phi_1(\psi^{x,y}) \Phi_2(\psi^y). \]  

(2.14)
The following reduces the work needed to check that the conditions for Lemma 2.12 hold. Let \( r_i = \sqrt{1 - p_i}, \) \( i = 1, 2. \)

**Lemma 2.15.** In the context of inequality (2.14),

\[
\mu_2^{\Lambda(\lambda, \psi^{-y})}(I_x) \leq \mu_2^{\Lambda(\lambda, \psi^y)}(I_x)r_2^f.
\]

**Corollary 2.16.** Condition (2.13) also implies (2.14) for all appropriate \( y. \)

**Proof of Lemma 2.15.** Let \( B \) and \( C \) be the sets of edges joining \( x \) and \( y \) to \( \psi \)-open vertices, respectively. Let \( F \) be the set of \( f \) edges with end vertices \( x, y. \) Let \( B_0, C_0, F_0 \) be the decreasing events that all the edges in \( B, C, F \) are closed, respectively.

By the monotonicity of the random-cluster model, the probability of \( I_x \) is increased if the edges in \( C \) are deleted from \( \Lambda(\lambda, \psi^{-y}) \),

\[
\mu_2^{\Lambda(\lambda, \psi^{-y})}(I_x) = \mu_2^{\Lambda(\lambda, \psi^{-y})}(B_0 \cap F_0) \leq \mu_2^{\Lambda(\lambda, \psi^{-y})\setminus C}(B_0 \cap F_0).
\]

Notice that

\[
\mu_2^{\Lambda(\lambda, \psi^{-y})\setminus C}(F_0) = \left\{ \begin{array}{ll}
1 & \text{if } f = 0 \\
\frac{q_2(1 - p_2)}{p_2(q_2 + q_2(1 - p_2))} & \text{if } f = 1
\end{array} \right\} \leq r_2^f.
\]

This is by the conditional probability property of the random-cluster model (1.2) and the elementary inequality,

\[
\frac{q(1 - p)}{p + q(1 - p)} \leq \sqrt{1 - p}, \quad p \in [0, 1], q \in [1, 2].
\]

Therefore, deleting edges that have been conditioned closed,

\[
\mu_2^{\Lambda(\lambda, \psi^{-y})\setminus C}(B_0 \cap F_0) \leq \mu_2^{\Lambda(\lambda, \psi^{-y})\setminus C}(B_0 \mid F_0)r_2^f = \mu_2^{\Lambda(\lambda, \psi^y)}(I_x)r_2^f.
\]

**Proof of Lemma 2.12.** We will only show part (i), that inequality (2.13) implies

\[
\Phi_1(\psi)\Phi_2(\psi^y) \geq \Phi_1(\psi^x)\Phi_2(\psi).
\]
2.5 The vertex DRC marginal measure

Part (ii) follows in the same way. Use (1.1) to rewrite (2.13),

\[
\frac{a_2}{1 - a_2} Z_{p_2,q_2}^{RC}(\Lambda(\lambda, \psi^x)) \leq \frac{a_1}{1 - a_1} Z_{p_1,q_1}^{RC}(\Lambda(\lambda, \psi)) \leq \frac{b}{1 - b}.
\]

Notice that \(|V_{\psi^x} - V_\psi| = 1\) and \(|E_{\psi^x} - E_\psi| = b\). Therefore by (2.11) the result follows.

The proofs of the Theorems 2.7 and 2.10 now follow from Lemma 2.12. In particular Corollary 2.16 means we only have to check that (2.13) holds. Let \(a_i = a, p_i = p\) and \(q_i = q\) for \(i = 1, 2\).

Proof of Theorem 2.7. Clearly (2.13) holds with \(\Phi = \Phi_1 = \Phi_2\). Now apply Theorem 2.1.

Proof of Theorem 2.10. It is sufficient to check (2.13) in each case. Let \(x\) be a vertex incident to \(b\) edges in \(\Lambda(\lambda, \psi^x)\). Let \(\mu_i = \mu_i^{\Lambda(\lambda, \psi^x)}\) for \(i = 1, 2\).

(a) For an edge \(e\), let \(J_e\) be the event that \(e\) is open. By [5] and positive association,

\[
\frac{d}{dp_2} \mu_2(I_x) = \frac{1}{p_2(1 - p_2)} \sum_{e \in E} \text{cov}(I_x, J_e) \leq \frac{1}{p_2(1 - p_2)} \sum_{e \in B} \text{cov}(I_x, J_e).
\]

For \(e \in B\), \(I_x\) and \(J_e\) are mutually exclusive,

\[
\frac{d}{dp_2} \log \mu_2(I_x) \leq -\frac{1}{p_2(1 - p_2)} \sum_{e \in B} \mu_2(J_e) \leq -\frac{b}{p_2(1 - p_2)} p_2 + q_2(1 - p_2) \leq -\frac{b}{2(1 - p_2)} \text{ as } q_2 \in [1, 2].
\]

Integrating from \(p_1\) to \(p_2\),

\[
\frac{\mu_2(I_x)}{\mu_1(I_x)} \leq \frac{(1 - p_2)^{b/2}}{(1 - p_1)^{b/2}}.
\]

As \(q_1 = q_2\) and \(a_1 \leq a_2\), (2.13) holds and the result follows.
2.5 The vertex DRC marginal measure

(b) By the conditional edge densities of the random-cluster measure, \( \mu_2(I_x) \leq (1 - p_2)^{b/2} \) and \( \mu_1(I_x) \geq (1 - p_1)^b \), so (2.13) holds.

(c) Under these conditions, \( \mu_2 \geq_{st} \mu_1 \) so \( \mu_2(I_x) \leq \mu_1(I_x) \).

(d) Again \( \mu_2 \geq_{st} \mu_1 \) so \( \mu_2(I_x) \leq \mu_1(I_x) \).

\[ \square \]

Proof of Theorem 2.8 Let \( \lambda_i = (\kappa_i, \rho_i) \), \( i = 1, 2 \). By the monotonicity of the vertex DRC measure we may assume \( \kappa_1 = \kappa_2 \). We can adapt the proof of Lemma 2.12. By the monotonicity of the random-cluster model, \( \mu_1^{\Lambda(\lambda_1, \psi^x)}(I_x) \geq \mu_2^{\Lambda(\lambda_2, \psi^x)}(I_x) \); apply Theorem 2.2.

\[ \square \]

Proof of Theorem 2.9. Let \( A \) be the decreasing event that all vertices in \( \Lambda_2 \setminus \Lambda_1 \) are closed. By the definition of the DRC measure,

\[ \Phi^0_{\Lambda_1, a, p, q}(\cdot | A) = \Phi^0_{\Lambda_2, a, p, q}(\cdot | A) \]

Event \( A \) is decreasing so by the monotonicity of \( \Phi_0^{\Lambda_2, a, p, q} \), \( \Phi_0^{\Lambda_1, a, p, q} \leq_{st} \Phi_0^{\Lambda_2, a, p, q} \). Let \( \lambda = (\kappa, \rho) \) be a configuration such that off \( \Lambda_2 \), \( \kappa = 1 \) and \( \rho = 1 \). Then

\[ \Phi^1_{\Lambda_1, a, p, q} \geq_{st} \Phi^\lambda_{\Lambda_1, a, p, q} \]

Notice \( \Phi^1_{\Lambda_2, a, p, q} \) can be written as a convex combination of terms on the right hand side above, so \( \Phi^1_{\Lambda_1, a, p, q} \geq_{st} \Phi^\lambda_{\Lambda_2, a, p, q} \).

\[ \square \]

Proof of Theorem 2.6. Since \( q \in [1, 2] \), \( \Phi^\lambda_{\Lambda_2, a, p, q} \) is monotonic by Theorem 2.7. As \( J_x \) is increasing, a lower bound for \( \Phi(J_x | T_x) \) is obtained by conditioning on all vertices other than \( x \) being closed. In that case, \( x \) contributes \( qa/(1 - a) \) when open and 1 when closed; the lower bound follows.

An upper bound is obtained by conditioning on all the vertices other that \( x \) being open, and assuming they are in the same open cluster. This time \( x \) contributes no more than

\[ r^\delta q \left( \frac{a}{1 - a} \right) \sum_{\omega \in \{0, 1\}^\delta} \prod_{i=1}^\delta \left( \frac{p}{1 - p} \right)^\omega(i) \]

when open, and 1 when closed.

\[ \square \]
2.6 Edge comparisons

In the previous section, we only dealt with the vertex states. In this section we will give results for the edge states and the full DRC measure. Let $\Lambda = (V, E)$ be a region in $\mathbb{G}$. Let $\lambda \in \Theta$, and let $\phi = \phi^\lambda_{\Lambda, a, p, q}$ be the diluted-random-cluster measure. Define the DRC edge measure by

$$\Upsilon(\omega) = \sum_{\psi \in \Psi} \phi(\psi, \omega), \quad \omega \in \Omega.$$ 

In the above sum, $\psi$ that are not compatible with edge configuration $\omega$ in $\Lambda$ and boundary conditions $\lambda$ on $\mathbb{G} \setminus \Lambda$ contribute nothing. Assume $\mathbb{G}$ is a simple graph with maximum vertex degree $\delta$. Let

$$w_j = \left[ r^j q \left( \frac{a}{1-a} \right) \right]^{-1}.$$ 

This can be thought of as the cost of closing a vertex under $\phi$. Let $(\psi, \omega)$ be a configuration such that $\psi(x) = 0$. If $j$ is the number of $\psi$-open neighbouring vertices of $x$,

$$\phi(\psi, \omega) = \phi(\psi^x, \omega) w_j, \quad w_j \leq w_\delta.$$ 

**Theorem 2.17.** Let $0 < a \leq a' = 1$, $p, p' \in (0, 1)$ and $q \in [1, \infty)$. Let $\Upsilon = \Upsilon^\lambda_{\Lambda, a, p, q}$ and $\Upsilon' = \Upsilon'^\lambda_{\Lambda, a', p', q'}$. If

$$\frac{p}{1-p} \geq \frac{p'}{1-p'} (1 + 2w_\delta + w_\delta w_{\delta-1})$$

then $\Upsilon \geq_{st} \Upsilon'$.

**Theorem 2.18.** Let $0 < a < 1$, $0 \leq p < 1$ and $q \in [1, 2]$. Let $\lambda \in \Theta$. Let $\Lambda$ be a region in $\mathbb{G}$.

(i) $\phi^\lambda_{\Lambda, a, p, q}$ is increasing in $a$, $p$ and $\lambda$.

(ii) If $\lambda = 0$, $\phi^\lambda_{\Lambda, a, p, q}$ is increasing in $\Lambda$

(iii) If $\lambda = 1$, $\phi^\lambda_{\Lambda, a, p, q}$ is decreasing in $\Lambda$. 

29
2.6 Edge comparisons

Proof of Theorem 2.17. \( \Upsilon' \) is monotonic, it is simply a random-cluster measure. To apply Theorem 2.2 we must show

\[
\Upsilon'(\omega) \Upsilon(\omega^e) \geq \Upsilon'(\omega^e) \Upsilon(\omega), \quad \forall \omega : \omega(e) = 0, \tag{2.19}
\]

\[
\Upsilon'(\omega) \Upsilon(\omega^{e,f}) \geq \Upsilon'(\omega^e) \Upsilon(\omega^f), \quad \forall \omega : \omega(e) = \omega(f) = 0. \tag{2.20}
\]

Let \( \omega \) be an edge configuration from (2.20). By the definition of the random-cluster measure,

\[
\frac{\Upsilon'(\omega)}{\Upsilon'(\omega^e)} = \left( \frac{1 - p'}{p'} \right) q^{k'}
\]

where

\[
k' = k(1, \omega) - k(1, \omega^e) = \begin{cases} 
1 & \text{if } e \text{ is an isthmus of the } (1, \omega^e)\text{-open graph,} \\
0 & \text{otherwise.} 
\end{cases}
\]

Let \( x \) and \( y \) be the end vertices of \( e \). Let \( A \) be the set of vertex configurations compatible with \( \omega^{e,f} \), \( A = \{ \psi : (\psi, \omega^{e,f}) \in \Theta \} \). Write \( A_x \) for the set of configurations in \( A \) with vertex \( x \) turned off, \( \{ \psi_x : \psi \in A \} \). Then (2.20) becomes,

\[
\left( \frac{1 - p'}{p'} \right) q^{k'} \sum_{\psi \in A} \phi(\psi, \omega^{e,f}) \geq \sum_{\psi \in A \cup A_x \cup A_y \cup A_{x,y}} \phi(\psi, \omega^f).
\]

This holds if

\[
\left( \frac{1 - p'}{p'} \right) q^{k'} \sum_{\psi \in A} \phi(\psi, \omega^f) \geq \sum_{\psi \in A} (1 + 2w_5 + w_{8}w_{8-1}) \phi(\psi, \omega^f) \left( \frac{1 - p}{p} \right) q^{k}.
\]

with

\[
k = k(\psi_i, \omega^f) - k(\psi_i, \omega^{e,f}) \leq k'.
\]

Splitting up the sum, we see that (2.20) holds. The proof of (2.19) is very similar. \( \square \)

Proof of Theorem 2.18. In each case the statement holds for the vertex component of the measure, see Theorems 2.8, 2.9 and 2.10 (a). Recall that \( \phi_{\Lambda,n,p,q}^\lambda \) has the same distribution as the random-cluster measure \( \mu_{\Lambda}^{\lambda,\psi} \) on the graph \( \Lambda(\lambda, \psi) \) generated by \( \psi \sim \Phi_{\Lambda,n,p,q}^\lambda \). Let \( X(\psi, \omega) \) be an increasing random variable defined
on the vertex and edge states of $G$. For fixed $\psi$, define a random-cluster random variable, $X_\psi(\omega) = X(\psi, \omega)$,

$$\phi(X) = \Phi(\mu_{p,q}^{A(\lambda,\psi)}(X_\psi)).$$

The random-cluster measure is increasing in $p$ and $A(\lambda, \psi)$. $X_\psi$ is an increasing random variable, and it is increasing in $\psi$. Therefore $\mu_{p,q}^{A(\lambda,\psi)}(X_\psi)$ is an increasing function of $\psi$.

2.7 Infinite-volume measures

We now are in a position to discuss infinite-volume limit measures. We will work on the hypercubic lattice, take $G = L^d$ with $d \geq 2$. The results easily generalize to other lattices. Let $0 < a, p < 1$ and $q \in (0, \infty)$. Let $\mathcal{F}_\Theta$ be the $\sigma$-algebra generated by the weak topology on $\Theta$. Call a probability measure $\phi$ on $(\Theta, \mathcal{F}_\Theta)$ a limit diluted-random-cluster measure with parameters $a, p, q$ and boundary conditions $\lambda \in \Theta$ if it is an accumulation point of a sequence of measures $(\phi_{\lambda,\Lambda_n}^{A(\lambda,\psi)}(X_\psi))$ with $\Lambda_n \rightarrow L^d$. Let $\mathcal{W}_{a,p,q}$ be the set of such measures; let $\co \mathcal{W}_{a,p,q}$ be the closure of the convex hull of $\mathcal{W}_{a,p,q}$.

The configuration space $\Theta$ is compact. The product $\Psi \times \Omega$ of discrete topological spaces is compact, and $\Theta$ is a closed subset of $\Psi \times \Omega$. It is standard by Prohorov’s theorem [2] that any sequence of measures on $\Theta$ must have a fixed point. Hence $\mathcal{W}_{a,p,q}$ is non-empty. Using the results of the previous three sections, we can show that many results for the $q \geq 1$ random-cluster model have analogues for the DRC model when $q \in [1,2]$. Let $0 < a, p < 1$, $q \in [1,2]$ and define the DRC model with regular boundary conditions, $\phi_{a,p,q}^b = \lim_{\Lambda \rightarrow L^d} \phi_{\lambda,a,p,q}^b$ for $b \in \{0,1\}$.

**Theorem 2.21.** Let $\phi$ be a limit DRC measure. Consider the vertex measure $\Phi$ obtained by projecting $\phi$ onto the set of vertex configurations.

(i) If $\phi \in \mathcal{W}_{a,p,q}$ then $\Phi$ is positively associated.

(ii) If $\phi \in \co \mathcal{W}_{a,p,q}$ then $\Phi$ has a finite energy property.
2.7 Infinite-volume measures

By the conditional probability properties of the random-cluster model, if $\Phi$ has a finite energy property then the corresponding edge measure $\Upsilon$ has a finite energy property. Given the states of all but one edge, that edge is open with probability bounded away from 0 and 1.

**Theorem 2.22.** Limit measures $\phi_{a,p,q}^b$ are

(i) well defined,

(ii) stochastically increasing in $a, p$ and $b$,

(iii) extremal in $\overline{\mathcal{W}}_{a,p,q}$,

(iv) tail trivial, and

(v) translation invariant.

Also $\phi_{a,p,q}^b$ have the 0/1-infinite cluster property with respect to both nearest-neighbour vertex clusters and edge clusters. The number of infinite clusters of each type is a.s. constant and equal to either 0 or 1.

**Theorem 2.23.** If $q \in \{1, 2\}$ and $s \in \{0, 1, \ldots, q\}$ then the limit Blume–Capel–Potts measure $\pi_{K,\Delta,q}^s = \lim_{\Lambda \to \Lambda^d} \pi_{\Lambda,a,p,q}^s$ exists.

We can also define DLR type DRC measures. Let $\mathcal{R}_{a,p,q}$ be the set of measures $\phi$ such that,

$$\phi(A | T_\Lambda)(\lambda) = \phi_{\Lambda,a,p,q}^\lambda(A), \quad A \in \mathcal{F}_\Theta, \quad \phi\text{-a.e. } \lambda.$$

Let $x, y \in \mathbb{Z}^d$ and write $1_{\{x \leftrightarrow y\}}$ for the indicator function of the event $x$ is connected to $y$. A consequence of Theorem 2.22 is that the set of discontinuities of $1_{\{x \leftrightarrow y\}}$ is $\phi_{a,p,q}^b$-a.s. null. By this we mean that for almost all configurations $\theta$, there is a region $\Lambda = \Lambda(\theta)$ such that $x \leftrightarrow y$ is decided inside $\Lambda$. If the event is not determined inside a region $\Lambda'$, then $x$ and $y$ must both be joined to the boundary $\partial \Lambda'$ by open paths that do not meet inside $\Lambda'$. This cannot be the case for arbitrarily large $\Lambda'$, or $x$ and $y$ are in distinct infinite edge clusters. It follows from the above that $\phi_{a,p,q}^b \in \mathcal{R}_{a,p,q}$.

The following theorem collects some standard results [34]. Positive association and stochastic orderings are preserved under weak limits.
2.7 Infinite-volume measures

**Theorem 2.24.** Let $I$ be a countable set. Let $\Omega = \{0, 1\}^I$ and let $\mathcal{F}$ be the $\sigma$-algebra generated by the product topology.

(a) Let $(\mu_n)$ and $(\nu_n)$ be two sequences of probability measure on $(\Omega, \mathcal{F})$, with weak limits $\mu, \nu$ respectively. If $\mu_n \leq_{st} \nu_n$ for all $n$ then $\mu \leq_{st} \nu$.

(b) If measures $(\mu_n)$ on $(\Omega, \mathcal{F})$ converges weakly to $\mu$, and each $\mu_n$ is positively associated, then so is $\mu$.

(c) Let $\mu, \nu$ be measures on $(\Omega, \mathcal{F})$. Let $J_i(\omega) = \omega(i)$. If $\mu \leq_{st} \nu$ and $\mu(J_i) = \nu(J_i)$ for all $i \in I$ then $\mu = \nu$.

**Corollary 2.25.** The vertex comparison results of Theorem 2.10 and the edge comparison result of Theorem 2.17 extend to $\mathbb{L}^d$.

The following lemma is needed in the proof of Theorem 2.23 for $s = 1, \ldots, q$.

**Lemma 2.26.** Let $A$ be an increasing closed event. Let $X$ be a $\phi^1$-a.s. continuous random variable. Then

$$\phi^1_{\Lambda}(X1_A) \to \phi^1(X1_A) \text{ as } \Lambda \to \mathbb{L}^d.$$ 

**Proof of Theorem 2.21.**

(i) Write $\Phi = \lim_{\Lambda \to \mathbb{L}^d} \Phi^\Lambda_{a,p,q}$. By Theorem 2.7, $\Phi$ is the limit of monotonic measures so the result follows by Theorem 2.24 (b).

(ii) We adapt the proof of Theorem 3.3 in [33]. Suppose first $\phi \in \mathcal{W}_{a,p,q}$, say

$$\phi = \lim_{\Lambda \to \mathbb{Z}^d} \phi^\Lambda_{\Lambda}.$$ 

We will write $\mathcal{F}_V$ for the $\sigma$-algebra generated by the states of vertices in $\Lambda = (V, E)$ and $\mathcal{T}_V$ for $\sigma$-algebra generated by the states of vertices in $\mathbb{Z}^d \setminus V$. Let $\Lambda' = (V', E')$. By the martingale convergence theorem [36] and weak convergence,

$$\phi^\Lambda(J_x \mid \mathcal{T}_x) = \lim_{\Lambda' \to \mathbb{Z}^d} \phi^\Lambda(J_x \mid \mathcal{F}_{V' \setminus x})$$ 

$$= \lim_{\Lambda' \to \mathbb{Z}^d} \lim_{\Lambda \to \mathbb{Z}^d} \phi^\Lambda_{\Lambda}(J_x \mid \mathcal{F}_{V \setminus x}).$$
By Theorem 2.6 this limit is bounded away from 0 and 1. The bounds depend only on \(a, p\) and \(q\) so the result extends to the convex hull of \(\mathcal{W}_{a,p,q}\) by linearity. If \(\phi \in \overline{\text{co}\mathcal{W}_{a,p,q}}\) we can write \(\phi = \lim_n \phi_n\) with each \(\phi_n\) in the convex hull of \(\mathcal{W}_{a,p,q}\). Write \(I_{\kappa,V\setminus x}\) for the cylinder event that the vertices in \(V\setminus x\) take the values from \(\kappa\). Again a.s. by martingale convergence,

\[
\phi(J_x | \mathcal{T}_x)(\kappa) = \lim_{\Lambda \rightarrow \mathbb{Z}^d} \phi(J_x | \mathcal{F}_{V\setminus x})(\kappa)
\]

\[
= \lim_{\Lambda \rightarrow \mathbb{Z}^d} \frac{\phi(J_x \cap I_{\kappa,V\setminus x})}{\phi(I_{\kappa,V\setminus x})}
\]

\[
= \lim_{\Lambda \rightarrow \mathbb{Z}^d} \lim_{n \rightarrow \infty} \frac{\phi_n(J_x \cap I_{\kappa,V\setminus x})}{\phi_n(I_{\kappa,V\setminus x})}.
\]

The vertex states of \(\phi\) have a finite energy property.

An event is called a cylinder event if it only depends on the state of a finite number of vertices and edges. The set of cylinder events is convergence determining \[6\]; the same is true for just the set of increasing cylinder events.

**Proof of Theorem 2.22**

(i) Let \(A\) be an increasing cylinder event. Let \(\Lambda_n, \Delta_n\) be two increasing sequences of regions such that \(\Lambda_n \rightarrow \mathbb{L}^d, \Delta_n \rightarrow \mathbb{L}^d\). Then by Theorem 2.18 the sequence \(\phi_{\Lambda_n,a,p,q}^b(A)\) is monotone. For every \(\Lambda_n\) there exists an \(m\) such that \(\Lambda_n \subset \Delta_m\), and vice versa. Hence \(\phi_{a,p,q}^b(A)\) is well-defined.

(ii) This follows from Theorem 2.18 take the limit \(\Lambda \rightarrow \mathbb{L}^d\) by Theorem 2.24 (a).

(iii) This follows from Theorem 2.18 take the limit \(\Lambda \rightarrow \mathbb{L}^d\) by Theorem 2.24 (a).

(iv) The following argument is standard \[11 34\]. First consider the case \(b = 0\). Let \(\Lambda \subset \Delta\) be regions in \(\mathbb{L}^d\). Let \(\mathcal{T}_\Lambda\) be the \(\sigma\)-field generated by the vertices and edges of \(\mathbb{L}^d\setminus \Lambda\). Let \(A\) be an increasing cylinder event on \(\Lambda\), and \(T\) an
event on $\Delta \setminus A$. Let $I_{0,\Delta \setminus A}$ be the decreasing event that vertices and edges in $\Delta \setminus A$ are closed. By Theorem 2.18 (ii),
\[
\phi_\Delta^0(A) = \phi_\Delta^0(A \mid I_{0,\Delta \setminus A}) \leq \phi_\Delta^0(A \mid T).
\]
Take the limits $\Delta \to \mathbb{Z}^d$ and then $\Lambda \to \mathbb{Z}^d$ to get
\[
\phi^0(A)\phi^0(T) \leq \phi^0(A \cap T)
\]
for $T$ in the tail $\sigma$-field $T = \bigcap_{\Lambda} \text{a region of } L^d \setminus T_\Lambda$. Replace $T$ with $T^c$,
\[
\phi^0(A)\phi^0(T^c) \leq \phi^0(A \cap T^c).
\]
Adding the two inequalities above we get $\phi^0(A) \leq \phi^0(A)$. Therefore the inequalities hold with equality,
\[
\phi^0(A)\phi^0(T) = \phi^0(A \cap T).
\]
This holds for all increasing cylinder events $A$, and so for all events. Let $A = T$, $\phi^0(T) = \phi^0(T)^2$, $\phi^0$ is tail-trivial. Similarly for $b = 1$.

(v) Let $A$ be an increasing cylinder event on $\Lambda$. Let $\tau : L^d \to L^d$ be a translation of the lattice. Then
\[
\phi^0(A) \geq \phi^0_\Lambda(A) = \phi^0_\tau(A) \to \phi^0(\tau A), \quad \Lambda \to \mathbb{Z}^d.
\]
The above inequality also holds for $\tau^{-1}$, so $\phi(A) = \phi(\tau A)$. Similarly for 1 boundary conditions with the inequalities reversed.

The 0/1-infinite cluster property follows from [13] as $\phi^b$ is translation invariant and the projected measures on vertices and edges have finite energy properties.

**Proof of Lemma 2.26.** Write $[1_A]_\Lambda^1(\theta)$ for the indicator function of $A$ evaluated at $\theta$ on $\Lambda$ with boundary condition 1. As $A$ is increasing and closed, for all $\theta \in \Theta$,
\[
[1_A]_\Lambda^1(\theta) \downarrow 1_A(\theta) \text{ as } \Lambda \to \mathbb{L}^d.
\]
Let regions $\Lambda, \Lambda', \Lambda''$ satisfy $\Lambda' \subset \Lambda \subset \Lambda''$. By monotonicity, as used in the proof of Theorem 2.9,

$$
\phi^1_\Lambda([1_A]_{\Lambda'}) \geq \phi^1_\Lambda(1_A) \geq \phi^1_\Lambda([1_A]_{\Lambda}).
$$

Let $\Lambda'' \to \mathbb{L}^d$. By weak convergence,

$$
\phi^1_\Lambda([1_A]_{\Lambda'}) \geq \phi^1_\Lambda(1_A) \geq \phi^1([1_A]_{\Lambda}).
$$

Take the limit $\Lambda \to \mathbb{L}^d$. By weak convergence and the monotone convergence theorem,

$$
\phi^1([1_A]_{\Lambda'}) \geq \lim_{\Lambda \to \mathbb{L}^d} \phi^1_\Lambda(1_A) \geq \phi^1(1_A).
$$

Now let $\Lambda' \to \mathbb{L}^d$. Again by the monotone convergence theorem,

$$
\phi^1(1_A) \geq \lim_{\Lambda' \to \mathbb{L}^d} \phi^1_\Lambda(1_A) \geq \phi^1(1_A).
$$

Let $(\Lambda_n)$ be a sequence of regions that tend to $\mathbb{L}^d$. The sequence of measures $(\phi^1_{\Lambda_n})$ is decreasing so there is a monotonic coupling. Let $\theta, (\theta_n)$ be random variables such that each $\theta_n$ has law $\phi^1_{\Lambda_n}, \theta$ has law $\phi^1$ and $\theta_n \downarrow \theta$ as $n \to \infty$. $X$ is $\phi^1$-a.s. continuous so by weak convergence $X(\theta_n) \to X(\theta), \phi^1$-a.s.. Also $1_A(\theta_n) \to 1_A(\theta), \phi^1$-a.s.. By the dominated convergence theorem the result follows.

Proof of Theorem 2.23. By the coupling of the finite BCP and DRC models, and the existence of $\lim_{\Lambda \to \mathbb{L}^d} \pi^b_\Lambda$, $b = 0, 1$, we can show that $\lim_{\Lambda \to \mathbb{L}^d} \pi^b_\Lambda$ exists. Let $\sigma \in \Sigma$ be a spin configuration. Let $A$ be the BCP cylinder event that the spins in a region $\Lambda = (V, E)$ agree with $\sigma$. Events such as $A$ are convergence determining. Let $B$ be the DRC event

(i) for all $x \in V$, $\psi(x) = 0$ if and only if $\sigma(x) = 0$, and

(ii) for all $x, y \in V$, $x \leftrightarrow y$ only if $\sigma(x) = \sigma(y)$.

We can partition event $B$ into disjoint events according to number of open clusters that intersect $V$. Let $B_i$ be the union of $B$ with the event $V$ intersects $i$ open clusters. Let $\Delta$ be a region containing $\Lambda$. Suppose first $b = 0$ and $s = 0$. By the properties of the coupling we can write $\pi^b_\Lambda(A) = \sum_{i=0}^{\lfloor |V|/q \rfloor} q^{-i} \phi^b_\Lambda(B_i)$. Take the limit $\Delta \to \mathbb{Z}^d$. 

36
In the case \( b = 1 \) and \( s = 1, \ldots, q \) we must modify the expression for \( \pi^s_\Delta(A) \) to take into account the fixed spin of the boundary cluster. Let

\[
A^+ = \{ \exists x \in V : \sigma(x) = s \text{ and } x \leftrightarrow \infty \}, \\
A^- = \{ \exists x \in V : \sigma(x) \neq s \text{ and } x \leftrightarrow \infty \}.
\]

By Lemma 2.26 we can take the limit \( \Delta \to \mathbb{L}^d \) below to give the result,

\[
\pi^s_\Delta(A) = \sum_{i=0}^{\lfloor \log_2 q \rfloor} q^{-i} \left[ \phi^1_\Delta(B_i) + (q-1)\phi^1_\Delta(B_i \cap A^+) - \phi^1_\Delta(B_i \cap A^-) \right].
\]

\[
\phi^\lambda = \phi^\lambda_{a,p,q} \text{ on } \mathbb{L}^d. \quad \text{By the coupling, we also get a uniqueness result for } q = 2 \text{ Blume–Capel–Potts limit measures } \pi^\xi = \pi^\xi_{K,\Delta,q}. \quad \text{In this section, let } (2.3) \text{ fix the relationship between } a, p \text{ and } K, \Delta.
\]

**Theorem 2.27.** Let \( q = 2 \). The set of points \( a, p \) at which \( \phi^0_{a,p,q} \neq \phi^1_{a,p,q} \), and \( \pi^0_{K,\Delta,q} \neq (\pi^1_{K,\Delta,q} + \pi^2_{K,\Delta,q})/2 \), can be covered by a countable collection of rectifiable curves of \([0,1]^2\).

Let \( \Lambda = (V,E) \) be a region in \( \mathbb{L}^d \). Take \( b \in \{0,1\} \) to be both a DRC boundary condition and a BCP boundary condition. By Theorem 2.4 we can define,

\[
G^b_\Lambda(K,\Delta) = \frac{1}{|V|} \log q^{-b} Z^{DRC}_{a,p,q}(b,\Lambda) = \frac{1}{|V|} \log Z^{BCP}_{K,\Delta,q}(b,\Lambda)
\]

\[
= \frac{1}{|V|} \log \sum_{\sigma} \exp \left[ -K|E_\sigma| + 2K \sum_{\langle x,y \rangle \in E_\sigma} \delta_{\sigma(x),\sigma(y)} - \Delta|V_\sigma| \right].
\]
2.8 Uniqueness by convexity of pressure

The sum is over BCP configurations $\sigma$ that are compatible with boundary condition $b$. Differentiating $G^b_\Lambda(K, \Delta)$ with respect to $\Delta$,

$$\frac{\partial}{\partial \Delta} G^b_\Lambda(K, \Delta) = - \frac{1}{|V|} \pi^b(|V_\sigma|)$$

$$= - \frac{1}{|V|} \phi^b(|V_\psi|).$$

Also

$$\frac{\partial}{\partial K} G^b_\Lambda(K, \Delta) = - \frac{1}{|V|} \pi^b \left( - |E_{\sigma}| + 2 \sum_{(x,y) \in E_{\sigma}} \delta_{\sigma(x),\sigma(y)} \right)$$

$$= - \frac{1}{|V|} \phi^b \left( - |E_{\psi}| + 2 \sum_{e \in E} \omega(e) \right).$$

Let $i = (i_1, i_2)$ be a unit vector in $\mathbb{R}^2$. Let Var denotes variance with respect to $\pi^b$. Differentiating in the direction $i$ on the $(K, \Delta)$-plane,

$$\frac{\partial^2}{\partial i^2} G^b_\Lambda(K, \Delta) = \frac{1}{|V|} \text{Var} \left( - i_1 |E_{\sigma}| + 2 i_1 \sum_{(x,y) \in E_{\sigma}} \delta_{\sigma(x),\sigma(y)} - i_2 |V_{\sigma}| \right) \geq 0.$$

$G^b_\Lambda$ is convex.

**Lemma 2.28.** Let $\Delta \in \mathbb{R}$, $K \in [0, \infty)$ and $b \in \{0, 1\}$. The limit,

$$G(K, \Delta) = \lim_{\Lambda \to \mathbb{R}} G^b_\Lambda(K, \Delta)$$

exists. It is a convex function of $K, \Delta$ and independent of $b$.

**Proof of Theorem 2.27.** By Theorem 8.18 of [23] and the convexity of $G$, the subset of $[0, \infty) \times \mathbb{R}$ where $G(K, \Delta)$ is not differentiable can be covered by a countable collection of rectifiable curves.

Let $J_x$ be the event that vertex $x$ is open. By translation invariance and the monotonicity of vertex DRC measures, for any region $\Lambda$ containing $x$,

$$\frac{1}{|V|} \sum_{y \in V} \phi^0_{\Lambda}(\psi(y)) \leq \phi^0(J_x) \leq \phi^1(J_x) \leq \frac{1}{|V|} \sum_{y \in V} \phi^1_{\Lambda}(\psi(y)). \quad (2.29)$$
2.8 Uniqueness by convexity of pressure

The left and right most terms in (2.29) are \( \frac{\partial}{\partial \Lambda} G^b_\Lambda, \) \( b = 0,1 \) respectively. If \( G \) is differentiable at \((a,p)\), \( \frac{\partial}{\partial \Lambda} G^b_\Lambda \to \frac{\partial}{\partial \Lambda} G \) as \( \Lambda \to \infty \); \( \phi^0(J_e) = \phi^1(J_e) \). By Theorem 2.24 (c), the corresponding vertex measures satisfy \( \Phi^0 = \Phi^1 \).

As \( \Phi^0 = \Phi^1 \), \( \lim_{\Lambda \to L_d} |V|^{-1} \phi^b(|E|) = \phi^1(|E|) \). By Theorem 2.24 (c) gives \( \phi^0 = \phi^1 \).

We will use the following lemmas in a subadditivity argument to show \( G^b_\Lambda \) converges as \( \Lambda \to \mathbb{Z}^d \). The first bounds the BCP partition function. The second bounds how the partition function changes when edges are removed from the graph. Write \( Z^{BCP}_G \) for the BCP partition function on a graph \( G \) with no boundary conditions.

**Lemma 2.30.** For graph \( G = (V, E) \),

\[ -K|E| - |V| \cdot |\Delta| + |V| \log 3 \leq \log Z^{BCP}_G \leq K|E| + |V| \cdot |\Delta| + |V| \log 3. \]

**Lemma 2.31.** Suppose removing edges \( F \subset E \) from a graph \( G = (V, E) \) creates disjoint subgraphs \( G_1 = (V_1, E_1), i=1,2 \). Then

\[ \log Z^{BCP}_{G_1} + \log Z^{BCP}_{G_2} - K|F| \leq \log Z^{BCP}_G \leq \log Z^{BCP}_{G_1} + \log Z^{BCP}_{G_2} + K|F|. \]

**Proof of Lemma 2.28.** First assume \( b = 0 \). For a region \( \Lambda = (V, E) \) the edges between \( V \) and the boundary \( \partial V \) make no contribution to \( Z^{BCP}_{K,\Delta,q}(b, \Lambda) \) and can be ignored. Let \( \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d \) and let \( \Lambda_n \) be the region defined by the \( |\mathbf{n}| = n_1 \ldots n_d \) vertices in \( \prod_{i=1}^d [1, n_i] \). Let \( \mathbf{k} \in \mathbb{N}^d \) and write

\[ (\mathbf{n}, \mathbf{k}) = \left( \left\lfloor \frac{n_i}{k_i} \right\rfloor : 1 \leq i \leq d \right), \quad \left\lfloor \frac{n}{k} \right\rfloor = \prod_{i=1}^d \left\lfloor \frac{n_i}{k_i} \right\rfloor. \]

Let \( \Delta \) be the region defined by the vertices in \( \Lambda_n \) but not in \( \Lambda_{(n,k)} \). Then by Lemma 2.31,

\[ \left\lfloor \frac{n}{k} \right\rfloor \left( \log Z^{BCP}_{\Lambda_k} - K \sum_{i=1}^d \frac{|k|}{k_i} \right) + \log Z^{BCP}_\Delta \leq \log Z^{BCP}_{\Lambda_n}. \]
2.9 Phases and comparison results

Apply Lemma 2.30 to $\Delta$,

$$\log Z_{\Delta}^{BCP} \geq (-dK - |\Delta| + \log 3) \sum_{i=1}^{d} \frac{|n|}{k_i}.$$ 

Divide by $|n|$ and let the $n_i$ tend to $\infty$,

$$\frac{1}{|k|} \log Z_{\Lambda_k}^{BCP} - K \sum_{i=1}^{d} \frac{1}{k_i} \leq \lim \inf_{n \to \infty} G_{\Lambda_n}^b.$$ 

Let $k \to \infty$,

$$\lim \sup_{k \to \infty} G_{\Lambda_k}^b \leq \lim \inf_{n \to \infty} G_{\Lambda_n}^b.$$ 

Hence $G = \lim_{n \to \infty} G_{\Lambda_n}^b$ exists. Each $G_{\Lambda_n}^b$ is a convex function of $K, \Delta$ and so $G$ is too. If $b = 1$ we can ignore the boundary conditions by Lemma 2.31; the proportion of vertices in $\Lambda_n$ adjacent to the external vertex boundary tends to 0 as $n \to \infty$.

Proof of Lemma 2.30: The partition function $Z_{G}^{BCP}$ is a sum over $3^{|V|}$ terms, each bounded between $\exp(-K|E| - |V| \cdot |\Delta|)$ and $\exp(K|E| + |V| \cdot |\Delta|)$.

Proof of Lemma 2.31: Notice $\log Z_{(V,E\setminus F)}^{BCP} = \log Z_{G_1}^{BCP} + \log Z_{G_2}^{BCP}$, the BCP measure is independent between disjoint graph components. For every edge you add to a graph, the change to $\log Z^{BCP}$ is bounded by $\pm K$.

2.9 Phases and comparison results

In this section we will discuss the phase diagrams of the BCP model and the DRC model. By the coupling of Theorem 2.4 and in particular equation (2.5), we can discuss them in parallel. For concreteness we will specify 1 boundary conditions for both models with $q = 2$ and $d = 2$.

There are two properties of the DRC model that we are interested in:

(i) $\phi^1$ has an infinite open edge cluster.

(ii) $\phi^1$ has an infinite closed vertex cluster.
For all \( a, p \) the probability of each property is either 0 or 1. The corresponding phases of the BCP model are:

(i) \( \pi^1(\sigma(0) = 1) > \frac{1}{2}\pi^1(\sigma(0) \in \{1, 2\}) \).

(ii) \( \pi^1 \) has an infinite 0 spin nearest-neighbour cluster.

Capel’s predictions for the Blume–Capel phase diagram are illustrated in Figures 2.1 on the \( K, \Delta \) plane and in Figure 2.2 on the \( a, p \) plane. The areas indicated are,

(a) where property (i) holds,

(b) where property (ii) holds,

(c) where neither property (i) nor property (ii) holds, there is neither type of long range order.

Recall the original parameterization of the Blume–Capel model, \( K = \beta J \) and \( \Delta = \beta D \). With \( J, D \) fixed, as temperature increases \((K, \Delta)\) moves towards \((0, 0)\) along a straight line, \((a, p)\) moves towards \((1/2, 0)\) along the path

\[
\frac{a}{1 - a} = (1 - p)^{D/(2J)}.
\]
Such paths are indicated in Figures 2.1 and 2.2 by grey lines.

A first order phase transition is one where a property of the model, such as the density of +1 spins, is discontinuous as a function of the parameters. A second order phase transition is one without this phenomenon. Capel predicted that a tri-critical point existed where the three phases meet, lying on the curve \( a/(1 - a) = (1 - p)^{\frac{2}{3} \log 4} \). The boundary between (a) and (b) is thought to be first-order and arrive at \((a, p) = (0, 1)\) at the same gradient as the curve \( a/(1 - a) = (1 - p) \). The boundary between (a) and (c) is thought to be second-order. By the comparison inequalities, if \((a, p)\) lies in region (a) then so does the rectangle bounded by \((a, p)\) and \((1, 1)\). Similarly, if \((a, p)\) lies in region (b) then
2.9 Phases and comparison results

so does the rectangle bounded by (0, 0) and (a, p).

We can use the comparison results with certain points of reference to draw more detailed inferences about the phase diagram. Three points are marked on Figure 2.3:

(i) \((a, p, q) = (1, \sqrt{2}/(1 + \sqrt{2}), 2)\): The \(a = 1\) DRC model is a random-cluster measure. The \(q = 2\) random-cluster measure on \(L^2\) has critical point \(p = \sqrt{2}/(1 + \sqrt{2}) \sim 0.57\). By Theorem 2.17, the \(q = 2\) DRC measure has an infinite open edge cluster to the right of the curve from \((a, p) = (1, 0.57)\) to \((a, p) = (1, 1)\). The corresponding BCP measure has a preponderance of 1 states. Below the line \(p = 0.57\), \(\phi\) has no infinite open edge clusters. In
the corresponding BCP measure, the 1-spins and the 2-spins have the same distribution.

(ii) \((a, p, q) = (0, (1 - p^\text{site})/(1 + p^\text{site}), 2)\): To the left of this point on the line \(p = 0\), \(\Phi\) is a product measure with an infinite cluster of closed vertices. Take \(p^\text{site}_c\), the critical probability for independent site percolation on \(\mathbb{L}^2\), to be 0.59. Site percolation is a nearest-neighbour vertex percolation model. Vertices are open with probability \(p\), adjacent open vertices are connected. By Theorem 2.10 (b) this point stochastically dominates the area to the left of the curve from \((0.25, 0)\) to \((0, 1)\). There is an infinite nearest-neighbour cluster of closed vertices. This corresponds to a BCP measure that has an infinite cluster of 0 states.

(iii) \((a, p, q) = (0, p^\text{site}_c/(2 - p^\text{site}_c), 2)\): To the right of this point on the line \(p = 0\), \(\Phi\) is a product measure with an infinite cluster of open vertices. By part (c) of Theorem 2.10, the region to the right of the line \(a = 0.41\) stochastically dominates \((0.41, 0)\), and therefore \(\Phi\) has an infinite open vertex cluster.

In addition, the DRC measure on the line \(a/(1 - a) = 1 - p\) with \(q = 1\) correspond to an Ising model with zero external magnetic field. On the line segment above \((a, p) = (1/((1 + \sqrt{2})^4 + 1), 1 - 1/(1 + \sqrt{2})^4)\), the Ising model is supercritical—the boundary conditions determine which spin is dominant. Below the supercritical line segment, the \(q = 1\) DRC model has a preponderance of 0 spins. Above the supercritical line segment, there is a preponderance of 1 spins.

By part (c) of Theorem 2.10, compare \((a_1, p_1, 2)\) and \((a_2, p_2, 1)\) with \((a_2, p_2)\) ‘just below’ the supercritical portion of \(a/(1 - a) = 1 - p\). We need

\[
(1 - p_2)^3 \geq 2 \left( \frac{a_1}{1 - a_1} \right) (1 - p_1)^2.
\]

This occurs if

\[
1 - p_2 \geq \left( \frac{2a_1}{1 - a_1} \right), \quad p_1 \geq p_2 \geq 1 - \left( 1 + \sqrt{2} \right)^{-4}.
\]

The area below the curve from \((a, p) = (0.0541, .986)\) to \((0, 1)\) is in region (b).
2.9 Phases and comparison results

By part (d) of Theorem 2.10, compare \((a_2, p_2, 2)\) and \((a_1, p_1, 1)\) with \((a_1, p_1)\) 'just above' the supercritical portion of \(a/(1 - a) = 1 - p\). We need,

\[
2 \left( \frac{a_2}{1 - a_2} \right) (1 - p_2)^2 \geq (1 - p_1)^3 \quad \text{and} \quad \frac{p_1}{1 - p_1} \leq \frac{p_2}{2(1 - p_2)}.
\]

Set \(p_1 = \frac{p_2}{2 - p_2}\). Then we need,

\[
\frac{2a_2}{1 - a_2} \geq \frac{8(1 - p_2)}{(2 - p_2)^3} \quad \text{and} \quad p_2 \geq \frac{2 - 2(1 + \sqrt{2})^4}{2 - (1 + \sqrt{2})^4}.
\]

The area to the right of the curve from \((a, p) = (.0145, .970)\) to \((a, p) = (0, 1)\) has an infinite open vertex cluster. Further, we can see that this region is in (a) as follows. Let \(S\) be the set of dominant spins of a super-critical Ising model with parameter \(\beta J = -\frac{1}{8} \log(1 - p_1)\). By the coupling of the Ising and the random-cluster model, the critical probability for bond percolation on \(S\) is a.s. less than \(\pi\) where \(\pi = 1 - \exp(-2\beta J)\). The set of 1 vertices under \(\phi_{a_2, p_2, 2}\) stochastically dominates \(S\), and the \((p_2, 2)\) random-cluster measure stochastically dominates a product measure with density \(\pi\). Hence \(\phi_{a_2, p_2, 2}\) has an infinite open edge cluster.
Chapter 3

Influence and monotonic measures

3.1 Influence beyond KKL

In the first chapter we introduced the topics of influence and sharp thresholds. We have so far only stated the influence result of KKL for Bernoulli product measure with density 1/2, the uniform measure on the Hamming space $\Omega = \{0, 1\}^N$. The theorem has been extended to more general product measures. A sharp threshold result follows by Russo’s formula. In this chapter we extend the notion of influence to monotonic measures. We prove influence and sharp threshold results for monotonic measures, in particular the random-cluster measure.

A classic result in percolation theory is due to Kesten; the critical value for bond percolation on the square planar lattice $L^2$ is one half [42]. His proof settled an old conjecture. The lower bound $p_c \geq 1/2$ was shown by Harris in 1960. Harris’s lower bound made use of the self duality of bond percolation on $L^2$ at $p = 1/2$. More recently, Bollobás and Riordan [9] produced a shorter proof using the newly developed sharp threshold technology.

The main motivation for this chapter is to explore the possibility of adapting their technique to the random-cluster setting. The random-cluster model on $L^2$ has a form of self duality at $p_{sd} = \sqrt{q}/(1 + \sqrt{q})$. We prove a sharp threshold result for short-ways rectangle crossings on a finite region of the square planar lattice.
with cyclic boundary conditions. This result is not strong enough to determine the critical value $p_c$. In Section 3.9 we show that a related but stronger result, a sharp threshold result for long-ways rectangle crossings on $\mathbb{L}^2$, would determine $p_c$.

Friedgut and Kalai extended the KKL influence result to $\mu_p$, product measure with density $p \in (0, 1)$. Recall that we write $\omega \in \Omega$ for a general configuration and $X_i$ for the random variable $X_i(\omega) = \omega(i)$. Let $i \in I = \{1, \ldots, N\}$; let $A \subset \Omega$ be an increasing event. The influence of coordinate $i$ is,

$$I_A(i) = \mu_p(A \mid X_i = 1) - \mu_p(A \mid X_i = 0) = \mu_p(\omega^i \in A, \omega_i \notin A). \quad (3.1)$$

Throughout the chapter, $C > 0$ is an absolute constant.

**Theorem 3.2 (Friedgut and Kalai [28]).** There exists $i \in I$ such that

$$I_A(i) \geq C \min\{\mu_p(A), \mu_p(A^c)\} \frac{\log N}{N}.$$ 

The following is the most general influence result for product measure. Let $E = [0, 1]^N$, the Euclidean cube, and let $(E, \mathcal{E}, \lambda)$ be the uniform (Lebesgue) measure space. $E$ has the coordinate-wise partial order. Let $A \subset E$ be increasing, if $x \leq y$ and $x \in A$ then $y \in A$. Let $U_i$ be the identity random variable for the $i$-th coordinate, $U_i(x) = x(i)$. A continuous version of influence is

$$I_A(i) = \lambda(A \mid U_i = 1) - \lambda(A \mid U_i = 0). \quad (3.3)$$

**Theorem 3.4 (BKKKL [10]).** For some $i \in I$,

$$I_A(i) \geq C \min\{\lambda(A), \lambda(A^c)\} \frac{\log N}{N}.$$ 

In Section 3.10 we extend this to an influence result for continuous monotonic measures.

Notice that the result of Friedgut and Kalai follows from the result of BKKKL by embedding. Let $f : E \rightarrow \Omega$ be defined by

$$f(x(1), \ldots, x(N)) = (\omega(1), \ldots, \omega(N))$$
3.2 Symmetric events

where for all $i \in I$,

$$\omega(i) = \begin{cases} 1 & x(i) \geq 1 - p, \\ 0 & \text{otherwise}. \end{cases}$$

Function $f$ is increasing, if $A \subset \Omega$ is increasing so is $B = f^{-1}(A) \subset E$. For all $i$, $I_A(i) = I_B(i)$.

The definition of influence for $(E, \mathcal{E}, \lambda)$ requires some clarification. It depends on the version of conditional expectation used. Interpret the conditional expectations in (3.3) as the $N-1$ dimensional Lebesgue measures on $\{1\} \times [0,1]^{N-1}$ and $\{0\} \times [0,1]^{N-1}$. We could have chosen a more robust definition of influence, let

$$\hat{I}_A(i) = \lim_{\varepsilon \to 0} \lambda(A \mid U_i \in [1-\varepsilon, 1]) - \lambda(A \mid U_i \in [0,\varepsilon]).$$

Consider the two events

$$A = \{ x \in [0,1]^2 : x(1) \geq 1/2 \}, \quad B = A \cup \{ x \in [0,1]^2 : x(2) = 1 \}.$$

They only differ on a null set, so $\hat{I}_A(2) = \hat{I}_B(2) = 0$. In contrast, $I_A(2) = 0$, $I_B(2) = 1/2$; the influence for $B$ seems exaggerated. However, Theorem 3.4 also holds with $\hat{I}$ substituted for $I$. Given an increasing set $B$, construct a set $A$ that only differs from $B$ on a null set. Remove points on the ‘upper’ faces such as $(1, x(2), \ldots, x(N))$ if $\lim_{x(1) \to 1-} I_A(x) = 0$, and add points on the ‘lower’ faces such as $(0, x(2), \ldots, x(N))$ if $\lim_{x(1) \to 0+} I_A(x) = 1$. Then $A$ is increasing and $\lambda(A) = \lambda(B)$. For all $i \in I$, $I_A(i) = \hat{I}_B(i)$; we can apply Theorem 3.4 to $A$.

3.2 Symmetric events

Let $S_N$ be the group of permutations of $I = \{1, \ldots, N\}$. Let $\Gamma$ be a subgroup of $S_N$ such that for all $i, j \in I$ there is a $\gamma \in \Gamma$ with $\gamma(i) = j$; $\Gamma$ is said to act transitively on $I$. For an event $A$ and a permutation $\gamma$, define event $A_\gamma$ by

$$(\omega(1), \ldots, \omega(N)) \in A \leftrightarrow (\omega(\gamma(1)), \ldots, \omega(\gamma(N))) \in A_\gamma.$$ 

An event $A$ is $\Gamma$-invariant if for all $\gamma \in \Gamma$, $A = A_\gamma$. An event is symmetric if it is $\Gamma$-invariant for a group $\Gamma$ that acts transitively on $I$. Take for instance the
3.3 Sharp thresholds for product measures

Majority event

\[ A = \left\{ \omega \in \Omega : \sum_{i \in I} \omega(i) \geq N/2 \right\}. \]

\( A \) is \( S_N \)-invariant. Similarly the tribe event with \( m \) tribes of \( n \) members is invariant with respect to the group of size \( m!(n!)^m \) generated by permutations within a tribe, and permutations of whole tribes. For a symmetric event \( A \), the influence is the same for each \( i \) so

\[ \sum_{i \in I} I_A(i) \geq C \min\{\mu_p(A), \mu_p(A^c)\} \log N. \]  \hspace{1cm} (3.5)

Contrast this with the dictatorship event \( A = \{\omega : \omega(1) = 1\}; \sum_{i \in I} I_A(i) = 1 \) for all \( p \).

### 3.3 Sharp thresholds for product measures

Let \( \omega \in \Omega \) and let \( A \) be an increasing event. If \( \omega_i \notin A, \omega^i \in A \), we say that \( i \) is critical for \( A \). The probability that coordinate \( i \) is critical is \( I_A(i) \). The higher the expected number of critical coordinates, the more sensitive \( \mu_p(A) \) is to small changes in \( p \). Russo’s formula \([51]\) quantifies this,

\[ \frac{d}{dp} \mu_p(A) = \sum_{i \in I} I_A(i). \]

Let \( A \) be a symmetric event, by Russo’s formula and (3.5),

\[ \frac{d}{dp} \mu_p(A) \geq C \min\{\mu_p(A), \mu_p(A^c)\} \log N. \]

When \( \mu_p(A) \leq 1/2 \), dividing both sides above by \( \mu_p(A) \) gives,

\[ \frac{d}{dp} \log \mu_p(A) \geq C \log N. \]

Similarly for \( \mu_p(A) \geq 1/2 \),

\[ \frac{d}{dp} \log \mu_p(A^c) \leq -C \log N. \]

Integrating over \( p \) gives a sharp threshold result \([8, 28]\).
3.4 Self duality for bond percolation on \( L^2 \)

**Theorem 3.6.** Let \( A \notin \{\emptyset, \Omega\} \) be an increasing symmetric event. Let \( \varepsilon \in [0, 1/2] \). If \( p, p' \in (0, 1) \) are such that \( \mu_p(A) \geq \varepsilon \) and

\[
p' - p = \frac{2\log 1/(2\varepsilon)}{C \log N}
\]

then \( \mu_{p'}(A) \geq 1 - \varepsilon \).

This technique was first used in [28] to prove a sharp threshold theorem for graph properties on Erdős–Rényi graphs. The same technique can be used on graphs with less symmetry. Let \( \mathbb{T}^d = (\mathbb{Z}/n\mathbb{Z})^d \) be the \( d \)-dimensional torus; the nearest neighbour graph on \( \mathbb{T}^d \) can be thought of as the subgraph \([0, n]^d\) of \( \mathbb{L}^d \) with cyclic boundary conditions imposed. Let \( \Gamma \) be the group of permutations generated by the translations and rotations of \( \mathbb{T}^d \). Every \( \Gamma \)-invariant increasing event has a sharp threshold.

### 3.4 Self duality for bond percolation on \( L^2 \)

Let \( G = (V, E) \) be a simple graph (no loops or multiple edges) drawn on the Euclidean plane such that no edges cross. In every face, including in unbounded faces, mark a point; let \( V' \) be the set of marked points. These are the dual vertices. Every edge \( e \in E \) separates two faces; let \( e' \) be a dual edge between the corresponding dual vertices. Let \( E' \) be the set of dual edges; then \( G' = (V', E') \) is the planar dual of \( G \). \( G' \) may depend on the particular planar embedding of \( G \). Similarly, if a graph is embedded in a planar manifold, such as a torus, the dual with respect to that manifold can be constructed by following the same algorithm. In Figure 3.1 we show a small finite graph (edges marked as black lines) and the dual graph (dual vertices marked as black dots, dual edges marked as dotted lines).

Bond percolation with density \( p \) on \( G \) is naturally coupled to bond percolation with density \( 1 - p \) on the dual graph. Every edge \( e \in E \) crosses exactly one dual edge \( e' \in E' \), and vice versa. Define \( e' \) open if \( e \) is closed, and \( e' \) closed if \( e \) is open.

The square planar lattice \( \mathbb{L}^2 = (\mathbb{Z}^2, \mathbb{E}^2) \) is self dual. In Figure 3.2 we show a subset of the lattice \( \mathbb{L}^2 \) and the isomorphic dual lattice. We have shaded part of
3.4 Self duality for bond percolation on $L^2$

Figure 3.1: Finite graph and dual

Figure 3.2: Square planar lattice and isomorphic dual
3.5 Influence for monotonic measures

the figure grey. We have left in black a $[0, n + 1] \times [0, n]$ rectangle $R$ ($n = 4$) and a similar rectangle $R'$ in the dual lattice, rotated through $90^\circ$. Write $\text{LW}(R)$ for the event that $R$ is crossed long-ways (horizontally in this case) by a path of open edges. Notice that exactly one of the following happens: either $\text{LW}(R)$, or $R'$ is crossed long-ways (vertically) by open dual edges, $\text{LW}(R')$. This is a consequence of the max-cut min-flow theorem. As $\mu_{1-p}$ is the measure on the dual lattice,

$$\mu_p(\text{LW}(R)) + \mu_{1-p}(\text{LW}(R)) = 1.$$ 

At $p = 1/2$, $p = 1 - p$ so bond percolation is self-dual, $\mu_{1/2}(\text{LW}(R)) = 1/2$. Write $R_{m,n}$ for a $[0, m] \times [0, n]$ rectangle. Seymour and Welsh [52] proved that for $k = 1, 2, \ldots$,

$$\liminf_{n \to \infty} \mu_{1/2}(\text{LW}(R_{n, kn})) > 0.$$ 

Kesten used an intricate path counting argument to show a sharp threshold for the event $\text{LW}(R_{n, 2n})$ for large $n$. It follows that $p_c = 1/2$. We will later give an argument for this involving a construction of rectangles. Bollobás and Riordan [9] used the sharp threshold result of Friedgut and Kalai to simplify Kesten’s result. Place an $R_{2n, 4n}$ rectangle in a discrete torus, large enough to prevent ‘wrapping round’, say $T_{6n}^2 = (\mathbb{Z}/6n\mathbb{Z})^2$. Let $A$ be the symmetric increasing event that there is a copy of $R_{n, 5n}$ somewhere in $T_{6n}^2$, allowing rotation and translation, that is crossed long-ways. The torus can be covered by a small number of copies of $R_{2n, 4n}$ such that $A$ implies one of the copies is crossed long-ways. The sharp threshold result for $A$ implies a sharp threshold result for $\text{LW}(R_{2n, 4n})$.

### 3.5 Influence for monotonic measures

Let $\mu$ be a positive measure on the Hamming space $\Omega = \{0, 1\}^N$; $\mu(\omega) > 0$ for all $\omega \in \Omega$. For $J \subset I$ and $\xi \in \Omega$, let

$$\mu^\xi_J(\eta) = \mu(X_j = \eta(j) \text{ for } j \in J \mid X_i = \xi(i) \text{ for } i \in I \setminus J).$$

Recall that the following are equivalent for $\mu$ positive [34].

(i) $\mu$ satisfies the FKG lattice condition [14].

52
(ii) \( \mu \) is monotonic: for \( J \subset I \) and \( A \) increasing, \( \mu_j^\xi(A) \) is increasing in \( \xi \in \Omega \).

(iii) \( \mu \) is strongly positively-associated: for \( J \subset I, \xi \in \Omega \) and \( A, B \) increasing, 
\[
\mu_j^\xi(A \cap B) \geq \mu_j^\xi(A)\mu_j^\xi(B).
\]

Let \( \mu \) be a positive monotonic measure on \((\Omega, \mathcal{F})\). For an increasing set \( A \subset \Omega \), define the conditional influence of coordinate \( i \) by,
\[
I_A(i) = \mu(A \mid X_i = 1) - \mu(A \mid X_i = 0).
\]

As \( \mu \) is monotonic, the influence is always non-negative. When \( \mu \) is a product measure, conditional influence is exactly the influence, see (3.1).

Equality (3.1) for product measures also motivates an alternative definition for influence,
\[
\overline{I}_A(i) = \mu(\omega^i \in A, \omega_i \notin A).
\]

Unlike with product measures, it is not generally the case that \( I_A(i) = \overline{I}_A(i) \).

Consider the following measure. Take \( N \) odd for convenience. Let \( \mu_{1/3}, \mu_{2/3} \) be product measures on \( \Omega \) with densities 1/3 and 2/3 respectively. Let \( \mu = (\mu_{1/3} + \mu_{2/3})/2 \); this corresponds to picking a coin at random, with bias 1/3 or 2/3, and then tossing it \( N \) times. Let \( A \) be the majority event, that more than half of the coordinates are 1. By symmetry \( \mu(A) = 1/2 \).

For each coordinate, the conditional influence \( I_A(i) = \Theta(1) \). The majority event is strongly correlated with the choice between the two coins. However, coordinate one is only critical if the other coordinates sum to exactly \((N - 1)/2\). The probability of this, \( \overline{I}_A(1) \), decay exponentially with \( N \). While there can be no influence theorem for the \( \overline{I}_A(i) \), we can extend the influence results for product measure to conditional influence results for monotonic measures.

**Theorem 3.7.** Let \( \mu \) be a positive monotonic measure on \((\Omega, \mathcal{F})\). Let \( A \subset \Omega \) be an increasing event. Then for some \( i \in I \),
\[
I_A(i) \geq C \min\{\mu(A), \mu(A^c)\} \frac{\log N}{N}.
\]
3.5 Influence for monotonic measures

Proof. Let \( \lambda \) be the uniform measure on \([0, 1]^N\). We will construct a function \( f : [0, 1]^N \rightarrow \{0, 1\}^N \). We will do this on a coordinate-wise basis, constructing \( f = (f_1, \ldots, f_N) \in \Omega \) for given \( x \in [0, 1]^N \). Function \( f \) will have the following properties.

(i) For all \( i \in I \), \( f_i \) only depends on \( x(1), \ldots, x(i) \).

(ii) For all \( \omega \in \Omega \), \( \lambda(f = \omega) = \mu(\omega) \).

(iii) Function \( f \) is increasing with respect to the corresponding partial orders on the domain and co-domain.

Let \( B = f^{-1}(A) = \{x \in [0, 1]^N : f(x) \in A\} \). The distribution of \( f \) under \( \lambda \) is equal to the distribution of the identity function under \( \mu \), so \( \lambda(B) = \mu(A) \).

The result then follows by showing that \( I_A(i) \geq I_B(i) \) for all \( i \in I \). Notice that \( I_B(i) \) is an influence with respect to the product measure on \([0, 1]^N\) and \( I_A(i) \) is a conditional influence with respect to \( \mu \).

We now construct \( f(x) \). Let

\[
    f_1 = \begin{cases} 
        1 & x(1) \geq 1 - \mu(X_1 = 1), \\
        0 & \text{otherwise}. 
    \end{cases}
\]

Notice \( f_1 \) only depends on \( x(1) \), \( f_1 \) is increasing in \( x(1) \) and \( \lambda(f_1 = 1) = \mu(X_1 = 1) \).

For \( i \in \{1, \ldots, N - 1\} \), assume inductively:

(i) \( f_i \) only depends on \( x(1), \ldots, x(i) \),

(ii) \( \mu(X_k = a_k \, \text{for} \, k = 1, \ldots, i) = \lambda(f_k = a_k \, \text{for} \, k = 1, \ldots, i) \) for all \( a_1, \ldots, a_i \in \{0, 1\} \), and

(iii) \( f_i \) is increasing in \( x(1), \ldots, x(i) \).

Let

\[
    f_{i+1} = \begin{cases} 
        1 & x(i + 1) \geq 1 - \mu(X_{i+1} = 1 \mid X_k = f_k \, \text{for} \, k = 1, \ldots, i), \\
        0 & \text{otherwise}. 
    \end{cases}
\]

Then clearly the first and second property follow immediately. The third property follows by the monotonicity of \( \mu \). The event \( X_{i+1} = 1 \) is increasing, so
3.5 Influence for monotonic measures

\[ \mu(X_{i+1} = 1 \mid X_k = f_k \text{ for } k = 1, \ldots, i) \] is increasing in \( f_1, \ldots, f_i \), which by assumption are increasing in \( x(1), \ldots, x(i) \).

By induction \( f \) has the properties claimed. In the construction of \( f \) we have made a choice—the order in which we construct the \( f_i \) based on the \( f_j \) already constructed—we arbitrarily took the natural order 1, \ldots, \( n \). Below we will consider a different order.

We now claim that \( I_A(i) \geq I_B(i) \) for \( i \in I \), where \( B = f^{-1}(A) \). The result follows from this claim by the BKKKL influence result applied to \( B \).

For \( i = 1 \), this is immediate. Measure \( \mu \) is positive, so the conditional distributions of \( f \) under \( \lambda \) given \( x(1) = 0 \) and \( x(1) = 1 \) are exactly the conditional distributions of \( \mu \) given \( X_1 = 0 \) and \( X_1 = 1 \), respectively. Let \( i > 1 \), we construct a function \( g : [0,1]^N \rightarrow \{0,1\}^N \) in a similar fashion to \( f \), except we construct coordinates \( g_j \) in the order \( i, 1, \ldots, i-1, i+1, \ldots, N \). Let

\[
g_i = \begin{cases} 1 & x(i) \geq 1 - \mu(X_i = 1) \\ 0 & \text{otherwise.} \end{cases}
\]

For \( j = 1, \ldots, i-1 \), let

\[
g_j = \begin{cases} 1 & x(j) \geq 1 - \mu(X_j = 1 \mid g_i, g_1, \ldots, g_{j-1}) \\ 0 & \text{otherwise.} \end{cases}
\]

Then for \( j = i+1, \ldots, N \), let

\[
g_j = \begin{cases} 1 & x(j) \geq 1 - \mu(X_j = 1 \mid g_i, \ldots, g_{j-1}) \\ 0 & \text{otherwise.} \end{cases}
\]

Let \( D = g^{-1}(A) \). Then \( I_D(i) = I_A(i) \), just as \( I_B(1) = I_A(1) \). All that remains is to show that \( I_D(i) \geq I_B(i) \). This follows from the monotonicity of \( \mu \). Let \( x \in [0,1]^N \) and write

\[ x^i = (x(1), \ldots, 1, \ldots, x(N)), \quad x_i = (x(1), \ldots, 0, \ldots, x(N)), \]

with the change in the \( i \)-th coordinate. We show that \( g(x^i) \geq f(x^i) \). Then combined with the converse result \( g(x_i) \leq f(x_i) \) the claim follows. The intuitive
reason why \( g(x^i) \geq f(x^i) \) is that \( \mu \) is positively associated, so the sooner we process the ‘good’ information that coordinate \( i \) takes the value 1, the greater the upward lift elsewhere. Firstly, \( f_i(x^i) = g_i(x^i) = 1 \). We can now check inductively that \( g_j(x^i) \geq f_j(x^i) \) for \( j = 1, \ldots, i - 1, i + 1, \ldots, N \). Look at the definitions for \( f_j \) and \( g_j \). At every step, the values \( f_k(x^i) \) already defined \( (k = 1, \ldots, j - 1) \) are smaller than the corresponding \( g_k(x^i) \), so by monotonicity \( g_j(x^i) \geq f_j(x^i) \). \( \square \)

3.6 Russo formula for monotonic measures

Let \( \mu \) be a positive measure on \( \Omega = \{0, 1\}^N \), it need not be a probability measure. Suppose that \( \mu \) satisfies the FKG lattice condition (1.4). Define a parameterized set of probability measures \( \mu_p \) for \( p \in (0, 1) \) by

\[
\mu_p(A) = \frac{1}{Z} \sum_{\omega \in A} \mu(\omega) \prod_{i \in I} p^{\omega(i)}(1 - p)^{1 - \omega(i)}
\]

with

\[
Z = \sum_{\omega \in \Omega} \mu(\omega) \prod_{i \in I} p^{\omega(i)}(1 - p)^{1 - \omega(i)}.
\]

Then it is immediate that \( \mu_p \) is positive and satisfies the FKG lattice condition; \( \mu_p \) is monotonic. The obvious example of such a family of measures is given by the random-cluster measures on a graph. Fix \( q \geq 1 \) and let \( I \) be the edge set of a graph. Take \( \mu(\omega) = q^{k(\omega)}; k(\omega) \) counts the number of open-clusters. Then \( \mu_p \) is the \((p, q)\) random-cluster measure.

Russo’s formula has been extended to cover random-cluster measures [5]. It extends naturally to this setting. For convenience, write \( A \) for the indicator function of an increasing event \( A \).

**Theorem 3.8.** With \( \text{cov}_p \) the covariance with respect to \( \mu_p \),

\[
\frac{d}{dp} \mu_p(A) = \frac{1}{p(1 - p)} \sum_{i \in I} \text{cov}_p(X_i, A).
\]
Proof. Differentiating $Z \times \mu_p(A)$,
\[
\frac{d}{dp} \sum_{\omega \in A} \mu(\omega) \prod_{i \in I} p^{\omega(i)}(1 - p)^{1 - \omega(i)} = \sum_{i \in I} \sum_{\omega \in A} \mu(\omega) \prod_{i \in I} p^{\omega(i)}(1 - p)^{1 - \omega(i)} \left[ \frac{\omega(i)}{p} - \frac{1 - \omega(i)}{1 - p} \right] = Z \sum_{i \in I} \frac{\mu_p(X_i A)}{p} - \frac{\mu_p((1 - X_i) A)}{1 - p}.
\]

The above with $A = \Omega$ gives
\[
\frac{d}{dp} Z = Z \sum_{i \in I} \frac{\mu_p(X_i)}{p} - \frac{\mu_p((1 - X_i))}{1 - p}.
\]

Therefore,
\[
\frac{d}{dp} \mu_p(A) = \sum_{i \in I} \frac{\mu_p(X_i A)}{p} - \frac{\mu_p((1 - X_i) A)}{1 - p} - \mu_p(A) \sum_{i \in I} \frac{\mu_p(X_i)}{p} - \frac{\mu_p((1 - X_i))}{1 - p} = \sum_{i \in I} \frac{1}{p(1 - p)} \text{cov}_p(X_i, A).
\]

Conditional influence and covariance are related,
\[
I_A(i) = \frac{\mu_p(AX_i)}{\mu_p(X_i)} - \frac{\mu_p(A(1 - X_i))}{\mu_p(1 - X_i)} = \frac{\mu_p(AX_i) \mu_p(1 - X_i) - \mu_p(A(1 - X_i)) \mu_p(X_i)}{\mu_p(X_i) \mu_p(1 - X_i)} = \frac{\text{cov}_p(X_i, A)}{\mu_p(X_i) \mu_p(1 - X_i)}.
\]

Let
\[
\xi_p = \min_{i \in I} \frac{\mu_p(X_i) \mu_p(1 - X_i)}{p(1 - p)},
\]
so
\[
\frac{d}{dp} \mu_p(A) \geq \xi_p \sum_{i \in I} I_A(i).
\]
Say that \( \mu \) is \( \Gamma \)-invariant if for all \( \gamma \in \Gamma \), for all events \( A \), \( \mu(A) = \mu(A \gamma) \). If increasing event \( A \) and measure \( \mu \) are both \( \Gamma \)-invariant for some transitive \( \Gamma \) then \( IA(i) \) is independent of \( i \). Let \( \mu \) be such a measure. Then

\[
\frac{d}{dp} \mu_p(A) \geq C \xi_p \min\{\mu_p(A), \mu_p(A^c)\} \log N.
\]

There is an additional complication compared with the product case. If we want to prove the existence of a sharp threshold in the neighbourhood of \( p \), then we need to bound \( \xi_p \) away from zero.

### 3.7 Sharp thresholds for the random-cluster measure

Let \( G = (V, E) \) be a finite simple graph. A graph automorphism is a permutation \( \pi \) on the vertex set \( V \) such that

\[
\forall x, y \in V, \ (x, y) \in E \leftrightarrow (\pi(x), \pi(y)) \in E.
\]

Let \( \gamma \) be the induced permutation on edge set \( E \),

\[
\gamma(e) = (\pi(x), \pi(y)), \quad e = (x, y) \in E.
\]

Call \( \gamma \) a graph edge automorphism. Let \( \Pi \) be the group of graph automorphisms of \( G \). Let \( \Gamma \) be the corresponding group of graph edge automorphisms. We then say that \( G \) is \( \Gamma \)-invariant. If \( \Gamma \) acts transitively on \( E \) we say that \( G \) is symmetric.

Let \( \phi_{p,q} \) be the \((p, q)\) random-cluster measure on \( \{0, 1\}^E \) with \( p \in (0, 1), q \geq 1 \),

\[
\phi_{p,q}(\omega) = \frac{1}{Z} q^{k(\omega)} \prod_{e \in E} p^{\omega(e)(1 - p)} 1^{1 - \omega(e)}.
\]

If \( \Gamma \) is the group of graph edge automorphisms, \( \phi_{p,q} \) is \( \Gamma \)-invariant.

Recall that \( \phi_{p,q} \) has a finite energy property. Let \( C_e \) be the event that the end vertices of edge \( e = (x, y) \) are connected by an open path of edges in \( E \setminus \{e\} \). Then

\[
\phi_{p,q}(X_e \mid C_e) = p, \quad \phi_{p,q}(E_e \mid C_e^c) = \frac{p}{p + q(1 - p)}.
\]
3.8 Self duality for the random-cluster model

As \( t \to t(1-t) \) is concave,

\[
\min\{\phi_{p,q}(X_e = 1)\phi_{p,q}(X_e = 0)\} \\
\geq \min\left\{ t(1-t) : t \in \left\{ \frac{p}{p+q(1-p)}, p \right\} \right\}
\]

so

\[
\xi_p \geq \frac{\min\{p(1-p), p(1-p)q/(p+q(1-p))\}}{p(1-p)} \geq 1/q.
\]

If graph \( G \) and increasing event \( A \) are symmetric,

\[
\frac{d}{dp} \phi_{p,q}(A) \geq \frac{C}{q} \min\{\phi_{p,q}(A), \phi_{p,q}(A^c)\} \log N.
\]

As in Section 3.3 this gives a sharp threshold result.

**Theorem 3.9.** Let \( A \not\in \{\emptyset, \Omega\} \) be an increasing symmetric event on symmetric graph \( G \). Let \( \varepsilon \in [0, 1/2] \). If \( p, p' \in (0, 1) \) are such that \( \phi_{p,q}(A) \geq \varepsilon \) and

\[
p' - p = \frac{2q \log 1/(2\varepsilon)}{C \log N}
\]

then \( \phi_{p',q}(A) \geq 1 - \varepsilon \).

### 3.8 Self duality for the random-cluster model

The square planar lattice \( \mathbb{L}^2 \) and the torus \( \mathbb{T}_n^2 \) are self dual graphs. Bond percolation with density \( p \) is dual to bond percolation with density \( 1-p \) on the same graph. Let \( q \geq 1 \) and let \( p, p' \) satisfy

\[
\frac{p'}{1-p'} = \frac{1-p}{p}.
\]

The \((p, q)\) random-cluster model on a finite graph is dual to the \((p', q)\) random-cluster model on the dual graph. Let \( \phi_{n,p} \) be the \((p, q)\) random-cluster model on \( \mathbb{T}_n^2 \), \( \phi_{n,p} \) is dual to \( \phi_{n,p'} \). The self dual point is \( p_{sd} = \sqrt{q}/(1+\sqrt{q}) \). On \( \mathbb{L}^2 \) we have to take into account boundary conditions. The \((p, q)\) random-cluster model with free boundary conditions \( \phi^0_{p,q} \) is dual to the wired \((p', q)\) random-cluster measure \( \phi^1_{p',q} \), \( \phi^1_{p_{sd},q} \) is dual to \( \phi^0_{p_{sd},q} \).
3.8 Self duality for the random-cluster model

Let $k \geq 2$, $n \geq 1$. Let $R = R_{n,n+1}$ be the $[0,n] \times [0,n+1]$ rectangle in $\mathbb{T}^2_{kn}$ and let $LW(R)$ be the event that $R$ is crossed long-ways. By duality, $\phi_{kn,p_{sd}}(LW(R)) = 1/2$. Let $A$ be the event that $\mathbb{T}^2_{kn}$ contains a copy of $R$, up to rotation and translation, that is crossed long-ways. With $\Gamma$ the group of graph edge automorphisms of $\mathbb{T}^2_{kn} = (V,E)$,

$$A = \{ \omega \in \Omega : \exists \gamma \in \Gamma, \gamma(\omega) \in LW(R) \}, \quad \Omega = \{0,1\}^E.$$

The increasing event $A$ and measure $\phi_{kn,p}$ are $\Gamma$-invariant. Therefore, for $p \geq p_{sd}$, as $|E| = 2(kn)^2$,

$$\phi_{kn,p}(A) \geq 1 - \frac{1}{2} |E|^{-C(p-p_{sd})/q} \geq 1 - (kn)^{-2C(p-p_{sd})/q}.$$

We will now use the ‘square-root trick’ \[9\] to produce a sharp threshold result for the asymmetric event that a given rectangle is crossed. Let $A_1, \ldots, A_M$ be a number of increasing events with equal probability. Then by positive association

$$1 - \phi_{kn,p} \left( \bigcup_{i=1}^M A_i \right) = \phi_{kn,p} \left( \bigcap_{i=1}^M A_i^c \right) \geq \prod_{i=1}^M \phi_{kn,p}(A_i^c) = (1 - \phi_{kn,p}(A_1))^M$$

so

$$\phi_{kn,p}(A_1) \geq 1 - M \sqrt{1 - \phi_{kn,p} \left( \bigcup_{i=1}^M A_i \right)}.$$

Let $1 < \alpha < k$, let $H_{n,\alpha} = [0,\alpha n] \times [0,n/\alpha]$ and let $V_{n,\alpha} = [0,n/\alpha] \times [0,\alpha n]$. We will cover $\mathbb{T}^2_{kn}$ with copies of $H_{n,\alpha}$ and $V_{n,\alpha}$ so that $A$ implies that one of the copies is crossed short-ways. Let

$$h_{n,\alpha} = \{(l_1 | n(\alpha - 1)|, l_2 | n(1 - \alpha^{-1})| : \lambda_1, \lambda_2 \in \mathbb{Z} \} \cap [0, kn]^2,$$

$$v_{n,\alpha} = \{(l_1 | n(1 - \alpha^{-1})|, l_2 | n(\alpha - 1)| : \lambda_1, \lambda_2 \in \mathbb{Z} \} \cap [0, kn]^2.$$

Let $\mathcal{H}$ be the set of translations of $H_{n,\alpha}$ by the vectors in $h_{n,\alpha}$. Similarly let $\mathcal{V}$ be the set of translations of $V_{n,\alpha}$ by the vectors $v_{n,\alpha}$. Let $M = |\mathcal{H} \cup \mathcal{V}|$. If $A$ occurs, then an element of $\mathcal{H} \cup \mathcal{V}$ is crossed short-ways. By the square-root trick, the probability $H_{n,\alpha}$ is crossed short-ways,

$$\phi_{kn,p}(SW(H_{n,\alpha})) \geq 1 - \sqrt{M} \sqrt{1 - \phi_{kn,p}(A)}.$$
M = |h_{n,\alpha}| + |v_{n,\alpha}|
= 2 \left[ \begin{array}{c} \frac{kn}{n[\alpha - 1]} \\ \frac{kn}{n[1 - \alpha^{-1}]} \end{array} \right] 
\leq 2 \left( 1 + \frac{k}{\alpha - 1 - n^{-1}} \right) \left( 1 + \frac{k}{1 - \alpha^{-1} - n^{-1}} \right),

so \( M \approx 2k^2\alpha/(\alpha - 1)^2 \) for \( n, k \) large.

**Theorem 3.10.** Let \( k \geq 2, n \geq 1, \) and \( \alpha \in (1, k) \). For \( p \in [p_{sd}, 1] \),
\[
\phi_{kn,p}(SW([0, n\alpha] \times [0, n\alpha^{-1}])) \geq 1 - e^{-g(p-p_{sd})}
\]
where
\[
g = g(k, n, \alpha, q) = \frac{2C}{Mq} \log(kn).
\]

### 3.9 Long-ways crossings and the critical point

Let \( q \geq 1 \). The critical point for the random-cluster measure on \( \mathbb{L}^2 \) is,
\[
p_c = \inf \{ p : \phi^1_{p,q}(0 \leftrightarrow \infty) > 0 \}.
\]

It is known that \( p_c \geq p_{sd} \), and it is conjectured that \( p_c = p_{sd} \) [34].

For completeness, we will now give a construction connecting rectangle crossings to percolation. The starting premise is stronger than the conclusion of Theorem 3.10. It involves long-ways crossings rather than short-ways crossings, and it is set on the square planar lattice rather than on a finite torus. Let \( \phi^1_{p,q} \) be the \((p,q)\) random-cluster measure on \( \mathbb{L}^2 \) with wired boundary conditions.

**Theorem 3.11.** Let \( q \geq 1 \). Let \( p_k = \phi^1_{p,q}(LW(R_{n,2n})) \) with \( n = 2^k \). If for all \( p \geq p_{sd} \),
\[
\prod_{k=1}^{\infty} p_k > 0,
\]
the critical point \( p_c = p_{sd} \).

**Proof.** We use a construction from [17]. Let
\[
A_k = \begin{cases} 
\text{LW([0, 2^k] \times [0, 2^{k+1}]}) & \text{k odd,} \\
\text{LW([0, 2^{k+1}] \times [0, 2^k])} & \text{k even.}
\end{cases}
\]
3.10 Influence for continuous random variables

By positive association and symmetry, for \( p > p_{sd} \),

\[
\phi^1_{p,q} \left( \bigcap_k A_k \right) \geq \prod_k \phi^1_{p,q}(A_k) > 0.
\]

The intersection of the \( A_k \) implies the existence of an infinite path; a path along the length of \( A_k \) crosses a path along the length of \( A_{k+1} \).

Let \( p > p_{sd} > p' \) satisfy the duality condition \( p'/(1-p') = q(1-p)/p \). By the duality of \( \phi^1_{p,q} \) and \( \phi^0_{p',q} \), \( 1 - p_k \) is the probability of a short-ways crossing,

\[
1 - p_k = \phi^0_{p',q}(\text{SW}(R_2^{k+1}-1,2^{k+1})).
\]

Let \( \text{rad}(C) \) be the maximum \( n \) such that 0 is joined to the boundary of the box \([-n,n]^2\) by a path of open edges.

\[
\sum_{k=1}^{\infty} 1 - p_k \leq \sum_{k=1}^{\infty} 2^{k+1} \phi^0_{p',q}(\text{rad}(C) \geq 2^k + 1) \leq 4 \sum_{n=1}^{\infty} \phi^0_{p',q}(\text{rad}(C) \geq n) = 4 \phi^0_{p',q}(\text{rad}(C)).
\]

If \( \phi^0_{p',q}(\text{rad}(C)) < \infty \) then \( \prod_k p_k > 0 \). If the expectation of \( \text{rad}(C) \) is finite for all \( p' < p_{sd} \) then \( p_c = p_{sd} \).

3.10 Influence for continuous random variables

We have extended Friedgut and Kalai’s influence result from product measure to monotonic measures on \( \{0,1\}^N \). The most general influence result for product measures is the result of BKKKL on the Euclidean cube \([0,1]^N\). The definitions of monotonicity and positive association have been extended to this setting. Holley’s theorem and the FKG theorem also have continuous versions [1, 50]. It seems natural to ask how Theorem 3.7 can be extended beyond discrete monotonic measures.

With \( E = [0,1]^N \), let \((E,\mathcal{E},\lambda)\) be the uniform (Lebesgue) measure on \( E \). Let \( U_i \) be the projection random variable from \( E \) onto the \( i \)-th coordinate,
3.10 Influence for continuous random variables

\[ U_i(x) = x(i), \text{ for } i \in I = \{1, \ldots, N\}. \]  While it is not generally the case that increasing subsets of \( E \) are Borel measurable, the following is presumably well known.

**Lemma 3.12.** Increasing subsets of \( E \) are Lebesgue measurable.

Let \( \rho \) be a density function \( \rho : E \to [0, \infty) \) with \( \lambda(\rho) = 1 \). Let \( \mu_\rho \) be the corresponding probability measure,

\[ \mu_\rho(A) = \int_A \rho(x) \, dx, \quad A \in \mathcal{E}. \]

We will state results in this section in terms of density functions, though of course they are results for the corresponding measures. We say that \( \rho \) is stochastically dominated by \( \nu \), written \( \rho \leq_{st} \nu \), if for all increasing sets \( A \),

\[ \mu_\rho(A) \leq \mu_\nu(A). \]

The following theorem is an extension of Holley’s theorem.

**Theorem 3.13 (Preston [50]).** Let \( \rho, \nu \) be density functions on \( E \). If

\[ \forall x, y \in E, \quad \rho(x \wedge y)\nu(x \vee y) \geq \rho(x)\nu(y) \]

then \( \rho \leq_{st} \nu \).

We say that \( \rho \) is positively associated if for all \( A, B \) increasing

\[ \mu_\rho(A \cap B) \geq \mu_\rho(A)\mu_\rho(B). \]

Let \( J \subset I \) and let \( \xi \in E \). Define

\[ E_\xi^J = \{ x \in E : x(i) = \xi(i) \text{ for } i \in I \setminus J \}. \]

Now define the conditional measure \( \mu^\xi_{\rho,J}(A) \) by the density

\[ \rho^\xi_J(x) \propto \rho(x)1_{E_\xi^J}(x). \]

Call \( \rho \) monotonic if for all \( J \subset I, \rho^\xi_J \) is stochastically increasing in \( \xi \). We say that density function \( \rho \) satisfies the FKG lattice condition if

\[ \forall x, y \in E, \quad \rho(x \wedge y)\rho(x \vee y) \geq \rho(x)\rho(y). \]
As in the discrete case, the FKG lattice condition for \( \rho \) positive implies strong positive-association [1]. This follows from Preston’s Theorem. Let \( J \subset I \), let \( \xi \in \mathbb{E} \), and let \( A, B \) be increasing subsets of \( E^\xi_J \). For all \( x, y \in E^\xi_J \),

\[
\rho^\xi_J(x \land y)\rho^\xi_J(x \lor y)1_A(x \lor y) \geq \rho^\xi_J(x)\rho^\xi_J(y)1_A(y).
\]

Therefore \( \mu^\xi_J(\cdot) \leq_{st} \mu^\xi_J(\cdot | A) \), so

\[
\mu^\xi_J(A \cap B) \geq \mu^\xi_J(A)\mu^\xi_J(B).
\]

The FKG lattice condition also implies monotonicity. Let \( J \subset I \) and \( \xi \leq \zeta \in \mathbb{E} \). Let \( x, y \in [0,1]^J \), and write \( x^\xi \) for the configuration in \( E^\xi_J \) that agrees with \( x \) on \( J \). Putting \( x^\xi \) and \( y^\zeta \) into the FKG lattice condition gives

\[
\rho^\xi_J((x \land y)^\xi)\rho^\zeta_J((x \lor y)^\zeta) \geq \rho^\xi_J(x^\xi)\rho^\zeta_J(y^\zeta),
\]

so \( \rho^\xi_J \leq_{st} \rho^\zeta_J \). We can summarize these results as follows.

**Theorem 3.14.** If strictly positive density function \( \rho \) satisfies the FKG lattice condition,

(i) \( \rho \) is strongly positively-associated, and

(ii) \( \rho \) is monotonic.

Define the conditional influence \( I_A^\rho(i) \) for monotonic density \( \rho \) and increasing set \( A \),

\[
I_A^\rho(i) = \mu^\rho(A | U_i = 1) - \mu^\rho(A | U_i = 0).
\]

Though it only depends on \( \rho \) on a null set, by monotonicity it satisfies

\[
I_A^\rho(i) \geq \hat{I}_A^\rho(i) := \lim_{\varepsilon \to 0} \mu^\rho(A | U_i \in [1 - \varepsilon, 1]) - \mu^\rho(A | U_i \in [0, \varepsilon]).
\]

**Theorem 3.15.** Let \( \rho \) be a strictly positive monotonic density. Then

\[
\exists i \in I, \ I_A^\rho(i) \geq C \min\{\mu^\rho(A), \mu^\rho(A^c)\}\frac{\log N}{N}.
\]
3.10 Influence for continuous random variables

However, there does not seem to be a corresponding sharp threshold result for $E = [0, 1]^N$. Let $\rho$ be a monotonic density function. As in the discrete case, we can define an increasing family of monotonic density functions,

$$\rho_p(x) = \frac{1}{Z} \rho(x) \prod_{i=1}^{N} \rho^{x(i)}(1 - p)^{1 - x(i)},$$

with normalising constant $Z$. Russo’s formula extends to this setting,

$$\frac{d}{dp} \mu_p(A) = \frac{1}{p(1 - p)} \sum_{i=1}^{N} \text{cov}_{\rho_p,U_i}(1_A).$$

However, we cannot bound $\text{cov}_{\rho_p,U_i}(1_A)$ in terms of $I_A(i)$. Let $\rho = 1$, so $\rho_p$ is a product density for all $p \in (0, 1)$. Let $\pi = p/(1 - p)$, then

$$\rho_p(x) = \begin{cases} 
\frac{\log \pi}{2p-1} \prod_{i=1}^{N} \rho^{x(i)}(1 - p)^{1 - x(i)} & p \neq 1/2, \\
1 & p = 1/2.
\end{cases}$$

Let $A = [N^{-1}, 1]^N$. For all $i \in I$, $I_A(i) = 1$, but as $N \to \infty$,

$$\mu_p(A) \to \begin{cases} 
\pi^{-1/(\pi-1)} & p \neq 1/2, \\
e^{-1} & p = 1/2.
\end{cases}$$

There is no sharp threshold.

**Proof of Theorem 3.15.** The proof is similar to the discrete case. We construct an increasing function $f : E \to E$ such that:

(i) The law of the identity function $U$ under $\mu_\rho$ is the same as the law of $f(U)$ under $\lambda$.

(ii) Given $f_1, \ldots, f_i$ increasing functions of $U_1, \ldots, U_i$, we construct $f_{i+1}$ as an increasing function of $U_1, \ldots, U_{i+1}$ such that $(f_1, \ldots, f_{i+1})$ under $\lambda$ has the same distribution as $(U_1, \ldots, U_{i+1})$ under $\mu_\rho$.

(iii) With $B = f^{-1}(A)$, $I_A(1) = I_B(1)$, and $I_A(i) \geq I_B(i)$ for all $i \in I$.

We first choose a strictly increasing function $f_1 : E \to [0, 1]$ such that $f_1$ only depends on the first coordinate. For convenience we will also write $f_1$ as a function of one variable. We need,

$$\lambda(f_1(U_1) \leq x(1)) = f_1^{-1}(x(1)) = \mu_\rho(U_1 \leq x(1)).$$
Let $\phi : [0, 1] \rightarrow [0, 1]$ be the continuous, increasing bijection given by

$\phi(x(1)) = \mu_\rho(U_1 \leq x(1)) = \int_0^{x(1)} Z(t) \, dt,$

$Z(t) = \int_{[0,1]\setminus\{1\}} \rho(t, y) \, dy.$

Take $f_1 = \phi^{-1}$. For an integrable function $h : [0, 1] \rightarrow \mathbb{R}$, by change of variable,

$\lambda(h(f_1(U_1))) = \int_0^1 h(\phi^{-1}(u(1))) \, du(1) = \int_0^1 h(x(1)) Z(x(1)) \, dx(1).$

For every $x$, the density $\rho$ conditional on the first coordinate is

$\rho_{\{1\}}(y) = \frac{\rho(x(1), y)}{Z(x(1))}.$

The integrability of $\rho_{\{1\}}$ follows by Tonelli’s theorem. For any increasing event $A$, by the change of variable formula,

$\mu_\rho(A) = \int_0^1 \int_{[0,1]\setminus\{1\}} \rho(x(1), y) 1_A(x(1), y) \, dy \, dx(1)$

$= \int_0^1 \int_{[0,1]\setminus\{1\}} \rho_{\{1\}}(y) 1_A(x(1), y) \, dy \, Z(x(1)) \, dx(1)$

$= \int_0^1 \int_{[0,1]\setminus\{1\}} \rho_{\{1\}}^{f(x)}(y) 1_A(f_1(u(1)), y) \, dy \, du(1).$

We have ‘simplified’ sampling from the $N$ dimensional density $\rho$. Take the first coordinate to have distribution $\lambda(h(f_1(U_1)))$ and then sample from the $N - 1$ dimensional density $\rho_{\{1\}}^{f(x)}$. Repeat this process a further $N - 1$ times for coordinates $2, 3, \ldots, N$; $\rho_{\{1,\ldots,N\}}^{f(x)} \equiv 1$ and

$\mu_\rho(A) = \int_{[0,1]^N} 1_A(f(u)) \, du.$

Let $B = f^{-1}(A)$. As $f_1(x)$ only depends on $x(1)$, $I_A(1) = I_B(1)$.

Let $i > 1$. Construct $g$ as we constructed $f$ but, as in the proof of Theorem 3.7, modify the order in which that coordinates are ‘sampled’. Process the coordinates in the order $i, 1, \ldots, i - 1, i + 1, \ldots, N$ instead $1, \ldots, N$. We must show that $f(x) \leq g(x)$ when $x(i) = 1$, and similarly that $f(x) \geq g(x)$ when $x(i) = 0$. Given
3.10 Influence for continuous random variables

\( x(i) = 1, \ f_1, \ g_1 \) only depend on \( x(1) \). We must show that \( f_1(x(1)) \leq g_1(x(1)) \), the result then follows by an inductive argument. Let

\[
\phi^f(x) = \mu_\rho(U_1 \leq x), \quad \phi^g(x) = \mu_\rho(U_1 \leq x \mid U_i = 1). 
\]

By monotonicity \( \mu_\rho(U_1 \leq x \mid U_i = a) \) is decreasing in \( a \), so \( \phi^f(t) \geq \phi^g(t) \). As \( f_1 = (\phi^f)^{-1} \) and \( g_1 = (\phi^g)^{-1} \), the result follows. \( \square \)

**Proof of Lemma 3.12**. This is immediate if \( N = 1 \). Suppose inductively that increasing subsets of \([0, 1]^N\) are Lebesgue measurable. Let \( A \) be an increasing subset on \([0, 1]^N \times [0, 1] \). Define a function \( f : [0, 1]^N \to [0, 1] \cup \infty \) by

\[
f(x) = \inf\{ y \in [0, 1] : f(x, y) \in A \}.
\]

The set \( A \) is increasing, so \( f \) is decreasing. By the inductive hypothesis, the sets \( \{ x : f(x) < c \} \) are Lebesgue measurable; \( f \) is a measurable function and the graph \( G_f \) is null by Fubini’s theorem. Consider the subset of \([0, 1]^{N+1}\) consisting of the ‘area above the graph’ \( G_f \),

\[
\overline{A} = \{(x, y) \in [0, 1]^{N+1} : y \geq f(x)\}.
\]

\( \overline{A} \) is measurable by the construction of the Lebesgue integral. Now \( A \subseteq \overline{A} \) and \( \overline{A} \setminus A \subset G_f \) is null, so \( A \) is measurable. The result follows by induction. \( \square \)
Chapter 4

The mean-field zero-range process

4.1 Balls and boxes

In this chapter we consider a mean-field zero-range process (ZRP) motivated by a microcanonical ensemble. By considering the entropy of the system, we show that the empirical distribution rapidly converges—in the sense of Kullback–Leibler divergence—to a geometric distribution. The proof utilizes arguments of spectral gap and log Sobolev type.

Suppose there are a number of boxes, $N$, each containing $R$ indistinguishable balls. Iteratively, pick a source box and a sink box, both uniformly at random; if the source box is not empty then move a ball to the sink box. This is a Markov chain on the set

$$B_N = \{ b \in \mathbb{N}^N : b_1 + \cdots + b_N = NR \}.$$

Here $b_i$ represents the number of balls in the $i$-th box. With probability $N^{-1}$ the source box and the sink box are the same so no change occurs.

The transition probability between neighbouring elements of $B_N$ is $N^{-2}$. The Markov chain is reversible with respect to the uniform distribution on $B_N$. If we regard the boxes as distinguishable particles, and the balls as quanta of energy, the system is said to be a microcanonical ensemble with Maxwell–Boltzmann
4.1 Balls and boxes

One can also consider the balls to be particles; the process can be said to have Bose–Einstein statistics. Let $N_k$ be the number of boxes containing $k$ balls, $\sum N_k = N$. At equilibrium, the probability that $(N_0, N_1, \ldots) = (n_0, n_1, \ldots)$ is proportional to the number of ball configurations $b \in B_N$ compatible with $(n_0, n_1, \ldots)$. The number of ways of putting $b$ balls into $N$ boxes is $\binom{NR+N-1}{N-1}$. Fix $R$. Under the equilibrium measure $\pi$,

$$\pi(B_1 \geq k) = \frac{\binom{NR - k + N - 1}{N - 1}}{\binom{NR + N - 1}{N - 1}} = \left(\frac{R}{R + 1}\right)^k \left(1 + k^2 O\left(\frac{1}{N}\right)\right) \text{ as } N \to \infty.$$  

The law of $B_1$ converges to a geometric distribution with mean $R$. The geometric distribution can be thought of as a Gibbs measure with respect to a linear energy function, say $H : \mathbb{N} \to \mathbb{R}$ given by $H(n) = n$. Recall from Section 1.3 that Gibbs measures are maximum entropy distributions. The equilibrium law of $B_1$ converges to the maximum entropy distribution of $\mathbb{N}$ with mean $R$.

In Section 4.3 we give an example to motivate our use of entropy to study the mean-field ZRP. The Ehrenfest urn model is a simple example of a Markov chain that can be studied by looking at the entropy of the empirical distribution. The Ehrenfest model was proposed as a probabilistic model to help explain the deterministic nature of Boltzmann’s $H$ theory. In Section 4.5 we define an entropy for the mean-field ZRP and show how it is related to the equilibrium distribution. We also summarize our results for how the entropy evolves.

In Section 4.4 we discuss some general convergence techniques for reversible Markov chains. If $N_0$, the number of empty boxes, is ‘fixed’, any given box behaves like a biased random walk on $\mathbb{N}$. The equilibrium measure for the random walk is a geometric distribution. We study this random walk in detail in Section 4.7 using the techniques of Section 4.4. We will use these results to analyse the mean-field ZRP.

In Section 4.8 we adapt a martingale concentration result. We will prove bounds on the expected step-wise change of the mean-field ZRP entropy. The concentration result will be used to show that, with high probability, the entropy decreases as the bounds suggest.
Figure 4.1: A sample path: the empirical distribution for $N = 10^7$, $R = 20$ at intervals of $40N$ steps
In Section 4.9 we use coupling arguments to prove bounds on the empirical distribution of the mean-field ZRP. By coupling the process with a system of $N$ biased random walks on $\mathbb{N}$, we prove an upper bound on the tail of the empirical distribution. We show if a significant fraction of the boxes are empty, then after $O(R^2N\log N)$ steps, $\sum_{k \geq j} N_k$ decays exponentially in $j$. By a second coupling argument, we show that with high probability $N_0 \geq N(5R)^{-1}$ after $O(NR^2 \log(R + 1))$ steps.

In Sections 4.10 and 4.11 we use the results obtained from the biased random walk. We use a log Sobolev technique to prove that when the entropy is large, the entropy tends to halve over a period of $O(R^2N\log N)$. Once the entropy is $O(R^{-2})$, we use a spectral gap type technique to show that the entropy tends to halve over a period of $O(R^3N)$.

### 4.2 Notation

We make use of asymptotic notation: ‘order less than’, $O$; ‘order equal to’, $\Theta$; and ‘order greater than’, $\Omega$, to describe the behaviour of functions as $N \to \infty$. With $c_1, c_2, N_0$ positive constants,

if $\forall N \geq N_0$, $|f(N)| \leq c_2|g(N)|$ write $f(N) = O(g(N))$,

if $\forall N \geq N_0$, $c_1|g(N)| \leq |f(N)| \leq c_2|g(N)|$ write $f(N) = \Theta(g(N))$,

if $\forall N \geq N_0$, $c_1|g(N)| \leq |f(N)|$ write $f(N) = \Omega(g(N))$.

In Section 4.1 we made use of asymptotic notation with $R$ fixed and $N$ tending to infinity. We will want to express bounds that for all $R$, given $R$, hold for all $N$ sufficiently large. When we write

$$f(R, N) = O(g(R, N)) \text{ as } N \to \infty$$

we mean that there is a constant $c_2$ and a function $N_0$ such that,

$$\forall R, \forall N \geq N_0(R), \ |f(R, N)| \leq c_2|g(R, N)|.$$
The function log should be interpreted as the natural logarithm, unless indicated otherwise by a subscript; \( \log_a x = \log x / \log a \).

If \( X(0), X(1), X(2), \ldots \) is a sequence of real valued random variables, we will write \( \Delta X(i) \) for the step-wise change \( X(i + 1) - X(i) \).

For random variables \( X, Y \), we say that \( X \) is \textit{stochastically smaller} than \( Y \), written \( X \leq_{st} Y \), if there is a coupling \((X, Y)\) such that \( X \leq Y \) almost surely.

4.3 The Ehrenfest urn model

In Chapter [1] we introduced the lazy Ehrenfest urn model [20]. An urn contains \( N \) balls, red or green. Start with the urn full of green balls. Each step pick a ball uniformly at random, and with probability one half replace it with a ball of the other colour. Assuming the balls are ordered, the process is time reversible with respect to the uniform distribution on the set \( \{\text{red, green}\}^N \). Let \( N_R \) be the number of red balls, and let \( X_R = N_R/N \) be the fraction of red balls. After \( i \) steps, each ball has been picked binomial \( \text{Bin}(i, N^{-1}) \) times. Each ball is red with probability

\[
p_R(i) = (1 - [1 - N^{-1}]^i)/2.
\]

The Markov chain is of course just a random walk on the Hamming cube. The mixing properties have been studied in great detail [19, 41, 48]. The fraction of red balls is close to \( 1/2 \) after \( O(N \log N) \) steps.

An alternative approach to the problem is to take a hydrodynamic limit [43]. Condition on \( X_R = x \). The next step, \( X_R \) increases by \( N^{-1} \) with probability \((1 - x)/2 \) and decreases by \( N^{-1} \) with probability \( x/2 \); the expected change is \( \mathbb{E} (\Delta X_R) = N^{-1} (1/2 - x) \). The hydrodynamic limit \( X_R^f \) is the solution to the equation

\[
\frac{d}{dt} X_R^f = 1/2 - X_R^f, \quad X_R^f(0) = 0.
\]

In this simple case, we can easily find the hydrodynamic limit in closed form,

\[
X_R^f(t) = (1 - e^{-t})/2.
\]
The hydrodynamic limit can also be studied by looking at the behaviour of a certain statistic. Let $X^f_R = 1 - X^f_G$ be fraction of green balls, and define entropy

$$S(X^f_R, Xf_G) = -X^f_R \log X^f_R - X^f_G \log X^f_G.$$ 

Entropy is initially 0; maximum entropy $S_{\text{max}} = \log 2$ is obtained in the limit $t \to \infty$ when $X^f_R = X^f_G = 1/2$. Again, we use the function $\delta$ introduced in Section 1.3

$$\delta(x, y) = y \log y - x \log x - (y - x) \left[ \frac{d}{dt} t \log t \big|_{t=x} \right] = y \log \frac{y}{x} - (y - x).$$

The function is sketched for arbitrary $x$ in Figure 4.3. If $x = y$, $\delta(x, y) = 0$. By the convexity of $t \mapsto t \log t$, $\delta(x, y) \geq 0$ for all $x, y$. The function is bounded between $(x - y)^2/(3x)$ and $(x - y)^2/x$ when $0 \leq y \leq 2x$. The upper bound is also
valid for all $y$. We can easily check that when $X^f_R = x$
\[
\frac{d}{dt} S = (1/2 - x) \log \frac{1-x}{x} \geq S_{\text{max}} - S.
\]
Without having to solve the hydrodynamic limit differential equation, we can see that $S_{\text{max}} - S$ decays exponentially,
\[
\frac{d}{dt} (S_{\text{max}} - S) \leq -(S_{\text{max}} - S).
\]
We can also perform an analogous calculation before taking the hydrodynamic limit,
\[
\mathbb{E}(\Delta S \mid X_R = k/N) \geq \frac{S_{\text{max}} - S}{N} + O(N^{-2}).
\]
We will use this approach to analyse the mean-field zero-range process. There are two additional complications over the continuous case. Firstly, we have an ‘error’ term of order $N^{-2}$. Secondly, the Markov chain is ergodic; over large time scales we cannot simply say that $S$ is close to $S_{\text{max}}$. 

Figure 4.3: Graph of $\delta(x,y)$ as a function of $y$
4.4 Markov chain convergence

Let $\Omega$ be a finite set. Let $K : \Omega \times \Omega \to [0,1]$ be a Markov kernel: for all $x$, $\sum_y K(x,y) = 1$. If $\mu$ is a distribution on $\Omega$, $\mu K^i$ gives the distribution after $i$ steps. We can also run the Markov chain in continuous time at rate 1, $K - I$ is the corresponding Q-matrix. We say that the Markov chain is lazy if $K(x,x) \geq 1/2$ for all $x \in \Omega$. We will assume that $K$ is irreducible, lazy and time reversible with respect to equilibrium probability measure $\pi$; for all $x, y$, $\pi(x) K(x,y) = \pi(y) K(y,x)$.

It is then standard that the left eigenvectors of $K$, suitably normalized, form a basis for $\mathbb{R}^\Omega$ orthonormal with respect to the inner product

$$\langle f, g \rangle_\pi = \langle f, \pi^{-1} g \rangle_{\mathbb{R}^\Omega} = \sum_{x \in \Omega} f(x)g(x)/\pi(x).$$

Label the eigenvectors $F_1, \ldots, F_n$, with eigenvalues $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq 0$ respectively. The eigenvalues are all non-negative as the Markov Chain is lazy. The first eigenvector $F_1 = \pi$, the stationary distribution.

For a probability measure $\mu$ on $\Omega$,

$$\mu(x) = \pi(x) + \sum_{i \geq 2} \alpha_i F_i(x), \quad \alpha_i = \langle \mu, F_i \rangle_\pi.$$

The $\chi^2$-distance from $\mu$ to $\pi$, $\|\mu/\pi - 1\|_2^2$, can be written

$$\sum_{x \in \Omega} \frac{(\mu(x) - \pi(x))^2}{\pi(x)} = \langle \mu - \pi, \mu - \pi \rangle_\pi = \sum_{i \geq 2} \alpha_i^2.$$

The quantity $1 - \lambda_2$ is called the spectral gap. More generally, if the Markov chain is not lazy, the spectral gap is $1 - \max_{i \geq 2} |\lambda_i|$. Let $\mu_t = \mu e^{-t(I-K)}$. The eigenvalue of $F_i$ with respect to $e^{-t(I-K)}$ is $e^{-t(1-\lambda_i)}$ so,

$$\frac{d}{dt} \sum_x \frac{(\mu_t(x) - \pi(x))^2}{\pi(x)} = -\sum_{x,y} \pi(x)K(x,y) \left[ \frac{\mu_t(x)}{\pi(x)} - \frac{\mu_t(y)}{\pi(y)} \right]^2 \leq -2(1 - \lambda_2) \sum_x \frac{(\mu_t(x) - \pi(x))^2}{\pi(x)}.$$ (4.1)
4.4 Markov chain convergence

Define the total variation distance between $\mu$ and $\pi$,

$$\|\mu - \pi\|_{TV} = \sup_{A \subset \Omega} |\mu(A) - \pi(A)| = \frac{1}{2} \sum_x |\mu(x) - \pi(x)|.$$ 

Write $\tau_1(\varepsilon)$ for the time to convergence within $\varepsilon$ in total variation,

$$\tau_1(\varepsilon) = \inf \{ t : \forall \mu, \|\mu_t - \pi\|_{TV} \leq \varepsilon \}.$$ 

Inequality (4.1) can be used to show the following bound [48],

$$\tau_1(\varepsilon) \leq \frac{\log(\varepsilon \pi_{\text{min}})}{\log \lambda_2}, \quad \pi_{\text{min}} = \min_x \pi(x). \quad (4.2)$$

We will also use a log Sobolev technique. Log Sobolev inequalities for discrete Markov chains are described in [19]. Define the Dirichlet form, and the Laplacian for $f, g : \Omega \rightarrow \mathbb{R}$,

$$\mathcal{E}(f, g) = \frac{1}{2} \sum_{x,y} \pi(x) K(x,y) [f(x) - f(y)][g(x) - g(y)],$$

and

$$\mathcal{L}(f) = \sum_x |f(x)|^2 \log \left( \frac{|f(x)|^2}{\|f\|_{2,\pi}^2} \right) \pi(x).$$

The log Sobolev constant is defined,

$$\nu = \min_{f} \left\{ \frac{\mathcal{E}(f, f)}{\mathcal{L}(f)} : \mathcal{L}(f) \neq 0 \right\}.$$ 

This is in a superficial way similar to the spectral gap; we can write the spectral gap in term of $\mathcal{E}$,

$$1 - \lambda_2 = \min_{f} \left\{ \frac{\mathcal{E}(f, f)}{\text{var}_\pi(f)} : \text{var}_\pi(f) \neq 0 \right\}.$$ 

The entropy of $\mu$ relative to $\pi$, also called the Kullback–Leibler divergence from $\mu$ to $\pi$, is defined by

$$D_{KL}(\mu || \pi) = \sum_x \mu(x) \log \frac{\mu(x)}{\pi(x)} = \sum_x \delta(\pi(x), \mu(x)) \geq 0.$$ 

It is not generally the case that $D_{KL}(\mu || \pi) = D_{KL}(\pi || \mu)$. The log Sobolev constant can be used to show convergence in Kullback–Leibler divergence under the action of $e^{-t(I-K)}$,

$$\frac{d}{dt} D_{KL}(\mu_t || \pi) = -\mathcal{E}\left( \frac{\mu_t}{\pi}, \log \frac{\mu_t}{\pi} \right) \leq -4\nu D_{KL}(\mu_t || \pi). \quad (4.3)$$

76
Corollary A.4 of [19] gives a general bound for the log Sobolev constant in terms of the spectral gap and $\pi_{\text{min}}$,
\[
\nu \geq \frac{(1-2\pi_{\text{min}})(1-\lambda_2)}{\log(1/\pi_{\text{min}} - 1)}.
\]

### 4.5 Entropy for the mean-field ZRP

With $X_k = N_k/N$ the fraction of boxes with $k$ balls, the empirical distribution is the random variable $X = (X_0, X_1, \ldots)$. Let $D$ be the set of distributions with mean $R$; $x \in [0, 1]^N$ such that $\sum x_k = 1$ and $\sum kx_k = R$. Define the entropy function $S : D \to [0, (R+1) \log(R+1) - R \log R]$ by
\[
S(x) = -\sum_{k=0}^{\infty} x_k \log x_k, \quad 0 \log 0 = 0.
\]

Up to a factor of $\log 2$ this is the standard information-theoretic entropy. It is related to the thermodynamic entropy $\log N!/ (N_0!N_1! \ldots)$. The multinomial coefficient $N!/ (N_0!N_1! \ldots)$ counts the permutations of $B_1, \ldots, B_N$; under the equilibrium measure $\pi$,
\[
\pi(X = x) = |B_N|^{-1} \frac{N!}{n_0!n_1! \ldots}, \quad x_k = n_k/N.
\]

The entropy and thermodynamic entropy are related. Provided each box contains $O(R \log N)$ balls,
\[
\left| S(X) - \frac{1}{N} \log \frac{N!}{N_0!N_1! \ldots} \right| = O(RN^{-1} \log^2 N).
\]

This is simply by Stirling’s approximation [11],
\[
\log \left( \frac{n!}{(n/e)^n} \right) = O(\log n).
\]

We can estimate the expected change in entropy in terms of the expected changes $\mathbb{E}(\Delta X_k \mid X = x)$. This calculation is meant only to motivate two rather
lengthy calculations, the proofs of Lemmas 4.16 and 4.20.

\[ \mathbb{E}(\Delta X_k \mid X = x) \approx \frac{1}{N} \sum_{k=0}^{\infty} \frac{(x_k(1-x_0) - x_{k+1})}{k > 0} \]

\[ \mathbb{E}(\Delta S \mid X = x) \approx -\sum_{k=0}^{\infty} \frac{d}{dt}(t \log t) \bigg|_{t=x_k} \mathbb{E}(\Delta X_k \mid X = x) \]

\[ \approx \frac{1}{N} \sum_{k=0}^{\infty} \frac{(x_k(1-x_0) - x_{k+1}) \log x_k(1-x_0)}{x_{k+1}} \]

\[ \geq \frac{1}{N} \sum_{k=0}^{\infty} \frac{(x_k(1-x_0) - x_{k+1})^2}{\max\{x_k(1-x_0), x_{k+1}\}} \geq 0 \]

It is interesting to consider the hydrodynamic limit for this process. It is the solution \( X^f : [0, \infty) \rightarrow [0, 1]^N \) to the differential equation,

\[ \frac{d}{dt} X^f_k = \begin{cases} - (X^f_k(1-X^f_0) - X^f_{k+1}) & k = 0, \\ (X^f_{k-1}(1-X^f_0) - X^f_k) - (X^f_k(1-X^f_0) - X^f_{k+1}) & k > 0, \end{cases} \]

with initial condition,

\[ X^f_k(0) = \begin{cases} 1 & k = R, \\ 0 & \text{otherwise}. \end{cases} \]

The solution \( X^f(t) \) is also the marginal probability distribution of a certain continuous time random walk on \( \mathbb{N} \). Let the walk start \( C(0) = R \). At time \( t \), if \( C(t) > 0 \) step left (decrement \( C \) by 1) at rate 1; step right (increment \( C \) by 1) at rate \( 1 - P(C(t) = 0) \). Then \( P(C(t) = k) = X^f_k(t) \).

The distribution \( x \) with maximum entropy for mean \( R \) is the geometric distribution \( x_k = R^k/(R+1)^{k+1} \). Call this distribution \( G^R \). Define \( S \) to be the gap between the maximum entropy and the entropy,

\[ \mathcal{S}(x) = S(G^R) - S(x). \]

Note that \( \Delta S(X) = -\Delta \mathcal{S}(X) \), and we expect \( \mathcal{S}(X) \) to decrease to almost zero. We can now hope for inequalities of the form

\[ \mathbb{E}(\Delta \mathcal{S} \mid X = x) \leq -c \mathcal{S}(x), \quad c \text{ constant}. \]

We also have the following.
4.5 Entropy for the mean-field ZRP

Lemma 4.4. $\overline{S}(x)$ is the relative entropy of $x$ with respect to $G^R$, that is the Kullback–Leibler divergence from $x$ to $G^R$. Also, $\overline{S}(x)$ is less than Kullback–Leibler divergence from $x$ to the geometric distribution with first term $x_0$ and mean $(1 - x_0)/x_0$.

\[
\overline{S}(x) = \sum x_k \log \frac{x_k}{R^k/(R + 1)^{k+1}} = \sum \delta(R^k/(R + 1)^{k+1}, x_k)
\]

\[
\leq \sum x_k \log \frac{x_k}{x_0(1 - x_0)^k} = \sum \delta(x_0(1 - x_0)^k, x_k).
\]

The main results of this chapter are in Sections 4.9, 4.10 and 4.11. They can be summarized as follows. Consider the evolution of the Markov chain up to $N^{21/20}$ steps. With probability $1 - O(N^{-1/3})$ the following occurs.

(i) After time $O(NR^2 \log(R + 1))$, $X_0 \geq (5R)^{-1}$. As a consequence, we get a bound on the tail of the empirical distribution. After $O(R^2N \log N)$ steps, $\sum_{k\geq j} X_k \leq 8(1 - a)^j$ for $j$ such that $(1 - a)^j \geq 2N^{-1} \log N$.

(ii) From then on,

\[
E(\Delta \overline{S} \mid X = x) \leq -\frac{\overline{S}(x)}{500R^2N \log N} + \frac{250R \log N}{N^2}.
\]

After time $O(R^2N \log(R + 1) \log N)$ steps, $\overline{S} \leq \phi := 10^{-4}(R + 1)^{-2}$. However, when $\overline{S}$ is small, this bound on $E(\Delta \overline{S} \mid X = x)$ is rather weak. It does not imply that $\overline{S} = O(R^3N^{-1} \log^2 N)$ until $\Omega(R^2N \log^2 N)$ steps.

(iii) The bound on $E(\Delta \overline{S} \mid X = x)$ improves to,

\[
E(\Delta \overline{S} \mid X = x) \leq -\frac{\overline{S}(x)}{4000(R + 1)^3N} + \frac{100(R + 1) \log N}{N^2}.
\]

After $O(R^3N \log N)$ steps, $\overline{S} = O(N^{-1} \log^2 N)$.

In addition, we can show that $\overline{S}$ tends to decrease to $O(R^4N^{-1} \log N)$.

In order to prove the above, we first analyse a biased random walk on $\mathbb{N}$. We use the equilibrium distribution and mixing time in Section 4.9 to obtain the bound on the tail of the empirical distribution. In Section 4.10 we use the Markov chain’s log Sobolev constant. In Section 4.11 we use an inequality related to the spectral gap.
4.6 Total variation convergence

Let $K : \mathcal{B}_N \times \mathcal{B}_N \rightarrow [0,1]$ be the Markov kernel for the mean-field ZRP. A natural quantity of interest is the rate at which the total variation distance from equilibrium,

$$
\|\mu K^i - \pi\|_{TV} = \sup_{A \subset \mathcal{B}_N} |\mu K^i(A) - \pi(A)|,
$$

decays. The time to convergence within $\varepsilon$ in total variation is defined,

$$
\tau_1(\varepsilon) = \inf\{i : \forall \mu, \|\mu K^i - \pi\|_{TV} \leq \varepsilon\}.
$$

Notice that $\tau_1(\varepsilon)$ is a maximum over all starting distributions $\mu$. It is standard to call $\tau_1(1/4)$ the total variation mixing time; it can be used to control the Markov chain convergence to an arbitrary degree of accuracy,

$$
\tau_1(2^{-k}) \leq k\tau_1(1/4), \quad k = 2, 3, \ldots.
$$

The results for the entropy described in Section 4.5 assume that the Markov chain starts from a particular state, with $R$ balls in every box. Under the equilibrium measure, the maximum number of balls per box is $O(R \log N)$ with high probability. The results immediately generalize to allow any starting distribution such that every box contains $O(R \log N)$ balls. However, if we start with all $RN$ balls in one box, the process clearly does not reach equilibrium before time $RN^2/2$—one box will still contain half the balls. Trying to calculate the total variation mixing time would clearly be the wrong approach.

One could consider a truncated version of the mean-field ZRP. The mean-field ZRP is defined on the set $\mathcal{B}_N$; $\mathcal{B}_N$ can be thought of as a graph, with edges between the elements of $\mathcal{B}_N$ that differ by the location of one ball. The mean-field ZRP is then the random walk on $\mathcal{B}_N$ with uniform transition probability $N^{-2}$: each step, move to each of the adjacent states with probability $N^{-2}$. The equilibrium distribution is uniform.

The process can be truncated as follows. Remove certain states, and any incident edges, from the graph $\mathcal{B}_N$. Then consider the random walk on the new graph with uniform transition probability $N^{-2}$. The equilibrium distribution is again uniform. Provided the fraction of the vertices of $\mathcal{B}_N$ that have been removed is small, the change to the equilibrium measure is small.
One such truncation is to remove all the ball configuration where any box holds more than $O(R \log N)$ balls. Rough calculations suggest that under the equilibrium measure, $\mathcal{S}(X)$ has order at least $RN^{-1} \log N$ with high probability. Given that we show $\mathcal{S}(X) = O(R^4 N^{-1} \log N)$ with high probability after $O(R^3 N \log N)$ steps, it is tempting to conjecture that for the restricted Markov chain, $\tau_1(1/4) = O(R^3 N^{-1} \log N)$.

Our results deal with the process when it is ‘far’ from equilibrium. Proving total variation convergence might well require a rather different technique. The empirical distribution is a Markov chain on

$$D' = \left\{ x \in \{0, N^{-1}, 2N^{-1}, \ldots, 1\}^N : \sum_k x_k = 1, \sum_k kx_k = R \right\}.$$ 

The Markov chain on $D'$ is similar to the one analysed by Frieze, Kannan and Polson [29]. They looked at the problem of efficiently sampling from a log-concave distribution on a convex subset of Euclidean space. They considered a reversible Markov chain on a convex subset of the hypercubic lattice, chosen so that the equilibrium distribution approximates the log-concave distribution. The mixing time in bounded in terms of the dimension and the diameter of the state space.

Our results do seem to be a step in the right direction. We know that the empirical distribution is soon effectively confined to a subset of $D'$,

$$D'' = \left\{ x \in D' : \mathcal{S}(x) = O(N^{-1} \log^2 N) \text{ and } x_k = 0 \text{ for } k = \Omega(R \log N) \right\}.$$ 

It would therefore be interesting to know the total variation mixing time for the process when truncated such that the support of empirical distribution is $D''$. The diameter of $D''$ is much smaller than the diameter of $D'$.

### 4.7 A biased random walk

To study the evolution of a single box of the mean-field ZRP, you must know something about the state of the rest of the system. It is sufficient to know $X_0$, the fraction of empty boxes. Consider the step from time $i$ to $i+1$. If $B_1(i) > 0$, box one loses a ball with probability approximately $1/N$. Box one gains a ball with probability approximately $(1 - X_0(i))/N$. We consider the effect of ‘fixing’
4.7 A biased random walk

Let \( X_0 = a \). Over a short period of time \( X_0 \) does not change much. The empirical distribution evolves almost as if each box is an independent copy of the following biased random walk when ‘\( n \)’ is sufficiently large.

Let \( a \in (0, 1) \) and let \( n \) be a positive integer. Define a discrete Markov chain \( C \) on state space \( \Omega = \{0, \ldots, n - 1\} \). Each step,

(i) decrease \( C \) by one with probability \( 1/N \) (unless \( C = 0 \)),

(ii) increase \( C \) by one with probability \( (1 - a)/N \) (unless \( C = n - 1 \)).

That is, \( C \) has Markov kernel \( K : \Omega \times \Omega \to [0, 1] \),

\[
K(j, k) = \begin{cases} 
\frac{(1 - a)}{N} & k = j + 1, \\
\frac{1}{N} & k = j - 1, \\
0 & |k - j| > 1,
\end{cases}
\]

and \( \sum_k K(j, k) = 1 \). The Markov chain has stationary distribution \( \pi(k) = \hat{a}(1 - a)^k \), with \( \hat{a} = a/(1 - (1 - a)^n) \). Of course, \( \pi(j)K(j, k) = \pi(k)K(k, j) \); the Markov chain is reversible. Unless \( N \leq 3 \), the Markov chain is lazy.

**Lemma 4.5.**

(i) Suppose \( C(0) < n/2 \). The probability that \( C \) hits the right boundary, \( n - 1 \), within \( N^{21/20} \) steps is \( O(N^{21/20}(1 - a)^{n/2}) \).

(ii) The spectral gap \( 1 - \lambda_2 \geq a^2/(4N) \). The Markov chain \( C \) mixes, in total variation to within \( \varepsilon \) of \( \pi \), in time \( \tau_1(\varepsilon) \leq 4Na^{-2}\log(\varepsilon\pi(n - 1)) \).

(iii) Let \( \mu \) be the law of \( C(0) \). Choose \( j \leq l \) such that \( \mu(C > l) \leq \pi(j)/2 \). Then \( (\mu K^i)(k) \leq 2\pi(k) \) for \( k \leq j \) when \( i \geq 4Na^{-2}\log 2/\pi(l) \).

Arguably it would be more natural to take \( n = \infty \) and consider a random walk on \( \mathbb{N} \). However, with \( n \) finite, the analysis is much simpler. Fortunately, provided \( a = \Omega(R^{-1}) \), we can take \( n \) large enough that \( C \) is very unlikely to reach \( n - 1 \) over the first \( N^{21/20} \) steps, while still having \( C \) mix in time \( O(R^2N\log N) \).

We can also run the Markov chain in continuous time. The inequalities of Section 4.4 can be simplified by \( \pi(k + 1) = (1 - a)\pi(k) \) and

\[
N \frac{d}{dt} \mu_t(k) \big|_{t=0} = \mathbb{1}_{(k > 0)}[\mu(k - 1)(1 - a) - \mu(k)] \\
- \mathbb{1}_{(k < n-1)}[\mu(k)(1 - a) - \mu(k + 1)].
\]
The spectral gap inequality (4.1) becomes,

\[ \frac{1}{N} \sum_{k=0}^{n-2} \frac{(\mu(k)(1-a) - \mu(k+1))^2}{\pi(k+1)} \geq (1 - \lambda_2) \frac{1}{n} \sum_{k=0}^{n-1} \frac{(\mu(k) - \pi(k))^2}{\pi(k)}. \]

The left and right hand sides of the above inequality are very sensitive to changes in the tail of \( \mu \). Obviously it is not generally possible to just ignore these terms. However, we can do exactly that if \( \mu(0) = a \). We will state this in terms of the empirical distribution, so take \( a = X_0 \).

Lemma 4.6. For \( j = 1, 2, \ldots \), the empirical distribution satisfies

\[ \sum_{k=0}^{j} \frac{(X_k(1-X_0) - X_{k+1})^2}{X_0(1-X_0)^{k+1}} \geq \frac{X_0^2}{4} \sum_{k=1}^{j+1} \frac{(X_k - X_0(1-X_0)^k)^2}{X_0(1-X_0)^k}. \]

Inequality (4.3), containing the log Sobolev constant, is equivalent to

\[ \frac{1}{N} \sum_{k=0}^{n-2} (\mu(k)(1-a) - \mu(k+1)) \log \frac{\mu(k)(1-a)}{\mu(k+1)} \geq 4\nu \sum_{k=0}^{n-1} \mu(k) \log \frac{\mu(k)}{\pi(k)}. \]

We can adapt this to apply to the empirical distribution. As we may have \( X_k = 0 \), set \( X'_k = X_k \vee (e^{-1}N^{-1}) \) to avoid dividing by zero. Assume \( N \geq n \geq 10 \).

Lemma 4.7. If \( X_0 > 0 \) and \( X_k = 0 \) for \( k \geq n - 1 \),

\[ \frac{1}{N} \sum_{k=0}^{n-2} \frac{(X'_k(1-X_0) - X'_{k+1}) \log \frac{X'_k(1-X_0)}{X'_{k+1}}}{X_0(1-X_0)^k} \geq \frac{X_k}{X_0(1-X_0)^k} - 4\nu \frac{n \log N}{N}. \]

For fixed \( a \), the log Sobolev constant

\[ \nu \geq \frac{(1 - 2\pi(n-1))(1 - \lambda_2)}{\log(1/\pi(n-1) - 1)} = \Omega(N^{-1}n^{-1}) \]

by Corollary A.4 of [19]. It is clear that the lower bound \( \nu = \Omega(N^{-1}n^{-1}) \) cannot be greatly improved. Consider a probability distribution that places its mass far away from zero, and diffusely. Then the rate of change of the relative entropy is approximately \(-a^{-1}N^{-1} \log(1-a)\), due to the downward bias of the random walk, while the entropy can have order \(-n \log(1-a)\).
4.8 Concentration results

We will use a concentration result, Theorem 3.15 from [46], strengthened as mentioned in their Section 3.5.

**Theorem 4.8.** Let \( X = (0 = X(0), X(1), \ldots, X(m)) \) be a martingale with respect to filtration \( \{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \cdots \subset \mathcal{F}_m \). Let \( b \) be the maximum step size, \( |\Delta X(i)| \leq b \). Let \( \sigma_i^2 \) be the conditional variance of \( \Delta X(i) \) given \( \mathcal{F}_i \), and let \( \hat{v} \) be an upper bound on \( \sum_{i=0}^{m-1} \sigma_i^2 \). For any \( \lambda \geq 0 \),

\[
\mathbb{P} \left( \max_{i \leq m} X(i) \geq \lambda \right) \leq \exp \left( \frac{-\lambda^2}{2\hat{v} + 2b\lambda/3} \right).
\]

The above result can be applied to the following class of processes by constructing a suitable martingale. Let \( (W(i) : 0 \leq i \leq N^{21/20}) \) be a process adapted to filtration \( (\mathcal{F}_i) \) with \( \left| \Delta W(i) - \mathbb{E}(\Delta W(i) \mid \mathcal{F}_i) \right| \leq b \). Suppose that with probability \( 1 - q \) from time 0 to \( N^{21/20} \),

(a) \( W(i) \in [0, \alpha] \), and

(b) while \( W(i) \in [\alpha/8, \alpha] \), \( \mathbb{E}(\Delta W(i) \mid \mathcal{F}_i) \leq -\gamma \).

Our corollary will be applied to \( \overline{S} \) once we have bounded the expected change \( \mathbb{E}(\Delta S \mid X = x) \). The quantity \( \hat{v} \) is defined in the proof. It is an upper bound on sum of the step-wise variances of the constructed martingale \( X \).

**Corollary 4.9.** If

\[
\frac{\alpha}{16} \geq 2b \log N + \sqrt{6\hat{v} \log N}
\]

then \( \mathbb{P}(\forall i \geq 9\alpha/(16\gamma), \ W(i) \leq \alpha/2) \geq 1 - q - N^{-1} \).

**Proof.** Assume \( W(0) \geq 3\alpha/8 \). Let \( M \) be the stopping time triggered by any of the following:

(i) \( W(i) \leq \alpha/8 \),

(ii) \( \sum_{j=0}^{i-1} \mathbb{E}(\Delta W(j) \mid \mathcal{F}_j) \leq 3\alpha/16 - W(0) \),

(iii) \( W(i) - W(0) - \sum_{j=0}^{i-1} \mathbb{E}(\Delta W(j) \mid \mathcal{F}_j) \geq \alpha/16 \),
4.9 Coupling results

(iv) \( \mathbb{E}(\Delta W(i) \mid \mathcal{F}_i) > -\gamma \), or

(v) \( W(i) \geq \alpha \).

\( M \) is less than \( m := \lceil 13\alpha/(16\gamma) \rceil \). Let \( W^M \) denote \( W \) stopped at \( M \); define a martingale \( X \) by \( X(0) = 0 \) and \( \Delta X(i) = \Delta W^M(i) - \mathbb{E}(\Delta W^M(i) \mid \mathcal{F}_i) \). Let \( \hat{v} \) be an upper bound on \( \sum_{i=0}^{m-1} \text{var}(\Delta X(i) \mid \mathcal{F}_i) \). Let \( c \in (0,1) \) and suppose

\[
\mathbb{P} \left( \sup_i X^M(i) \geq \alpha/16 \right) \leq c. \tag{4.11}
\]

Then with probability \( 1 - q - c \), \( W(i) \) first leaves the interval \([\alpha/4, \alpha]\) before time \( 13\alpha/(16\gamma) \), having decreased to less than \( \alpha/4 \). In particular, \( W(i) \leq \alpha/2 \) after \( 9\alpha/16 \) steps. Let \( c = N^{-3} \), so (4.10) implies (4.11) by Theorem 4.8. If the process stops because of (i) or (ii), or if initially \( W \leq 3\alpha/16 \), wait for \( W \geq 3\alpha/8 \). Then start the construction of \( X \) over again. The number of times that the process is restarted is less than \( N^{21/20} \). Therefore, the probability that \( W \geq \alpha/2 \) at any time between \( 9\alpha/(16\gamma) \) and \( N^{21/20} \) is less than \( q + N^{21/20}c \). \( \square \)

We have two other uses for Theorem 4.8. If \( X \sim \text{Bin}(n,p) \),

\[
\mathbb{P}(\lvert X - np \rvert \geq \lambda) \leq 2 \exp \left( \frac{-\lambda^2}{2np + 2\lambda/3} \right).
\]

Firstly, if \( \mathbb{E}(X) \geq 8 \log N \) then \( \mathbb{P}(X \geq 2\mathbb{E}(X)) \leq 2N^{-3} \). We use this to bound the tail of the empirical distribution. Secondly, if \( X \) is a Poisson random variable with parameter \( \mu \leq N^{1/4} \),

\[
\mathbb{P}(\lvert X - \mu \rvert \geq 3N^{1/8} \sqrt{\log N}) \leq 2N^{-3}.
\]

We use this to show that the coupling defined in the proof of Theorem 4.12 has the properties claimed. Over the next three sections we give the main results.

### 4.9 Coupling results

We have described, and stated mixing results for, a biased random walk. The random walk was motivated by the evolution of a single box. We will now give coupling results that make the relationship with the mean-field ZRP more precise.
4.9 Coupling results

We first show that if \( X_0 \), the fraction of empty boxes, is bounded below then we can couple the individual boxes to independently evolving copies of the biased random walk. Choose \( n = n(a) \) maximal such that \( \pi(n - 1) \geq 2N^{-10} \). This is an assumed upper bound on the number of balls any box will contain. Let \( A_{a,s} \) be the event

(i) from time \( s \) until time \( N^{21/20} \), \( X_0 \geq a + 6N^{-1/8}\sqrt{\log N} \), and
(ii) at time \( s \), \( X_k = 0 \) for all \( k \geq n/2 \).

Starting from time \( s \), define a process \( C = (C_i)_{i=1}^N \). Let \( C_i(s) = B_i(s) \) for \( i = 1, \ldots, N \). Then let the \( C_i \) evolve independently, each according to the continuous time biased random walk in Section 4.7 on \( \{0, 1, \ldots, n - 1\} \) with bias \( a \). We then want to find a coupling of \( B \) and \( C \) such that the following event has high probability: for all integer \( t \) from \( s \) to \( N^{21/20} \), there exists \( t' \in [t - N, t + N] \) such that,

(i) \( B_i(t) \leq C_i(t') + 2 \) for all \( i \), and
(ii) \( C_i(t') < n - 2 \) for all \( i \).

Call this event \( A'_{a,s} \).

**Theorem 4.12.** There exists a coupling \( (B, C) \) with \( \mathbb{P}(A'_{a,s}) \geq \mathbb{P}(A_{a,s}) - O(N^{-2/5}) \).

Recall that by Lemma 4.5 (ii), the \( C_i \) mix to within \( N^{-10} \) in total variation in time \( \tau_1(N^{-10}) \). For convenience write

\[
\tau(a) = \tau_1(N^{-10}) + N = O(a^{-2}N \log N).
\]

After discrete time \( s + \tau(a) \), we can assume the \( B_i \) are bounded by \( C_i \) close to equilibrium. Let \( C_{a,s} \) be the event that from time \( s + \tau(a) \) to \( N^{21/20} \),

(i) \( \sum_{k \geq j} X_k \leq 8(1 - a)^j \) for all \( j \) with \( (1 - a)^j \geq 2N^{-1} \log N \), and
(ii) \( \sum_{k \geq j} X_k = 0 \) for \( j \) such that \( (1 - a)^j \leq N^{-5} \).

By the concentration calculations at the end of Section 4.8, \( \mathbb{P}(A'_{a,s} \setminus C_{a,s}) = O(N^{-2/5}) \), so \( \mathbb{P}(C_{a,s}) \geq \mathbb{P}(A_{a,s}) - O(N^{-2/5}) \).

Before we can make use of Theorem 4.12 we need to know that \( A_{a,s} \) occurs with high probability for some \( a, s \). This follows from another coupling argument, this time involving an unbiased random walk on \( \{0, 1, \ldots, 2R - 1\} \).
4.9 Coupling results

**Theorem 4.13.** Let $a = (5R)^{-1}$ and $s = 10NR^2 \log(R + 1)$. Then $\mathbb{P}(A_{a,s}) \geq 1 - O(N^{-2/5})$.

With $a \geq (5R)^{-1}$, the mixing time $\tau_1(N^{-10}) \leq 2000R^2N \log N$.

Under the event $C_{a,s}$ we can perform some calculations that we will need later. So far we only have a lower bound on $X_0$. We can calculate an upper bound on $X_0$ using our upper bound on the tail of the empirical distribution and the identities $\sum X_k = 1$, $\sum kX_k = R$. We will also need a bound on the contribution of the tail of the distribution $\sum_{k \geq j} \delta(X_0(1 - X_0)^k, X_k)$ to the upper bound $\sum \delta(X_0(1 - X_0)^k, X_k) \geq \mathbb{S}(X)$.

**Lemma 4.14.** Let $a \in [(5R)^{-1}, (R + 1)^{-1}]$. Under $C_{a,s}$ for integer $t$ from $s + \tau(a)$ to $N^{21/20}$:

(i) The fraction of empty boxes $X_0 \leq 0.99$ if $N$ is sufficiently large.

(ii) For all $j$,

$$\sum_{k \geq j} \delta(X_0(1 - X_0)^k, X_k) \leq 8(1 - a)^j \left[ \log 40R + (j + 5R) \log \frac{1 - a}{1 - X_0} \right] + \frac{80 \log^2 N}{N}.$$ 

**Proof of Theorem 4.12.** Define an alphabet of $2N$ letters, $U_i$ and $D_i$ for $i = 1, \ldots, N$. These should be understood to mean ‘move coordinate $i$ up’ and ‘move coordinate $i$ down’, respectively. We will use a Poisson process to produce a list of letters. Let $u = N^{1/4}$, do the following for time $u$ and independently for $i = 1, \ldots, N$: write $U_i$ at rate $(1 - a)/N$ and write $D_i$ at rate $1/N$. Use this list of letters to run $C$ for time $u$. Read through the list taking the appropriate actions. For each $U_i$ increment $C_i$ by one, for each $D_i$ decrement $C_i$ by one if $C_i > 0$.

We can use the same list to run $B$; each ‘$D_i$’ corresponds to one step of process $B$. Go through the ‘$D_i$’ symbols in the list in order:

(i) If $B_i = 0$ do nothing.

(ii) If $B_i > 0$, pick the first $U_j$ in the list that you have not already selected. If $i \neq j$, decrement $B_i$ and increment $B_j$. 

87
Discard any remaining \( 'U_i' \) letters. If we can follow this procedure, and if each coordinate \( i \) has appeared at most once (as either \( D_i \) or \( U_i \)), then the property \( B_i \leq C_i \) is preserved. However, the value of \( B_i \) during the processing of the list may be one greater than the final value of \( C_i \).

We require there are enough \( 'U_j' \) letters in the list for the letters \( 'D_i' \) that correspond to \( B_i > 0 \). The number of \( 'U_i' \) is a Poisson random variable with mean \( u(1 - a) \leq N^{1/4} \); by concentration, with probability \( 1 - 2N^{-3} \) there are at least \( u(1 - a) - 3N^{1/8} \sqrt{\log N} \) in the list.

Now condition on the \( 'U_j' \) in the list. The \( 'D_i' \) form a Poisson process with combined rate 1. We can run the process \( B \) as the \( D_i \) arrive. The total rate at which \( 'D_i' \) with \( B_i > 0 \) arrive is \( 1 - X_0 \). Provided the event \( A_{a,s} \) holds, \( X_0 \geq a + 6N^{-1/8} \sqrt{\log N} \); the rate at which \( 'D_i' \) with \( B_i > 0 \) arrive is bounded above by \( 1 - a - 6N^{-1/8} \sqrt{\log N} \). By another application of the concentration result, the number of \( D_i \) corresponding to \( B_i > 0 \) is less than \( u(1 - a) - 3N^{1/8} \sqrt{\log N} \).

Therefore with probability \( P(A_{a,s}) - 4N^{-3} \) we can run the two processes coupled in this way for a ‘block’ of time. We have run process \( C \) for time \( u \), and process \( B \) for as many steps as there are \( D_i \) in the list, with probability \( 1 - 2N^{-3} \) between \( u \pm 3N^{1/8} \sqrt{\log N} \) steps. Repeat the above block procedure \( b := 2\lfloor N^{4/5} \rfloor \) times; with probability \( P(A_{a,s}) - 6bN^{-3} \) we have run process \( B \) for at least \( N^{21/20} \) steps. When the construction fails, we can complete the coupling in an arbitrary manner.

The motivation for the choice of \( b \), the number of blocks, and \( u \), the size of each block, was to ensure that with high probability:

(i) \( B \) is run for at least order \( R^3N \log N \) steps,

(ii) \( (\forall i) \) (coordinate \( i \) appears at most twice per block), and

(iii) \( (\forall i) \) (coordinate \( i \) appears twice in no more than one block).

We can take \( P(A_{a,s}) - O(N^{-2/5}) \) as a lower bound on the probability of this event.

Let integer \( t \) in \( [s, N^{21/20}] \) be a discrete time for process \( B \), and let \( t' \) be the time of the end of the corresponding block for process \( C \). Then if the above event holds, \( B_i(t) \leq C_i(t') + 2 \) for all \( i \). At time \( s \), \( B_i \leq C_i + k_i \) with \( k_i = 0 \). If
coordinate $i$ does not appear twice is a block, and $B_i \leq C_i + k_i$ at the start of a block, then $B_i$ may be equal to $C_i + k_i + 1$ during the block but $B_i \leq C_i + k_i$ at the end of the block. If coordinate $i$ does appear twice in a block, from then on it may be necessary to take $k_i = 1$. Suppose $C_i$ starts a block at 0, is acted on by a $U_i$ and then by a $D_i$; $C_i = 0$ at the end of the block. Also starting from 0, $B_i$ may be acted on by $D_i$, $U_i$, ending up at 1.

Proof of Lemma 4.14. The empirical distribution under $C_{a,s}$ satisfies,

$$\sum_{k \geq j} X_k \leq 8(1-a)^j \quad \text{if} \quad (1-a)^j \geq 2N^{-1}\log N,$$

and

$$\sum_{k \geq j} X_k = 0 \quad \text{if} \quad (1-a)^j \leq N^{-5}.$$

(i) The upper bounds on $\sum_{k \geq j} X_k$ imply upper bounds on $\sum_{k \geq j} kX_k$. We choose $j$ sufficiently large, so that $\sum_{k \geq j} kX_k < R$, and sufficiently small, so that the upper bound $8(1-a)^j \geq \sum_{k \geq j} X_k$ is greater than 0.01. Then $X_0 \leq 1 - 8(1-a)^j$ or we get a contradiction, $\sum kX_k < R$.

We can assume $a = (5R)^{-1}$. Take $j = 29R$ so $8(1-a)^j \geq 0.01$. Also,

$$\sum_{k \geq 29R} kX_k \leq \sum_{k \geq 29R} 8ak(1-a)^k + 16N^{-1}\log N \times \log_{1-a} N^{-5} \leq 0.9R + O(RN^{-1} \log^2 N) < R$$

if $N$ is sufficiently large.

(ii) Notice that $\delta(x,y)$ is linear in the sense that $\delta(x,y) = x\delta(1,y/x)$. Check that

$$\delta(x,y) \leq \begin{cases} x & \text{if } y < x, \\ y \log y/x & \text{if } y \geq x. \end{cases}$$
Therefore

\[ \sum_{k \geq j} \delta(X_0(1 - X_0)^k, X_k) \leq \sum_{k \geq j} 8a(1 - a)^k \left[ \log \frac{8}{X_0} + k \log \frac{1 - a}{1 - X_0} \right] + \delta(N^{5}, 16N^{-1} \log N) \leq 8(1 - a)^j \left[ \log \frac{8}{X_0} + \left( j + \frac{1 - a}{a} \right) \log \frac{1 - a}{1 - X_0} \right] + \frac{80 \log^2 N}{N}. \]

**Proof of Theorem 4.13.** Apply Markov’s inequality to the empirical distribution \( X \). The mean is \( R \), so at most half the boxes have \( 2R \) balls in, \( \sum_{k \leq 2R^*} X_k \geq 1/2 \). Think of the boxes as being indistinguishable, but with labels \( 1, \ldots, N \) attached to indicate how they correspond to \( B_1, \ldots, B_N \). Let \( l = \left\lceil N/2 \right\rceil \). Initially \( B_i < 2R \) for \( i = 1, \ldots, l \). Run the Markov chain step by step. If at the completion of a step \( B_i = 2R \) for \( i \leq l \),

(i) pick \( j > l \) with \( B_j < 2R \),

(ii) swap the labels on the boxes \( i \) and \( j \), so \( B_i < 2R \).

Following this procedure, the property \( B_i < 2R \) for \( i = 1, \ldots, l \) is preserved. With each step and any subsequent relabelling:

(a) \( B_1 \) increases by one with probability less than \( (N - 1)/N^2 \).

(b) \( B_1 \) decreases by at least one with probability at least \( (N - 1)/N^2 \), unless \( B_1 = 0 \). Box one may be picked as the source box, or it may get too full and be subject to a relabelling.

We introduced notation for a Markov chain on the set \( \{0, 1, \ldots, n - 1\} \) in Section 4.7. Set \( n = 2R \) and \( a = 0 \), so the walk is unbiased. The spectral gap, see the proof of Lemma 4.5 is

\[ 1 - \lambda_2 = \frac{2}{N} \left( 1 - \cos \frac{\pi}{2R} \right) \geq \frac{1}{R^2 N}. \]
The stationary distribution $\pi$ is uniform on \{0, 1, \ldots, 2R - 1\}. The time needed to mix to within $(20R)^{-1}$ of $\pi$ in total variation is

$$\log \left(\frac{(20R)^{-1} \cdot (2R)^{-1}}{\log(\lambda_2)}\right) \leq NR^2 \log(40R^2) \leq 10NR^2 \log(R + 1).$$

Let $C_1, \ldots, C_l$ be $l$ copies of the continuous time Markov chain on state space \{0, 1, \ldots, 2R - 1\}. Let $C_i = B_i = R$ for $i = 1, \ldots, l$ at time 0, and let the $C_i$ evolve independently. Once the $C_i$ have mixed sufficiently, $|\{i : C_i = 0\}|$ is stochastically greater than a binomial Bin($l$, $(2R)^{-1} - (20R)^{-1}$) distribution. With high probability $|\{i : C_i = 0\}| \geq 9N/(40R) - 2N^{1/2} \sqrt{\log N}$.

Coupling $B$ and $C$ is much easier here than in the proof of Theorem 4.12. The result follows if we simply show that $B_i \leq C_i$ for ‘almost every’ $i$. Write letters $U_i$ and $D_i$ at rate $N^{-1}$, independently for $i = 1, \ldots, N$ for a block of time $u = N^{1/4}$. Use the $U_i$ and $D_i$, where $i \leq l$, to run $C$ for time $u$. We will also use the letters to run $B$. The number of ‘$U_i$’ in the block can be assumed to be within $N^{1/4} \pm 3N^{1/8} \sqrt{\log N}$; similarly for the ‘$D_i$’. Run $B$ for $N^{1/4} - 3N^{1/8} \sqrt{\log N}$ steps. Pick pairs $(D_i, U_j)$ from the list of letters, without replacement. For each pair move a ball from box $i$ to box $j$ if possible. Repeat $b := 2\lfloor N^{4/5} \rfloor$ times.

Under the concentration bound, the total number of letters left unused is order $O(N^{37/40} \sqrt{\log N})$. The probability that coordinate $i$ appears twice in at least one of the $b$ blocks is $O(N^{-7/10})$. It easily follows by concentration that with high probability

$$|\{i : B_i = 0\}| \geq |\{i : C_i = 0\}| + 8N^{7/8} \sqrt{\log N} - N/(40R).$$

To bound the number of balls per box at $s = 10NR^2 \log(R + 1)$, consider only the number of balls box one receives up to time $s$. This is bounded by a binomial Bin($s$, $1/N$) distribution. Again we use concentration,

$$B_1(s) \leq R + 2 \log N + \sqrt{60R^2 \log(R + 1) \log N} + 10R^2 \log(R + 1)$$

with probability $1 - N^{-3}$. \hfill \Box
4.10 Slow initial decrease in $\overline{S}$

By Section 4.9 we may assume $C_{a,s}$ with $a = (5R)^{-1}$ and $s+\tau(a) = O(R^2N \log N)$. Let $\phi = 10^{-4}(R+1)^{-2}$. In this section we use the log Sobolev result from Lemma 4.7 to show that $\overline{S}$ decreases to $\phi$. This method could be pushed further, to show that $\overline{S}$ decreases to $O(R^3N^{-1}\log^2 N)$. However, the bound on the time needed would be $\Omega(R^2N\log^2 N)$. In the next section we show that once $\overline{S} \leq \phi$, $\overline{S}$ decreases more quickly.

**Theorem 4.15.** With probability $1 - O(N^{-1/3})$, between $O(R^2N \log(R+1) \log N)$ and $N^{21/20}$, $\overline{S} \leq \phi$.

We need the following technical lemma to deal with the discrete nature of the process. Let $x'_k = x_k \vee (e^{-1}N^{-1})$.

**Lemma 4.16.** Assume $x_k = 0$ for $k \geq n' \geq 5 \log N$. If $x_0 \leq 1 - 10^{-9}$,

$$\mathbb{E}(\Delta S \mid X = x) \geq \frac{1}{2N} \sum_{k=0}^{n'-1} (x'_k(1-x_0)-x'_{k+1}) \log \frac{x'_k(1-x_0)}{x'_{k+1}} - \frac{10n'}{N^2}.$$  

**Proof of Theorem 4.15.** The inequality in Lemma 4.7 contains $\nu$, the log Sobolev constant for the biased random walk. Recall Corollary A.4 of [19], the general bound on log Sobolev constants in terms of the spectral gap and the minimum of the equilibrium distribution,

$$\nu \geq \frac{(1-2\pi(n-1))(1-\lambda_2)}{\log(1/\pi(n-1) - 1)}.$$  

By Lemma 4.5 (ii) with $a = (5R)^{-1}$, spectral gap $1 - \lambda_2 \geq (100R^2N)^{-1}$. With $n = n(a)$, the maximum $n$ such that $\pi(n-1) \geq 2N^{-10}$,

$$\nu \geq (1000R^2N \log N)^{-1}.$$  

By Lemma 4.14 (i) we may apply Lemma 4.16 with $n' = n(a)/2$. Combining the inequalities,

$$\mathbb{E}(\Delta \overline{S} \mid X = x) \leq -2\nu \overline{S} + \frac{250R \log N}{N^2}.$$  

(4.17)
We can now apply Corollary 4.9 to show that $\overline{S}$ decreases as inequality (4.17) suggests. Corollary 4.9 only gives a condition under which the upper bound on a process $(W(i) : 0 \leq i \leq N^{21/20})$ can be halved, so we need to apply it a number of times. Suppose inductively that with probability $1 - q_l$ from time $t_l$ to $N^{21/20}$,

$$\overline{S} \leq \alpha = S(G^R) \cdot 2^{-l}.$$ 

Let

$$W(i) = \overline{S}(X((i + t_l) \wedge \lfloor N^{21/20} \rfloor)).$$

We need a lower bound $\gamma$ on the expected decrease of $W$ when $W \in [\alpha/8, \alpha]$. This is obtained by substituting $\overline{S} = \alpha/8$ into inequality (4.17); we can take $\gamma = \alpha(8000R^2N\log N)^{-1}$. The maximum time before $W$ hits $\alpha/4$, or the concentration bound fails, is $m = 13\alpha/(16\gamma) = 6500R^2N\log N$. The step size $\Delta W$ is bounded by $\pm 2N^{-1}(1 + \log N)$ so take $b = 4N^{-1}(1 + \log N)$ and $\hat{v} = mb^2/4$. Inequality (4.10) is satisfied if $\alpha \geq 10^5R N^{-1/2} \log^2 N$.

Take $1 - q_0 = \mathbb{P}(C_{a,s})$, with $a = (5R)^{-1}$ and $s = 10NR^2\log(R + 1)$. Let $t_0 = s + \tau(a)$. Then we can take $q_{l+1} = q_l + N^{-1}$ and $t_{l+1} = t_l + m$. If $N$ is sufficiently large,

$$\phi \geq 10^5RN^{-1/2} \log^2 N$$

so the result follows. \hfill \Box

We delay the proof of Lemma 4.16.

4.11 Rapid decrease in $\overline{S}$

In this section we use the spectral gap related inequality, Lemma 4.6, to prove $\overline{S}$ decreases rapidly.

**Theorem 4.18.** With probability $1 - O(N^{-1/3})$, from $O(R^5N \log N)$ until $N^{21/20}$,

$$\overline{S}(X) = O(N^{-1} \log^2 N).$$

By Section 4.10 we may assume that $\overline{S} \leq \phi = 10^{-4}(R + 1)^{-2}$ from steps $O(R^2N \log(R + 1) \log N)$ to $N^{21/20}$. When the entropy is close to the maximum
entropy, $X_0$ must be close to $(R+1)^{-1}$. In Lemma 4.19 we bound the difference. This allows us to use Theorem 4.12 from Section 4.9 again, this time with an improved lower bound $a$. We also need two bounds related to $\Delta S$.

**Lemma 4.19.** If $\overline{S}(x) \leq (R+1)/(3R^2) \text{ then}$

$$
\left( x_0 - \frac{1}{R+1} \right)^2 \leq \frac{3R^2 \overline{S}}{(R+1)^3},
$$

**Lemma 4.20.** Assume $x_k = 0$ for $k \geq n' \geq 5 \log N$. Then

$$
\mathbb{E}(\Delta S \mid X = x) \geq \frac{1}{N} \sum_{k=0}^{n'-1} \left( x_k(1-x_k) - x_{k+1} \right)^2 \max\{x_k(1-x_k), x_{k+1} \} - \frac{10n'}{N^2}.
$$

**Lemma 4.21.** Suppose $x_0 \leq 2/3$ and $x_k = 0$ for $k \geq n'$. Then the variance of $\Delta S$ satisfies

$$
\mathrm{var}(\Delta S \mid X = x) \leq \frac{12 \log N}{N} \mathbb{E}(\Delta S \mid X = x) + \frac{120n' \log N + 200n'}{N^3}.
$$

**Proof of Theorem 4.18.** The proof is similar to the proof of Theorem 4.15. We start off with the assumption that $\overline{S} \leq \phi$ between $O(R^2 N \log(R+1) \log N)$ and $N^{21/20}$ steps. This allows us to bound $\mathbb{E}(\Delta \overline{S} \mid X = x)$ when $\overline{S}$ is in a suitable range. We then use Corollary 4.9 inductively.

However there are some added complications. One problem is that the bound on $\mathbb{E}(\Delta \overline{S} \mid X = x)$ will not apply ‘instantly’. Once $\overline{S} \leq \phi^{-l}$, $l = 0, 1, \ldots$, we will have to wait before we can assume an improved bound on $\mathbb{E}(\Delta \overline{S} \mid X = x)$. Secondly, we must use Lemma 4.21 to bound the variance of $\Delta S$. We split the proof of Theorem 4.18 into two parts. In the first part we will show that $\overline{S} \leq \psi = (R+1)/(3 \log^2_R(R+1)N)$. In the second, that $\overline{S} \leq 10^5 N^{-1} \log^2 N$.

We collect and ‘daisy-chain’ (i) Lemma 4.20, (ii) an elementary observation, (iii) Lemma 4.6, (iv) the bound $\delta(x, y) \leq (x - y)^2/x$ for function $\delta$, and (v) Lemma 4.4.

(i) $\mathbb{E}(\Delta S \mid X = x) \geq N^{-1} \sum_{k=0}^{j-2} \frac{(x_k(1-x_k) - x_{k+1})^2}{\max\{x_k(1-x_k), x_{k+1} \}} - 100(R+1)N^{-2} \log N$,

(ii) $\sum_{k=0}^{j-2} \frac{(x_k(1-x_k) - x_{k+1})^2}{\max\{x_k(1-x_k), x_{k+1} \}} \geq \sum_{k=0}^{j-2} \frac{(x_k(1-x_k) - x_{k+1})^2}{x_0(1-x_0)^{k+1}} \inf_{k \leq j-1} \frac{x_0(1-x_0)^k}{x_k}$,

(iii) $\sum_{k=0}^{j-2} \frac{(x_k(1-x_k) - x_{k+1})^2}{x_0(1-x_0)^{k+1}} \geq (x_0^2/4) \sum_{k=1}^{j-1} \frac{(x_k - x_0(1-x_0))}{x_0(1-x_0)^k}$,
4.11 Rapid decrease in $\mathcal{S}$

\[ (iv) \sum_{k=1}^{j-1} \frac{(x_k-x_0(1-x_0)^k)^2}{x_0(1-x_0)^k} \geq \sum_{k=1}^{j-1} \delta(x_0(1-x_0)^k, x_k), \]

\[ (v) \sum_{k=1}^{\infty} \delta(x_0(1-x_0)^k, x_k) \geq \mathcal{S}(x). \]

Therefore,

\[
\mathbb{E}(\Delta \mathcal{S} \mid X = x) \leq -\frac{x_0^3}{4N} \left( \mathcal{S} - \sum_{k \geq j} \delta(x_0(1-x_0)^k, x_k) \right) \inf_{k \leq j-1} \frac{(1-x_0)^k}{x_k} + \frac{100(R+1)\log N}{N^2}.
\]

The result now hinges on controlling, for some $j$,

\[ (a) \sum_{k \geq j} \delta(x_0(1-x_0)^k, x_k) \quad \text{and} \quad (b) \inf_{k \leq j} \frac{(1-x_0)^k}{x_k}. \]

We need $j$ sufficiently large to make quantity (a) small, but not too large or quantity (b) becomes too small.

**Part One:**

For the first induction argument, we take $\alpha = \phi 2^{-l}$, $l = 0, 1, \ldots, \lfloor \log_2 \phi/\psi \rfloor$. Assume inductively $\mathcal{S} \leq \alpha \in [\psi, \phi]$. By Lemma 4.19

\[ \left| X_0 - \frac{1}{R+1} \right| \leq \frac{R}{R+1} \sqrt{\frac{3\alpha}{R+1}}. \]

Then, provided $N$ is sufficiently large, we find that $X_0, a$ satisfy the requirements of Theorem 4.12 if we take,

\[ a = a(\alpha) = \frac{1}{R+1} - \frac{R}{R+1} \varepsilon(\alpha), \quad \varepsilon(\alpha) = 2\sqrt{\frac{3\alpha}{R+1}}. \quad (4.22) \]

Once $\mathcal{S} \leq \psi$, we will not need to keep improving our bound $a \leq X_0$ in this way.

For each value $a = a(\alpha)$, we apply Theorem 4.12 producing a corresponding set of coupled biased random walks $C^a = (C^{a}_i)_{i=1}^{N}$. Once we have $\mathcal{S} \leq \alpha$ with high probability, we must wait for the $C^a_i$ to mix, though not necessarily for as long as $\tau(a)$ steps. Choose $j \in \mathbb{N},$

\[ j = j(\alpha) = \lfloor \varepsilon(\alpha)^{-1} \rfloor, \quad \alpha \in [\psi, \phi]. \]
4.11 Rapid decrease in $S$

(a) When $\alpha = \phi$ wait $\tau(a)$ steps; by Lemma 4.5 (ii) the $C_i^a$ mix to within $N^{-10}$ in total variation. Then by concentration we can assume $X_k \leq 8(1 - a(\phi))^k$ for $k$ such that $(1 - a(\phi))^k \geq 2N^{-1}\log N$.

(b) When $\alpha = \phi^2 - l$, $l = 1, \ldots, \lfloor \log \phi/\psi \rfloor$, wait $O(RNj(\alpha))$ steps so we can bound $X_k$ for $k \leq j(\alpha)$ by Lemma 4.5 (iii). Then $X_k \leq 16(1 - a(\alpha))^k$ for $k \leq j(\alpha)$.

Note that if $k \leq j(\alpha)$ by $(4.22)$,

$$\left(\frac{R}{R+1}\right)^k \leq (1 - a(\alpha))^k \leq e(\frac{R}{R+1})^k.$$

By (a) and Lemma 4.14, if $\alpha \in [\psi, \phi]$ and $S \in [\alpha/8, \alpha]$, 

$$\sum_{k \geq j} \delta(X_0(1 - X_0)^k, X_k) \leq 200j(1 - a(\phi))^2 \leq \alpha/16.$$

By (b),

$$\inf_{k \leq j} \frac{(1 - X_0)^k}{X_k} \geq \inf_{k \leq j} \frac{e^{-1}(\frac{R}{R+1})^k}{16e(\frac{R}{R+1})^k} \geq \frac{1}{16\epsilon^2}.$$

Therefore when $S \in [\alpha/8, \alpha]$, $\alpha \in [\psi, \phi]$ and the $C_i^a(\alpha)$ have mixed,

$$\mathbb{E}(\Delta S \mid X = x) \leq -\frac{\alpha}{10^4(R+1)^3N}.$$

**Part Two:**

From now on fix

$$a = \frac{1}{R+1} - \frac{R}{R+1}\varepsilon, \quad \varepsilon = 2\sqrt{\frac{3\psi}{R+1}}.$$

We apply Theorem 4.12 only once more; wait $\tau(a)$ for the $C_i^a$ to mix. For $k$ such that $(1 - a)^k \geq 2N^{-1}\log N$, $X_k \leq 8(1 - a)^k$. Let

$$j = \lfloor \log_{1-a} 2N^{-1}\log N \rfloor.$$

If $k \leq j \leq 2\varepsilon^{-1}$, $(\frac{R}{R+1})^k \leq (1 - a)^k \leq e^2(\frac{R}{R+1})^k$. Therefore

$$\inf_{k \leq j} \frac{(1 - X_0)^k}{X_k} \leq \frac{1}{8\epsilon^4}.$$
By Lemma 4.14 with $N$ large,

$$\sum_{k \geq j} \delta(X_0(1 - X_0)^k, X_k) \leq 160N^{-1}\log^2 N.$$ 

For the second induction argument take $\alpha = \psi 2^{-l}$, $l = 0, 1, \ldots$. When $S \in [\alpha/8, \alpha]$ and $\alpha \in [10^5N^{-1}\log^2 N, \psi]$,

$$E(\Delta S \mid X = x) \leq -\frac{\alpha}{3 \cdot 10^4(R + 1)^3N}.$$ 

Using the bound on variance from Lemma 4.21,

$$\hat{v} \leq 12 \log N \frac{\alpha}{N} + \frac{10^8(R + 1)^4\log N}{N^2}.$$ 

Inequality (4.10) holds for $\alpha \geq 10^5N^{-1}\log^2 N$ if $N$ is sufficiently large.

**Proof of Lemma 4.19.** We prove this by maximizing $S(x)$ with $x_0$ fixed. We can assume that $x_1, x_2, x_3, \ldots$ form a geometric series, $a, ar, ar^2, \ldots$. Then

$$x_0 + \sum ar^k = 1, \quad \sum (k + 1)ar^k = R.$$ 

Solve to get,

$$a = \frac{(1 - x_0)^2}{R}, \quad r = \frac{R - 1 + x_0}{R}.$$ 

Therefore the maximum value for $S$ is

$$-x_0 \log x_0 - (1 - x_0) \log \frac{(1 - x_0)^2}{R} - (R - 1 + x_0) \log \frac{R - 1 + x_0}{R}.$$ 

Check that for $0 \leq y \leq 2x$, $(x - y)^2/(3x) \leq \delta(x, y)$. If $x_0 \leq 2(R + 1)^{-1}$,

$$S \geq S(G^R) - S(x_0, a, ar, \ldots) \geq \delta \left( \frac{1}{R + 1}, x_0 \right) + 2\sqrt{R} \delta \left( \frac{\sqrt{R}}{R + 1}, \frac{1 - x_0}{\sqrt{R}} \right) + R \delta \left( \frac{R}{R + 1}, \frac{R - 1 + x_0}{R} \right) \geq \frac{1}{3} \left( x_0 - \frac{1}{R + 1} \right)^2 \frac{(R + 1)^3}{R^2}.$$ 

This lower bound on $\bar{S}$ is a convex function of $x_0$. \qed
4.12 Final observations

We have shown that the event $\mathcal{S} = O(N^{-1} \log^2 N)$ from time $O(R^3 N \log N)$ to $N^{21/20}$ has high probability. In this section we modify our argument to show that $\mathcal{S}$ decreases to $O(R^4 N^{-1} \log N)$. Rough calculations suggest that under the equilibrium measure, the expected value of $\mathcal{S}$ is $\Omega(RN^{-1} \log N)$; our result may in some sense be tight.

As in part two of the proof of Theorem 4.18, take

$$a = \frac{1}{R+1} - \sqrt{\frac{12R^2 \psi}{(R+1)^3}}, \quad j = \lfloor \log_{1-a} 2N^{-1} \log N \rfloor.$$  

We can then assume that

$$E(\Delta \mathcal{S} \mid X = x) \leq -\frac{\mathcal{S} - \sum_{k \geq j} \delta(x_0(1-x_0)^k, x_k)}{2000(R+1)^3 N} + \frac{100(R+1) \log N}{N^2}.$$  

We sharpen Lemma 4.14. With $s = O(R^3 N \log N)$, let $E$ be the subset of $C_{a,s}$ where $S = O(N^{-1} \log^2 N)$ from $s$ to $N^{21/20}$. Let $E'$ be the intersection of events $E$ and

$$\left\{ \forall i \in [s, N^{21/20}], \sum_{k \geq j} \delta(X_0(1-X_0)^k, X_k) \geq \frac{1000(R+1)^2 \log N}{N} \right\}.$$  

**Lemma 4.23.** $\mathbb{P}(E') \geq 1 - O(N^{-1/3}).$

We will now consider a modification of the entropy gap process $\mathcal{S}$. Let $\mathcal{F}_i$ be the $\sigma$-algebra generated by the mean field ZRP up to step $i$. Let $W$ be a random process defined from $O(R^3 N \log N)$ to $N^{21/20}$,

$$W(i) = \begin{cases} \mathcal{S}(X(i)) & \text{if } \mathbb{E}(E' \mid \mathcal{F}_i) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $n = \lfloor \log \log N \rfloor$. Let $C$ be the discrete time biased random walk on $\Omega = \{0, 1, \ldots, n - 1\}$ that decreases by one with probability $2/3$ if $C > 0$ and increases by one with probability $1/3$ if $C < n - 1$. $C$ is a faster version of the biased random walk from Section 4.7 with bias $1/2$. The total variation mixing time for $C$ is $O(n)$ steps.
4.12 Final observations

By the definition of $E'$, $W = O(N^{-1} \log^2 N)$. Say that $\log_2(W/\alpha)$ is dominated by $C$ with maximum step time $m$ if it has the following property. Whenever $2^kW$ is within a small interval of $\alpha/2$ with positive integer $k$, then with probability $2/3$, $2^kW$ decreases to below $\alpha/4$ within $m$ steps without having reached $\alpha$.

**Theorem 4.24.** $\log_2(W/\alpha)$ is dominated by $C$ with maximum step time $O(R^3N)$.

**Proof.** We use the construction from Section 4.8. In Corollary 4.9 the construction is applied repeatedly with $c$, the probability of ‘error’, very small each time. We now apply the construction once at a time and with relaxed error probability; let $c = 1/3$. Then for inequality (4.11), we need a weaker condition than (4.10), only that

\[
\frac{\alpha}{16} \geq (2/3)b \log 3 + \sqrt{2\hat{v}\log 3}.
\]

Then if $W$ is in a small neighbourhood of $\alpha/2$, with probability $2/3$, $W$ enters $[0, \alpha/4]$ without increasing beyond $\alpha$. If $\alpha \geq 10^7(R + 1)^4N^{-1}\log N$ and $W \in [\alpha/8, \alpha]$ then

\[
\mathbb{E}(\Delta W \mid X = x) \leq -\frac{\alpha}{2 \cdot 10^4(R + 1)^3N}
\]

and

\[
\hat{v} \leq 12 \log N \frac{\alpha}{N} + \frac{10^8(R + 1)^4\log N}{N^2}.
\]

Inequality (4.25) holds for $N$ sufficiently large. \qed

**Proof of Lemma 4.23.** As $j \geq (1/2) \log_{1-a} N^{-1}$, under $C_{a,s}$, 

\[
\sum_{k \geq 10j} \delta(X_0(1 - X_0)^k, X_k) = \sum_{k \geq 10j} \delta(X_0(1 - X_0)^k, 0) = O(N^{-5}).
\]

Similarly, we must bound the sum from $j$ to $10j$. Check,

\[
\delta(\lambda, x) \leq \lambda(e^2 + 1) + [x \log x/(\lambda e) - \lambda e^2]_+.
\]

Suppose $X \sim \text{Bin}(N, \lambda/N)$ and consider $Y = [X \log X/(\lambda e) - e^2\lambda]_+$. Let $Z$ be the geometric random variable with distribution $((1 - e^{-1})e^{-n})_{n=0}^\infty$; $P(Z \geq n) =$
We will show that $P(Y \geq n) \leq \exp(-n)$, so $Y \leq_{st} 1 + Z$. We can write,

$$\sum_{k=j}^{10j} \delta(X_0(1 - X_0)^k, X_k) \leq \sum_{k=j}^{10j} \delta(4(1 - a)^k, X_k) + X_k \log \frac{4(1 - a)^k}{X_0(1 - X_0)^k} \leq X_k \log \frac{X_k}{4(1 - a)^k e} - 4(1 - a)^k e^2 + 4(1 - a)^k (e^2 + 1) + \left[ X_k \log \frac{X_k}{4(1 - a)^k e} - 4(1 - a)^k e^2 \right].$$

For $k \leq 10j$, $(1 - X_0)^k \leq (1 - a)^k \leq e^{20}(1 - X_0)^k$. Up to modification off $E'$, $N_k = N X_k \leq_{st} \text{Bin}(N, 4(1 - a)^k)$. With $Z_i$ independent random variables equal to $Z$ in distribution,

$$\sum_{k=j}^{10j} \delta(X_0(1 - X_0)^k, X_k) \leq_{st} N^{-1} \sum_{k=1}^{9j} Z_i + \frac{500(R + 1)^2 \log N}{N}.$$

We can bound the probability that the sum of $l := 9j$ independent random variables $Z_i$ is greater than $B := le^{-1/4}/(1 - e^{-1/4}) \leq 4l$. Let $\hat{Z}_i$ be independent geometric $((1 - e^{-1/4})e^{-k/4})_{k=0}^\infty$ random variables. By an elementary comparison,

$$P\left( \sum_{i=1}^{l} Z_i \geq B \right) \leq P\left( \sum_{i=1}^{l} \hat{Z}_i \geq B \right) \left( \frac{1 - e^{-1}}{1 - e^{-1/4}} \right)^l \exp(-3/4)B \leq \exp(-l) = O(N^{-1}).$$

It remains to show that $P(Y \geq n) \leq \exp(-n)$. The set of positive values $Y$ can take is,

$$A_\lambda = \left\{ k \log \frac{k}{\lambda e} - e^2 \lambda : k \in \mathbb{N} \right\} \cap [0, \infty)$$

is sparser than $\mathbb{N} = \{0, 1, \ldots\}$; by this we mean that there is a one to one function $f : \mathbb{N} \to A_\lambda$ such that $n \leq f(n)$ for all $n$. Let $y = k \log k/(\lambda e) \in A_\lambda$,

$$P(Y = y) = \binom{N}{k} \left( \frac{\lambda}{N} \right)^k \left( 1 - \frac{\lambda}{N} \right)^{N-k} \leq \left( \frac{k/e}{k!} \right)^k \left( \frac{\lambda}{k/e} \right)^k \leq (1 - e^{-1}) \exp(-y).$$

Therefore $P(Y \geq n) \leq \exp(-n)$.
4.13 Proofs relating to entropy and the biased random walk

Proof of Lemma 4.4. If \( x_k = t(1 - t)^k \), then \( x \) has mean \( \sum kx_k = (1 - t)t^{-1} \) and entropy \( S(x) = -\log t - (1 - t)t^{-1} \log(1 - t) \). By definition,

\[
S(x) = S(G^R) - \sum x_k \log x_k, \quad G^R_k = R^k / (R + 1)^{k+1}.
\]

Using \( \sum x_k = 1 \) and \( \sum kx_k = R \),

\[
S(x) = \sum x_k \log \frac{x_k}{R^k / (R + 1)^{k+1}} = \log \left[ \frac{t(1 - t)^R}{\frac{1}{R+1} (\frac{R}{R+1})^R} \right] + \sum x_k \log \frac{x_k}{t(1 - t)^R}.
\]

For \( t \in [0, 1] \), \( t(1 - t)^R \) is maximized by \( t = (R + 1)^{-1} \). The relative entropy of \( x \) with respect to a geometric distribution is minimised by \( G^R \). The relative entropy can also be written using the function \( \delta(x, y) = y \log y/x + (y - x) \),

\[
\sum \delta(R^k / (R + 1)^{k+1}, x_k) = \sum x_k \log \frac{x_k}{R^k / (R + 1)^{k+1}}, \text{ and}
\]

\[
\sum \delta(x_0(1 - x_0)^k, x_k) = \sum x_k \log \frac{x_k}{x_0(1 - x_0)^k}.
\]

Proof of Lemma 4.5. We will work with the left eigenvalues of \( K \). The equilibrium distribution \( \pi(k) = F_1(k) = \hat{a}(1 - a)^k \) is the unique (up to scalar multiples) eigenvector of \( K \) with eigenvalue 1. The other \( n - 1 \) eigenvalues of the Markov chain can be written as the imaginary parts of a geometric sequences. Let \( A_j = \sqrt{1 - a} \exp(i\pi(j - 1)/n) \) for \( j = 2, \ldots, n \). Let \( F_j(k) = R_j^{-1/2} \Im[(1 - A_j)A_j^k] \) with \( R_j \) the normalizing constant,

\[
R_j = \frac{n}{\hat{a}} \left[ 1 - \frac{a}{2} - \sqrt{1 - a} \cos \frac{\pi(j - 1)}{n} \right].
\]
For convenience, allow the definition of $F_j(k)$ to extend to $F_j(-1)$ and $F_j(n)$; note that $F_j(-1)(1 - a) = F_j(0)$ and $F_j(n - 1)(1 - a) = F_j(n)$. Check,

$$(F_j K)(k) = F_j(k) + \frac{1}{N} \Delta [(1 - a)F_j(k - 1) - (1 + (1 - a))F_j(k) + F_j(k + 1)]$$

$$(F_j K)(k) = F_j(k) - \frac{1}{N} \Delta [(1 - A_j)A_j^k(1 - (1 - a)A_j^{-1})(1 - A_j)]$$

$$= F_j(k) (1 - N^{-1}|1 - A_j|^2).$$

Therefore, for $j \geq 2$, the $j$-th eigenvalue is

$$\lambda_j = 1 - N^{-1}|1 - A_j|^2$$

$$= 1 - 2N^{-1} \left( 1 - a/2 - \sqrt{1 - a} \cos \frac{(j - 1)\pi}{n} \right)$$

$$\leq 1 - 2N^{-1}(1 - a/2 - \sqrt{1 - a})$$

$$\leq 1 - a^2/(4N).$$

(i) Suppose the Markov chain $C$ starts at $k$ with probability 1. $C$ can be restricted to the subset \{k, k + 1, \ldots, n - 1\}. It is then a translation of the random walk with \(\prime n' = n - k\). The Markov chain on \{0, \ldots, n - 1\} started at $k$ is stochastically smaller than the MC restricted to \{k, \ldots, n - 1\} and started at $k$, which in turn is stochastically smaller than the equilibrium distribution of the Markov chain restricted to \{k, \ldots, n - 1\}. Therefore for all $i$, $(\mu K^i)(C \geq j + k) \leq (1 - a)^i$.

(ii) With $\tau_1(\varepsilon)$ the time to convergence within $\varepsilon$ in the sense of total variation, we will show the standard bound [48],

$$\tau_1(\varepsilon) \leq \frac{\log(\varepsilon \pi_{\text{min}})}{\log \lambda_2}.$$

Write $K^i(x, y)$ for the $i$-step transition probability from $x$ to $y$. Then

$$K^i(x, y) = \sum_j \langle \delta_x, F_j \rangle \pi F_j(y) \lambda_j^i$$

$$= \pi(y) + \sum_{j \geq 2} \pi(x)^{-1} F_j(x) F_j(y) \lambda_j^i.$$
4.13 Proofs relating to entropy and the biased random walk

As $1 = K^0(x, x) \geq \pi(x)^{-1} \sum_{j \geq 2} F_j(x)^2$, $\sum_{j \geq 2} F_j(x)^2 \leq \pi(x)$.

By the Cauchy-Schwarz inequality, with $\pi_{\min} = \min_x \pi(x)$,

$$|K^i(x, y) - \pi(y)| = \pi(x)^{-1} \left| \sum_{j \geq 2} F_j(x)F_j(y)\lambda_j \right|$$

$$\leq \lambda_2^i \pi(x)^{-1} \sqrt{\sum_{j \geq 2} F_j(x)^2 \sum_{j \geq 2} F_j(y)^2}$$

$$\leq \lambda_2^i \pi(x)^{-1} \sqrt{\pi(x)\pi(y)}$$

$$\leq \pi(y) \frac{\lambda_2^i}{\sqrt{\pi(x)\pi(y)}} \leq \pi(y) \frac{\lambda_2^i}{\pi_{\min}}.$$ 

With $1 - \lambda_2 \geq a^2/(4N)$,

$$\tau_1(\varepsilon) \leq \log_{\lambda_2}(\varepsilon\pi(n-1)) \leq -4Na^{-2}\log(\varepsilon\pi(n-1)).$$

(iii) Suppose first that $\mu(C = l) = 1$. The general result follows by linearity.

From part (ii),

$$\left| \frac{K^i(l, k)}{\pi(k)} - 1 \right| \leq \frac{\lambda_2^i}{\sqrt{\pi(k)\pi(l)}}$$

$$\leq \frac{\lambda_2^i}{\pi(l)}$$

$$\leq 1/2 \text{ if } i \geq 4Na^{-2}\log 2/\pi(l).$$

Proof of Lemma 4.6. For convenience, change variables as follows. Let $u_k = (X_k - X_0(1 - X_0)^k)/\sqrt{X_0(1 - X_0)^k}$ for $k = 0, 1, \ldots, j + 1, u_0 = 0$. First consider the problem: minimize $f$ for fixed $g$ and $u_{j+1}$,

$$f = \sum_{k=0}^j (u_k\sqrt{1 - X_0} - u_{k+1})^2, \quad g = \sum_{k=1}^{j+1} u_k^2.$$ 

Let $c = 1 - \sqrt{1 - X_0} \geq X_0/2$. We will prove the lemma by showing that $f \geq c^2 g$.

Introduce a Lagrange multiplier $\hat{\mu}$,

$$\hat{\mu} \frac{\partial}{\partial u_k} g = \frac{\partial}{\partial u_k} f, \quad k = 1, \ldots, j.$$ 

103
4.13 Proofs relating to entropy and the biased random walk

With a further change of variables this is simply,

\[ u_{k+1} - 2\mu u_k + u_{k-1} = 0, \quad k = 1, \ldots, j. \]

When the equation \( x^2 - 2\mu x + 1 = 0 \) has distinct roots, the solution of the difference equation can be expressed in terms of powers of the roots, that is \( u_k = c_1 \left( \mu - \sqrt{\mu^2 - 1} \right)^k + c_2 \left( \mu + \sqrt{\mu^2 - 1} \right)^k \). The solution, up to a multiplicative factor, can be written as follows with \( \Theta \) a function of \( \mu \).

<table>
<thead>
<tr>
<th>( \mu )</th>
<th>( u_k )</th>
<th>( X_k - X_0(1 - X_0)^k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( &lt;-1 )</td>
<td>( (-1)^k \sin \Theta k )</td>
<td>( \sqrt{X_0(1 - X_0)^k} (-1)^k \sin \Theta k )</td>
</tr>
<tr>
<td>(-1 )</td>
<td>( (-1)^k k )</td>
<td>( \sqrt{X_0(1 - X_0)^k} (-1)^k k )</td>
</tr>
<tr>
<td>((-1, 1) )</td>
<td>( \sin \Theta k )</td>
<td>( \sqrt{X_0(1 - X_0)^k} \sin \Theta k )</td>
</tr>
<tr>
<td>( 1 )</td>
<td>( k )</td>
<td>( \sqrt{X_0(1 - X_0)^k} k )</td>
</tr>
<tr>
<td>( &gt;1 )</td>
<td>( \sin \Theta k )</td>
<td>( \sqrt{X_0(1 - X_0)^k} \sin \Theta k )</td>
</tr>
</tbody>
</table>

If \( u_k = k \), \( f \geq c^2 g \) follows by evaluating the quadratic sums \( f \) and \( g \).

\[
\begin{align*}
    f & = \sum_{k=0}^{j} (k\sqrt{1 - X_0} - (k + 1))^2 \\
    & = \sum_{k=0}^{j} (kc + 1)^2 \\
    & = c^2 j(j + 1)(2j + 1) + 2c j(j + 1) + (j + 1) \\
    g & = \sum_{k=1}^{j+1} k^2 = \frac{(j + 1)(j + 2)(2j + 3)}{6}
\end{align*}
\]

If \( u_k = \sin \Theta k \) we can use the following identity to expand \( f - c^2 g \).

\[
\begin{align*}
    \sum_{k=0}^{j} \sin \Theta (k + a) \sin \Theta k & \equiv \sum_{k=0}^{j} \frac{\cos \Theta a - \cos \Theta (2k + a)}{2} \\
    & \equiv \frac{j + 1}{2} \cos \Theta a - \frac{\sin \Theta (2j + 1 + a) - \sin \Theta (a - 1)}{4 \sin \Theta}
\end{align*}
\]
4.13 Proofs relating to entropy and the biased random walk

Therefore

\[ f - c^2 g = 2\sqrt{1 - X_0} \left( \sum_{k=0}^{j} \sin^2 \Theta(k + 1) - \sum_{k=0}^{j} \sin \Theta(k + 1) \sin \Theta k \right) \]

\[ - (1 - X_0) \left( \sum_{k=0}^{j} \sin^2 \Theta(k + 1) - \sum_{k=0}^{j} \sin^2 \Theta(k) \right) \]

\[ = (j + 1)\sqrt{1 - X_0}(1 - \cos \Theta) + \frac{1}{2}\sqrt{1 - X_0} \tan \frac{\Theta}{2} \sin \Theta(2j + 2) \]

\[ + \frac{1}{2}(\sqrt{1 - X_0} - (1 - X_0))(1 - \cos \Theta(2j + 2)) \]

\[ \geq \sqrt{1 - X_0} \left[ (j + 1)(1 - \cos \Theta) + \frac{1}{2} \tan \frac{\Theta}{2} \sin \Theta(2j + 2) \right] \]

\[ = \sqrt{1 - X_0} \sin^2 \Theta \left[ (2j + 2) + \frac{\sin \Theta(2j + 2)}{\sin \Theta} \right] \geq 0. \]

For \( u_k = \sinh \Theta k \), we can apply Osborne’s rule to convert the previous trigonometric identity to a hyperbolic one,

\[ -f + c^2 g = (j + 1)\sqrt{1 - X_0}(1 - \cosh \Theta) \]

\[ - \frac{1}{2}\sqrt{1 - X_0} \tanh \frac{\Theta}{2} \sinh \Theta(2j + 2) \]

\[ + \frac{1}{2}(\sqrt{1 - X_0} - (1 - X_0))(1 - \cosh \Theta(2j + 2)) \]

\[ \leq - \sqrt{1 - X_0} \sinh^2 \Theta \left[ (2j + 2) + \frac{\sinh \Theta(2j + 2)}{\sinh \Theta} \right] \leq 0. \]

We can ignore the solutions \((u_k)\) corresponding to \( \mu \leq -1 \), they are clearly never optimal.

**Proof of Lemma 4.7.** Define \( f \) by \( f(k) = X'_k / \pi(k) \) where \( X'_k = X_k \vee (e^{-1}/N) \).
4.14 Technical lemmas relating to $\Delta S$

We use the log Sobolev constant, following the proof of Theorem 3.6 in [19].

$$\frac{1}{N} \sum_k (x_k'(1 - a) - x_k'(1 - a) \log \frac{x_k'(1 - a)}{x_{k+1}}$$

$$= \frac{1}{N} \sum_k \pi(k + 1)[f(k) - f(k + 1)][\log f(k) - \log f(k + 1)]$$

$$= \sum_{j,k} \pi(j)K(j,k)f(j)[\log f(j) - \log f(k)]$$

$$= \mathcal{E}(f, \log f)$$

$$\geq 4\mathcal{E}(\sqrt{f}, \sqrt{f})$$

$$\geq 4\nu\mathcal{L}(\sqrt{f}).$$

We can now switch back to the empirical distribution $X = (X_k)$ by admitting a small error term.

$$\mathcal{L}(\sqrt{f}) = \sum X_k' \log \frac{X_k'}{\|\sqrt{f}\|_{\frac{1}{2}, \pi}(k)}$$

$$= \sum X_k \log \frac{X_k}{\pi(k)} + \sum_{k \notin \Omega, X_k = 0} \frac{e^{-1}}{N} \log \frac{e^{-1}N^{-1}}{\pi(k)} - \sum X_k' \log \sum X_k'$$

$$\geq \sum X_k \log \frac{X_k}{\pi(k)} - \left[\frac{ne^{-1}}{N} \log eN + (1 + ne^{-1}N^{-1}) \log(1 + ne^{-1}N^{-1})\right]$$

$$\geq \sum X_k \log \frac{X_k}{\pi(k)} - \left[\frac{ne^{-1}}{N} \log eN + (1 + ne^{-1}N^{-1})ne^{-1}N^{-1}\right]$$

$$\geq \sum X_k \log \frac{X_k}{X_0(1 - X_0)^k} - \frac{n \log N}{N}.$$

\[\square\]

4.14 Technical lemmas relating to $\Delta S$

All that remains now is to prove Lemmas 4.16, 4.20 and 4.21. The proofs of Lemmas 4.16 and 4.20 are very similar. We express $\mathbb{E}(\Delta S \mid X = x)$ as a sum, and then carefully bound the terms. The proof of Lemma 4.21 is similarly technical. To make the calculations more manageable we define two functions. Let $f : [0, \infty) \to \mathbb{R}$ be given by

$$f(x) = x \log x, \quad f(0) = 0,$$
and let $g : [0, \infty) \to \mathbb{R}$ be the ‘discrete-derivative’ of $f$,

$$g(x) = f(x + 1/N) - f(x).$$

If $x = 0$, take $x \times g(x - 1/N)$ to be equal to zero.

**Proof of Lemma 4.16.** Consider one step of the Markov chain starting from $X = x$. Let $R_k$ be the random variable, taking values in $\{-1/N, 0, +1/N\}$, measuring the movement from $X_k$ to $X_{k+1}$; that is the fraction of boxes that hold $k$ balls and gain a ball, minus the fraction of boxes that hold $k + 1$ balls and lose a ball. Let $m_k = x_k(1 - x_0) - x_{k+1}$ so $\mathbb{E}(R_k) = m_kN^{-1} + O(N^{-2})$. Define $R_{-1} = m_{-1} = 0$. As $\sum_{k=0}^{n'-1} m_k = 0$ we can write:

$$\mathbb{E}(\Delta S) = \sum_{k=0}^{\infty} \mathbb{E}[f(x_k) - f(x_k + R_{k-1} - R_k)]$$

$$= \sum_{k=0}^{\infty} \mathbb{E}[f(x_k) - f(x_k + R_{k-1}) + f(x_k + R_{k-1}) - f(x_k + R_{k-1} - R_k)]$$

$$= \sum_{k=0}^{n'-1} \mathbb{E}[f(x_k + R_{k-1}) - f(x_k + R_{k-1} - R_k) + f(x_{k+1}) - f(x_{k+1} + R_k)]$$

$$= \sum_{k=0}^{n'-1} \mathbb{E}[f(x_k + R_{k-1}) - f(x_k + R_{k-1} - R_k) + m_kN^{-1}\log(1 - x_0)$$

$$+ f(x_{k+1}) - f(x_{k+1} + R_k)].$$

We must then bound the difference, for $k = 0, 1, \ldots, n' - 1$, between terms

$$Z_k := \mathbb{E}[f(x_k + R_{k-1}) - f(x_k + R_{k-1} - R_k) + m_kN^{-1}\log(1 - x_0)$$

$$+ f(x_{k+1}) - f(x_{k+1} + R_k)]$$

and

$$\frac{1}{2N}(x'_k(1 - x_0) - x'_{k+1})\log \frac{x'_k(1 - x_0)}{x'_{k+1}}.$$

For $k \geq 1$, we list the possible values of the pair $(R_{k-1}, R_k)$ such that $R_k \neq 0$.

<table>
<thead>
<tr>
<th>$r_{k-1}$</th>
<th>$r_k$</th>
<th>$\mathbb{P}((R_{k-1}, R_k) = (r_{k-1}, r_k) \mid X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$+1/N$</td>
<td>$-1/N$</td>
<td>$x_{k-1}x_{k+1}$</td>
</tr>
<tr>
<td>$-1/N$</td>
<td>$+1/N$</td>
<td>$x_k(x_k - 1/N)$</td>
</tr>
<tr>
<td>$0$</td>
<td>$-1/N$</td>
<td>$x_{k+1}(1 - x_{k-1} - x_k - 1/N)$</td>
</tr>
<tr>
<td>$0$</td>
<td>$+1/N$</td>
<td>$x_k(1 - x_0 - x_k - x_{k+1})$</td>
</tr>
</tbody>
</table>
4.14 Technical lemmas relating to $\Delta S$

Using function $g(x) = f(x + 1/N) - f(x)$ we split $Z_k$ into larger and smaller terms.

$$Z_k = m_k N^{-1} \log(1 - x_0)$$
$$+ [g(x_{k+1} - 1/N) - g(x_k + 1/N)]x_{k-1}x_{k+1}$$
$$+ [g(x_k - 2/N) - g(x_{k+1})]x_k(x_k - 1/N)$$
$$+ [g(x_{k+1} - 1/N) - g(x_k)]x_{k+1}(1 - x_{k-1} - x_k - 1/N)$$
$$+ [g(x_k - 1/N) - g(x_{k+1})]x_k(1 - x_0 - x_k - x_{k+1})$$

$$= Z_k^A + Z_k^B$$

$$Z_k^A = m_k [N^{-1} \log(1 - x_0) + g(x_k) - g(x_{k+1})]$$
$$- x_{k-1}x_{k+1}[g(x_k + 1/N) - g(x_k)]$$

$$Z_k^B = - x_k(1 - x_0 - x_{k+1})[g(x_k) - g(x_k - 1/N)]$$
$$- x_{k+1}(1 - x_k)[g(x_{k+1}) - g(x_{k+1} - 1/N)]$$
$$- x_k(x_k - 1/N)[g(x_k - 1/N) - g(x_k - 2/N)]$$
$$- (x_k/N)[g(x_k - 1/N) - g(x_{k+1})]$$
$$- (x_{k+1}/N)[g(x_{k+1} - 1/N) - g(x_k)]$$

We write the equivalent terms for $k = 0$.

<table>
<thead>
<tr>
<th>$r_0$</th>
<th>$P(R_0 = r_0 \mid X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1/N$</td>
<td>$x_1(1 - x_0 - 1/N)$</td>
</tr>
<tr>
<td>$+1/N$</td>
<td>$x_0(1 - x_0 - x_1)$</td>
</tr>
</tbody>
</table>

$$Z_k = m_k N^{-1} \log(1 - x_0)$$
$$+ [g(x_1 - 1/N) - g(x_0)]x_1(1 - x_0 - 1/N)$$
$$+ [g(x_0 - 1/N) - g(x_1)]x_0(1 - x_0 - x_1)$$

$$= Z_0^A + Z_0^B$$

108
4.14 Technical lemmas relating to $\Delta S$

\[
Z_0^A = m_k [N^{-1} \log(1 - x_0) + g(x_0) - g(x_1)] \\
Z_0^B = -x_0 (1 - x_0 - x_1) [g(x_0) - g(x_0 - 1/N)] \\
\quad - x_1 (1 - x_0) [g(x_1) - g(x_1 - 1/N)] \\
\quad - (x_1/N) [g(x_1 - 1/N) - g(x_0)]
\]

So by definition, $\mathbb{E}(\Delta S \mid X = x) = \sum_k Z_k^A + Z_k^B$. We will first bound $\sum_k Z_k^B$. For $x, y \in \{0, 1/N, \ldots, 1\}$,

(i) $x \ [g(x) - g(x - 1/N)] \leq 2N^{-2} \log 2$, and

(ii) $(x/N)(g(x - 1/N) - g(y)) \leq xN^{-2}(1 + \log N)$.

Therefore $\sum Z_k^B \geq -6n'N^{-2} \log 2 + 2N^{-2}(1 + \log N) \geq -5n'N^{-2}$. We will now show that if $x_0 \leq 1 - 10^{-9}$,

\[
m_k (N^{-1} \log(1 - x_0) + g(x_k) - g(x_{k+1})) - x_{k-1} x_{k+1} [g(x_k + 1/N) - g(x_k)] \\
\geq (2N)^{-1} (x'_k (1 - x_0) - x'_{k+1}) \log \frac{x'_k (1 - x_0)}{x'_{k+1}} - 5N^{-2}.
\]  

(4.26)

Write $x_k = a/N$, $x_{k+1} = b/N$. Let $a' = a \lor e^{-1}$, $b' = b \lor e^{-1}$. It is now sufficient to show $A \geq B/2 - 5$,

\[
A = (a(1 - x_0) - b)[\log(1 - x_0) + f(a + 1) - f(a) - f(b + 1) + f(b)] \\
\quad - b[f(a + 2) - 2f(a + 1) + f(a)], \\
B = (a' (1 - x_0) - b') \log \frac{a'(1 - x_0)}{b'}.
\]

We will use the following simple facts in our case by case analysis.

(a) If $k \geq 1$,

\[
1 + \log k \leq f(k + 1) - f(k) \leq 1 + \log k + (2k)^{-1}.
\]

(b) For all $c, x \geq 0$, $x \log x - cx \geq -\exp(c - 1)$.

(c) For all $k \in \mathbb{N}$, $0 \leq k[f(k + 2) - 2f(k + 1) + f(k)] \leq 1$.

We now consider cases (i) $a = b = 0$, (ii) $0 = a < b$, (iii) $0 = b < a$, (iv) $0 < b < a(1 - x_0)$ and (v) $0 < a(1 - x_0) \leq b$. 

109
(i) If $a = b = 0$:

$$A = 0$$
$$B = -x_0 e^{-1} \log(1 - x_0) \leq 10$$
$$A \geq B/2 - 5$$

(ii) If $0 = a < b$:

$$A = b \left[ \log \frac{b}{1 - x_0} + (b + 1) \log(1 + b^{-1}) \right] - b[2 \log 2]$$
$$\geq b \left[ \log \frac{b}{1 - x_0} + 1 - 2 \log 2 \right]$$
$$B = (b - e^{-1}) \left[ \log \frac{b}{1 - x_0} + 1 \right]$$
$$A - B/2 \geq b [\log b + 1 - 4 \log 2]/2$$
$$\geq -5$$

(iii) If $0 = b < a$, set $t = a(1 - x_0)$.

$$A \geq t[\log t + 1]$$
$$B = (t - e^{-1})[\log t + 1]$$
$$A - B/2 \geq -5$$

(iv) If $0 < b < a(1 - x_0)$, set $t = a(1 - x_0)/b \geq 1$.

$$A \geq b(t - 1) \left[ \log t - \frac{1}{2b} \right] - 1$$
$$B = b(t - 1) \log t$$
$$A - B/2 \geq (t - 1)[\log t - 1]/2 - 1$$
$$\geq -5$$

(v) If $0 < a(1 - x_0) \leq b$, set $t = b/a(1 - x_0) \geq 1$.

$$A \geq a(1 - x_0)(t - 1) \left[ \log t - \frac{1}{2a} \right] - t(1 - x_0)$$
$$B \geq a(1 - x_0)(t - 1) \log t$$
$$A - B/2 \geq \frac{1}{2}(1 - x_0)(t - 1)[\log t - 1] - t(1 - x_0) \geq -5$$
4.14 Technical lemmas relating to $\Delta S$

Proof of Lemma 4.20. The proof is similar to the proof of Lemma 4.16. In place of inequality (4.26) we must show,

$$m_k(N^{-1} \log(1 - x_0) + g(x_k) - g(x_{k+1})) - x_{k-1} x_{k+1}[g(x_k + 1/N) - g(x_k)]$$

$$\geq \frac{1}{N} \frac{m_k^2}{\max\{x_k(1 - x_0), x_{k+1}\}} - \frac{5}{N^2}.$$ 

This is equivalent to showing $A \geq C - 5$, with

$$A = (a(1 - x_0) - b)(\log(1 - x_0) + f(a + 1) - f(a) - f(b + 1) + f(b))$$

$$- b[f(a + 2) - 2f(a + 1) + f(a)],$$

$$C = (a(1 - x_0) - b)^2 / \max\{a(1 - x_0), b\}.$$ 

We consider the following cases: (i) $a = b = 0$, (ii) $0 = b < a$ (iii) $0 = a < b$,

(iv) $0 < b < a(1 - x_0)$ and (v) $0 < a(1 - x_0) \leq b$.

(i) If $a = b = 0$ then $A = C = 0$.

(ii) If $0 = b < a$, set $t = a(1 - x_0)$.

$$A - C \geq t \log t - t \geq -5$$ 

(iii) If $b > a = 0$:

$$A - C \geq b \log \frac{b}{1 - x_0} - b (1 + 2 \log 2)$$

$$\geq -5$$

(iv) If $0 < b < a(1 - x_0)$, set $t = a(1 - x_0)/b > 1$.

$$A \geq b(t - 1) \left[ \log t - \frac{1}{2b} \right] - 1$$

$$C = b(t - 1) \left[ 1 - \frac{1}{t} \right]$$

$$A - C \geq (t - 1) \left[ \log t - 3/2 \right] - 1 \geq -5$$

111
(v) If $0 < a(1 - x_0) \leq b$, set $t = b/(a(1 - x_0))$.

$$A \geq a(1 - x_0)(t - 1) \left[ \log t - \frac{1}{2a} \right] - t(1 - x_0)$$

$$C = a(1 - x_0)(t - 1) \left[ 1 - \frac{1}{t} \right]$$

$$A - C \geq (1 - x_0)[(t - 1)(\log t - 5/2) - 1] \geq -5$$

\[\Box\]

Proof of Lemma 4.21: Of course $\text{var}(\Delta S | X = x) \leq \mathbb{E}((\Delta S)^2 | X = x)$. Once we have bounded $\mathbb{E}((\Delta S)^2 | X = x)$ the result follows by Lemma 4.16. Sum over the number of balls in the source box, $i$, and the sink boxes, $j$.

\[
\sum_{i>0} \sum_{j \geq 0} x_i (x_j - \delta_{ij}/N) \left[ g(x_i - N^{-1}) - g(x_{i-1}) + g(x_j - N^{-1}) - g(x_{j+1}) \right]^2 \\
\leq \sum_{i>0} \sum_{j \geq 0} x_i (x_j - \delta_{ij}/N) \left[ -g(x_{i-1}) - N^{-1} \log(1 - x_0) + g(x_i - N^{-1}) \right. \\
+ \left. g(x_j - N^{-1}) + N^{-1} \log(1 - x_0) - g(x_{j+1}) \right]^2 \\
\leq 2 \sum_{i>0} \sum_{j \geq 0} x_i x_j \left[ (g(x_{i-1}) + N^{-1} \log(1 - x_0) - g(x_i - N^{-1}))^2 \right. \\
+ \left. (g(x_j - N^{-1}) + N^{-1} \log(1 - x_0) - g(x_{j+1}))^2 \right] \\
\leq 2 \sum_{k=0}^{n'-1} x_k (1 - x_0) (g(x_k - N^{-1}) + N^{-1} \log(1 - x_0) - g(x_{k+1}))^2 \\
+ x_{k+1} (g(x_k) + N^{-1} \log(1 - x_0) - g(x_{k+1} - N^{-1}))^2 \\
\]

Let

\[
A = x_k (1 - x_0) \left[ g(x_k - N^{-1}) + N^{-1} \log(1 - x_0) - g(x_{k+1}) \right]^2 \\
+ x_{k+1} \left[ g(x_k) + N^{-1} \log(1 - x_0) - g(x_{k+1} - N^{-1}) \right]^2, \\
B = (x_k' (1 - x_0) - x_{k+1}') \log x_k' (1 - x_0) / x_{k+1}'.
\]

With $x_0 \leq 2/3$ it is straightforward to check $N^3 A \leq 3NB \log N + 100$; consider the cases $x_k = x_{k+1} = 0, x_{k+1} > x_k = 0, x_k > x_{k+1} = 0$ and $x_k, x_{k+1} > 0$.  \[\Box\]
References


