Interacting Particle Systems

J.M. Swart

May 14, 2009

2

Contents

Preface

Interacting particle systems, in the sense we will be using the word in these lecture notes, are countable systems of locally interacting Markov processes. More precisely, in an interacting particle system one has a lattice -the canonical choice is the d-dimensional integer lattice $\mathbb{Z}^d$- such that on each site in the lattice, there sits a continuous-time Markov process with a finite state space whose jump rates depend on the states of the Markov processes situated at near-by sites. Interacting particle systems are often used as extremely simplified 'toy models' for certain stochastic phenomena that involve a spatial structure.

Yet, the apparent simplicity of these models is treacherous. Although it is usually easy to define an interacting particle system, it is often much harder to prove anything nontrivial about its behavior. With a few exceptions, explicit calculations tend not to be feasible, hence one has to be satisfied with qualitative statements and some explicit bounds. Despite intensive research over more than thirty years, some simple statements about easy-to-formulate problems still remain open and others that have been solved required nontrivial and lengthy proofs.

As a reward for all this, on the other hand, it turns out that despite their simple rules, interacting particle systems are often remarkably subtle models that capture the sort of phenomena one is interested in much better than might initially be expected. Thus, while it may seem outrageous to assume that "Plants of a certain type occupy points in the square lattice $\mathbb{Z}^2$, live for an exponential time with mean one, and place seeds on unoccupied neighboring sites with rate λ" it turns out that making the model more realistic often does not change much in its overall behavior. Indeed, there is a general philosophy in the field, that is still unsufficiently understood, which says that interacting particle systems come in 'universality classes' with the property that all models in one class have roughly the same behavior.

As a mathematical discipline, the subject of interacting particle systems is still relatively young. It started around 1970 with the work of R.L. Dobrushin and F. Spitzer, and many other authors joining in the next few years. By 1975, general existence and uniqueness questions had been settled, four classic models had been introduced (the exclusion process, the stochastic Ising model, the voter model and the contact process), and elementary (and less elementary) properties of these models had been proved. In 1985, when Liggett's published his famous book [Lig85], the subject had established itself as a mature field of study. Since then, it has continued to grow rapidly, to the point where it is impossible to accurately capture the state of the art in a single book. Indeed, it would be possible to write a book on each of the four classic models mentioned above, while many new models

have been introduced and studied.

While interacting particle systems, in the narrow sense indicated above, have apparently not been the subject of mathematical study before 1970, the subject has close links to some problems that are considerably older. In particular, the Ising model (without time evolution) has been studied since 1925 while both the Ising model and the contact process have close connections to percolation, which has been studied since 1957. In recent years, more links between interacting particle systems and other, older subjects of mathematical research have been established, and the field continues to recieve new impulses not only from the applied, but also from the more theoretical side.

Chapter 1

Markov processes

1.1 Introduction

Interacting particle systems are countable collections of interacting continuous-time Markov processes with finite state spaces. Such Markov processes are sometimes called 'continuous-time Markov chains' or even more confusing 'Markov chains with continuous parameter'. Of course, they are not chains, but stochastic processes with a continuous time parameter. Nevertheless, because of their finite state space, they are effectively equivalent to certain 'embedded' Markov chains and their analysis is very similar to that of finite-state Markov chains.

In this chapter we review some elementary theory of Markov processes. We focus mainly on continuous-time, finite-state Markov processes. One is tempted to say that these processes are easy, but, of course, this is only true as long as their state space is small. If one is interested in limiting properties as the size of the state space is sent to infinity, then one may run into difficult problems, some of which are closely related to the interacting particle systems we are interested in.

1.2 Conditional probabilities

By definition, a *Polish space* is a separable topological space E on which there exists a complete metric generating the topology. Polish spaces are particularly nice for doing probability theory on. We equip a Polish space E standardly with the *Borel-σ-field* $\mathcal{B}(E)$ generated by the open subsets of E. We let $B(E)$ denote space of bounded, real, $\mathcal{B}(E)$-measurable functions on E. Polish spaces have nice reproducing properties; for example, if E is a Polish space and F is a closed or an open subset of E, then the space F is also Polish (in the embedded topology).

Also, if $E_1, E_2, \ldots$ is a finite or countably infinite sequence of Polish spaces, then the product space $E_1 \times E_2 \times \cdots$ equipped with the product topology is again Polish, and the Borel-σ-field on the product space coincides with the product-σ-field of the Borel-σ-fields on the individual spaces.

Let E, F be Polish spaces. By definition, a *probability kernel* from E to F is a function $K : E \times \mathcal{B}(F) \to \mathbb{R}$ such that

    (i) $K(x, \cdot)$ is a probability measure on F for each $x \in E$,

    (ii) $K(\cdot, A)$ is a real measurable function on E for each $A \in \mathcal{B}(F)$.

Proposition 1.1 (Decomposition of probability measures) *Let E, F be Polish spaces and let μ be a probability measure on $E \times F$. Then there exist a (unique) probability measure ν on E and a (in general not unique) probability kernel K from E to F such that*

$$\int f \mathrm{d}\mu = \int_E \nu(\mathrm{d}x) \int_F K(x, \mathrm{d}y) f(x, y) \qquad \big(f \in B(E \times F)\big). \qquad (1.1)$$

If K, K' are probability kernels from E to F such that (1.1) holds, then there exists a set $N \in \mathcal{B}(E)$ with $\nu(N) = 0$ such that $K(x, \cdot,) = K(x, \cdot,)$ for all $x \in E \backslash N$. Conversely, if ν is a probability measure on E and K is a probability kernel from E to F, then formula (1.1) defines a unique probability measure on $E \times F$.

Note that it follows obviously from (1.1) that

$$\nu(A) = \mu(A \times F) \qquad \big(A \in \mathcal{B}(E)\big),$$

i.e., ν is the first *marginal* of the probability measure μ.

If X and Y are random variables, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and taking values in E and F, respectively, then setting

$$\mu(A) := \mathbb{P}[(X, Y) \in A] \qquad \big(A \in \mathcal{B}(E \times F)\big)$$

defines a probability law on $E \times F$ which is called the joint law of X and Y. By Proposition 1.1, we may write μ in the form (1.1) for some probability law ν on E and probability kernel K from E to F. We observe that

$$\nu(A) = \mathbb{P}[X \in A] \qquad \big(A \in \mathcal{B}(E)\big),$$

i.e., ν is the law of X. We will often denote the law of X by $\mathbb{P}[X \in \cdot]$. Moreover, we introduce the notation

$$\mathbb{P}\big[Y \in A \,\big|\, X = x\big] := K(x, A) \qquad \big(x \in E, \ A \in \mathcal{B}(F)\big),$$

where $K(x, A)$ is the probability kernel from E to F defined in terms of μ as in (1.1). Note that $K(x, A)$ is defined uniquely for a.e. x with respect to the law of X. We call $\mathbb{P}[Y \in \cdot \,|\, X = x]$ the conditional law of Y given X. Note that with the notation we have just introduced, formula (1.1) takes the form

$$\mathbb{E}[f(X, Y)] = \int_E \mathbb{P}[X \in \mathrm{d}x] \int_F \mathbb{P}\big[Y \in \mathrm{d}y \,\big|\, X = x\big] f(x, y). \qquad (1.2)$$

Closely related to this, one also defines

$$\mathbb{P}\big[Y \in A \,\big|\, X\big] := K(X, A) \qquad \big(A \in \mathcal{B}(F)\big).$$

Note that this is the random variable (defined on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$) obtained by plugging X into the function $x \mapsto K(x, A)$.

If $f : F \to \mathbb{R}$ is a measurable function such that $E[|f(Y)|] < \infty$, then we let

$$\mathbb{E}\big[f(Y) \,\big|\, X = x\big] := \int_F \mathbb{P}\big[Y \in \mathrm{d}y \,\big|\, X = x\big] f(y)$$

denote the conditional expectation of $f(Y)$ given X. Note that for fixed f and Y, the map $x \mapsto \mathbb{E}\big[f(Y) \,\big|\, X = x\big]$ is a measurable real function on E. Plugging X into this function yields a random variable which we denote by $\mathbb{E}[f(Y)|X]$. We observe that for each $g \in B(E)$, one has

$$\begin{aligned}
\mathbb{E}\big[g(X)\mathbb{E}[f(Y)|X]\big] &= \int_E \mathbb{P}[X \in \mathrm{d}x] g(x) \mathbb{E}[f(Y)|X = x] \\
&= \int_E \mathbb{P}[X \in \mathrm{d}x] g(x) \int_F \mathbb{P}[Y \in \mathrm{d}y | X = x] f(y) \\
&= \int_E \mathbb{P}[X \in \mathrm{d}x] \int_F \mathbb{P}[Y \in \mathrm{d}y | X = x] g(x) f(y) \\
&= \int_{E \times F} \mathbb{P}[(X, Y) \in \mathrm{d}(x, y)] g(x) f(y) = \mathbb{E}[g(X) f(Y)].
\end{aligned}$$

Moreover, since $E[f(Y)|X]$ can be written as a function of X, it is easy to check that $E[f(Y)|X]$ is measurable with respect to the σ-field generated by X. One may take these properties as an alternative definition of $\mathbb{E}[f(Y)|X]$. More generally, if R is a real-valued random variable with $\mathbb{E}[|R|] < \infty$, defined on some probability

space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-field, then there exists an a.s. (with respect to the underlying probability measure $\mathbb{P}$) unique random variable $\mathbb{E}[R|\mathcal{G}]$ such that $E[R|\mathcal{G}]$ is $\mathcal{G}$-measurable and

$$\mathbb{E}\big[G\,\mathbb{E}[R|\mathcal{G}]\big] = \mathbb{E}[GR] \qquad \forall \text{ bounded } \mathcal{G}\text{-measurable } G.$$

In the special case that $R = f(Y)$ and $\mathcal{G}$ is the σ-field generated by X one recovers $\mathbb{E}[f(Y)|X] = \mathbb{E}[R|\mathcal{G}]$.

1.3   Poisson point processes

Let E be a Polish space. Recall that a sequence of finite measures μ_n *converges weakly* to a limit μ, denoted as $\mu_n \Rightarrow \mu$, if and only if

$$\int f \mathrm{d}\mu_n \underset{n\to\infty}{\longrightarrow} \int f \mathrm{d}\mu \qquad \big(f \in \mathcal{C}_{\mathrm{b}}(E)\big),$$

where $\mathcal{C}_{\mathrm{b}}(E)$ denotes the space of bounded continuous real functions on E. We let $\mathcal{M}(E)$ denote the space of finite measures on E, equipped with the topology of weak convergence. It can be shown that $\mathcal{M}(E)$ is Polish. We let

$$\mathcal{N}(E) := \Big\{\nu \in \mathcal{M}(E) : \exists n \geq 0,\ x_1, \ldots, x_n \in E \text{ s.t. } \nu = \sum_{i=1}^{n} \delta_{x_i}\Big\}$$

denote the space of all *counting measures* on E, i.e., all measures that can be written as a finite sum of delta-measures. Being a closed subset of $\mathcal{M}(E)$, the space $\mathcal{N}(E)$ is again Polish.

For any counting measure $\nu \in \mathcal{N}(E)$ and $f \in B(E)$ we introduce the notation

$$f^\nu := \prod_{i=1}^{n} f(x_i) \qquad \text{where} \quad \nu = \sum_{i=1}^{n} \delta_{x_i},$$

with $f^0 := 1$ (where 0 denotes the counting measure that is identically zero). It is easy to see that $f^\nu f^{\nu'} = f^{\nu+\nu'}$. Let $\nu = \sum_{i=1}^{n} \delta_{x_i}$ be a counting measure, let $\phi \in B(E)$ satisfy $0 \leq \phi \leq 1$, and let $\chi_1, \ldots, \chi_n$ be independent *Bernoulli random variables* (i.e., random variables with values in $\{0,1\}$) with $\mathbb{P}[\chi_i = 1] = \phi(x_i)$. Then the random counting measure

$$\nu' := \sum_{i=1}^{n} \chi_i \delta_{x_i}$$

is called a ϕ-*thinning* of the counting measure ν. Note that

$$\mathbb{P}[\nu' = 0] = \prod_{i=1}^{n} \mathbb{P}[\chi_i = 0] = (1 - \phi)^{\nu}.$$

More generally, one has

$$\mathbb{E}\big[(1 - f)^{\nu'}\big] = (1 - f\phi)^{\nu} \qquad \big(f \in B(E),\ 0 \leq f \leq 1\big). \tag{1.3}$$

(Setting $f = 0$ here yields the previous formula.) To see this, note that if $\chi'_1, \ldots, \chi'_n$ are Bernoulli random variables with $\mathbb{P}[\chi'_i = 1] = f(x_i)$, independent of each other and of the χ_i's, and

$$\nu'' := \sum_{i=1}^{n} \chi'_i \chi_i \delta_{x_i},$$

then, since conditional on ν', the measure ν'' is distributed as an f-thinning of ν', one has

$$\mathbb{P}\big[\nu'' = 0\big] = \mathbb{E}\big[(1 - f)^{\nu'}\big],$$

while on the other hand, since ν'' is an $f\phi$-thinning of ν, one has $\mathbb{P}[\nu'' = 0] = (1 - f\phi)^{\nu}$. One can prove that (1.3) charcterizes the law of the random counting measure ν' uniquely.

Proposition 1.2 (Poisson counting measure) *Let E be a Polish space and let μ be a finite measure on E. Then there exists a random counting measure ν on E whose law is uniquely characterized by*

$$\mathbb{E}\big[(1 - f)^{\nu}\big] = e^{-\int f \mathrm{d}\mu} \qquad \big(f \in B(E),\ 0 \leq f \leq 1\big). \tag{1.4}$$

If $A_1, \ldots, A_n$ are disjoint measurable subsets of E, then $\nu(A_1), \ldots, \nu(A_n)$ are independent Poisson distributed random variables with mean $\mathbb{E}[\nu(A_i)] = \mu(A_i)$ $(i = 1, \ldots, n)$.

Proof (sketch) We can find counting measures $\nu_n := \sum_{i=1}^{N_n} \delta_{x_{n,i}}$ such that

$$\mu_n := \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{x_{n,i}} \underset{n \to \infty}{\Longrightarrow} \mu.$$

Let ν'_n be a thinning of ν_n with the constant function $1/N_n$. Then, for any $f \in \mathcal{C}(E)$ satisfying $\varepsilon \leq f \leq 1$ for some $\varepsilon > 0$, one has

$$\mathbb{E}\big[(1 - f)^{\nu'_n}\big] = (1 - \tfrac{1}{N_n} f)^{\nu_n} = e^{\int \log(1 - \frac{1}{N_n} f) \mathrm{d}\nu_n} = e^{\int N_n \log(1 - \frac{1}{N_n} f) \mathrm{d}\mu_n} \underset{n \to \infty}{\longrightarrow} e^{-\int f \mathrm{d}\mu},$$

where we have used that $N_n \log(1 - \frac{1}{N_n} f) \to -f$ and $\mu_n \Rightarrow \mu$. This can be used to show that the random counting measures ν'_n converge weakly in law to a limiting random counting measure ν which satisfies (1.4) for all any $f \in \mathcal{C}(E)$ satisfying $\varepsilon \leq f \leq 1$ for some $\varepsilon > 0$. The generalization to general $f \in B(E)$ with $0 \leq f \leq 1$ follows by approximation. Finally, (1.4) can be used to show that if $A_1, \ldots, A_n$ are disjoint measurable sets, then $\nu(A_1), \ldots, \nu(A_n)$ are independent Poisson distributed random variables with means as indicated in the theorem. ∎

The random measure ν whose law is defined in Proposition 1.2 is called a *Poisson counting measure* with *intensity* μ.

Lemma 1.3 (Sum of independent Poisson counting measures) *Let E be a Polish space and let ν_1, ν_2 be independent Poisson counting measures on E with intensities μ_1, μ_2, respectively. Then $\nu_1 + \nu_2$ is a Poisson counting measure with intensity $\mu_1 + \mu_2$.*

Proof One can straightforwardly check this from (1.4). Note that thinnings have a similar property, so the statement is also rather obvious from our approximation of Poisson counting measures with thinnings. ∎

Set

$$\mathcal{N}_1(E) := \big\{\nu \in \mathcal{N}(E) : \nu(\{x\}) \in \{0, 1\} \ \forall x \in E\big\}.$$

Since $\mathcal{N}_1(E)$ is an open subset of $\mathcal{N}(E)$, it is a Polish space. We can identify elements of $\mathcal{N}_1(E)$ with finite subsets of E; indeed, $\nu \in \mathcal{N}_1(E)$ if and only if $\nu = \sum_{x \in \Delta} \delta_x$ for some finite $\Delta \subset E$. We skip the proof of the following lemma.

Lemma 1.4 (Poisson point set) *Let μ be a finite measure on a Polish space E and let ν be a Poisson counting measure with intensity μ. Then $\mathbb{P}[\nu \in \mathcal{N}_1(E)] = 1$ if and only if μ is* nonatomic, *i.e., $\mu(\{x\}) = 0$ for all $x \in E$.*

If μ is a nonatomic measure on some Polish space, ν is a Poisson counting measure with intensity μ, and Δ is the random finite set associated with ν, then we call Δ a *Poisson point set* with intensity μ.

If E is a locally compact space and μ is a locally compact measure on E (i.e., a measure such that $\mu(K) < \infty$ for each compact $K < \infty$), then Poisson counting measures and Poisson point sets with intensity μ are defined analogously to the finite measure case. We will in particular be interested in the case that $E = [0, \infty)$ and μ is a multiple of Lebesgue measure.

Lemma 1.5 (Exponential times) *Let $r > 0$ be a constant and let $(\sigma_k)_{k \geq 1}$ be i.i.d. exponentially distributed random variables with mean $\mathbb{E}[\sigma_k] = 1/r$ ($k \geq 1$). Set $\tau_n := \sum_{k=1}^{n} \sigma_k$ ($n \geq 1$). Then $\{\tau_n : n \geq 1\}$ is a Poisson point set on $[0, \infty)$ with indensity $r\mathrm{d}t$, where $\mathrm{d}t$ denotes Lebesgue measure.*

Proof Discrete approximation. ∎

1.4 Markov chains

Let E be a Polish space. By definition, a *Markov chain* with *state space* E is a discrete-time stochastic process $(X_k)_{k \geq 0}$ such that for all $0 \leq l \leq m \leq n$

$$\mathbb{P}\big[(X_l, \ldots, X_m) \in A, \ (X_m, \ldots, X_n) \in B \,\big|\, X_m\big]$$
$$= \mathbb{P}\big[(X_l, \ldots, X_m) \in A \,|\, X_m\big]\,\mathbb{P}(X_m, \ldots, X_n) \in B \,\big|\, X_m\big] \quad \text{a.s.} \tag{1.5}$$

for each $A \in \mathcal{B}(E^{m-l+1})$ and $B \in \mathcal{B}(E^{n-m+1})$. In words, formula (1.5) says that the past and the future are conditionally independent given the present. A similar definition applies to Markov chains $(X_k)_{k \in I}$ where $I \subset \mathbb{Z}$ is some interval (possibly unbounded on either side). It can be shown that (1.5) is equivalent to the statement that

$$\mathbb{P}\big[X_k \in A \,\big|\, (X_0, \ldots, X_{k-1})\big] = \mathbb{P}\big[X_k \in A \,\big|\, X_{k-1}\big] \quad \text{a.s.} \tag{1.6}$$

for each $k \geq 1$ and $A \in \mathcal{B}(E)$. For any sequence $(X_k)_{k \geq 0}$ of E-valued random variables, repeated application of (1.2) gives

$$\mathbb{E}\big[f(X_0, \ldots, X_n)\big] = \int \mathbb{P}\big[(X_0, \ldots, X_{n-1}) \in \mathrm{d}(x_0, \ldots, x_{n-1})\big]$$
$$\times \int \mathbb{P}\big[X_n \in \mathrm{d}x_n \,\big|\, (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})\big] f(x_0, \ldots, x_n)$$
$$= \int \mathbb{P}\big[(X_0, \ldots, X_{n-2}) \in \mathrm{d}(x_0, \ldots, x_{n-2})\big]$$
$$\times \int \mathbb{P}\big[X_{n-1} \in \mathrm{d}x_{n-1} \,\big|\, (X_0, \ldots, X_{n-2}) = (x_0, \ldots, x_{n-2})\big]$$
$$\times \int \mathbb{P}\big[X_n \in \mathrm{d}x_n \,\big|\, (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})\big] f(x_0, \ldots, x_n)$$
$$= \int \mathbb{P}\big[X_0 \in \mathrm{d}x_0\big] \int \mathbb{P}\big[X_1 \in \mathrm{d}x_1 \,\big|\, X_0 = x_0\big] \int \mathbb{P}\big[X_2 \in \mathrm{d}x_2 \,\big|\, (X_0, X_1) = (x_0, x_1)\big]$$
$$\times \cdots \times \int \mathbb{P}\big[X_n \in \mathrm{d}x_n \,\big|\, (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})\big] f(x_0, \ldots, x_n).$$

If $(X_k)_{k \geq 0}$ is a Markov chain, then by (1.6) this simplifies to

$$\mathbb{E}\big[f(X_0, \ldots, X_n)\big]$$
$$= \int \mathbb{P}\big[X_0 \in \mathrm{d}x_0\big] \int \mathbb{P}\big[X_1 \in \mathrm{d}x_1 \,\big|\, X_0 = x_0\big]$$
$$\times \cdots \times \int \mathbb{P}\big[X_n \in \mathrm{d}x_n \,\big|\, X_{n-1} = x_{n-1}\big] f(x_0, \ldots, x_n).$$

As this formula shows, the law of a Markov chain $(X_k)_{k \geq 0}$ is uniquely determined by its *initial law* $\mathbb{P}[X_0 \in \cdot\,]$ and its *transition probabilities* $\mathbb{P}[X_n \in \mathrm{d}x_n \,|\, X_{n-1} = x_{n-1}]$ ($k \geq 1$). By definition, a Markov chain is *time-homogeneous* if its transitition probabilities are the same in each time step, more precisely, if there exists a probability kernel $P(x, \mathrm{d}y)$ on E such that

$$\mathbb{P}[X_n \in \cdot \,|\, X_{n-1} = x] = P(x, \cdot) \quad \text{for a.e. } x \text{ w.r.t. } \mathbb{P}[X_{n-1} \in \cdot\,],$$

which is equivalent to

$$\mathbb{P}[X_n \in \cdot \,|\, X_{n-1}] = P(X_{n-1}, \cdot) \quad \text{a.s.} \tag{1.7}$$

We will usually be interested in time-homogeneous Markov chains only. In fact, we will often fix a probability kernel $P(x, \mathrm{d}y)$ on E and use the word 'Markov chain' when we in fact mean the collection of all (time-homogeneous) Markov chains whose transition probabilities are given by $P(x, \mathrm{d}y)$ in each time step. Note that we can combine (1.6) and (1.7) in a single condition: a sequence $(X_k)_{k \geq 0}$ of E-valued random variables is a Markov chain with transition probability $P(x, \mathrm{d}y)$ (and arbitrary initial law) if and only if

$$\mathbb{P}[X_k \in \cdot \,|\, (X_0, \ldots, X_{k-1})] = P(X_{k-1}, \cdot) \quad \text{a.s.} \quad (k \geq 1), \tag{1.8}$$

which is equivalent to

$$\mathbb{E}[f(X_k) \,|\, (X_0, \ldots, X_{k-1})] = Pf(X_{k-1}) \quad \text{a.s.} \quad (k \geq 1, \ f \in B(E)). \tag{1.9}$$

If $K(x, \mathrm{d}y)$ is a probability kernel on a Polish space E, then setting

$$Kf(x) := \int_E K(x, \mathrm{d}y) f(y) \qquad (x \in E \ \ f \in B(E))$$

defines a linear operator $K : B(E) \to B(E)$. We sometimes use this notation also if f is not a bounded function, as long as the integral is well-defined for every x. If K, L are probability kernels on E, then we define the *composition* of K and L as

$$(KL)(x, A) := \int_E K(x, \mathrm{d}y) L(y, A) \qquad (x \in E \ \ f \in B(E)).$$

It is straightforward to check that this formula defines a probability kernel on E. If $K : B(E) \to B(E)$ and $L : B(E) \to B(E)$ are the linear operators associated with the probability kernels $K(x, \mathrm{d}y)$ and $L(x, \mathrm{d}y)$, then the linear operator associated

with the composed kernel $(KL)(x, \mathrm{d}y)$ is just KL, the composition of the linear operators K and L.

If $(X_k)_{k\geq 0}$ is a Markov chain with transition probability $P(x, \mathrm{d}y)$, then we let P denote the linear operator from $B(E)$ to $B(E)$ associated with $P(x, \mathrm{d}y)$ and we let P^n denote the n-fold composition of the kernel / linear operator P with itself, where $P^0(x, \mathrm{d}y) := \delta_x(\mathrm{d}y)$ (the delta measure in x). With this notation we may generalize (1.8) to

$$\mathbb{P}\big[X_{k+n} \in \cdot \,\big|\, (X_0, \ldots, X_k)\big] = P^n(X_k, \cdot) \quad \text{a.s.} \quad (k, n \geq 0), \qquad (1.10)$$

which is equivalent to

$$\mathbb{E}\big[f(X_{k+n}) \,\big|\, (X_0, \ldots, X_k)\big] = P^n f(X_k) \quad \text{a.s.} \quad \big(k, n \geq 0, \ f \in B(E)\big). \qquad (1.11)$$

1.5 Finite-state Markov processes

As opposed to Markov chains, dealing with Markov processes with continuous time is in general technically more difficult. In the present section we will restrict ourselves to Markov processes with continuous time but finite state space. As we will see, these processes are very similar to Markov chains. In fact, they are in many ways 'effectively' equivalent to certain 'embedded' Markov chains.

Let S be a finite set. If μ is a probability measure on S and $i \in S$, then we write $\mu(i) := \mu(\{i\})$ and likewise, if $K(i, A)$ is a probability kernel on S, then we write $K(i, j) := K(i, \{j\})$. Now probability kernels correspond to matrices and the composition of two kernels corresponds to the usual matrix product. If $K(i, j)$ is a probability kernel on S and K is the associated linear operator on $B(S)$, then

(i) $f \geq 0$ implies $Kf \geq 0$ $\big(f \in B(S)\big)$,

(ii) $K1 = 1$,

where we write 1 to denote the constant function $1(i) := 1$ $(i \in S)$. Conversely, it is easy to see that each linear operator $K : B(S) \to B(S)$ that satisfies the properties (i) and (ii) corresponds to a probability kernel on S.

By definition, a (continuous) *transition probability* on S is a collection $(P_t(i, j))_{t\geq 0}$ of probability kernels on S such that

(i) $\displaystyle\lim_{t \downarrow 0} P_t(i, j) = P_0(i, j) = \delta_i(j)$ $\qquad (i, j \in S)$,

(ii) $\displaystyle\sum_j P_s(i, j)P_t(j, k) = P_{s+t}(i, k)$ $\qquad (s, t \geq 0, \ i, k \in S)$.

In terms of the associated linear operators, this says that

$$
\begin{aligned}
&\text{(i)} \quad \lim_{t \downarrow 0} P_t f = P_0 f = f \qquad \big(f \in B(S) \big), \\
&\text{(ii)} \quad P_s P_t = P_{s+t} \qquad\qquad (s, t \geq 0),
\end{aligned}
$$

i.e., the operators $(P_t)_{t \geq 0}$ form a (continuous) semigroup. Such a semigroup is called a *Markov semigroup.*

By definition, we let $\mathcal{D}_S[0, \infty)$ denote the space of all piecewise constant right-continuous functions from $[0, \infty)$ to S. Equivalently, this is the space of functions $w : [0, \infty) \to S$ such that

$$
\begin{aligned}
&\text{(i)} \quad \lim_{t \downarrow s} w_t = w_s \qquad\qquad\quad (s \geq 0), \\
&\text{(ii)} \quad \lim_{t \uparrow s} w_t =: w_{s-} \text{ exists} \qquad (s > 0).
\end{aligned}
$$

We call $\mathcal{D}_S[0, \infty)$ the space of *cadlag* functions from $[0, \infty)$ to S. (After the French **c**ontinue **à d**roit, **l**imite **à g**auche.) It is possible to equip this space with a (rather natural) topology such that $\mathcal{D}_S[0, \infty)$ is a Polish space; we will skip the details. By definition, we say that an S-valued stochastic process $(X_t)_{t \geq 0}$ defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has *cadlag sample paths* if for every $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is cadlag. We may view such a stochastic process as a single random variable, taking values in the Polish space $\mathcal{D}_S[0, \infty)$. Now

$$
\mathbb{P}\big[(X_t)_{t \geq 0} \in A\big] \qquad \big(A \in \mathcal{B}(\mathcal{D}_S[0, \infty)) \big)
$$

is a probability law on $\mathcal{D}_S[0, \infty)$ called the *law* of the process $(X_t)_{t \geq 0}$. It is not difficult to prove that this law is uniquely determined by the *finite dimensional distributions*

$$
\mathbb{P}\big[(X_{t_1}, \ldots, X_{t_n}) \in A\big] \qquad (A \subset S^n).
$$

We recall that a *filtration* is a collection $(\mathcal{F}_t)_{t \geq 0}$ of σ-fields such that $s \leq t$ implies $\mathcal{F}_s \subset \mathcal{F}_t$. If $(X_t)_{t \geq 0}$ is a stochastic process, then the *filtration generated by* $(X_t)_{t \geq 0}$ is defined as

$$
\mathcal{F}_t := \sigma\big(X_s : 0 \leq s \leq t\big) \qquad (t \geq 0),
$$

i.e., $\mathcal{F}_t$ is the σ-field generated by the random variables $(X_s)_{0 \leq s \leq t}$. The next definition generalizes (1.11) to the continuous-time setting.

Definition 1.6 (Finite-state Markov process) *Let S be a finite set, let $(X_t)_{t \geq 0}$ be a stochastic process with values in S and let $(P_t)_{t \geq 0}$ be a continuous transition probability on S. Then, by definition, we say that $(X_t)_{t \geq 0}$ is a (time-homogeneous,*

continuous-time) Markov process corresponding to the transition probability $(P_t)_{t\geq0}$ if $(X_t)_{t\geq0}$ has cadlag sample paths and

$$\mathbb{E}\big[f(X_t)\,\big|\,\mathcal{F}_s\big] = P_{t-s}f(X_s) \quad \text{a.s.} \quad \big(0 \leq s \leq t, \ f \in B(S)\big), \qquad (1.12)$$

where $(\mathcal{F}_t)_{t\geq0}$ is the filtration generated by $(X_t)_{t\geq0}$.

As we will see below, for a given continuous transition probability $(P_t)_{t\geq0}$ and probability law μ on S there exists a unique (in distribution) Markov process $(X_t)_{t\geq0}$ with *initial law* $\mathbb{P}[X_0 \in \,\cdot\,] = \mu$ such that (1.12) holds.

So far our approach has been very abstract, since we do not know what a general transition semigroup $(P_t)_{t\geq0}$ (or, equivalently, Markov semigroup) on S looks like.

Proposition 1.7 (Markov semigroups on finite sets) *Let $(P_t)_{t\geq0}$ be a Markov semigroup on a finite set S. Then there exist nonnegative constants $r(i,j)$ $(i,j \in S, \ i \neq j)$ such that*

$$\begin{aligned} Gf(i) &:= \lim_{t\downarrow0} t^{-1}\big(P_tf(i) - f(i)\big) \\ &= \sum_{j\in S} r(i,j)\big(f(j) - f(i)\big) \qquad \big(i \in S, \ f \in B(S)\big). \end{aligned} \qquad (1.13)$$

Conversely, each collection of nonnegative constants $\{r(i,j) : i,j \in S, \ i \neq j\}$ defines a unique Markov semigroup on S through (1.13) and the formula

$$P_tf(i) = e^{tG}f := \sum_{n=0}^{\infty} \frac{t^n}{n!}G^nf(i) \qquad \big(t \geq 0, \ i \in S, \ f \in B(S)\big).$$

We call G the *generator* of the Markov semigroup $(P_t)_{t\geq0}$. If $(X_t)_{t\geq0}$ is an S-valued process with cadlag sample paths such that (1.12) holds, then we also say that $(X_t)_{t\geq0}$ is a Markov process with generator G. We call $r(i,j)$ the *rate* of *jumps* from i to j. By applying (1.13) and (1.12) to functions of the form $f(i) := 1_{\{i=j\}}$, we see that if $(X_t^i)_{t\geq0}$ denotes the Markov process started in the initial law $X_0^i := 1$, then

$$\mathbb{P}[X_t^i = j] = r(i,j)t + O(t^2) \quad \text{as } t \to 0 \quad (i,j \in S, \ i \neq j).$$

This says that if we start the process in the state i, then for small t, the probability that we jump from i to j somewhere in the interval $(0,t)$ is $tr(i,j)$ plus a term of order t^2.

Proof of Proposition 1.7 We will skip the proof that each continuous semigroup of linear operators on a finite-dimensional real linear space is of the form $(e^{tG})_{t\geq0}$,

where G is called the *generator* of the semigroup. We will prove, however, that if this semigroup is a Markov semigroup, then G must have the special form in (1.13). Let $G(i,j)$ be the matrix associated with G, i.e., $Gf(i) = \sum_j G(i,j)f(j)$. Then

$$P_t f(i) = f(i) + t \sum_j G(i,j)f(j) + O(t^2) \quad \text{as } t \to 0 \qquad \big(i \in S,\ f \in B(S)\big).$$

Now the condition that $P_t f \geq 0$ for all $f \geq 0$ implies that $Gf(i) \geq 0$ whenever $f(i) = 0$, hence $G(i,j) \geq 0$ for each $i \neq j$. Moreover, the condition that $P_t 1 = 1$ implies that

$$1 = 1 + t \sum_j G(i,j) + O(t^2) \quad \text{as } t \to 0 \qquad (i \in S),$$

which shows that $\sum_j G(i,j) = 0$ for each i. Setting $r(i,j) := G(i,j)$ for $i \neq j$ and using the fact that $G(i,i) = -\sum_{j \neq i} r(i,j)$, we see that G can be cast in the form (1.13). The fact that conversely, each generator of this form defines a Markov semigroup will follow from our explicit construction of the associated Markov process below. ∎

We are now ready to state the most important theorem of this chapter, which tells us how to construct finite-state Markov processes based on a collection of Poisson point processes.

Theorem 1.8 (Poisson construction of Markov process) *Let S be a finite set and let $\mathcal{M}$ be a finite set whose elements are functions $m : S \to S$. Let $(r_m)_{m \in \mathcal{M}}$ be nonnegative constants and let Δ be a Poisson point set on $\mathcal{M} \times [0,\infty) = \{(m,t) : m \in \mathcal{M},\ t \geq 0\}$ with intensity $r_m \mathrm{d}t$, where $\mathrm{d}t$ denotes Lebesgue measure. Order the elements of Δ as $(m_1, t_1), (m_2, t_2), \dots$ where $t_1 < t_2 < \cdots$ and set $t_0 := 0$. Let X_0 be an S-valued random variable, independent of Δ. Define*

$$X_t := m_n \cdots m_1(X_0) \qquad \big(t \in [t_n, t_{n+1}),\ n \geq 0\big). \tag{1.14}$$

Then $(X_t)_{t \geq 0}$ is a Markov process with generator

$$Gf(i) = \sum_{m \in \mathcal{M}} r_m \big(f(m(i)) - f(i)\big) \qquad \big(i \in S,\ f \in B(S)\big). \tag{1.15}$$

Proof We will use approximation with Markov chains. We will be a bit sloppy concerning the precise sense of the convergence and concentrate on the main idea.

Set $R := \sum_{m \in \mathcal{M}} r_m$ and choose $\varepsilon > 0$ such that $\varepsilon R \leq 1$. Let $(\rho_k)_{k \geq 1}$ be i.i.d. $\mathcal{M}$-valued random variables with

$$\mathbb{P}[\rho_k = m] = \frac{r_m}{R} \qquad (k \geq 1, \; m \in \mathcal{M}).$$

Let $(\chi_k)_{k \geq 1}$ be i.i.d. Bernoulli random variables with $\mathbb{P}[\chi_k = 1] = \varepsilon R$, independent of the $(\rho_k)_{k \geq 1}$, and define a random set $\Delta_\varepsilon \subset \{(m, \varepsilon k) : m \in \mathcal{M}, \; k \geq 1\}$ by

$$\Delta_\varepsilon := \big\{(\rho_k, \varepsilon k) : \chi_k = 1\big\}.$$

Set

$$m_k^\varepsilon(i) := \begin{cases} \rho_k(i) & \text{if } \chi_k = 1, \\ i & \text{if } \chi_k = 0, \end{cases} \qquad (i \in S),$$

and define a discrete-time process $(X_{\varepsilon k})_{k \geq 0}$ by

$$X_{\varepsilon k} := m_k^\varepsilon \cdots m_1^\varepsilon(X_0) \qquad (k \geq 0).$$

Then $(X_{\varepsilon k})_{k \geq 0}$ is a Markov chain in S that jumps in each time step from the state i to the state $m(i)$ with probability $\varepsilon \frac{r_m}{R}$ and stays in the state i with probability $1 - \varepsilon R$. It is not hard to see that for $\varepsilon \to 0$, the random sets Δ_ε converge in an appropriate sense to a Poisson point set on $\mathcal{M} \times [0, \infty) = \{(m, t) : m \in \mathcal{M}, \; t \geq 0\}$ with intensity $r_m dt$ and hence the process $(X_{\varepsilon k})_{k \geq 0}$ converges in an appropriate sense to the continuous-time process $(X_t)_{t \geq 0}$ defined in (1.14). The process $(X_t)_{t \geq 0}$ inherits the Markov property from the approximating processes $(X_{\varepsilon k})_{k \geq 0}$. Moreover, if $(X_t^i)_{t \geq 0}$ denotes the process defined as in (1.14) with $X_0^i := i$, then we claim that

$$\mathbb{E}\big[f(X_t^i)\big] = (1 - tR)f(i) + t \sum_{m \in \mathcal{M}} r_m f(m(i)) + O(t^2) \quad \text{as } t \to 0.$$

Indeed, if t is small, then the probability that $(m, s) \in \Delta$ for some $s \in (0, t)$ is approximately $t r_m$ while the probability that $\Delta \cap (\mathcal{M} \times (0, t))$ contains more than one element is of order $O(t^2)$. This explains (rather than proves) formula (1.15). ∎

Theorem 1.8 shows that for each finite collection $\mathcal{M}$ of maps $m : S \to S$ and nonnegative rates $(r_m)_{m \in \mathcal{M}}$, we can construct a unique associated Markov process $(X_t)_{t \geq 0}$ in S. We note that the inverse problem is far from unique; with this we mean that for a given Markov process, or rather, for a given Markov semigroup $(P_t)_{t \geq 0}$, there may be many different ways of writing the generator G in the form (1.15). Once we have chosen a particular way of writing G in the form (1.15), Theorem 1.8 provides us with a natural way of *coupling* processes started in different

initial states. Indeed, using the same Poisson point set $\Delta = \{(m_1, t_1), (m_2, t_2), \ldots\}$, setting

$$X_t^i := m_n \cdots m_1(i) \qquad \big(t \in [t_n, t_{n+1}),\ n \geq 0,\ i \in S\big)$$

defines for each $i \in S$ a Markov process $(X_t^i)_{t \geq 0}$ with generator G started in the initial state $X_0^i = i$, and all these processes (for different i) are in a natural way defined on one and the same underlying probability space (i.e., they are *coupled*). Such couplings are very important in the theory of interacting particle systems. We conclude this section with the next result, which is similar to Theorem 1.8, but not quite the same.

Proposition 1.9 (Embedded Markov chain) *Let S be a finite set, let $r(i,j)$ $(i, j \in S,\ i \neq j)$ be nonnegative constants, and assume that*

$$R(i) := \sum_{j:j\neq i} r(i,j) > 0 \qquad (i \in S).$$

Let $(\sigma_k^i)_{k \geq 1}^{i \in S}$ be independent exponentially distributed random variables with mean $\mathbb{E}[\sigma_k^i] = 1/R(i)$. Define a probability kernel $P(i,j)$ on S by

$$P(i,j) := \begin{cases} \dfrac{r(i,j)}{R(i)} & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \qquad (i, j \in S).$$

Let $(Y_k)_{k \geq 0}$ be a Markov chain in S with transition kernel $P(i,j)$ and let $l_n(i) := |\{k : 0 \leq k \leq n,\ Y_k = i\}|$ denote the number of times $(Y_k)_{k \geq 0}$ has visited i by time n. Define inductively times $(t_k)_{k \geq 0}$ by $t_0 := 0$ and

$$t_{n+1} := t_n + \sigma_{l_n(Y_n)}^{Y_n} \qquad (n \geq 0),$$

and define a continuous-time process $(X_t)_{t \geq 0}$ by

$$X_t := Y_n \qquad \text{for } t \in [t_n, t_{n+1}).$$

Then $(X_t)_{t \geq 0}$ is the Markov process with generator

$$Gf(i) = \sum_j r(i,j)\big(f(j) - f(i)\big) \qquad \big(i \in S,\ f \in B(S)\big).$$

Proof (sketch) This can be proved by approximation with Markov chains in the spirit of the proof of Theorem 1.8. ∎

We note that Proposition 1.9 can be extended to the case that $R(i) = 0$ for some points $i \in S$ by allowing that $\sigma_1^i := \infty$ (these points are then *traps* for

the continuous-time Markov process). It is not hard to see that conversely, given the continuous-time process $(X_t)_{t\geq 0}$, one can uniquely construct the associated *embedded Markov chain* $(Y_k)_{k\geq 0}$. Because of Proposition 1.9, many properties of the continuous-time Markov process $(X_t)_{t\geq 0}$ can be translated in problems for the embedded Markov chain $(Y_k)_{k\geq 0}$ and vice versa. One should not make the mistake of thinking that all problems will become easier when translated into the language of the Markov chain $(Y_k)_{k\geq 0}$. Indeed, continuous-time Markov processes have some nice properties that discrete chains lack and it has happened more than once that problems for discrete Markov chains could be solved by looking at a cleverly chosen associated Markov processes in continuous time.

1.6 Invariant laws

Let $(X_t)_{t\geq 0}$ be a Markov process with finite state space S, Markov semigroup $(P_t)_{t\geq 0}$, generator G and jump rates $\{r(i,j) : i,j \in S, \ i \neq j\}$. By definition, an *invariant law* for $(P_t)_{t\geq 0}$ is a probability measure μ on S such that

$$\sum_i \mu(i)P_t(i,j) = \mu(j) \qquad (t \geq 0, \ j \in S).$$

If μ is an invariant law for a Markov semigroup $(P_t)_{t\geq 0}$ then we can construct a process $(X_t)_{t\in\mathbb{R}}$ that is also defined for negative times, such that (compare (1.12)) $\mathbb{P}[X_t \in \cdot] = \mu$ for all $t \in \mathbb{R}$ and

$$\mathbb{E}\big[f(X_t)\,\big|\,\mathcal{F}_s\big] = P_{t-s}f(X_s) \quad \text{a.s.} \quad \big(s \leq t, \ f \in B(S)\big), \tag{1.16}$$

where $\mathcal{F}_s := \sigma(X_u : -\infty < u \leq s)$. Such a Markov process $(X_t)_{t\in\mathbb{R}}$ is *stationary*, i.e.,

$$\mathbb{P}\big[(X_{s+t})_{t\in\mathbb{R}} \in \cdot\big] = \mathbb{P}\big[(X_t)_{t\in\mathbb{R}} \in \cdot\big] \qquad (s \in \mathbb{R}).$$

In view of this, invariant laws are sometimes also called *stationary laws*. By definition, a Markov process with given jump rates $\{r(i,j) : i,j \in S, \ i \neq j\}$ is *irreducible* if

$$\forall S' \subset S \text{ with } S' \neq \emptyset, S \ \ \exists i \in S', \ j \notin S' \text{ such that } r(i,j) > 0.$$

Proposition 1.10 (Ergodicity) *Consider a Markov process on a finite state space S with jump rates $\{r(i,j) : i,j \in S, \ i \neq j\}$. If the jump rates are irreducible, then the Markov process has a unique invariant law μ and the process $(X_t)_{t\geq 0}$ started in any initial law satisfies*

$$\mathbb{P}[X_t = i] \xrightarrow[t\to\infty]{} \mu(i) \qquad (i \in S). \tag{1.17}$$

Proof (sketch) Let $(X_t^i)_{t \geq 0}$ denote the process started in the initial state $X_0^i = i$ and let

$$\tau_j^i := \inf\{t \geq 0 : X_t^i = j\}$$

denote the first arrival time of $(X_t^i)_{t \geq 0}$ in j and let

$$\lambda_i := \sum_j r(i, j)\mathbb{E}[\tau_i^j]$$

denote the *mean return time* to the state i, which is finite for each $i \in S$ by irreducibility. Then one can check that

$$\mu(i) := \frac{\lambda_i^{-1}}{\sum_j \lambda_j^{-1}} \qquad (i \in S)$$

is an invariant law for the process $(X_t)_{t \geq 0}$. To prove (1.17), let $(X_t)_{t \geq 0}$ be started in any initial law and let $(X_t')_{t \geq 0}$ be an independent Markov process with the same jump rates, started in the initial law μ. Set

$$\sigma := \inf\{t \geq 0 : X_t = X_t'\}$$

and define

$$X_t'' := \begin{cases} X_t & (t < \sigma), \\ X_t' & (\sigma \leq t). \end{cases}$$

Then $(X_t'')_{t \geq 0}$ is equal in law to $(X_t)_{t \geq 0}$. Using irreducibility, one can prove that $\sigma < \infty$ a.s. It follows that

$$\left|\mathbb{P}[X_t = i] - \mu(i)\right| \leq \mathbb{P}[X_t'' = i, \ X_t' \neq i] + \mathbb{P}[X_t'' \neq i, \ X_t' = i] \leq \mathbb{P}[\sigma < t] \xrightarrow[t \to \infty]{} 0,$$

which implies (1.17).   ∎

Formula (1.17) is often described in words by saying that the Markov process $(X_t)_{t \geq 0}$ is *ergodic*, although this is not entirely correct terminology. (In fact, the statement that (1.17) holds for each initial law is stronger than the statement that the stationary process defined in (1.16) is ergodic.)

We recall from (1.5) that the Markov property is symmetric with respect to time reversal. Thus, if $(X_1, \ldots, X_n)$ is a (finite) Markov chain, then so is $(X_n, \ldots, X_1)$; similar statements hold for continuous-time processes. However, if a Markov process is time-homogeneous, then the same need not be true for the time-reversed process. An exception are stationary Markov processes: reversing the time in a stationary Markov process yields a stationary, hence time-homogeneous Markov process. The transition probabilities of this time-reversed process need not be the same as those of the original process, however. This leads to the following definition.

Definition 1.11 (Reversibility) *Let S be a finite set and let $(P_t)_{t\geq 0}$ be a Markov semigroup on S. Then, by definition, we say that an invariant law μ of $(P_t)_{t\geq 0}$ is reversible if the stationary process in (1.16) satisfies*

$$\mathbb{P}\big[(X_{-t})_{t\in\mathbb{R}}\in\,\cdot\,\big]=\mathbb{P}\big[(X_{t-})_{t\in\mathbb{R}}\in\,\cdot\,\big]. \tag{1.18}$$

Note that $(X_{-t})_{t\in\mathbb{R}}$ has *left*-continuous sample paths, which is why we compare this in (1.18) with $(X_{t-})_{t\in\mathbb{R}}$, the left-continuous modification of $(X_t)_{t\in\mathbb{R}}$. We state the following fact without proof.

Proposition 1.12 (Detailed balance) *A probability law μ on a finite set S is a reversible invariant law for a Markov process in S with jump rates $\{r(i,j):i,j\in S,\ i\neq j\}$ if and only if*

$$\mu(i)r(i,j)=\mu(j)r(j,i) \qquad (i,j\in S). \tag{1.19}$$

Condition (1.19) is called *detailed balance*. Note that this says that in equilibrium, jumps from i to j happen with the same frequency as jumps from j to i.

1.7 Feller processes

Let E be a compact metrizable space. (Such spaces are always separable.) We let $\mathcal{C}(E)$ denote the space of continuous real functions on E, equipped with the supremumnorm

$$\|f\|:=\sup_{x\in E}|f(x)| \qquad (f\in\mathcal{C}(E)).$$

We let $\mathcal{M}_1(E)$ denote the space of probability measures on E (equipped with the topology of weak convergence). We note that $\mathcal{C}(E)$ is a separable Banach space and that $\mathcal{M}_1(E)$ is a compact metrizable space.

By definition, a *continuous transition probability* on E is a collection $(P_t(x,\mathrm{d}y))_{t\geq 0}$ of probability kernels on E such that

(i) $(x,t)\mapsto P_t(x,\,\cdot\,)$ is a continuous map from $E\times[0,\infty)$ into $\mathcal{M}_1(E)$,

(ii) $\displaystyle\int_E P_s(x,\mathrm{d}y)P_t(y,\mathrm{d}z)=P_{s+t}(x,\mathrm{d}z)$ and $P_0(x,\,\cdot\,)=\delta_x$ $(x\in E,\ s,t\geq 0)$.

Each continuous transition probability defines a Markov semigroup $(P_t)_{t\geq 0}$ by

$$P_t f(x):=\int_E P_t(x,\mathrm{d}y)f(y) \qquad \big(f\in B(E)\big). \tag{1.20}$$

It follows from the continuity of the transition probability that the operators P_t map the space $\mathcal{C}(E)$ into itself. Conversely, each collection of linear operators $P_t : \mathcal{C}(E) \to \mathcal{C}(E)$ such that

$$\text{(i)} \quad \lim_{t \to 0} \|P_t f - f\| = 0,$$
$$\text{(ii)} \quad P_s P_t f = P_{s+t} f \quad \text{and} \quad P_0 f = f,$$
$$\text{(iii)} \quad f \geq 0 \text{ implies } P_t f \geq 0,$$
$$\text{(iv)} \quad P_t 1 = 1,$$

corresponds to a unique continuous transition probability on E. Such a collection of linear operators $P_t : \mathcal{C}(E) \to \mathcal{C}(E)$ is called a *Feller semigroup*.

By definition, the *generator* is the operator

$$Gf := \lim_{t \to 0} t^{-1} (P_t f - f),$$

which is defined only for functions $f \in \mathcal{D}(G)$, where

$$\mathcal{D}(G) := \left\{ f \in \mathcal{C}(E) : \text{the limit } \lim_{t \to 0} t^{-1} (P_t f - f) \text{ exists} \right\}.$$

Here, when we say that the limit exists, we mean the limit in the topology on $\mathcal{C}(E)$, which is defined by the supremumnorm $\| \cdot \|$. It can be shown that a Feller semigroup is determined uniquely by its generator.

We say that an operator A on $\mathcal{C}(E)$ with domain $\mathcal{D}(A)$ satisfies the maximum principle if, whenever a function $f \in \mathcal{D}(A)$ assumes its maximum over E in a point $x \in E$, we have $Af(x) \leq 0$. We say that a linear operator A with domain $\mathcal{D}(A)$ acting on a Banach space $\mathcal{V}$ (in our example the space $\mathcal{C}(E)$ equipped with the supremunorm) is *closed* if and only if its graph $\{(f, Af) : f \in \mathcal{D}(A)\}$ is a closed subset of $\mathcal{V} \times \mathcal{V}$. The following is a version of the Hille-Yosida theorem:

Proposition 1.13 (Hille-Yosida, first version) *A linear operator G on $\mathcal{C}(E)$ is the generator of a Feller semigroup if and only if*

(i) $1 \in \mathcal{D}(G)$ *and* $G1 = 0$.

(ii) G *satisfies the maximum principle.*

(iii) $\mathcal{D}(G)$ *is dense in* $\mathcal{C}(E)$.

(iv) *For every $f \in \mathcal{D}(G)$ there exists a continuously differentiable function $t \mapsto u_t$ such that $u_0 = f$, $u_t \in \mathcal{D}(G)$, and $\frac{\partial}{\partial t} u_t = G f_t$ for each $t \geq 0$.*

(v) *G is closed.*

Here the differentiation with respect to t is in the Banach space $\mathcal{C}(E)$.

In practice, it us usually not feasible to explicitly write down the full domain of the generator of a Feller semigroup. Instead, one often first defines a 'pregenerator' wich is defined for a smaller class of functions, and then constructs the 'full generator' by taking the closure of the pregenerator.

By definition, a linear operator A with domain $\mathcal{D}(A)$ on a Banach space $\mathcal{V}$ is *closable* if the closure of its graph (as a subset of $\mathcal{V} \times \mathcal{V}$) is the graph of a linear operator $\overline{A}$ with domain $\mathcal{D}(\overline{A})$, called the *closure* of A. The next version of the Hille-Yosida theorem is often usueful.

Proposition 1.14 (Hille-Yosida, second version) *A linear operator G on $\mathcal{C}(E)$ with domain $\mathcal{D}(G)$ is closable and its closure $\overline{G}$ is the generator of a Feller semigroup if and only if*

(i) *$(1, 0) \in \overline{\{(f, Gf) : f \in \mathcal{D}(G)\}}$ (i.e., $(1, 0)$ is in the closure of the graph of G).*

(ii) *G satisfies the maximum principle.*

(iii) *$\mathcal{D}(G)$ is dense in $\mathcal{C}(E)$.*

(iv) *There exists an $r \in (0, \infty)$ and a dense subspace $\mathcal{D} \subset \mathcal{C}(E)$ with the property that for every $f \in \mathcal{D}$ there exists a $p_r \in \mathcal{D}(G)$ such that $(1 - rG)p_r = f$.*

By definition, a *Feller process* associated to a given Feller semigroup $(P_t)_{t \geq 0}$ is a stochastic process $(X_t)_{t \geq 0}$ with values in E and cadlag sample paths, such that (compare (1.16))

$$\mathbb{E}\big[f(X_t) \,\big|\, \mathcal{F}_s\big] = P_{t-s}f(X_s) \quad \text{a.s.} \quad \big(s \leq t, \ f \in \mathcal{C}(E)\big), \tag{1.21}$$

where $(\mathcal{F}_t)_{t \geq 0}$ is the filtration generated by $(X_t)_{t \geq 0}$. It can be shown that if $(P_t)_{t \geq 0}$ is a Feller semigroup, then for each probability law μ on E there exists a unique (in law) Feller process associated to $(P_t)_{t \geq 0}$ with initial law $\mathbb{P}[X_0 \in \cdot] = \mu$. Feller processes have many nice properties, such as the strong Markov property.

1.8   Tightness and weak convergence

Since we will need this on some occasions, we recall a few facts from basic measure theory. Let $\mathbb{E}$ be a Polish space and let $\mathcal{M}_1(E)$ be the space of probability measures

on E, equipped with the topology of weak convergence. By definition, a set $\mathcal{R} \subset \mathcal{M}_1(E)$ is *tight* if

$$\forall \varepsilon > 0 \exists K \subset E \text{ s.t. } K \text{ is compact and } \mu(E\backslash K) \leq \varepsilon \; \forall \mu \in \mathcal{R}.$$

A well-known result says that the closure of $\mathcal{R}$ is compact (i.e., $\mathcal{R}$ is 'precompact') as a subset of $\mathcal{M}_1(E)$ if and only if $\mathcal{R}$ is tight. In particular, if $(\mu_n)_{n \geq 0}$ is a sequence of probability measures on E then we say that such a sequence is tight if the set $\{\mu_n : n \geq 0\} \subset \mathcal{M}_1(E)$ is tight. Note that each tight sequence of probabability measures has a weakly convergent subsequence. Recall that a *cluster point* of a sequence is a limit of some subsequence of the sequence. We sometimes say 'weak cluster point' when we mean a 'cluster point in the topology of weak convergence'. One often needs tightness because of the following simple fact.

Lemma 1.15 (Tightness and weak convergence) *Let $(\mu_n)_{n \geq 0}$ be a tight sequence of probability measures on a Polish space E and assume that $(\mu_n)_{n \geq 0}$ has only one weak cluster point μ. Then μ_n converges weakly to μ.*

Note that if E is compact, then tightness comes for free, i.e., every sequence of probability measures on E is tight and $\mathcal{M}_1(E)$ is itself a compact space.

Chapter 2

The contact process

2.1 Introduction

In this chapter, we study the contact process. The contact process is one of the most basic and most intensively studied interacting particle systems. It was introduced in the mathematical literature by Harris in 1974 [Har74] and a few years later independently in the high-energy physics literature as the 'reggeon spin model'. Many important questions about the behavior of the nearest-neighbour contact process on $\mathbb{Z}^d$ were solved by Bezuidenhout and Grimmett in 1990–1991 (building, of course, on the work of many others) [BG90, BG91]. Physicists, not hindered by the burden of rigorous proof, proceeded must faster. In fact, the only statements about the contact process that physicists consider nontrivial -concerning its critical behavior in low dimensions- remain largely unproved by mathematicians up to date. The contact process continues to be the subject of intense study in the mathematical literature. Questions about its critical behavior in high dimensions were recently answered in [HS05]. In addition, all kind of variations on the original process such as contact processes in a random environment [Lig92, Rem08] or contact processes on more general lattices have recieved a lot of attention.

2.2 Definition of the model

We start by introducing the most basic contact process: the nearest-neighbour contact process on $\mathbb{Z}^d$. Recall that

$$\mathbb{Z}^d := \left\{ i = (i_1, \ldots, i_d) : i_k \in \mathbb{Z} \ \forall k = 1, \ldots, d \right\}$$

is the *d-dimensional integer lattice.*   Points $i \in \mathbb{Z}^d$ are often called *sites.* The *nearest-neighbour* contact process on $\mathbb{Z}^d$ is a collection

$$\big(X_t(i)\big)_{t \geq 0,\ i \in \mathbb{Z}^d}$$

of $\{0,1\}$-valued random variables, such that for each $i \in \mathbb{Z}^d$, the process

$$\big(X_t\big)_{t \geq 0}$$

is a continuous-time Markov process with state space $\{0,1\}$ whose jump rates depend on the states of its *nearest neighbours*, i.e., the processes $X_t(j)$ with $|i-j| = 1$. More precisely, if at some time t the state of the process is $x = (x(i))_{i \in \mathbb{Z}^d} \in \{0,1\}^{\mathbb{Z}^d}$, then the state at site i jumps as follows:

$x(i)$ jumps:

$\quad 0 \mapsto 1 \qquad$ with rate $\lambda \sum_{j:|i-j|=1} x(j)$,

$\quad 1 \mapsto 0 \qquad$ with rate 1.

Here $\lambda > 0$ is a fixed constant, called the *infection rate.*

For example, if $d = 2$, then we can imagine that each site $i \in \mathbb{Z}^2$ represents a tree in an infinite orchard. If $X_t(i) = 1$ then we say that the tree i is at time t *infected* with a certain disease, while if $X_t(i) = 0$ we say that the tree i is *healthy.* Healthy trees become infected with a rate that is proportional to their number of infected neighbours, and infected trees get healthy with constant *recovery rate* 1.

Although we say informally that the process $(X_t(i))_{t \geq 0}$ at the site i is a continuous-time Markov process, because of the fact that its jump rates depend on the states of its neighbours, the process $(X_t(i))_{t \geq 0}$, on its own, actually does not have the Markov property. On the other hand, if we let

$$X_t := \big(X_t(i)\big)_{i \in \mathbb{Z}^d} \qquad (t \geq 0)$$

denote the whole collection of processes $X_t(i)$ indexed by sites $i \in \mathbb{Z}^d$, then the $\{0,1\}^{\mathbb{Z}^d}$ valued process $(X_t)_{t \geq 0}$ *does* have the Markov property. We may formally write the generator of this Markov process as

$$\begin{aligned}
Gf(x) = {}& \lambda \sum_i 1_{\{x(i)=0\}} \sum_{j:|i-j|=1} x(j)\big(f(x + \delta_i) - f(x)\big) \\
& + \sum_i 1_{\{x(i)=1\}}\big(f(x - \delta_i) - f(x)\big),
\end{aligned} \qquad (2.1)$$

$(x \in \{0, 1\}^{\mathbb{Z}^d})$, where we define $\delta_i \in \{0, 1\}^{\mathbb{Z}^d}$ by

$$\delta_i(j) := 1_{\{i=j\}} \qquad (i, j \in \mathbb{Z}^d).$$

Note that we do not say precisely for which functions f the expression in (2.1) is defined. At least for functions that depend on finitely many coordinates only, one can check that the infinite sums reduce to finite sums and hence Gf is well-defined. It can be shown that if one first defines Gf for such functions and then takes the closure of the operator G, then this closure generates a Feller semigroup on $\{0, 1\}^{\mathbb{Z}^d}$. We may then define $(X_t)_{t \geq 0}$ as the Feller process associated with this semigroup. We will follow this approach in Chapter A. In the present chapter, we will use an alternative approach for constructing $(X_t)_{t \geq 0}$, which is based on Theorem 1.8.

2.3 The graphical representation

For each pair of neighbouring sites $i, j \in \mathbb{Z}^d$, $|i - j| = 1$, let us define a map $m_{ij} : \{0, 1\}^{\mathbb{Z}^d} \to \{0, 1\}^{\mathbb{Z}^d}$ by

$$(m_{ij}x)(k) := \begin{cases} 1 & \text{if } k = i, \ x(j) = 1, \\ x(k) & \text{otherwise,} \end{cases} \qquad (i, j, k \in \mathbb{Z}^d, \ x \in \{0, 1\}^{\mathbb{Z}^d}).$$

Moreover, for each $i \in \mathbb{Z}^d$, let us define $p_i : \{0, 1\}^{\mathbb{Z}^d} \to \{0, 1\}^{\mathbb{Z}^d}$ by

$$(p_i x)(j) := \begin{cases} 0 & \text{if } j = i, \\ x(j) & \text{if } j \neq i, \end{cases} \qquad (i, j \in \mathbb{Z}^d, \ x \in \{0, 1\}^{\mathbb{Z}^d}).$$

Then we may rewrite the generator G in (2.1) in the form

$$Gf(x) = \lambda \sum_{i,j: |i-j|=1} \big(f(m_{ij}(x)) - f(x)\big) + \sum_{i} \big(f(p_i(x)) - f(x)\big). \qquad (2.2)$$

$(x \in \{0, 1\}^{\mathbb{Z}^d})$. This form of the generator reminds us of Theorem 1.8. The only difference is that there, we considered a finite set of maps $\mathcal{M}$, while at present we have to deal with infinitely many maps m_{ij} and p_i. However, since our maps act only locally, we may still hope that the same approach works.

In view of these considerations and inspired by Theorem 1.8, we let

$$\mathcal{E} := \big\{(i, j) : i, j \in \mathbb{Z}^d, \ |i - j| = 1\big\}$$

denote the set of all ordered nearest-neighbour pairs and we let Δ^i and Δ^r be independent Poisson point sets on $\mathcal{E} \times [0, \infty)$ and $\mathbb{Z}^d \times [0, \infty)$ with intensities $\lambda\, \mathrm{d}t$ and $1\, \mathrm{d}t$, respectively. We interpret a point $(i, j, t) \in \Delta^i$ as the event that at time t, the state of our contact process jumps from x to $m_{ij}(x)$. Likewise, a point $(i, t) \in \Delta^r$ indicates that at time t, the state of our contact process jumps from x to $p_i(x)$.

Unlike the situation in Theorem 1.8, we can no longer order the elements of Δ^i and Δ^r by their time coordinates; indeed, for each time interval (s, t) with $s < t$ the sets $\Delta^i \cap (\mathcal{E} \times (s, t))$ and $\Delta^r \cap (\mathbb{Z}^d \times (s, t))$ contain infinitely many points. However, since most of these points refer to events that take place somewhere far away, we can still hope that by concatenating all maps m_{ij} and p_i that matter locally in the right order, we obtain a well-defined process.

In Figure 2.1 we have drawn a finite piece of the sets Δ^i and Δ^r for the process on $\mathbb{Z}$. We have drawn space horizontally and time vertically. Points $(i, j, t) \in \Delta^i$ have been indicated by drawing an arrow from (i, t) to (j, t) while points $(i, t) \in \Delta^r$ have been indicated with a black box. The arrows indicate potential infections: if at time t there is an arrow from i to j and just before this time, the site i is infected and j is healthy, then after this time both i and j should be infected; otherwise nothing happens. Likewise, a black box at a point (i, t) indicates a potential recovery: if just before time t, the site i is infected, then after time t it should be healthy; otherwise nothing happens.

With this in mind, we make the following definitions. We start by noting that if $[s, u]$ is a time interval and $\gamma : [s, u] \to \mathbb{Z}^d$ is a cadlag function, then it is possible that γ makes a jump at time u (i.e., $\gamma_u \neq \gamma_{u-}$), but because of right-continuity, it cannot happen that γ makes a jump at time s. In view of this and since in what follows, it will be inconvenient to have a definition that is not symmetric with respect to time reversal, we start with a definition that may seem a little cumbersome at first sight. By definition, by a *path* in $\mathbb{Z}^d$ we will mean a pair of functions (γ_{t-}, γ_t) defined on some time interval $[s, u]$ with $s \leq u$ and taking values in $\mathbb{Z}^d$, such that

$$
\begin{aligned}
\lim_{t \downarrow t_0} \gamma_{t-} &= \gamma_{t_0} && \big(t_0 \in [s, u)\big), \\
\lim_{t \uparrow t_0} \gamma_t &= \gamma_{t_0-} && \big(t_0 \in (s, u]\big).
\end{aligned}
\tag{2.3}
$$

Note that this definition allows for the case that $\gamma_{s-} \neq \gamma_s$, which is what we wanted. We will often identify a path, as we have just defined it, with the set $\gamma \subset \mathbb{Z}^d \times \mathbb{R}$ defined by

$$
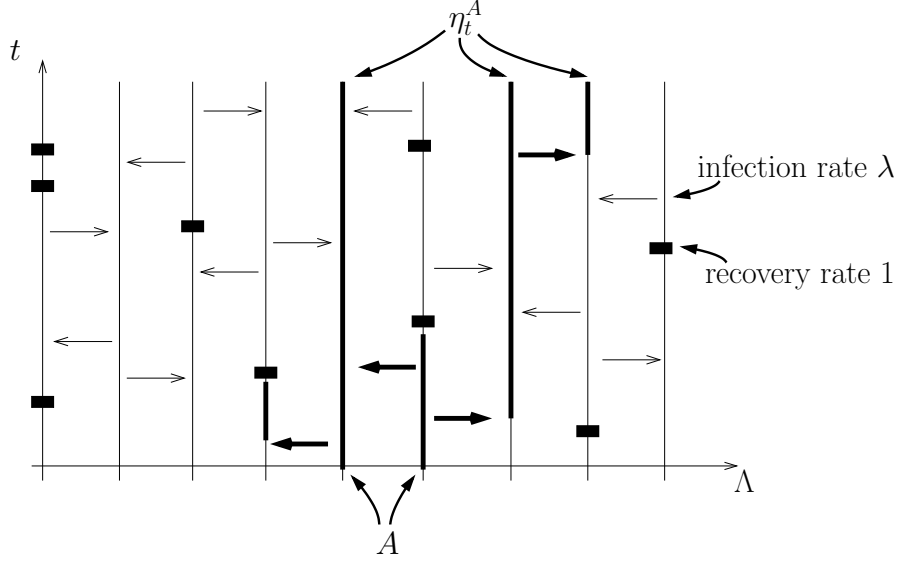\gamma := \big\{ \gamma_{t-} : t \in [s, u] \big\} \cup \big\{ \gamma_t : t \in [s, u] \big\}.
$$

Figure 2.1: Graphical representation of the contact process.

Note that both the functions γ_t and γ_{t-}, as well as the *starting time s* and *final time u* can be read off from the set γ.

Next, for points $(i, s), (j, u) \in \mathbb{Z}^d \times [0, \infty)$, we say that there is an *open path* from to (i, s) to (j, u), denoted as $(i, s) \rightsquigarrow (j, u)$, if and only if there exists a path γ in $\mathbb{Z}^d$ with starting time s and final time t such that $\gamma_{s-} = i$, $\gamma_u = j$, and

$$(\gamma_{t-}, \gamma_t, t) \in \Delta^{\mathrm{i}} \quad \text{for all } t \in [s, u] \text{ s.t. } \gamma_{t-} \neq \gamma_t,$$
$$\gamma \cap \Delta^{\mathrm{r}} = \emptyset.$$

In words, this says that an open path must walk upwards in time, may use infection arrows, but must avoid recovery symbols. Next, for each $x \in \{0, 1\}^{\mathbb{Z}^d}$, we define

$$X_t^x(i) := \begin{cases} 1 & \text{if } \exists j \text{ s.t. } x(j) = 1, \ (j, 0) \rightsquigarrow (i, t), \\ 0 & \text{otherwise}, \end{cases}$$

$(t \geq 0, \ i \in \mathbb{Z}^d, \ x \in \{0, 1\}^{\mathbb{Z}^d})$. A little thinking convinces us that $(X_t^x)_{t \geq 0}$ is the contact process we wanted to construct. (Note that we have arranged our definitions in such a way that $(X_t^x(i))_{t \geq 0}$ has cadlag sample paths for each $i \in \mathbb{Z}^d$.)

Often, it is convenient to use slightly different notation. By identifying a set with its indicator function, we see that the space $\{0, 1\}^{\mathbb{Z}^d}$ is in a natural way isomorphic

to the space of all subsets of $\mathbb{Z}^d$. With this in mind, for any $A \subset \mathbb{Z}^d$, we define

$$\eta_t^A := \big\{i : A \times \{0\} \rightsquigarrow (i,t)\big\} \qquad \big(t \geq 0, \ A \subset \mathbb{Z}^d\big), \tag{2.4}$$

where $A \times \{0\} \rightsquigarrow (i,t)$ indicates the event that $(j,0) \rightsquigarrow (i,t)$ for some $j \in A$. With this notation, we see that

$$\{i : x(i) = 1\} = A \quad \text{implies} \quad \{i : X_t^x(i) = 1\} = \eta_t^A,$$

i.e., if A is the set of sites that are infected at time 0 then η_t^A is the set of sites that are infected at time t.

Exercise 2.1 Invent graphical representations for the interacting particle systems on $\mathbb{Z}$ with generators (compare (2.1))

$$\begin{aligned}
G'f(x) = \lambda \sum_i &1_{\{x(i)=0\}} 1_{\{x(i-1)+x(i+1)>0\}} \big(f(x+\delta_i) - f(x)\big) \\
&+ \sum_i 1_{\{x(i)=1\}} \big(f(x-\delta_i) - f(x)\big)
\end{aligned}$$

and

$$\begin{aligned}
G''f(x) = \lambda \sum_i &1_{\{x(i)=0\}} \big(x(i-1) + x(i+1)\big)^2 \big(f(x+\delta_i) - f(x)\big) \\
&+ \sum_i 1_{\{x(i)=1\}} \big(f(x-\delta_i) - f(x)\big).
\end{aligned}$$

2.4   The survival probability

By definition, we say that the nearest-neighbor contact process on $\mathbb{Z}^d$ with infection rate λ *survives* if

$$\theta(\lambda, d) = \theta(\lambda) := \mathbb{P}\big[\eta_t^{\{0\}} \neq \emptyset \ \forall t \geq 0\big] > 0.$$

If this probability is zero, then we say that the contact process *dies out* or *gets extinct*.

By a combination of rigorous mathematics, nonrigourous methods, and computer simulations, theoretical physicists have discovered the following properties of the function θ. There exists a *critical value* $\lambda_c = \lambda_c(d)$ with $0 < \lambda_c < \infty$ such that $\theta(\lambda) = 0$ for $\lambda \leq \lambda_c$ and $\theta(\lambda) > 0$ for $\lambda > \lambda_c$. The function θ is continuous, strictly increasing and concave on $[\lambda_c, \infty)$ and satisfies $\lim_{\lambda \to \infty} \theta(\lambda) = 1$. One has
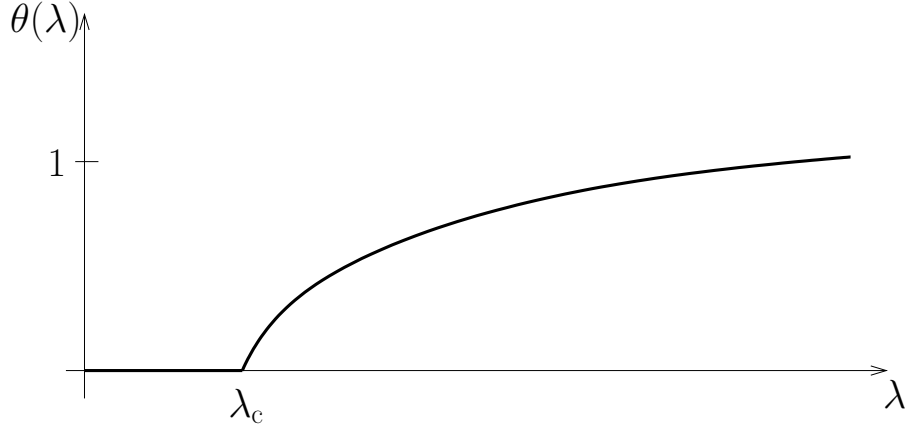
$$\lambda_c(1) = 1.6489 \pm 0.0002.$$

Figure 2.2: Survival probability.

Moreover, $\lambda_c(d)$ is decreasing in d and satisfies

$$\lambda_c(d) \approx \frac{1}{2d} \qquad \text{as } d \to \infty, \tag{2.5}$$

where the notation $f(z) \approx g(z)$ as $z \to z_0$ means that

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = 1 \qquad \text{as } z \to z_0.$$

The behavior of θ near the critical point is very interesting. One has

$$\theta(\lambda) \sim (\lambda - \lambda_c)^\beta \qquad \text{as } \lambda \downarrow \lambda_c, \tag{2.6}$$

where we write $f(z) \sim g(z)$ as $z \to z_0$ if

$$\lim_{z \to z_0} \frac{\log(f(z))}{\log(g(z))} = 1.$$

The constant $\beta = \beta(d)$ is a *critical exponent*, approximately given by

$$\begin{aligned} \beta(1) &\cong 0.276487, \\ \beta(2) &\cong 0.584, \\ \beta(3) &\cong 0.81, \\ \beta(d) &= 1 \qquad (d \geq 4). \end{aligned}$$

In dimensions $d \neq 4$, it is believed that (2.6) can be strengthened to $\theta(\lambda) \approx c(\lambda - \lambda_c)^\beta$ for some $0 < c < \infty$.

Below, we will prove some of the easier properties of the function θ, such as monotonicity, the existence of a critical parameter λ_c, and the fact that θ is right-continuous everywhere and left-continuous everywhere except possibly at the critical point λ_c. Proving that θ is left-continuous at λ_c, which by our previous remarks is equivalent to the statement that $\theta(\lambda_c) = 0$, kept probabilists occupied for some 15 years, till Bezuidenhout and Grimmett proved this in their celebrated paper [BG90]. Quite recently, it has been proved that (2.6) holds with $\beta = 1$ if the dimension d is sufficiently large. The critical behavior in dimensions $d = 1, 2, 3$ remains very much an unsolved problem. Physicists come to their prediction (2.6) using (nonrigorous) renormalization group arguments, where critical exponents can be related to eigenvectors of linearized renormalization transformations near a fixed point. Mathematically, there are big problems even *defining* these renormalization transformations rigorously, let alone studying them.

In dimension $d = 1$ it is known rigorously that $1.539 < \lambda_c < 1.943$ [ZG85, Lig95]. For bounds in higher dimensions (including a proof of (2.5)), see [Lig85]. As far as I know, nobody has any idea how to prove that θ is concave on $[\lambda_c, \infty)$.

2.5   Extinction

Lemma 2.2 (Survival versus extinction) *If the contact process survives, then*

$$\mathbb{P}\big[\eta_t^A \neq \emptyset \ \forall t \geq 0\big] > 0 \tag{2.7}$$

for each finite nonempty $A \subset \mathbb{Z}^d$. If the contact process dies out, then this probability is zero for each finite nonempty $A \subset \mathbb{Z}^d$.

Proof Let A be finite and nonempty. For obvious reasons we also denote the probability in (2.7) by

$$\mathbb{P}\big[(A \times \{0\}) \rightsquigarrow \infty\big].$$

Now choose any $i \in A$. Then

$$\mathbb{P}\big[(0,0) \rightsquigarrow \infty\big] = \mathbb{P}\big[(i,0) \rightsquigarrow \infty\big] \leq \mathbb{P}\big[(A \times \{0\}) \rightsquigarrow \infty\big]$$
$$= \mathbb{P}\big[\exists j \in A \text{ s.t. } (j,0) \rightsquigarrow \infty\big] \leq \sum_{j \in A} \mathbb{P}\big[(j,0) \rightsquigarrow \infty\big] = |A|\mathbb{P}\big[(0,0) \rightsquigarrow \infty\big],$$

where we have used translation invariance and $|A|$ denotes the number of elements in A.   ∎

In this and the next section, we will prove the following result.

Theorem 2.3 (Critical infection rate) *For each $d \geq 1$ there exists a $\lambda_c = \lambda_c(d)$ with $0 < \lambda_c < \infty$ such that the nearest-neighbour contact process on $\mathbb{Z}^d$ with infection rate λ survives for $\lambda > \lambda_c$ and dies out for $\lambda < \lambda_c$.*

Note that this theorem says nothing about survival or extinction if $\lambda = \lambda_c(d)$.

As a first step towards Theorem 2.3, we prove the following fact.

Lemma 2.4 (Monotone coupling) *Let $(\eta_t)_{t \geq 0}$ and $(\eta'_t)_{t \geq 0}$ be countact processes on $\mathbb{Z}^d$ with infection rates $0 \leq \lambda \leq \lambda'$ and deterministic initial states $\eta_0 = A$ and $\eta'_0 = A'$ satisfying $A \subset A'$. Then $(\eta_t)_{t \geq 0}$ and $(\eta'_t)_{t \geq 0}$ can be coupled such that*

$$\eta_t \subset \eta'_t \qquad (t \geq 0).$$

In particular, survival of the contact process with infection rate λ implies survival of the contact process with infection rate λ'.

Proof Let $0 \leq \lambda \leq \lambda'$. Let Δ^i and $\tilde{\Delta}^i$ be independent Poisson point sets on $\mathcal{E} \times [0, \infty)$ with intensities $\lambda \, dt$ and $(\lambda' - \lambda) \, dt$, respectively, and let $\tilde{\Delta}^r$ be a Poisson point set on $\mathbb{Z}^d \times [0, \infty)$ with intensity $1 \, dt$, independent of Δ^i and $\tilde{\Delta}^i$. Then $\Delta^i \cup \tilde{\Delta}^i$ is a Poisson point set on $\mathcal{E} \times [0, \infty)$ with intensity $\lambda' \, dt$. We interpret points in Δ^i and $\tilde{\Delta}^i$ as infection arrows and points in $\tilde{\Delta}^r$ as recovery symbols. We let $\rightsquigarrow$ indicate the presence of an open path that may use infection arrows from Δ^i only and we write $\rightsquigarrow'$ to indicate the presence of an open path that may use infection arrows from $\Delta^i \cup \tilde{\Delta}^i$. Then

$$\eta_t = \{i : A \times \{0\} \rightsquigarrow (i, t)\} \subset \{i : A' \times \{0\} \rightsquigarrow' (i, t)\} = \eta'_t \qquad (t \geq 0)$$

since $A \subset A'$ and the process $(\eta'_t)_{t \geq 0}$ has more arrows at its disposal. ∎

It follows from Lemma 2.4 that the function $\lambda \mapsto \theta(\lambda)$ is nondecreasing and hence, for each $d \geq 1$, there exists a $0 \leq \lambda_c(d) \leq \infty$ such that the nearest-neighbour contact process on $\mathbb{Z}^d$ with infection rate λ survives for $\lambda > \lambda_c$ and dies out for $\lambda < \lambda_c$. To prove Theorem 2.3, we must show that $0 < \lambda_c(d) < \infty$. We start by proving a the lower bound on λ_c, which is easiest.

Proposition 2.5 (Exponential bound) *Let $(\eta_t^A)_{t \geq 0}$ be the nearest-neighbour contact process on $\mathbb{Z}^d$ with infection rate λ, started in a deterministic initial state $\eta_0^A = A$ with $|A| < \infty$. Then $|\eta_t^A| < \infty$ for all $t \geq 0$ a.s., and*

$$\mathbb{E}[|\eta_t^A|] \leq |A| e^{(2d\lambda - 1)t} \qquad (t \geq 0). \tag{2.8}$$

Proof We start by giving a heuristic argument. Each infected site is bordered by at most $2d$ uninfected sites. Therefore, the expected number of new sites that are infected by an already infected site in a small time interval of length dt is less than $2d\lambda\,dt$. On the other hand, each infected site has a probability of $1\,dt$ to die in the same time interval. Therefore, the expected number of infected sites as a function of time should satisfy

$$\tfrac{\partial}{\partial t}\mathbb{E}\big[|\eta_t^A|\big] \le \big(2d\lambda - 1\big)\mathbb{E}\big[|\eta_t^A|\big],$$

which implies (2.8).

To make this rigorous, we need to work a bit. We start by translating our argument into the language of semigroups and generators. Let $(P_t)_{t\ge 0}$ denote the Markov semigroup of the contact process and let f denote the function $f(A) := |A|$. Then (2.8) says that

$$P_t f(A) \le e^{(2d\lambda-1)t} f(A) \qquad (t \ge 0).$$

Let G be the generator of $(P_t)_{t\ge 0}$. In our present notation,

$$Gf(A) = \lambda \sum_{i\in A} \sum_{\substack{j\notin A \\ |i-j|=1}} \big(f(A\cup\{j\}) - f(A)\big) + \sum_{i\in A} \big(f(A\setminus\{i\}) - f(A)\big)$$

$$= \lambda|\{(i,j) \in \mathcal{E} : i \in A,\ j \notin A\}| - |A| \le (2d\lambda - 1)\,|A|.$$

This says that $Gf \le (2d\lambda - 1)f$. Assuming that we can differentiate semigroups as in the finite-dimensional case, we have

$$\tfrac{\partial}{\partial t}P_t f = \lim_{\varepsilon\downarrow 0}\varepsilon^{-1}(P_\varepsilon f - f)P_t = GP_t f \qquad (t \ge 0),$$

hence

$$\tfrac{\partial}{\partial t}\big(e^{(1-2d\lambda)t}P_t f\big) = (1 - 2d\lambda)e^{(1-2d\lambda)t}P_t f + e^{(1-2d\lambda)t}GP_t f \le 0$$

and therefore $e^{(1-2d\lambda)t}P_t f \le e^{(1-2d\lambda)0}P_0 f$, which implies (2.8).

Although our argument is now more formal, it is still not rigorous, since we do not know if we can differentiate our semigroup, which acts on functions in the infinite-dimensional space $B(\{0,1\}^{\mathbb{Z}^d})$, in the same way as we would do for a Markov process with a finite state space. To fix this, we use an approximation argument. For $N \ge 1$, let $\Lambda_N := \{-N, l\ldots, N\}^d$ denote a box of length $2N + 1$ centered around the origin. Using the graphical representation of the contact process, for each $A \subset \mathbb{Z}^d$, we set

$$\eta_t^{(N),A} := \big\{i \in \Lambda_N : (A \cap \Lambda_N) \times \{0\} \rightsquigarrow_N (i,t)\big\},$$

where $\rightsquigarrow_N$ to indicates the presence of an open path that stays in Λ_N. It is easy to see that

$$\eta_t^{(N),A} \uparrow \eta_t^A \quad \text{as } N \uparrow \infty \quad (t \geq 0).$$

By Theorem 1.8, $(\eta_t^{(N),A})_{t \geq 0}$ is a Markov process with finite state space. By essentially the same generator calculation as above, which is now fully justified since our semigroup acts on the finite-dimensional space $B(\{0,1\}^{\Lambda_N})$, we find that

$$\mathbb{E}[|\eta_t^{(N),A}|] \leq |A| e^{(2d\lambda-1)t} \qquad (t \geq 0, \ N \geq 1).$$

Letting $N \uparrow \infty$ we arrive at (2.8).

It follows from (2.8) that $\mathbb{P}[|\eta_t^A| < \infty] = 1$ for all $t \geq 0$. In order to complete the proof of Proposition 2.5, we need to sharpen this to the statement that $\mathbb{P}[|\eta_t^A| < \infty \ \forall t \geq 0] = 1$. Let $(\zeta_t^A)_{t \geq 0}$ be a contact process on $\mathbb{Z}^d$ without recoveries, i.e.,

$$\zeta_t^A := \{i \in \Lambda_N : A \times \{0\} \rightsquigarrow' (i,t)\},$$

where $\rightsquigarrow'$ is defined in the same way as $\rightsquigarrow$, with the exception that this time, paths need not avoid recovery symbols. Then obviously $\eta_t^A \subset \zeta_t^A$ for all $t \geq 0$, while the same reasoning as before gives

$$\mathbb{E}[|\zeta_t^A|] \leq |A| e^{2d\lambda t} \qquad (t \geq 0).$$

Since $|\zeta_t^A|$ is a nondecreasing function of time, we have

$$\mathbb{P}[|\eta_s^A| < \infty \ \forall s \in [0,t]] \geq \mathbb{P}[|\zeta_s^A| < \infty \ \forall s \in [0,t]] \geq \mathbb{P}[|\zeta_t^A| < \infty] = 1 \quad (t > 0).$$

Letting $t \uparrow \infty$ we find that $|\eta_t^A| < \infty$ for all $t \geq 0$ a.s. ∎

Proposition 2.5 has the following consequence.

Corollary 2.6 (Lower bound on critical infection rate) *The critical infection rate of the nearest-neighbour contact process on $\mathbb{Z}^d$ satisfies $\frac{1}{2d} \leq \lambda_c$.*

Proof By (2.8), for each $\lambda < \frac{1}{2d}$,

$$\mathbb{P}[\eta_t^A \neq \emptyset] \leq \mathbb{E}[|\eta_t^A|] \xrightarrow[t \to \infty]{} 0$$

for each finite $A \subset \mathbb{Z}^d$. ∎

In order to finish the proof of Theorem 2.3 we need to show that $\lambda_c < \infty$. As a preparation for this, in the next section, we will start by studying a closely related problem. Before we do this, we apply the tecnhiques developed so far to prove the following facts about the function $\theta(\lambda, d)$.

Proposition 2.7 (Monotonicty and right-continuity) *The survival probability $\theta(\lambda, d)$ is nondecreasing and right-continuous in λ, and nondecreasing in d.*

Proof The fact that $\theta(\lambda, d)$ is nondecreasing in λ follows from Lemma 2.4. The fact that $\theta(\lambda, d)$ is nondecreasing in d can be proved in a similar way, since if $d \leq d'$, then we may view $\mathbb{Z}^d$ as a subset of $\mathbb{Z}^{d'}$ and observe that if there is an open path that stays in $\mathbb{Z}^d$, then certainly there is an open path in $\mathbb{Z}^{d'}$.

To prove right continuity of $\theta(\lambda, d)$ in λ, we will improve the coupling used in the proof of Lemma 2.4 in such a way that we can define contact processes for any value of the infection rate on the same probability space. To this aim, consider the space $\mathcal{E} \times [0, \infty) \times [0, \infty)$ whose elements are triples $((i, j), t, \kappa)$ with $(i, j) \in \mathcal{E}$ and $t, \kappa \geq 0$, and let $\overline{\Delta}^{\mathrm{i}}$ be a Poisson point set on this set with indensity $\mathrm{d}t\mathrm{d}\kappa$. Then, for each $\lambda \geq 0$,

$$\Delta_\lambda^{\mathrm{i}} := \big(((i, j), t) : \exists \big((i, j), t, \kappa\big) \in \overline{\Delta}^{\mathrm{i}} \text{ with } \kappa \leq \lambda\big\}.$$

is a Poisson point sets on $\mathcal{E} \times [0, \infty)$ with intensity $\lambda \mathrm{d}t$. Let Δ^{r} be an independent Poisson point set on $\mathbb{Z}^d \times [0, \infty)$ with intensity $\mathrm{d}t$ and write $\rightsquigarrow_\lambda$ to indicate the presence of an open path in the graphical representations defined by $(\Delta_\lambda^{\mathrm{i}}, \Delta^{\mathrm{r}})$. Another way of saying this is that a point $((i, j), t, \kappa) \in \overline{\Delta}^{\mathrm{i}}$ indicates the presence of an arrow which has a value κ attached to it, and $\rightsquigarrow_\lambda$ indicates the presence of a path that may use only arrows with values $\kappa \leq \lambda$. Then we claim that

$$\lim_{\lambda \downarrow \lambda_0} \theta(\lambda_n) = \lim_{\lambda \downarrow \lambda_0} \mathbb{P}[(0, 0) \rightsquigarrow_\lambda \infty] = \mathbb{P}[(0, 0) \rightsquigarrow_\lambda \infty \; \forall \lambda > \lambda_0]$$
$$\overset{!}{=} \mathbb{P}[(0, 0) \rightsquigarrow_{\lambda_0} \infty] = \theta(\lambda_0).$$

The equality $\overset{!}{=}$ needs some explanation. It is obvious that $(0, 0) \rightsquigarrow_{\lambda_0} \infty$ implies $(0, 0) \rightsquigarrow_\lambda \infty \; \forall \lambda > \lambda_0$. On the other hand, if $(0, 0) \not\rightsquigarrow_{\lambda_0} \infty$ then by Proposition 2.5

$$\mathcal{I} := \big\{(i, t) \in \mathbb{Z}^d \times [0, \infty) : (0, 0) \rightsquigarrow_\lambda (i, t)\big\}$$

is a compact subset of $\mathbb{Z}^d \times [0, \infty)$, such that each each infection ends somewhere in a recovery sign, and all infection arrows starting in $\mathcal{I}$ and ending somewhere outside $\mathcal{I}$ have a value strictly larger than λ. Since there are only finitely many arrows with values $\kappa \in (\lambda, 2\lambda)$ starting in $\mathcal{I}$ and ending somewhere outside $\mathcal{I}$, we know that there is some $\lambda' > \lambda$ such that all arrows starting in $\mathcal{I}$ and ending somewhere outside $\mathcal{I}$ have a value larger than λ', i.e., we know that $(0, 0) \not\rightsquigarrow_{\lambda'} \infty$ for some $\lambda' > \lambda$. ∎

2.6   Oriented percolation

In order to prepare for the proof that the critical infection rate of the contact process is finite, in the present section, we will study *oriented* (or *directed*) *bond percolation* on $\mathbb{Z}^d$.   For $i, j \in \mathbb{Z}^d$, we write $i \leq j$ if $i = (i_1, \ldots, i_d)$ and $j = (j_1, \ldots, j_d)$ satisfy $i_k \leq j_k$ for all $k = 1, \ldots, d$. Let

$$\mathcal{A} := \big\{(i,j) : i,j \in \mathbb{Z}^d, \ i \leq j, \ |i - j| = 1\big\}.$$

We view $\mathbb{Z}^d$ as an infinite directed graph, where elements $(i, j) \in \mathcal{A}$ represent arrows (or directed *bonds*) between neighbouring sites. Note that all arrows point 'upwards' in the sense of the natural order on $\mathbb{Z}^d$.

Now fix some *percolation parameter* $p \in [0, 1]$ and let $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ be a collection of i.i.d. Bernoulli random variables with $\mathbb{P}[\omega_{(i,j)} = 1] = p$. We say that there is an *open path* from a site $i \in \mathbb{Z}^d$ to $j \in \mathbb{Z}^d$ if there exist $n \geq 0$ and a function $\gamma : \{0, \ldots, n\} \to \mathbb{Z}^d$ such that

$$(\gamma(k-1), \gamma(k)) \in \mathcal{A} \quad \text{and} \quad \omega_{(\gamma(k-1),\gamma(k))} = 1 \qquad (k = 1, \ldots, n).$$

We denote the presence of an open path by $\rightsquigarrow$. Note that open paths must walk upwards in the sense of the order on $\mathbb{Z}^d$. We write $0 \rightsquigarrow \infty$ to indicate the existence of an infinite open path starting at the origin $0 \in \mathbb{Z}^d$.

Theorem 2.8 (Critical percolation parameter) *For oriented percolation in dimensions $d \geq 2$ there exists a critical parameter $p_c = p_c(d)$ with $0 < p_c < 1$ such that $\mathbb{P}[0 \rightsquigarrow \infty] = 0$ for $p < p_c$ and $\mathbb{P}[0 \rightsquigarrow \infty] > 0$ for $p > p_c$.*

Proof The existence of a critical parameter $p_c \in [0, 1]$ follows from a monotone coupling argument like the one we used in the proof of Lemma 2.4. To prove that $0 < p_c$, let N_n denote the number of open paths of length n starting in 0. Since there are d^n different upward paths of length n starting at the origin, and each path has probability p^n to be open, we see that

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} N_n\Big] = \sum_{n=1}^{\infty} d^n p^n < \infty \qquad (p < 1/d)$$

This shows that $\sum_{n=1}^{\infty} N_n < \infty$ a.s., hence $\mathbb{P}[0 \rightsquigarrow \infty] = 0$ if $p < 1/d$, and therefore

$$\frac{1}{d} \leq p_c(d).$$

To prove that $p_c(d) < 1$ for $d \geq 2$ it suffices to consider the case $d = 2$, for we may view $\mathbb{Z}^2$ as a subset of $\mathbb{Z}^d$ $(d \geq 3)$ and then, if there is an open path that stays in
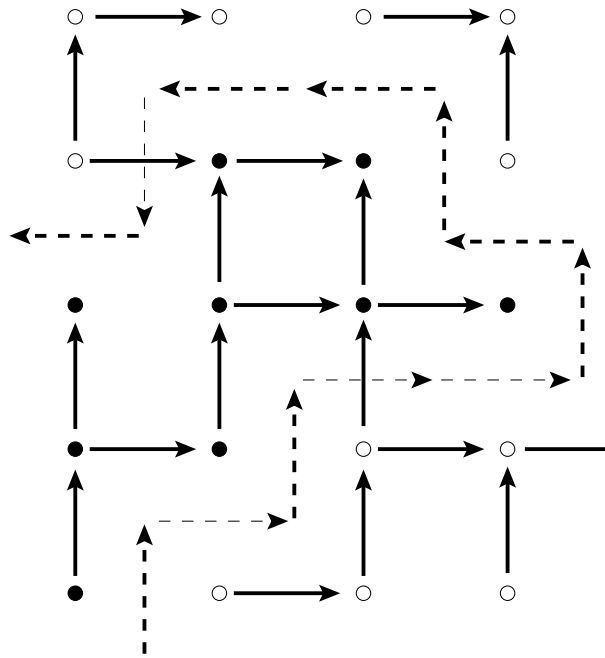
Figure 2.3: Peierls argument for oriented percolation.

$\mathbb{Z}^2$, then certainly there is an open path in $\mathbb{Z}^d$. (Note, by the way, that in $d = 1$ one has $\mathbb{P}[0 \rightsquigarrow \infty] = 0$ for all $p < 1$ hence $p_c(1) = 1$.)

We claim that

$$p_c(2) \leq \frac{8}{9}. \tag{2.9}$$

To prove this, we use a Peierls argument, named after R. Peierls who used a similar argument in 1936 for the Ising model [Pei36]. In Figure 2.3, we have drawn a piece of $\mathbb{Z}^2$. Open arrows are drawn in black; closed arrows are not drawn. Sites $i \in \mathbb{Z}^2$ such that $0 \rightsquigarrow i$ are indicated in black. These sites are called *wet*. Consider the *dual lattice*

$$\hat{\mathbb{Z}}^2 := \{(n + \tfrac{1}{2}, m + \tfrac{1}{2}) : (n, m) \in \mathbb{Z}^2\}.$$

If there are only finitely many wet sites, then the set of all non-wet sites contains one infinite connected component. (Here 'connected' is to be interpreted in terms of the unoriented graph $\mathbb{N}^2$ with nearest-neighbor edges.) Let γ be the boundary of this infinite component. Then γ is a nearest-neighbor path in $\hat{\mathbb{Z}}^2$, starting in some point $(n + \tfrac{1}{2}, -\tfrac{1}{2})$ and ending in some point $(-\tfrac{1}{2}, m + \tfrac{1}{2})$ with $n, m \geq 0$, such that all sites immediately to the left of γ are wet, and no open arrows starting at these sites cross γ. In Figure 2.3, we have indicated γ with dashed arrows.

From these considerations, we see that the following statement is true: one has $0 \nrightarrow \infty$ if and only if there exists a path in $\hat{\mathbb{Z}}^2$, starting in some point $(n + \tfrac{1}{2}, -\tfrac{1}{2})$ ($n \geq 0$), ending in some point $(-\tfrac{1}{2}, m + \tfrac{1}{2})$ ($m \geq 0$), and passing to the northeast of the origin, such that all arrows of γ in the north and west directions (indicated in bold in the figure) are not be crossed by an open arrow. Let M_n be the number of paths of length n with these properties. Since there are n dual sites from where such a path of length n can start, and since in each step, there are three directions where it can go, there are at most $n3^n$ paths of length n with these properties. Since each path must make at least half of its steps in the north and east directions, the expected number of these paths satisfies

$$\mathbb{E}\Big[\sum_{n=2}^{\infty} M_n\Big] \leq \sum_{n=2}^{\infty} n3^n(1-p)^{n/2} < \infty \qquad (p > \tfrac{8}{9})$$

and therefore

$$\mathbb{P}[0 \nrightarrow \infty] \leq \mathbb{P}\Big[\sum_{n=2}^{\infty} M_n \geq 1\Big] \leq \mathbb{E}\Big[\sum_{n=2}^{\infty} M_n\Big] < \infty.$$

This does not quite prove what we want yet, since we need the right-hand side of this equation to be less than one. To fix this, set $D_m := \{0, \ldots, m\}^2$. Then, by

the same arguments as before

$$\mathbb{P}[D_m \not\leadsto \infty] \leq \mathbb{P}\big[\sum_{n=2m}^{\infty} M_n \geq 1\big] \leq \mathbb{E}\big[\sum_{n=2m}^{\infty} M_n\big] \leq \sum_{n=2m}^{\infty} n3^n(1-p)^{n/2},$$

which in case $p > \frac{8}{9}$ can be made arbitrarily small by choosing m suffiently large. It follows that $\mathbb{P}[D_m \leadsto \infty] > 0$ for some m, hence $\mathbb{P}[i \leadsto \infty] > 0$ for some $i \in D_m$, and therefore, by translation invariance, also $\mathbb{P}[0 \leadsto \infty] > 0$. ∎

2.7 Survival

In the present section, we will complete the proof of Theorem 2.3 by showing that $\lambda_c < \infty$. The method we will use is comparison with oriented percolation. This neither leads to a particularly short proof nor does it yield a very good upper bound on λ_c, but it has the advantage that it is a very robust method that can be applied to many other interacting particle systems.

Let $\lambda_c(d)$ be the critical infection rate of the nearest-neighbour contact process on $\mathbb{Z}^d$. If $d \leq d'$, then we may view $\mathbb{Z}^d$ as a subset of $\mathbb{Z}^{d'}$, so by an obvious monotone coupling we see that

$$\lambda_c(d) \geq \lambda_c(d') \qquad (d \leq d').$$

In view of this, in order to finish the proof of Theorem 2.3, it suffices to show that $\lambda_c(1) < \infty$.

For notational convenience, we extend (without any harm done) our graphical representation to negative times, i.e., we let Δ^i and Δ^r be Poisson point subsets of $\mathcal{E} \times \mathbb{R}$ and $\mathbb{Z}^d \times \mathbb{R}$, respectively. We fix $T > 0$ and define a map from $\mathbb{Z}^2$ to $\mathbb{Z} \times \mathbb{R}$ by

$$\big(\kappa_i, \sigma_i\big) := \big(i_1 - i_2, T(i_1 + i_2)\big) \qquad \big(i = (i_1, i_2) \in \mathbb{Z}^2\big).$$

Recall from the previous section the definition of the set $\mathcal{A}$ of arrows on $\mathbb{Z}^d$. We wish to define a collection $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ of Bernoulli random variables such that

$$\omega_{(i,j)} = 1 \quad \text{implies} \quad (\kappa_i, \sigma_i) \leadsto (\kappa_j, \sigma_j) \qquad \big((i,j) \in \mathcal{A}\big).$$

For each $i \in \mathbb{Z}^2$ we define

$$\tau_i^- := \inf\big\{t \geq \sigma_i : (\kappa_i, \kappa_i - 1, t) \in \Delta^i\big\},$$
$$\tau_i^+ := \inf\big\{t \geq \sigma_i : (\kappa_i, \kappa_i + 1, t) \in \Delta^i\big\},$$

and we define the 'good events'

$$\mathcal{G}_i^- := \big\{ \tau_i^- < \sigma_i + T, \ \big(\{\kappa_i\} \times (\sigma_i, \tau_i^-)\big) \cap \Delta^{\mathrm{r}} = \emptyset,$$
$$\big(\{\kappa_i - 1\} \times (\tau_i^-, \sigma_i + T)\big) \cap \Delta^{\mathrm{r}} = \emptyset \big\},$$
$$\mathcal{G}_i^+ := \big\{ \tau_i^+ < \sigma_i + T, \ \big(\{\kappa_i\} \times (\sigma_i, \tau_i^+)\big) \cap \Delta^{\mathrm{r}} = \emptyset,$$
$$\big(\{\kappa_i + 1\} \times (\tau_i^+, \sigma_i + T)\big) \cap \Delta^{\mathrm{r}} = \emptyset \big\}.$$

We observe that the event $\mathcal{G}_i^\pm$ implies that $(\kappa_i, \sigma_i) \rightsquigarrow (\kappa_i \pm 1, \sigma_i + T)$ via an open path that stays in $\{\kappa_i, \kappa_i \pm 1\}$. In view of this, we set

$$\omega\big((i_1, i_2), (i_1 + 1, i_2)\big) := 1_{\mathcal{G}_i^+},$$
$$\omega\big((i_1, i_2), (i_1, i_2 + 1)\big) := 1_{\mathcal{G}_i^-}.$$

Then $\omega_{(i,j)} = 1$ implies the existence of an open path in the graphical representation for the contact process from (κ_i, σ_i) to (κ_j, σ_j) (with $(i,j) \in \mathcal{A}$), hence if we use the random variables $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ to define oriented percolation on $\mathbb{Z}^2$ in the usual way, then:

 > $i \rightsquigarrow j$ in the oriented percolation on $\mathbb{Z}^2$ defined by the random variables $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ implies $(\kappa_i, \sigma_i) \rightsquigarrow (\kappa_i, \sigma_i)$ in the graphical representation for the contact process.

We observe that

$$p := \mathbb{P}[\omega_{(i,j)} = 1] = \mathbb{P}(\mathcal{G}_i^\pm) = (1 - e^{-\lambda T}) e^{-T} \qquad \big((i,j) \in \mathcal{A}\big). \tag{2.10}$$

For λ sufficiently large, by a suitable choice of T, we can make p as close to one as we wish. We would like to conclude from this that $\mathbb{P}[(0,0) \rightsquigarrow \infty] > 0$ for the oriented percolation defined by the $\omega_{(i,j)}$'s, and therefore also $\mathbb{P}[(0,0) \rightsquigarrow \infty] > 0$ for the contact process. Unfortunately, the random variables $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ are not independent, and therefore Theorem 2.8 is not applicable. Luckily, the $(\omega_{(i,j)})_{(i,j) \in \mathcal{A}}$ are k-dependent for some suitable k, so by applying Theorem A.1 we can estimate them from below by an independent collection of Bernoulli random variables $(\tilde{\omega}_{(i,j)})_{(i,j) \in \mathcal{A}}$ whose succes probability $\tilde{p}$ can be made arbitrarily close to one, so we are done. ∎

Exercise 2.9 Combine formulas (2.9), (2.10) and (A.1) to derive an explicit upper bound on the critical infection rate λ_{c} of the one-dimensional contact process.

2.8 Duality and the Feller property

We extend our graphical representation to negative times and, generalizing (2.4), we define

$$
\begin{aligned}
\eta_t^{A,s} &:= \{i : A \times \{s\} \rightsquigarrow (i, s+t)\}, \\
\eta_t^{\dagger A,s} &:= \{i : (i, s-t) \rightsquigarrow A \times \{s\}\},
\end{aligned}
\tag{2.11}
$$

$(A \subset \mathbb{Z}^d,\ s \in \mathbb{R},\ t \geq 0)$. Since it is defined by a graphical representation obtained by turning our original graphical representation upside down, we see that

$$
(\eta_t^{\dagger A,s})_{t \geq 0}
$$

is a contact process with the same infection rate as our original process. Moreover, we observe that for any $t \geq 0$ and $A, B \subset \mathbb{Z}^d$,

$$
\mathbb{P}\big[\eta_t^{A,0} \cap B \neq \emptyset\big] = \mathbb{P}\big[A \times \{0\} \rightsquigarrow B \times \{t\}\big] = \mathbb{P}\big[A \cap \eta_t^{\dagger B,t} \neq \emptyset\big].
$$

This formula remains true if A and B are random sets, independent of the Poisson point processes of our graphical representation. This means that we have proved the following result.

Proposition 2.10 (Duality) *Let $(\eta_t)_{t \geq 0}$ and $(\eta_t')_{t \geq 0}$ be independent contact processes on $\mathbb{Z}^d$ with infection rate λ. Then*

$$
\mathbb{P}\big[\eta_t \cap \eta_0' \neq \emptyset\big] = \mathbb{P}\big[\eta_0 \cap \eta_t' \neq \emptyset\big] \qquad (t \geq 0).
$$

This result is especially useful in view of the following fact. Below and in what follows, $\mathcal{P}(\mathbb{Z}^d) := \{A : A \subset \mathbb{Z}^d\}$ denotes the set of all subsets of $\mathbb{Z}^d$, which is in a natural way isomorphic to $\{0,1\}^{\mathbb{Z}^d}$ (equipped with the product topology).

Lemma 2.11 (Distribution determining functions) *Let μ, ν be probability laws on $\mathcal{P}(\mathbb{Z}^d)$ such that*

$$
\int \mu(\mathrm{d}A) 1_{\{A \cap B \neq \emptyset\}} = \int \nu(\mathrm{d}A) 1_{\{A \cap B \neq \emptyset\}}
$$

for all finite nonempty $B \subset \mathbb{Z}^d$. Then $\mu = \nu$.

Proof We start by recalling the *Stone-Weierstrass theorem*. Let E be a compact metrizable set. By definition, a subset $\mathcal{F}$ of $\mathcal{C}(E)$ is an *algebra* if $\mathcal{F}$ is a linear space, $\mathcal{F}$ contains the constant function 1, and $f, g \in \mathcal{F}$ implies $fg \in \mathcal{F}$. We say that $\mathcal{F}$ *separates points* if for every $x, y \in E$ with $x \neq y$ there exists an $f \in \mathcal{F}$ with

$f(x) \neq f(y)$. The Stone-Weierstrass theorem says that if subset $\mathcal{F}$ of $\mathcal{C}(E)$ is an algebra that separates points, then $\mathcal{F}$ is dense in $\mathcal{C}(E)$.

Let $\mathcal{F}$ be the linear span of all functions of the form $A \mapsto 1_{\{A \cap B = \emptyset\}}$ with B a finite subset of $\mathbb{Z}^d$. It follows from our assumptions that

$$\int \mu(\mathrm{d}A) 1_{\{A \cap B = \emptyset\}} = \int \nu(\mathrm{d}A) 1_{\{A \cap B = \emptyset\}}$$

for each finite $B \subset \mathbb{Z}^d$, hence $\int \mu(\mathrm{d}A) f(A) = \int \nu(\mathrm{d}A) f(A)$ for all $f \in \mathcal{F}$ and therefore, by the Stone-Weierstrass theorem, $\int \mu(\mathrm{d}A) f(A) = \int \nu(\mathrm{d}A) f(A)$ for all $f \in \mathcal{C}(E)$, which implies $\mu = \nu$. $\blacksquare$

We define a transition probability on $\mathcal{P}(\mathbb{Z}^d)$ by

$$P_t(A, \cdot) := \mathbb{P}[\eta_t^A \in \cdot] \qquad (A \in \mathcal{P}(\mathbb{Z}^d), \ t \geq 0).$$

Lemma 2.12 (Feller property) *The $(P_t)_{t \geq 0}$ form a continuous transition probability on $\mathcal{P}(\mathbb{Z}^d)$.*

Proof The fact that

$$\int P_s(A, \mathrm{d}B) P_t(B, \mathrm{d}C) = P_{s+t}(A, \mathrm{d}C) \qquad (A \in \mathcal{P}(\mathbb{Z}^d), \ s, t \geq 0)$$

follows easily from our graphical representation and the fact that Poisson point sets restricted to disjoint parts of space are independent. To see that

$$(A, t) \mapsto P_t(A, \cdot)$$

is continuous, assume that $A_n \to A$ and $0 \leq t_n \to t$. Here $A_n \to A$ means that $1_{A_n}(i) \to 1_A(i)$ for each $i \in \mathbb{Z}^d$, or equivalently,

$$\forall i \in \mathbb{Z}^d \ \exists N \text{ s.t. } 1_{A_n}(i) = 1_A(i) \ \forall n \geq N.$$

We observe that

$$P_{t_n}(A_N, \cdot) = \mathbb{P}\big[\eta_{t_n}^{A_n, -t_n} \in \cdot\big].$$

We claim that $(A_n, t_n) \to (A, t)$ implies

$$\eta_{t_n}^{A_n, -t_n} \xrightarrow[n \to \infty]{} \eta_t^{A, -t} \quad \text{a.s.} \tag{2.12}$$

Indeed,

$$1_{\eta_{t_n}^{A_n, -t_n}}(i) = 1_{\{A_n \cap \eta_{t_n}^{\dagger \{i\}, 0} \neq \emptyset\}}.$$

By Proposition 2.5, $|\eta_s^{\dagger\,\{i\},0}| < \infty$ for all $s \geq 0$ a.s. It follows that there exists some (random) N such that $\eta_{t_n}^{\dagger\,\{i\},0} = \eta_t^{\dagger\,\{i\},0}$ for all $n \geq N$. Using this and the fact that $A_n \to A$, it is easy to see that

$$1_{\{A_n \,\cap\, \eta_{t_n}^{\dagger\,\{i\},0} \,\neq\, \emptyset\}} \xrightarrow[n\to\infty]{} 1_{\{A \,\cap\, \eta_t^{\dagger\,\{i\},0} \,\neq\, \emptyset\}} \quad \text{a.s.,}$$

which implies (2.12). As a consequence, we have

$$\mathbb{E}\big[f(\eta_{t_n}^{A_n,-t_n})\big] \xrightarrow[n\to\infty]{} \mathbb{E}\big[f(\eta_t^{A,-t})\big] \qquad \big(f \in \mathcal{C}(\mathcal{P}(\mathbb{Z}^d))\big),$$

which proves the desired continuity.   ∎

2.9   The upper invariant law

Extending our graphical representation to negative times as before, we define

$$\overline{\eta}_t := \big\{i \in \mathbb{Z}^d : -\infty \rightsquigarrow (i,t)\big\} \qquad (t \in \mathbb{R}).$$

Using the independence of restrictions of Poisson point processes to disjoint parts of space, we see that

$$\mathbb{P}\big[\overline{\eta}_u \in \cdot \,\big|\, (\overline{\eta}_s)_{s\leq t}\big] = P_{u-t}(\overline{\eta}_t, \cdot) \quad \text{a.s.} \quad (t \leq u),$$

hence $(\overline{\eta}_t)_{t\in\mathbb{R}}$ is a stationary contact process and its law at any given time

$$\overline{\nu} := \mathbb{P}\big[\overline{\eta}_t \in \cdot\,\big] \qquad (t \in \mathbb{R})$$

is an invariant law of the contact process. We call $\overline{\nu}$ the *upper invariant law* of the contact process (with given infection rate). As we will see in a moment, in a certain sense, it is the 'largest' invariant law of our process.

By definition, we say that a function $f : \{0,1\}^{\mathbb{Z}^d}$ is *monotone* if $f(x) \leq f(y)$ for all $x \leq y$.

Proposition 2.13 (Stochastic order) *Let μ, ν be probability laws on $\{0,1\}^{\mathbb{Z}^d}$. Then the following statements are equivalent:*

(i) $\int \mu(\mathrm{d}x)f(x) \leq \int \mu(\mathrm{d}x)f(x) \;\forall$ *monotone* $f \in \mathcal{C}(\{0,1\}^{\mathbb{Z}^d})$,

(ii) $\int \mu(\mathrm{d}x)f(x) \leq \int \mu(\mathrm{d}x)f(x) \;\forall$ *monotone* $f \in B(\{0,1\}^{\mathbb{Z}^d})$,

(iii) *It is possible to couple random variables X, Y with laws $\mu = P[X \in \cdot\,]$ and $\nu = P[Y \in \cdot\,]$ in such a way that $X \leq Y$.*

Proof The implications (iii)⇒(ii)⇒(i) are trivial. For the nontrivial implication (i)⇒(iii) (which we will never actually need to use), see [Lig85, Theorem II.2.4]. ∎

If probability laws μ, ν on $\{0,1\}^{\mathbb{Z}^d}$ satisfy the equivalent conditions (i)–(iii) from Proposition 2.13, then we say that μ and ν are *stochastically ordered* and we write[1] $\mu \leq \nu$.

Lemma 2.14 (Upper invariant law) *Let $\overline{\nu}$ be the upper invariant law of the contact process and let ν be any other invariant law. Then $\nu \leq \overline{\nu}$ in the stochastic order.*

Proof Let A be a random variable, taking values in $\mathcal{P}(\mathbb{Z}^d)$, with law $\mathbb{P}[A \in \cdot] = \nu$, and assume that A is independent of the Poisson point processes used in our graphical representation. Then, since ν is an invariant law, we have $\nu = \mathbb{P}[\eta_t^A \in \cdot]$ for all $t \geq 0$. Since the random variables

$$(\eta_t^A, \overline{\eta}_t),$$

take values in the compact space $\mathcal{P}(\mathbb{Z}^d)^2$, their laws are automatically tight, hence we can select a subsequence $t_n \to \infty$ such that the $(\eta_{t_n}^A, \overline{\eta}_{t_n})$ converge weakly in law to some limiting random variable (η^1, η^2), say, where η^1 has the law ν, η^2 has the law $\overline{\nu}$, and moreover

$$\mathbb{P}[i \in \eta^1, \ i \notin \eta^2] = \lim_{n \to \infty} \mathbb{P}[i \in \eta_{t_n}^A, \ i \notin \overline{\eta}_{t_n}] = \lim_{n \to \infty} \mathbb{P}[i \in \eta_0^{A, -t_n}, \ i \notin \overline{\eta}_0]$$
$$\leq \lim_{n \to \infty} \mathbb{P}[\mathbb{Z}^d \times \{-t_n\} \rightsquigarrow (i,0), \ -\infty \not\rightsquigarrow (i,0)] = 0 \qquad (i \in \mathbb{Z}^d),$$

where we have used that the events $\mathbb{Z}^d \times \{-t_n\} \rightsquigarrow (i,0)$ decrease monotonically to the event $-\infty \rightsquigarrow (i,0)$, hence the events $\mathbb{Z}^d \times \{-t_n\} \rightsquigarrow (i,0), \ -\infty \not\rightsquigarrow (i,0)$ decrease monotonically to the empty set. We conclude that $\eta^1 \subset \eta^2$ a.s., hence $\nu \leq \overline{\nu}$ in the stochastic order. ∎

By definition, we say that a probability law μ on $\mathcal{P}(\mathbb{Z}^d)$ is *nontrivial* if

$$\mu(\{\emptyset\}) = 0,$$

i.e., if μ gives zero probability to the configuration in which all sites are healthy.

---

[1]This notation may look a bit confusing at first sight, since, if μ, ν are probability measures on any measurable space $(\Omega, \mathcal{F})$, then one might interpret $\mu \leq \nu$ in a pointwise sense, i.e., in the sense that $\mu(A) \leq \nu(A)$ for all $A \in \mathcal{F}$. In practice, this does not lead to confusion, since pointwise inequality for probability measures is a nonsensical property. Indeed, it is easy to check that probability measures μ, ν satisfy $\mu \leq \nu$ in a pointwise sense if and only if $\mu = \nu$.

Lemma 2.15 (Survival and the upper invariant law) *For the contact process on $\mathbb{Z}^d$ with infection rate $\lambda \geq 0$, the following statements are equivalent:*

(i) *The contact process survives, i.e., $\theta(\lambda, d) > 0$.*

(ii) *The upper invariant law $\overline{\nu}$ is nontrivial.*

(iii) *There exists a nontrivial invariant law.*

Moreover, if the contact process dies out, then $\overline{\nu} = \delta_\emptyset$.

Proof The implication (ii)⇒(iii) is trivial and (iii)⇒(ii) follows from Lemma 2.14. To see that (i) and (ii) are equivalent, we start by observing that by duality, for each finite $B \subset \mathbb{Z}^d$

$$
\begin{aligned}
\int \overline{\nu}(\mathrm{d}A) 1_{\{A \cap B \neq \emptyset\}} &= \mathbb{P}[\overline{\eta}_0 \cap B \neq \emptyset] \\
&= \mathbb{P}[\eta_t^{\dagger\, B,0} \neq \emptyset \ \forall t \geq 0] = \mathbb{P}[\eta_t^B \neq \emptyset \ \forall t \geq 0].
\end{aligned}
\tag{2.13}
$$

Note that by Lemma 2.11, this formula determines the law $\overline{\nu}$ uniquely. In particular, we see that $\overline{\nu} = \delta_\emptyset$ if the contact process dies out. On the other hand, if the contact process survives, then $\overline{\nu} \neq \delta_\emptyset$. This is not quite the same as saying that $\overline{\nu}$ is nontrivial, but at least it tells us that $\mathbb{P}[\overline{\eta}_0 \neq \emptyset] > 0$. We observe that

$$
\mathbb{P}[\overline{\eta}_t \neq \emptyset] = \mathbb{P}[\overline{\eta}_t \neq \emptyset \,|\, \overline{\eta}_0 \neq \emptyset] \mathbb{P}[\overline{\eta}_0 \neq \emptyset],
$$

which by the stationarity of $\overline{\eta}$ implies that

$$
\mathbb{P}[\overline{\eta}_t \neq \emptyset \,|\, \overline{\eta}_0 \neq \emptyset] = 1 \qquad (t \geq 0).
$$

It follows that the conditioned law

$$
\overline{\nu}(\,\cdot\,|\{A : A \neq \emptyset\})
$$

is a nontrivial invariant law for the contact process, hence by the equivalence of (ii) and (iii), we must have that $\overline{\nu}$ is nontrivial. ∎

2.10   Ergodic behavior

We define translation operators $T_i : \mathcal{P}(\mathbb{Z}^d) \to \mathcal{P}(\mathbb{Z}^d)$ by

$$
T_i(A) := \{j + i : j \in A\} \qquad (i \in \mathbb{Z}^d).
$$

We say that a probability law μ on $\mathcal{P}(\mathbb{Z}^d)$ is *homogeneous* or *translation invariant* if $\mu \circ T_i^{-1} = \mu$ for all $i \in \mathbb{Z}^d$. A lot of work in the theory of interacting particle systems is concerned with classifying all invariant laws of a given system, and proving that the system started from certain initial laws converges in law to a certain invariant law. In the context of interacting particle systems, invariant laws are sometimes also called *equilibria* or *equilibrium laws*. If an interacting particle system has a unique invariant law, which is the limit law of the process started in any initial state, then it is often said that the system is *ergodic*.[2]

The main aim of the present section is to prove the following result.

Theorem 2.16 (Convergence to upper invariant law) *Let $(\eta_t)_{t \geq 0}$ be a contact process started in a homogeneus nontrivial initial law $\mathbb{P}[\eta_0 \in \cdot]$. Then*

$$\mathbb{P}[\eta_t \in \cdot] \underset{t \to \infty}{\Longrightarrow} \overline{\nu},$$

where $\overline{\nu}$ is the upper invariant law.

We start with two preparatory lemmas.

Lemma 2.17 (Extinction versus unbounded growth) *For each finite $A \subset \mathbb{Z}^d$, one has*

$$\eta_t^A = \emptyset \text{ for some } t \geq 0 \quad \text{or} \quad |\eta_t^A| \underset{t \to \infty}{\longrightarrow} \infty \quad \text{a.s.} \tag{2.14}$$

Proof Define

$$\rho(A) := \mathbb{P}\big[\eta_t^A \neq \emptyset \; \forall t \geq 0\big] \qquad (A \subset \mathbb{Z}^2, \; |A| < \infty).$$

It is not hard to see that for each $N \geq 0$ there exists an $\varepsilon > 0$ such that

$$|A| \leq N \quad \text{implies} \quad \rho(A) \leq 1 - \varepsilon. \tag{2.15}$$

We first argue why it is plausible that this implies (2.14) and then give a rigorous proof. Imagine that $|\eta_t^A| \nrightarrow \infty$. Then, in view of (2.15), the process infinitely often gets a chance of at least ε to die out, hence eventually it should die out. To make this rigorous, let

$$\mathcal{A}_A := \{\eta_t^A \neq \emptyset \; \forall t \geq 0\} \qquad (A \subset \mathbb{Z}^2, \; |A| < \infty).$$

---

[2]This is not very good terminology since it may lead to confusion with another, more usual concept of ergodicity. If $(X_t)_{t \in \mathbb{R}}$ is a stationary process, for example an interacting particle system in equilibrium, then by definition $(X_t)_{t \in \mathbb{R}}$ is *ergodic* if the law of $(X_t)_{t \in \mathbb{R}}$ gives probability zero or one to all events that are invariant under time shifts.

denote the event that the process $(\eta_t^A)_{t\geq 0}$ survives and let $\mathcal{F}_t$ be the σ-field gener-
ated by the Poisson point processes used in our graphical representation till time t.
Then

$$\rho(\eta_t^A) = \mathbb{P}\big[\mathcal{A}_A \,\big|\, \mathcal{F}_t\big] \xrightarrow[t\to\infty]{} 1_{\mathcal{A}_A} \quad \text{a.s.,} \tag{2.16}$$

where we have used an elementary result from probability theory which says that if
$\mathcal{F}_n$ is an increasing sequence of σ-fields and $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$, then $\lim_n \mathbb{P}[A|\mathcal{F}_n] =$
$\mathbb{P}[A|\mathcal{F}_\infty]$ a.s. for each measurable event A. (See [Loe63, § 29, Complement 10 (b)].)
In view of (2.15), formula (2.16) implies (2.14).   ■

Lemma 2.18 (Nonzero intersection) *Let* $(\eta_t)_{t\geq 0}$ *be a contact process started
in a homogeneus nontrivial initial law* $\mathbb{P}[\eta_0 \in \cdot]$. *Then for each* $s, \varepsilon > 0$ *there
exists an* $N \geq 1$ *such that for any subset* $A \subset \mathbb{Z}^d$

$$|A| \geq N \quad \text{implies} \quad \mathbb{P}\big[A \cap \eta_s \neq \emptyset\big] \geq 1 - \varepsilon.$$

Proof By duality (Lemma 2.10)

$$\mathbb{P}\big[A \cap \eta_s \neq \emptyset\big] = \mathbb{P}\big[\eta_s^A \cap \eta_0 \neq \emptyset\big]$$

where η_0 is independent of the graphical representation used to define η_s^A. Set
$\Lambda_M := \{-M, \dots, M\}^d$. It is not hard to see that for set $A \subset \mathbb{Z}^d$ with $|A| \geq N$
contains a subset $A' \subset A$ with $|A'| \geq N/|\Lambda_M|$ such that the sets

$$\big\{i + \Lambda_M : i \in A'\big\}$$

are disjoint, where as before we define $i + \Lambda_M = \{i + j : j \in \Lambda_M\}$. Write $\leadsto_{i+\Lambda_M}$
to indicate the presence of an open path that stays in $i + \Lambda_M$ and set

$$\eta_s^{\{i\}(M)} := \big\{j \in \mathbb{Z}^d : (i,0) \leadsto_{i+\Lambda_M} (j,s)\big\}.$$

Then, using Hölder's inequality[3] in the inequality marked with an exclamation

---

[3]Recall that Hölder's inequality says that $1/p + 1/q = 1$ implies $\|fg\|_1 \leq \|f\|_p\|g\|_q$, where
$\|f\|_p := (\int |f|^p \mathrm{d}\mu)^{1/p}$. By induction, this gives $\|\prod_{i=1}^n f_i\|_1 \leq \prod_{i=1}^n \|f_i\|_n$.

mark, we have

$$
\begin{aligned}
\mathbb{P}\big[\eta_s^A \cap \eta_0 = \emptyset\big] &= \int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\eta_s^A \cap B = \emptyset\big] \\
&\leq \int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\bigcup_{i \in A'} \eta_s^{\{i\}\,(M)} \cap B = \emptyset\big] \\
&= \int \mathbb{P}[\eta_0 \in \mathrm{d}B] \prod_{i \in A'} \mathbb{P}\big[\eta_s^{\{i\}\,(M)} \cap B = \emptyset\big] \\
&\overset{!}{\leq} \prod_{i \in A'} \Big(\int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\eta_s^{\{i\}\,(M)} \cap B = \emptyset\big]^{|A'|}\Big)^{1/|A'|} \\
&= \prod_{i \in A'} \Big(\int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\eta_s^{\{0\}\,(M)} \cap B = \emptyset\big]^{|A'|}\Big)^{1/|A'|} \\
&= \int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\eta_s^{\{0\}\,(M)} \cap B = \emptyset\big]^{|A'|}.
\end{aligned}
$$

Our arguments so far show that $|A| \geq N$ implies that

$$
\mathbb{P}\big[A \cap \eta_s = \emptyset\big] \leq \int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbb{P}\big[\eta_s^{\{0\}\,(M)} \cap B = \emptyset\big]^{N/|\Lambda_M|} =: f(N, M).
$$

Here, using the fact that

$$
\mathbb{P}\big[\eta_s^{\{0\}\,(M)} \cap B = \emptyset\big] < 1 \quad \text{if } B \cap \Lambda_M \neq \emptyset,
$$

we see that

$$
\lim_{N \uparrow \infty} f(N, M) = \int \mathbb{P}[\eta_0 \in \mathrm{d}B]\mathbf{1}_{\{B \cap \Lambda_M = \emptyset\}} = \mathbb{P}[\eta_0 \cap \Lambda_M = \emptyset].
$$

Since $\mathbb{P}[\eta_0 \in \cdot\,]$ is nontrivial, we have moreover

$$
\lim_{M \uparrow \infty} \mathbb{P}[\eta_0 \cap \Lambda_M = \emptyset] = \mathbb{P}[\eta_0 = \emptyset] = 0.
$$

Thus, we have shown that

$$
\lim_{M \to \infty} \lim_{N \to \infty} f(N, M) = 0.
$$

By a diagonal argument, for each $\varepsilon > 0$ we can choose N and M_N such that $f(N, M_N) \leq \varepsilon$, proving our claim. ∎

Exercise 2.19 Show by counterexample that the statement of Lemma 2.18 is false for $s = 0$.

Proof of Theorem 2.16 Since the space $\mathcal{P}(\mathbb{Z}^d)$ is compact, the laws of the η_t with $t \geq 0$ are tight, hence by Lemma 1.15 it suffices to prove that $\overline{\nu}$ is the only weak cluster point. By Lemma 2.11 and formula (2.13), it suffices to show that

$$\lim_{t\to\infty} \mathbb{P}\big[A \cap \eta_t \neq \emptyset\big] = \mathbb{P}\big[A \cap \overline{\eta}_0 \neq \emptyset\big] = \mathbb{P}\big[\eta_u^A \neq \emptyset \ \forall u \geq 0\big] =: \rho(A)$$

for all finite $A \subset \mathbb{Z}^d$. By duality (Lemma 2.10), this is equivalent to showing that

$$\lim_{t\to\infty} \mathbb{P}\big[\eta_{t-s}^A \cap \eta_s \neq \emptyset\big] = \rho(A) \qquad \big(A \subset \mathbb{Z}^d, \ |A| < \infty\big),$$

where $(\eta_t^A)_{t\geq 0}$ and $(\eta_t)_{t\geq 0}$ are independent and $s > 0$ is some fixed constant. For each $\varepsilon > 0$, we can choose N as in Lemma 2.18, and write

$$\begin{aligned}
\mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset\big] = {} & \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset \,\big|\, |\eta_t^A| = 0\big]\, \mathbb{P}\big[|\eta_t^A| = 0\big] \\
& + \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset \,\big|\, 0 < |\eta_t^A| < N\big]\, \mathbb{P}\big[0 < |\eta_t^A| < N\big] \\
& + \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset \,\big|\, |\eta_t^A| \geq N\big]\, \mathbb{P}\big[|\eta_t^A| \geq N\big].
\end{aligned}$$

Here, by Lemma 2.17 and our choice of N,

$$\begin{aligned}
&\text{(i)} \quad \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset \,\big|\, |\eta_t^A| = 0\big] = 0, \\
&\text{(ii)} \quad \lim_{t\to\infty} \mathbb{P}\big[0 < |\eta_t^A| < N\big] = 0, \\
&\text{(iii)} \quad \liminf_{t\to\infty} \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset \,\big|\, |\eta_t^A| \geq N\big] \geq 1 - \varepsilon, \\
&\text{(iv)} \quad \lim_{t\to\infty} \mathbb{P}\big[|\eta_t^A| \geq N\big] = \rho(A),
\end{aligned}$$

from which we conclude that

$$(1 - \varepsilon)\rho(A) \leq \liminf_{t\to\infty} \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset\big] \leq \limsup_{t\to\infty} \mathbb{P}\big[\eta_t^A \cap \eta_s \neq \emptyset\big] \leq \rho(A).$$

Since $\varepsilon > 0$ is arbitrary, our proof is complete. ∎

Theorem 2.16 has a simple corollary.

Corollary 2.20 (Homogeneous invariant laws) *All homogeneous invariant laws of a contact process are convex combinations of $\delta_\emptyset$ and $\overline{\nu}$.*

Proof If μ is a nontrivial homogeneous invariant law and $(\eta_t)_{t\geq 0}$ is a contact process started in the initial law μ, then by Theorem 2.16

$$\mu = \mathbb{P}[\eta_t \in \,\cdot\,] \underset{t\to\infty}{\Longrightarrow} \overline{\nu}.$$

This shows that if the contact process survives, and hence $\bar{\nu}$ is nontrivial (recall Lemma 2.15), then $\bar{\nu}$ is the only nontrivial homogeneous invariant law. We recall from the proof of Lemma 2.15 that if μ is any homogeneous invariant law, then we may write

$$\mu = \mu(\{\emptyset\})\delta_\emptyset + \big(1 - \mu(\{\emptyset\})\big)\mu(\,\cdot\,|\{A : A \neq \emptyset\})$$

where $\mu(\,\cdot\,|\{A : A \neq \emptyset\})$ is a nontrivial homogeneous invariant law. From this we see that all homogeneous invariant laws are convex combinations of $\delta_\emptyset$ and $\bar{\nu}$. On the other hand, if the contact process dies out, then $\bar{\nu} = \delta_\emptyset$ is the largest invariant law with respect to the stochastic order, hence $\delta_\emptyset$ is the only invariant law. ∎

As an application of Theorem 2.16, we prove the following result.

Proposition 2.21 (Left-continuity) *The function $\lambda \mapsto \theta(\lambda)$ is left-continuous on $(\lambda_{\mathrm{c}}, \infty)$.*

We first prove two preparatory lemmas. By definition, we let $\mathcal{C}_{\mathrm{loc}}(\mathcal{P}(\mathbb{Z}^d))$ denote the space of real functions on $\mathcal{P}(\mathbb{Z}^d)$ that are 'local', in te sense that they depend on finitely many coordinates only, i.e., these are functions $f : \mathcal{P}(\mathbb{Z}^d) \to \mathbb{R}$ of the form

$$f(A) = f'(\Lambda \cap A) \qquad \big(A \in \mathcal{P}(\mathbb{Z}^d)\big), \tag{2.17}$$

where $\Lambda \subset \mathbb{Z}^d$ is some finite set and $f' : \mathcal{P}(\Lambda) \to \mathbb{R}$ is some function.

Lemma 2.22 (Convergence of semigroups) *Let $(P_t^\lambda)_{t \geq 0}$ be the Markov semigroup of the contact process on $\mathbb{Z}^d$ with infection rate λ. Then*

$$\|P_t^{\lambda_n} f - P_t^\lambda f\| \underset{\lambda_n \to \lambda}{\longrightarrow} 0 \qquad \big(t, \lambda \geq 0, \ f \in \mathcal{C}_{\mathrm{loc}}(\{0,1\}^{\mathbb{Z}^d})\big).$$

where $\|\cdot\|$ denotes the supremumnorm.

Proof We use the coupling from the proof of Proposition 2.7 and set

$$\eta_t^{A,\lambda} := \big\{i : A \times \{0\} \rightsquigarrow_\lambda (i,t)\big\}.$$

Let f be a local function that depends only on the coordinates in a finite set $\Lambda \subset \mathbb{Z}^d$. Let $0 \leq \lambda_n \to \lambda$ and choose some λ' such that $\lambda_n \leq \lambda'$ for all n. Let Γ be the collection of all infection arrows $((i,j), t, \kappa) \in \overline{\Delta}^{\mathrm{i}}$ that are used in some infection path along arrows with values $\kappa \leq \lambda'$ starting at time zero and ending somewhere in the finite set Λ. Then Γ contains all arrows that are relevant for deciding which points belong to the set $\Lambda \cap \eta_t^{A,\lambda}$. Let Γ_λ denote the set of arrows

in Γ that have a value $\kappa \le \lambda$. Since Γ is a.s. finite (by Proposition 2.5), there a.s. exists some random m such that $\Gamma_{\lambda_n} = \Gamma_\lambda$ for all $n \ge m$. It follows that

$$|P_t^{\lambda_n} f(A) - P_t^\lambda f(A)| = \left| \mathbb{E} \left[f'(\Lambda \cap \eta_t^{A,\lambda_n}) - f'(\Lambda \cap \eta_t^{A,\lambda}) \right] \right|$$
$$\le 2 \|f\| \, \mathbb{P}[\Gamma_{\lambda_n} \ne \Gamma_\lambda] \xrightarrow[n\to\infty]{} 0,$$

and this convergence is *uniform* in A, as claimed. $\blacksquare$

Lemma 2.23 (Convergence of invariant laws) *Let ν_n, ν be probability laws on $\mathcal{P}(\mathbb{Z}^d)$ such that $\nu_n \Rightarrow \nu$ and let $0 \le \lambda_n \to \lambda$. Assume that ν_n is an invariant law for the contact process with infection rate λ_n, for each n. Then ν is an invariant law for the contact process with infection rate λ.*

Proof We introduce the notation

$$\mu f := \int \mu(\mathrm{d}x) f(x).$$

With this notation, if $(X_t)_{t \ge 0}$ is a Markov process with Markov semigroup $(P_t)_{t \ge 0}$, started in the initial law $\mathbb{P}[X_0 \in \cdot] = \mu$, then

$$\mu P_t f = \int \mathbb{P}[X_0 \in \mathrm{d}x] \int P_t(x, \mathrm{d}y) f(y) = \mathbb{E}[f(X_t)].$$

We write

$$|\nu P_t^\lambda f - \nu f| \le |\nu P_t^\lambda f - \nu_n P_t^\lambda f| + |\nu_n P_t^\lambda f - \nu_n P_t^{\lambda_n} f| + |\nu_n P_t^{\lambda_n} f - \nu_n f| + |\nu_n f - \nu f|$$

where of course $|\nu_n P_t^{\lambda_n} f - \nu_n f| = 0$ since ν_n is an invariant law for the process with infection rate λ_n. It follows from the Feller property of the contact process (Lemma 2.12) that $\mathbb{P}_t^\lambda$ maps continuous functions into continuous functions, hence by our assumption that $\nu_n \Rightarrow \nu$ we have

$$|\nu_n P_t^\lambda f - \nu P_t^\lambda f| \xrightarrow[n\to\infty]{} 0 \quad \text{and} \quad |\nu_n f - \nu f| \xrightarrow[n\to\infty]{} 0$$

for each $f \in \mathcal{C}(\mathcal{P}(\mathbb{Z}^d))$ and $t \ge 0$. Assuming that moreover $f \in \mathcal{C}_{\mathrm{loc}}(\mathcal{P}(\mathbb{Z}^d))$, we have by Lemma 2.22 that

$$|\nu_n P_t^\lambda f - \nu_n P_t^{\lambda_n} f| \le \|P_t^\lambda f - P_t^{\lambda_n} f\| \xrightarrow[n\to\infty]{} 0.$$

It follows that

$$\nu P_t^\lambda f = \nu f \qquad \left(t \ge 0, \ f \in \mathcal{C}_{\mathrm{loc}}(\mathcal{P}(\mathbb{Z}^d)) \right),$$

hence ν is an invariant law for the contact process with infection rate λ. ∎

Proof of Proposition 2.21 Let $\overline{\nu}_\lambda$ denote the upper invariant law of the contact process with infection rate λ. Choose $\lambda_c < \lambda_n \uparrow \lambda$. Since the space $\{0,1\}^{\mathbb{Z}^d}$ is compact, the measures $\overline{\nu}_{\lambda_n}$ are tight. Since by Lemma 2.23, each weak cluster point of the $\overline{\nu}_{\lambda_n}$ is a nontrivial homogeneous invariant law of the contact process with infection rate λ, by Corollary 2.20, one has

$$\overline{\nu}_{\lambda_n} \underset{n\to\infty}{\Longrightarrow} \overline{\nu}_\lambda.$$

Since by (2.13),

$$\theta(\lambda) = \int \overline{\nu}_\lambda(\mathrm{d}A) 1_{\{0\in A\}},$$

this implies that $\theta(\lambda_n) \to \theta(\lambda)$. ∎

2.11 Other topics

Corollary 2.20 tells us that all homogeneous invariant laws of a contact process are convex combinations of $\delta_\emptyset$ and the upper invariant law. One may wonder if there exist inhomogeneous invariant laws. The answer to this question is known to be negative. This follows from the following theorem, that strengthens Theorem 2.16 quite a bit:

Theorem 2.24 (Complete convergence) *The contact process started in any initial state satisfies*

$$\mathbb{P}[\eta_t^A \in \cdot\,] \underset{t\to\infty}{\Longrightarrow} \rho(A)\overline{\nu} + (1 - \rho(A))\delta_\emptyset,$$

where $\rho(A) := \mathbb{P}\big[\eta_t^A \neq \emptyset \;\forall t \geq 0\big]$.

Complete convergence was proved first only for λ sufficiently large. In [BG90], this was extended to arbitrary $\lambda \geq 0$. In fact, more is known: it is known that if the process survives, then the infected area grows approximately linear and has a deterministic limiting shape. This result is known as the *shape theorem*.

The proof of complete convergence is quite a bit more involved than the proof of Theorem 2.16. To understand why this is so, it is useful to generalize a bit and consider contact processes on more general lattices, e.g., infinite graphs. As long as the graph has some sort of translation invariant structure, Theorem 2.16 still holds (and the proof basically carries through without a change). However, complete convergence does not hold in this generality. In particular, for processes

on trees, it is known that there exist two critical values $\lambda_c < \lambda_c'$ such that in the intermediate regime complete convergence does not hold and there exist inhomogeneous invariant laws. The study of contact processes on more general lattices is quite a lively modern subject with several nice open problems.

Chapter 3

The Ising model

3.1 Introduction

In this chapter, we study the Ising model. The Ising model model was introduced by E. Ising in 1925 [Isi25] as a simple model for a ferromagnetic material, based on the theory of Gibbs measures, which dated from the late nineteenth century when people like Boltzmann tried to find a microscopic basis for the laws of thermodynamics that had been discovered earlier in that century. In his Phd thesis, Ising showed that the one-dimensional model that now bears his name does not exhibit a phase transition, and based on this he incorrectly concluded that the same is true in any dimension. In 1936, Peierls [Pei36] used his famous argument (a variation on which we have already seen in the previous chapter) to prove that this is conjecture wrong in dimensions two and more. In 1944, Onsager showed that the two-dimensional model can, in a certain sense, be solved explicitly [Ons44]. (No explicit solutions are known or believed to exist in dimensions three and more.)

The Ising model as such, it should be pointed out, is not an interating particle system. Rather, it is a certain probability law (Gibbs measure) on spin configurations, depending on a certain parameter related to the temperature of the system. It is possible, however, and physically meaningful, to construct interacting particle systems whose invariant measures are these Gibbs measures. Such interacting particle systems are called *stochastic Ising models*. The first one to do so was Glauber [Gla63]. The subject was taken up again and studied more profoundly by Dobrushin in a series of papers starting with [Dob71]. Using the 'interacting particle systems approach', it is possible to give nice short proofs of certain properties of the Ising model. Conversely, the Ising model gives in a natural way rise to a number of interesting interacting particle systems which have sufficiently many pleasant properties to make it possible to prove things about them, while on the

other hand they are sufficiently 'difficult' to be interesting.

3.2   Definition, construction, and ergodicity

For definiteness, we will introduce one stochastic Ising model, i.e., an interacting particle system that has the Gibbs measures of the Ising model as its invariant law(s), that we will mostly focus on. As we will see later, there exist several ways to invent a dynamics for the Ising model, and many things that we will prove for our specific model are valid more generally.

The model that we will mostly focus our attention on is the interacting particle system with the following description. At each site $i \in \mathbb{Z}^d$, there is an atom which has a property called *spin* which makes it acts like a small magnet that can either point up, in which case we say the site i is in the state $+1$, or down, in which case we say the site i is in the state -1. Our stochastic Ising model is therefore a Markov process $(X_t)_{t \geq 0}$ with state space $\{-1, 1\}^{\mathbb{Z}^d}$. We will consider the following dynamics: if the process is in a state $x = (x(i))_{i \in \mathbb{Z}^d} \in \{-1, 1\}^{\mathbb{Z}^d}$, then the spin at site i jumps as:

$$x(i) \text{ jumps:}$$

$$-1 \mapsto 1 \qquad \text{with rate } e^{-\beta \sum_{j: \, |i-j|=1} 1_{\{x(j)=-1\}}},$$

$$1 \mapsto -1 \qquad \text{with rate } e^{-\beta \sum_{j: \, |i-j|=1} 1_{\{x(j)=1\}}}.$$

Here $\beta > 0$ is a parameter (loosely) called the *inverse temperature*. Indeed, in the physical interpretation of the model, $\beta = J/kT$ where T is the temperature, J is the energy difference between aligned and unaligned neighboring spins, and k is Boltzmann's constant. The motivation for our dynamics is roughly as follows: due to the constant motion of atoms, spins tend to *flip* in a random way between the $+1$ and -1 state. However, because of the magnetic interaction between neighboring atoms, neighboring spins like to be aligned (i.e., point in the same direction). This is expressed by making a spin less likely to flip when it has a lot of neighbors that point in the same direction. This effect is stronger when β is large (i.e., when the temperature is low). Note that (contrary to what we saw for the contact process) our dynamics treat the two values $-1, +1$ for the spins symmetrically.

In order to construct our process rigorously, we use a graphical representation. We first write down the formal generator of our process, which is

$$Gf(x) := \sum_{i \in \mathbb{Z}^d} e^{-\beta \sum_{j \in \mathcal{N}_i} 1_{\{x(j)=x(i)\}}} \big(f(x^{\{i\}}) - f(x) \big) \qquad (x \in \{-1, 1\}^{\mathbb{Z}^d}),$$

where we let

$$x^A(i) := \begin{cases} -x(i) & \text{if } i \in A, \\ x(i) & \text{if } i \notin A, \end{cases} \qquad (x \in \{-1,1\}^{\mathbb{Z}^d}, \ A \subset \mathbb{Z}^d)$$

denote the spin configuration obtained from x by flipping all spins in A, and

$$\mathcal{N}_i := \{j \in \mathbb{Z}^d : |i - j| = 1\} \qquad (i \in \mathbb{Z}^d)$$

denotes the set of neighbors of a site i. To invent a graphical representation, we need to rewrite our generator in terms of local maps. For each $i \in \mathbb{Z}^d$ and subset $L \subset \mathcal{N}_i$, let us define the maps $m^-_{i,L}, m^+_{i,L}$ by

$$(m^-_{i,L}x)(k) := \begin{cases} -1 & \text{if } k = i, \ x(j) = -1 \ \forall j \in L, \\ x(k) & \text{otherwise,} \end{cases}$$

$$(m^+_{i,L}x)(k) := \begin{cases} +1 & \text{if } k = i, \ x(j) = +1 \ \forall j \in L, \\ x(k) & \text{otherwise,} \end{cases}$$

Then we may write our generator in the form

$$Gf(x) := \sum_{i \in \mathbb{Z}^d} \sum_{L \subset \mathcal{N}_i} p^{|L|}(1-p)^{2d-|L|}\big(f(m^-_{i,L}x) - f(x)\big)$$
$$+ \sum_{i \in \mathbb{Z}^d} \sum_{L \subset \mathcal{N}_i} p^{|L|}(1-p)^{2d-|L|}\big(f(m^+_{i,L}x) - f(x)\big),$$

where

$$p := 1 - e^{-\beta}.$$

To see why this is correct, note that according to our new formulation of the generator, the spin at site i flips from -1 to $+1$ at rate

$$\sum_{L \subset \mathcal{N}_i} p^{|L|}(1-p)^{2d-|L|} 1_{\{x(j)=+1 \ \forall j \in L\}}. \tag{3.1}$$

Let $\mathcal{L}$ be a random subset of $\mathcal{N}_i$ such that independently for each neighbor j of i, one has $\mathbb{P}[j \in \mathcal{L}] = p$. Then the rate in (3.1) may be rewritten as

$$\mathbb{P}[x(j) = +1 \ \forall j \in \mathcal{L}] = \prod_{j \in \mathcal{N}_i : x(j) = -1} \mathbb{P}[j \notin \mathcal{L}]$$
$$= (1-p)^{\sum_{j \in \mathcal{N}_i} 1_{\{x(j)=-1\}}} = e^{-\beta \sum_{j \in \mathcal{N}_i} 1_{\{x(j)=-1\}}},$$

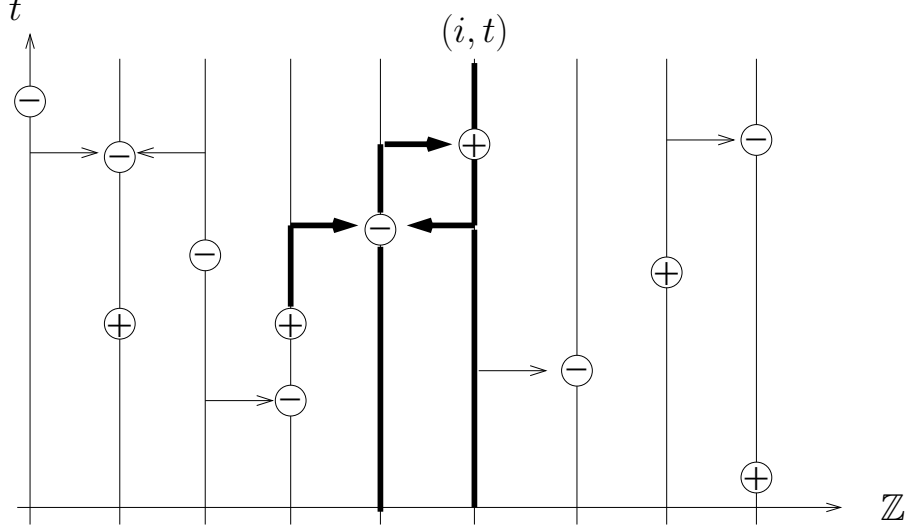as required. By symmetry, a similar argument holds for flips from $+1$ to -1.

Figure 3.1: Graphical representation of our stochastic Ising model.

Using these observations, we can define a graphical representation for our process as follows. Let $\mathcal{H}$ be the space of all triples of the form

$$(\sigma, i, L) \quad \text{with } \sigma \in \{-, +\}, \ i \in \mathbb{Z}^d, \ L \subset \mathcal{N}_i,$$

and let Δ be a Poisson point process on $\mathcal{H} \times \mathbb{R}$ with intensity $p^{|L|}(1-p)^{2d-|L|}\mathrm{d}t$. We interpret a point (σ, i, L, t) as saying that at time t, the state of of system changes according to the local map $m_{i,L}^{\sigma}$. To draw this in a picture, for each point $(\sigma, i, L, t) \in \Delta$, we draw a circle at the point $(i, t) \in \mathbb{Z}^d \times \mathbb{R}$ with the sign s in it, and we draw arrows starting at each point $j \in L$ and ending in i (see Figure 3.1). (Note that L may contain anything between zero and $2d$ elements.)

To see that this yields a well-defined process, we need to show that given the state of the system at time zero, there are a.s. only finitely many points (signifying local changes) in Δ that are relevant for deciding the state of a site i at some given time $t \geq 0$. In view of this, we make the following definition. For points $(i, s), (j, u) \in \mathbb{Z}^d \times \mathbb{R}$, we say that there is a *path of influence* from (i, s) to (j, u), denoted as $(i, s) \rightsquigarrow (j, u)$, if and only if there exists a path γ in $\mathbb{Z}^d$ with starting time s and final time t such that $\gamma_{s-} = i$, $\gamma_u = j$, and

(i) $\forall t \in [s, u]$ with $\gamma_{t-} \neq \gamma_t \ \exists (\sigma, i, L) \in \mathcal{H}$ s.t. $(\sigma, i, L, t) \in \Delta, \ \gamma_{t-} \in L, \ \gamma_t = i,$

(ii) $\nexists (\sigma, i, L, t) \in \Delta$ s.t. $|L| = \emptyset, \ t \in [s, u], \ \gamma_t = i.$

In our picture, this says that a path may use arrows but must avoid points (i, t), marked with an $\ominus$ or $\oplus$ where no arrows come in. Note that at such points, the spin at site i flips to the state -1 or $+1$, regardless of the state of the system prior to time t. For any $(i, s) \in \mathbb{Z}^d$, we set

$$\zeta_t^{(i,s)} := \big\{ j \in \mathbb{Z}^d : (j, s - t) \rightsquigarrow (i, s) \big\} \qquad (i \in \mathbb{Z}^d,\ s \in \mathbb{R},\ t \geq 0).$$

The next proposition is similar to Proposition 2.5.

Proposition 3.1 (Exponential bound) *For each $(i, u) \in \mathbb{Z}^d \times \mathbb{R}$, the process $(\zeta_t^{(i,u)})_{t \geq 0}$ satisfies $|\zeta_t^{(i,u)}| < \infty$ for all $t \geq 0$ a.s., and*

$$\mathbb{E}\big[|\zeta_t^{(i,u)}|\big] \leq e^{2(2dp - (1-p)^{2d})t} \qquad (t \geq 0).$$

Proof At each site i, points marked with an $\ominus$ or $\oplus$ occur at rate one each (hence rate 2 in total), and at each such point there are on average $2dp$ incoming arrows. Moreover, each such point has no incoming arrows with probability $(1-p)^{2d}$. Using these obvservations, the proof proceeds in exactly the same way as the proof of Proposition 2.5. ∎

Set
$$\mathcal{C}(j, u) := \overline{\{(i, t) : t \leq u,\ (i, t) \rightsquigarrow (j, u)\}} \qquad (j \in \mathbb{Z}^d,\ u \in \mathbb{R}),$$

where the bar means closure. For each $s \leq u$ and $j \in \mathbb{Z}^d$, we observe that the set

$$\Delta_s^{(j,u)} := \big\{ (\sigma, i, L, t) \in \Delta : s \leq t \leq u,\ (i, t) \in \mathcal{C}(j, u) \big\}$$

contains all local changes between time s and u that are relevant for the state of site j at time u. (In our definition of $\mathcal{C}(j, u)$ we have to take the closure because of our definition 'paths of influence' above, which was chosen with the aim of making the process $(\zeta_t^{(i,u)})_{t \geq 0}$ right-continuous.) By Proposition 3.1, the set $\Delta_s^{(j,u)}$ is a.s. finite for each $-\infty < s \leq u$ and $u \in \mathbb{R}$. In view of this, we define

$$X_t^{x,s}(i) := m_{i_n, L_n}^{\sigma_n} \cdots m_{i_1, L_1}^{\sigma_1}(x) \qquad (x \in \{-1, +1\}^{\mathbb{Z}^d},\ i \in \mathbb{Z}^d,\ s \in \mathbb{R},\ t \geq 0),$$

where

$$\Delta_s^{(i,s+t)} = \big\{ (\sigma_1, i_1, L_1, t_1), \ldots, (\sigma_n, i_n, L_n, t_n) \big\} \quad \text{with } t_1 < \cdots < t_n.$$

In particular, we put $X_t^x := X_t^{x,0}$. Then $(X_t^x)_{t \geq 0}$ is our stochastic Ising model started in the state x, constructed with the graphical representation above time s. Proposition 3.1 has an interesting corollary.
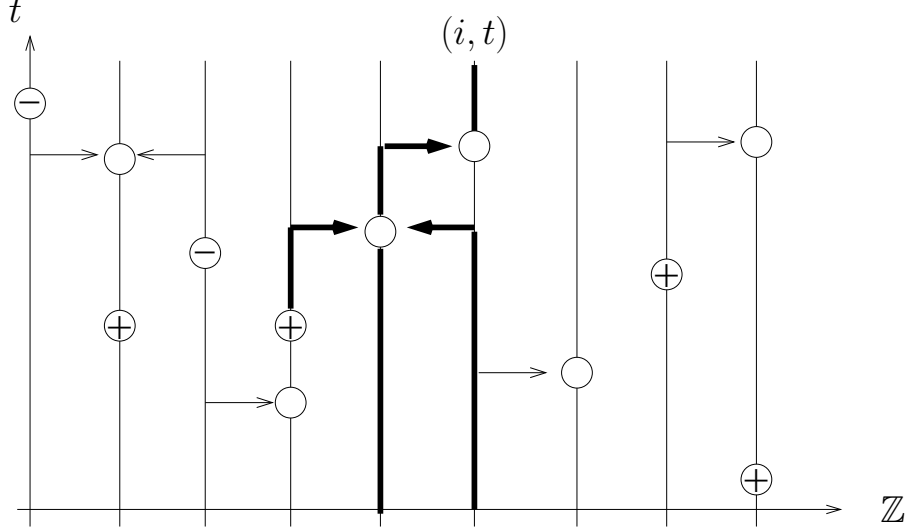
Figure 3.2: Alternative graphical representation of our stochastic Ising model.

Corollary 3.2 (Ergodicity for high temperature) *Let $\beta' := \sup\{\beta > 0 : 2d(1 - e^{-\beta}) - e^{-2d\beta} < 0\}$. Then, for each $\beta < \beta'$, our stochastic Ising model has a unique invariant measure ν and the process started from any initial law satisfies*

$$\mathbb{P}[X_t \in \cdot] \underset{t\to\infty}{\Longrightarrow} \nu. \tag{3.2}$$

Proof Since for each $\beta < \beta'$, one has $2d(1 - e^{-\beta}) - e^{-2d\beta} < 0$, by Proposition 3.1, the set $\Delta^{(i,t)}_{-\infty}$ is a.s. finite for each $(i,t) \in \mathbb{Z}^d \times \mathbb{R}$. It follows that the a.s. limit

$$X_t^{-\infty}(i) := \lim_{s\to-\infty} X^{x,s}_{s+t}(i) \tag{3.3}$$

exists for each $(i,t) \in \mathbb{Z}^d \times \mathbb{R}$ and does not depend on the choice of the initial state $x \in \{-1,+1\}^{\mathbb{Z}^d}$. Since $(X_t^{-\infty})t \in \mathbb{R}$ is a sationary stochastic Ising model, its law at any time is an invariant law, and the a.s. convergence in (3.3) implies the weak convergence in law in (3.2).   ∎

Exercise 3.3 (Alternative graphical representation) For each $i \in \mathbb{Z}^d$ let us define the maps

$$(m_i^- x)(k) := \begin{cases} -1 & \text{if } k = i, \\ x(k) & \text{otherwise}, \end{cases} \qquad (m_i^+ x)(k) := \begin{cases} +1 & \text{if } k = i, \\ x(k) & \text{otherwise}, \end{cases}$$

Moreover, for each $i \in \mathbb{Z}^d$ and nonempty subset $L \subset \mathcal{N}_i$, let us define

$$(m_{i,L}x)(k) := \begin{cases} -x(i) & \text{if } k = i, \ x(j) \neq x(i) \ \forall j \in L, \\ x(k) & \text{otherwise.} \end{cases}$$

Then we may rewrite our generator in the form

$$Gf(x) := (1-p)^{2d} \sum_{i \in \mathbb{Z}^d} \left(f(m_i^- x) + f(m_i^+ x) - 2f(x) \right)$$
$$+ \sum_{i \in \mathbb{Z}^d} \sum_{\emptyset \neq L \subset \mathcal{N}_i} p^{|L|} (1-p)^{2d-|L|} \left(f(m_{i,L}x) - f(x) \right).$$

Based on this, we may introduce an alternative graphical representation for our stochastic Ising model (see Figure 3.2). Use this to improve Corollary 3.2 by proving ergodicity for a larger range of the parameter.

3.3 Gibbs measures and finite systems

Let Λ be some finite set and let $H : \{-1, +1\}^\Lambda \to \mathbb{R}$ be some function. By definition, the *Gibbs measure* belonging to the *Hamiltonian* (or *energy function*) H and *inverse temperature* β is the probability measure on $\{-1, +1\}^\Lambda$ given by

$$\mu(\{x\}) = \frac{1}{Z} e^{-\beta H(x)} \qquad \left(x \in \{-1, +1\}^\Lambda \right), \tag{3.4}$$

where

$$Z := \sum_{x \in \{-1, +1\}^\Lambda} e^{-\beta H(x)} \tag{3.5}$$

is a normalization constant, also called the *partition sum*. Note that if H, H' are two energy functions that differ only by a constant, then the associated Gibbs measures are the same. Indeed, if $H(x) = H'(x) + c$ and μ, μ' are the associated Gibbs measures, then all probabilities in μ' get an extra factor $e^{-\beta c}$, but this disappears in the normalization. Indeed, we make the following simple observation.

Lemma 3.4 (Relative probabilities)
(a) *If Λ is a finite set and μ is the Gibbs measure on $\{-1, +1\}^\Lambda$ with Hamiltonian H and inverse temperature β, then*

$$\frac{\mu(\{x'\})}{\mu(\{x\})} = e^{-\beta(H(x') - H(x))} \qquad \left(x, x' \in \{-1, +1\}^\Lambda \right). \tag{3.6}$$

(b) *Conversely, if μ is a measure on $\{-1,+1\}^\Lambda$ and*

$$\frac{\mu(\{x^{\{i\}}\})}{\mu(\{x\})} = e^{-\beta(H(x^{\{i\}}) - H(x))} \qquad \left(i \in \Lambda, \ x \in \{-1,+1\}^\Lambda\right), \qquad (3.7)$$

then μ must be the Gibbs measure on $\{-1,+1\}^\Lambda$ associated with H and β.

Proof Part (a) is trivial. To prove part (b), we note that for each $x, x' \in \{-1,+1\}^{\mathbb{Z}^d}$ we can find $x_0, \ldots, x_n$ such that $x = x_0$, $x' = x_n$, and x_k differs only in one point from x_{k-1} $(k = 1, \ldots, n)$. In view of this, (3.7) implies (3.6). Choosing some arbitrary reference state x', we see that (3.6) determines all probabilities up to an overall multiplicative constant, which follows from the normalization. ∎

We need to introduce some notation. If S, R are disjoint sets, $x \in \{-1,+1\}^S$, and $y \in \{-1,+1\}^R$, then we define $x\&y \in \{-1,+1\}^{S\cup R}$ as $(x\&y)(i) := x(i)$ if $i \in S$ and $(x\&y)(i) := y(i)$ if $i \in R$. Now let Λ be a finite set, let $H : \{-1,+1\}^\Lambda \to \mathbb{R}$ be a function, and let $\mu^{\Lambda,\beta}$ be the Gibbs measure on $\{-1,+1\}^\Lambda$ with Hamiltonian H and inverse temperature β. For each $\Delta \subset \Lambda$ and $y \in \{-1,+1\}^{\Lambda\setminus\Delta}$, let $H_y^\Delta : \{-1,+1\}^\Delta \to \mathbb{R}$ be a function such that

$$H_y^\Delta(x) = H(x\&y) + c_y^\Delta \qquad \left(x \in \{-1,+1\}^\Delta\right),$$

where c_y^Δ is a constant that may depend on Δ and y but not on x. Let $\mu_y^{\Delta,\beta}$ be the Gibbs measure on $\{-1,+1\}^\Delta$ associated with H_y^Δ and β. (Note that this Gibbs measure is uniquely defined even though H_y^Δ is defined only up to a constant.) We make the following observations:

Lemma 3.5 (Conditional distributions)
(a) *If $(X(i))_{i\in\Lambda}$ is a random variable with law $\mu^{\Lambda,\beta}$, then for each $\Delta \subset \Lambda$, the conditional law of X inside Δ given its values outside Δ is given by*

$$\mathbb{P}\left[(X(i))_{i\in\Delta} \in \cdot \,\middle|\, (X(i))_{i\in\Lambda\setminus\Delta} = y\right] = \mu_y^{\Delta,\beta}. \qquad (3.8)$$

(b) *Conversely, if $(X(i))_{i\in\Lambda}$ is a random variable with values in $\{-1,+1\}^\Lambda$ and (3.8) holds for each $\Delta \subset \Lambda$ such that $|\Delta| = 1$, then the law of X must be equal to $\mu^{\Lambda,\beta}$.*

Proof We observe that

$$\frac{\mathbb{P}\left[(X(i))_{i\in\Delta} = x' \,\middle|\, (X(i))_{i\in\Lambda\setminus\Delta} = y\right]}{\mathbb{P}\left[(X(i))_{i\in\Delta} = x \,\middle|\, (X(i))_{i\in\Lambda\setminus\Delta} = y\right]}$$

$$= \frac{\mathbb{P}\left[(X(i))_{i\in\Delta} = x', \ (X(i))_{i\in\Lambda\setminus\Delta} = y\right]}{\mathbb{P}\left[(X(i))_{i\in\Delta} = x, \ (X(i))_{i\in\Lambda\setminus\Delta} = y\right]} = \frac{e^{-\beta H(x'\&y)}}{e^{-\beta H(x\&y)}} = e^{-\beta(H_y(x') - H_y(x))}.$$

In view of this, the statements follow from Lemma 3.4. ∎

The fact that we would like to prove is that Gibbs measures associated with the Hamiltonian [1]

$$H(x) := \sum_{\{i,j\} \in \mathcal{B}} 1_{\{x(i) \neq x(j)\}} \tag{3.9}$$

are reversible invariant measures for the stochastic Ising model constructed in the previous section. A "slight" problem with this statement is that the sum in this definition runs over the set

$$\mathcal{B} := \big\{ \{i,j\} : i,j \in \mathbb{Z}^d, \ |i-j| \big\}$$

of all (unordered) nearest neighbor pairs in $\mathbb{Z}^d$. As a consequence, for most x, the sum in (3.9) is actually infinite. In addition, the set $\{-1,+1\}^{\mathbb{Z}^d}$ is uncountable, so it is clear that we cannot define Gibbs measures on $\{-1,+1\}^{\mathbb{Z}^d}$ in the same way as we have done for finite lattices.

The solution to this problem is suggested by Lemma 3.5. Instead of looking at the absolute probability of one particular configuration x (which will typically be zero), we will look at conditional probabilities of finding certain configurations inside a finite set $\Lambda \subset \mathbb{Z}^d$, given what is outside.

To this aim, for each $\Lambda \subset \mathbb{Z}^d$, we define

$$\partial\Lambda := \{i \in \mathbb{Z}^d \backslash \Lambda : \exists j \in \Lambda \text{ s.t. } |i-j| = 1\}$$

and let

$$\mathcal{B}_\Lambda := \big\{ \{i,j\} : i,j \in \Lambda, \ |i-j| \big\} \quad \text{and} \quad \partial\mathcal{B}_\Lambda := \big\{ (i,j) : i \in \Lambda, \ j \in \partial\Lambda, \ |i-j| \big\}$$

denote the set of nearest-neighbor edges inside Λ and pointing out of Λ, respectively. For each finite $\Lambda \subset \mathbb{Z}^d$, $x \in \{-1,+1\}^\Lambda$ and $y \in \{-1,+1\}^{\mathbb{Z}^d \backslash \Lambda}$, we define

$$H_y^\Lambda(x) := \sum_{\{i,j\} \in \mathcal{B}_\Lambda} 1_{\{x(i) \neq x(j)\}} + \sum_{(i,j) \in \partial\mathcal{B}_\Lambda} 1_{\{x(i) \neq y(j)\}}.$$

We let $\mu_y^{\Lambda,\beta}$ denote the finite-volume Gibbs measure associated with H_y and β. We call this the finite-volume Gibbs measure with *boundary condition* y.

---

[1]Here I deviate from the usual definition of the Hamiltonian for the Ising model, which is

$$H'(x) := - \sum_{\{i,j\} \in \mathcal{B}} x(i)x(j) = \sum_{\{i,j\} \in \mathcal{B}} (2 \cdot 1_{\{x(i) \neq x(j)\}} - 1).$$

We observe that $H'(x) = 2H(x) + c$, where $c := |\mathcal{B}|$ is an irrelevant additive constant. In view of this, what is β in these lecture notes, is 2β in most of the literature on the Ising model.

Definition 3.6 (Infinite-volume Gibbs measures) *We say that the law μ of an $\{-1,+1\}^{\mathbb{Z}^d}$-valued random variable $(X(i))_{i\in\mathbb{Z}^d}$ is a Gibbs measure associated with the formal Hamiltonian (3.9) and inverse temperature β, if for each finite $\Lambda\subset\mathbb{Z}^d$, one has*

$$\mathbb{P}\big[(X(i))_{i\in\Lambda}\in\cdot\,\big|\,(X(i))_{i\in\mathbb{Z}^d\setminus\Lambda}=y\big]=\mu_y^{\Lambda,\beta}$$

for a.e. y w.r.t. μ.

We need to show that such infinite-volume Gibbs measures exist and are reversible invariant laws for the stochastic Ising model constructed in the previous section.

As a first step, we will study finite-volume Gibbs measures $\mu_y^{\Lambda,\beta}$ with fixed boundary conditions y. Our first result says that such finite-volume Gibbs measures are reversible invariant laws for a suitable finite-volume version of our stochastic Ising model.

Proposition 3.7 (Gibbs reversible law) *Let $\Lambda\subset\mathbb{Z}^d$ be a finite set, let $y\in\{-1,+1\}^{\mathbb{Z}^d\setminus\Lambda}$, and let $(X_t)_{t\geq0}$ be the finite state Markov process in $\{-1,+1\}^\Lambda$ that jumps as*

$$x\mapsto x^{\{i\}}\quad\text{with rate}\quad e^{-\beta\left(\sum_{j\in\mathcal{N}_i\cap\Lambda}1_{\{x(i)=x(j)\}}+\sum_{j\in\mathcal{N}_i\cap\partial\Lambda}1_{\{x(i)=y(j)\}}\right)}.$$

Then the Gibbs measure $\mu_y^{\Lambda,\beta}$ is a reversible invariant law for $(X_t)_{t\geq0}$. Moreover, the process $(X_t)_{t\geq0}$ started from any initial law satisfies $P[X_t\in\cdot]\underset{t\to\infty}{\Longrightarrow}\mu_y^{\Lambda,\beta}$.

Proof We must check detailed balance (1.19). Fix $i\in\Lambda$ and $x\in\{-1,+1\}^{\Lambda\setminus\{i\}}$, and define $x^-,x^+\in\{-1,+1\}^\Lambda$ by

$$x^-(j):=\begin{cases}-1&\text{if }j=i,\\x(j)&\text{otherwise,}\end{cases}\quad\text{and}\quad x^+(j):=\begin{cases}+1&\text{if }j=i,\\x(j)&\text{otherwise.}\end{cases}$$

We must check that

$$\mu_y^{\Lambda,\beta}(\{x^-\})r(x^-,x^+)=\mu_y^{\Lambda,\beta}(\{x^+\})r(x^+,x^-),\tag{3.10}$$

where $r(x^-,r^+)$ and $r(x^+,x^-)$ are the rates with which our process jumps from x^- to x^+ and back, respectively. Let

$$n_+:=\sum_{j\in\mathcal{N}_i\cap\Lambda}1_{\{x(j)=+1\}}+\sum_{j\in\mathcal{N}_i\cap\partial\Lambda}1_{\{y(j)=+1\}},$$
$$n_-:=\sum_{j\in\mathcal{N}_i\cap\Lambda}1_{\{x(j)=-1\}}+\sum_{j\in\mathcal{N}_i\cap\partial\Lambda}1_{\{y(j)=-1\}}.$$

Then

$$\frac{\mu_y^{\Lambda,\beta}(\{x^+\})}{\mu_y^{\Lambda,\beta}(\{x^-\})} = \frac{e^{-\beta n_-}}{e^{-\beta n_+}} = \frac{r(x^-, x^+)}{r(x^+, x^-)},$$

which implies (3.10). To check that the process $(X_t)_{t \geq 0}$ is ergodic, it suffices to check irreducibility and apply Proposition 1.10. ∎

3.4 The upper and lower invariant laws

We still need to show the existence of infinite-volume Gibbs measures, as well as the fact that these are reversible invariant laws for our infinite-volume stochastic Ising model. We will concentrate on two special infinite-volume Gibbs measures, which are the upper and lower invariant laws of our stochastic Ising model.

Proposition 3.8 (Upper and lower invariant laws) *Let $(X_t)_{t \geq 0}$ be the stochastic Ising model from Section 3.2, started in the initial state $X_0(i) = +1$ for all $i \in \mathbb{Z}^d$. Then*

$$\mathbb{P}\big[X_t \in \cdot\big] \underset{t \to \infty}{\Longrightarrow} \overline{\nu},$$

where $\overline{\nu}$ is an invariant law of the process with the property that if ν is any other invariant law, then $\nu \leq \overline{\nu}$ in the stochastic order. Likewise, if $(X_t)_{t \geq 0}$ is started in $X_0(i) = -1$ for all $i \in \mathbb{Z}^d$, then

$$\mathbb{P}\big[X_t \in \cdot\big] \underset{t \to \infty}{\Longrightarrow} \underline{\nu},$$

where $\underline{\nu}$ is an invariant law of the process with the property that if ν is any other invariant law, then $\underline{\nu} \leq \nu$ in the stochastic order.

Proof Our graphical representation shows that the Ising model is *monotone*, i.e., if X^x and $X^{x'}$ are processes started in initial states such that $x \leq x'$, then we can couple X^x and $X^{x'}$ such that $X_t^x \leq X_t^{x'}$ for all $t \geq 0$. In terms of the semigroup $(P_t)_{t \geq 0}$ of our process, this says that if μ, μ' are laws on $\{-1, +1\}^{\mathbb{Z}^d}$ such that $\mu \leq \mu'$ in the stochastic order, then $\mu P_t \leq \mu' P_t$ for all $t \geq 0$. Applying this to $\mu = \delta_{+1} P_{t-s}$ and $\mu' = \delta_{+1}$, where $+1$ denotes the all plus configuration, we see that $\delta_{+1} P_s \geq \delta_{+1} P_{t-s} P_s = \delta_{+1} P_t$ for all $0 \leq s \leq t$. This means that for each sequence of times $t_n \uparrow \infty$ we can couple the random variables X_{t_n} such that the X_{t_n} decrease to some a.s. limit. It is not hard to see that this implies that $\mathbb{P}[X_t \in \cdot]$ converges weakly to some limit law $\overline{\nu}$ as $t \to \infty$, and using this we can prove that $\overline{\nu}$ is an invariant law. (We skip the details.) The fact that $\overline{\nu}$ is the largest invariant law

in the stochastic order can be proved similar to the proof of Lemma 2.14. By symmetry, similar arguments apply to $\underline{\nu}$. ∎

For each finite $\Lambda \subset \mathbb{Z}^d$, we let $H_+^\Lambda(x)$ and $\mu_+^{\Lambda_n,\beta}$ denote the Hamiltonian $H_y^\Lambda(x)$ and finite-volume Gibbs measure, respectively, with boundary condition y given by $y(i) = +1$ for all $i \in \mathbb{Z}^d \backslash \Lambda$. We define $H_-^\Lambda(x)$ and $\mu_-^{\Lambda,\beta}$ similarly, with minus boundary conditions.

Proposition 3.9 (Limits of finite volume Gibbs measures) *Let $\Lambda_n \subset \mathbb{Z}^d$ be finite sets such that $\Lambda_n \uparrow \mathbb{Z}^d$. For each n, let $X^{\Lambda_n} = (X^{\Lambda_n}(i))_{i\in\mathbb{Z}^d}$ be a random variable such that $X(i) = +1$ for all $i \in \mathbb{Z}^d \backslash \Lambda_n$ and*

$$\mathbb{P}\big[(X^{\Lambda_n}(i))_{i\in\Lambda} \in \cdot\,\big] = \mu_+^{\Lambda_n,\beta}. \tag{3.11}$$

Then

$$\mathbb{P}\big[(X^{\Lambda_n}(i))_{i\in\mathbb{Z}^d} \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \overline{\nu}.$$

A similar statement holds for minus boundary conditions, in which case the limit is $\underline{\nu}$. Moreover, $\overline{\nu}$ and $\underline{\nu}$ are infinite-volume Gibbs measures in the sense of Definition 3.6.

Proof Let $(X_t)_{t\geq 0}$ be our infinite-volume stochastic Ising model started in $X_0 = \underline{+1}$ and for each n, let $(X_t^{\Lambda_n})_{t\geq 0}$ be a process such that $X_t^{\Lambda_n}(i) = +1$ for all $i \in \mathbb{Z}^d \backslash \Lambda_n$ and $t \geq 0$, while inside Λ, the process evolves as in Proposition 3.7, with plus boundary conditions and initial state $X_0^{\Lambda_n}(i) = +1$ for all i. Using the graphical representation, we see that we can couple our processes such that $X_t^{\Lambda_n} \geq X_t^{\Lambda_m} \geq X_t$ for all $t \geq 0$ and $n \leq m$. Taking the limit $t \to \infty$ we see that the random variables X^{Λ_n} from (3.11) can be coupled such that they decrease to an a.s. limit; in particular, this implies that their laws converge weakly to some limit ν. By using techniques similar to the proof of Lemma 2.23, we can prove that ν is an invariant law for the infinite-volume stochastic Ising model, while our coupling shows that $\nu \geq \overline{\nu}$. Since $\overline{\nu}$ is the largest invariant law, it follows that $\nu = \overline{\nu}$. The fact that $\overline{\nu}$ and $\underline{\nu}$ are infinite-volume Gibbs measures in the sense of Definition 3.6 is obvious, since the approximating finite-volume Gibbs measures have the right conditional distributions. ∎

3.5   The spontaneous magnetization

Let $\overline{\nu}$ be the upper invariant law of the Ising model. By definition, the quantity (which by translation invariance does not depend on $i \in \mathbb{Z}^d$)

$$m^*(\beta,d) = m^*(\beta) := \int \overline{\nu}(\mathrm{d}x)x(i) \qquad (\beta \geq 0)$$

is called the *spontaneous magnetization.* By symmetry, we have

$$\int \underline{\nu}(\mathrm{d}x)x(i) = -m^*(\beta).$$

Since moreover $\underline{\nu} \leq \overline{\nu}$, it follows that $\underline{\nu} \neq \overline{\nu}$ if and only if $m^*(\beta) > 0$. Since $\underline{\nu}$ and $\overline{\nu}$ are the lowest and highest invariant law in the stochastic order, this implies that our stochastic Ising model has a unique invariant law if and only if $m^*(\beta) = 0$. In this and the next section, we will prove the following theorem.

Theorem 3.10 (Phase transition of the Ising model) *The function $m^*(\beta, d)$ is nondecreasing and right-continuous in β and nondecreasing in d. In dimension $d = 1$ one has $m^*(\beta) = 0$ for all $\beta \geq 0$. On the other hand, for all dimensions $d \geq 2$, there exists a critical value $0 < \beta_c < \infty$ such that $m^*(\beta) = 0$ for $\beta < \beta_c$ and $m^*(\beta) > 0$ for $\beta > \beta_c$.*

In the present section, we will prove that $\beta \mapsto m^*(\beta, d)$ is nondecreasing and right-continuous and $d \mapsto m^*(\beta, d)$ is nondecreasing. In the next section, we will prove that $\beta_c = \infty$ in dimension $d = 1$ and $\beta_c < \infty$ in dimensions $d \geq 2$.

At first, one might think that monotonicity of the spontaneous magnetization in β and d can be proved by the same sort of monotonicity arguments that we have used so far, by coupling Markov processes (in our case, stochastic Ising models) with different values of β in such a way that one process 'stays above' the other. It seems, however, that this idea does not work. Indeed, increasing β means that spins 'like more to be aligned'. Since our dynamics treat pluses and minuses in a symmetric way, this means that pluses are more favored near other pluses and minuses are more favored near minuses, an effect that can work both ways. In view of this, we have to take a different approach. Our proof will be based on Griffiths' inequalities. An alternative proof (not given here) uses a representation of our Gibbs measures in terms of the so-called random cluster model. It can be shown that the latter is monotone in β and d in the usual sense, leading to the desired monotonicities for $m^*(\beta, d)$.

Let Λ be a finite set, let $\mathcal{P}(\Lambda)$ denote the set of all subsets of Λ, and let $\mathcal{P}(\Lambda) \ni A \mapsto J_A \in \mathbb{R}$ be any function. For any $A \in \mathcal{P}(\Lambda)$ and $x \in \{-1, +1\}^\Lambda$, we write

$$x_A := \prod_{i \in A} x(i),$$

where $x_\emptyset := +1$. We will be interested in Gibbs measures on $\{-1, +1\}^\Lambda$ of the form

$$\mu_J(\{x\}) := \frac{1}{Z_J} e^{\sum_{A \subset \Lambda} J_A x_A},$$

where Z_J is the normalization constant (also known as partition sum)

$$Z_J := \sum_x e^{\sum_A J_A x_A}.$$

We start by observing that

(i) $\frac{\partial}{\partial J_A} \log Z_J = \int \mu_J(\mathrm{d}x)x_A,$

(ii) $\frac{\partial^2}{\partial J_A \partial J_B} \log Z_J = \int \mu_J(\mathrm{d}x)x_A x_B - \int \mu_J(\mathrm{d}x)x_A \int \mu_J(\mathrm{d}x)x_B.$ (3.12)

To see this, just write

$$\frac{\partial}{\partial J_A} \log Z_J = \frac{\frac{\partial}{\partial J_A} Z_J}{Z_J}$$

and

$$\frac{\partial^2}{\partial J_A \partial J_B} \log Z_J = \frac{\partial}{\partial J_B} \frac{\frac{\partial}{\partial J_A} Z_J}{Z_J} = \frac{Z_J \frac{\partial^2}{\partial J_A \partial J_B} Z_J - (\frac{\partial}{\partial J_A} Z_J)(\frac{\partial}{\partial J_B} Z_J)}{Z_J^2},$$

where

$$\frac{\partial}{\partial J_A} Z_J = \frac{\partial}{\partial J_A} \sum_x e^{\sum_C J_C x_C} = \sum_x x_A e^{\sum_C J_C x_C}$$

and

$$\frac{\partial^2}{\partial J_A \partial J_B} Z_J = \frac{\partial}{\partial J_B} \sum_x x_A e^{\sum_C J_C x_C} = \sum_x x_A x_B e^{\sum_C J_C x_C}.$$

Proposition 3.11 (Griffiths' inequalities) *Assume that $J_A \geq 0$ for all $A \subset \Lambda$. Then*

(i) $\frac{\partial}{\partial J_A} \log Z_J \geq 0,$

(ii) $\frac{\partial^2}{\partial J_A \partial J_B} \log Z_J \geq 0$

for all $A, B \subset \Lambda$.

Proof We observe that

$$\begin{aligned}
Z_J &= \sum_x e^{\sum_A J_A x_A} \\
&= \sum_x \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_A J_A x_A \right)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{A_1} \cdots \sum_{A_n} \left(\prod_{k=1}^{n} J_{A_k} \right) \sum_x \prod_{k=1}^{n} x_{A_k}.
\end{aligned}$$

Since

$$x_A x_B = x_{A \triangle B},$$

where $A \triangle B$ denotes the symmetric difference of A and B, we see that

$$\sum_x \prod_{k=1}^n x_{A_k} = \sum_x x_{A_1 \triangle \cdots \triangle A_n} = \begin{cases} 2^{|\Lambda|} & \text{if } A_1 \triangle \cdots \triangle A_n = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$Z_J = 2^{|\Lambda|} \sum_{n=0}^\infty \frac{1}{n!} \sum_{A_1} \cdots \sum_{A_n} 1_{\{A_1 \triangle \cdots \triangle A_n = \emptyset\}} \prod_{k=1}^n J_{A_k}.$$

Likewise

$$\frac{\partial}{\partial J_A} \log Z_J = \frac{1}{Z_J} \sum_x x_A e^{\sum_C J_C x_C}$$

$$= \sum_{n=0}^\infty \frac{1}{n!} \sum_{A_1} \cdots \sum_{A_n} \left(\prod_{k=1}^n J_{A_k} \right) x_A \sum_x \prod_{k=1}^n x_{A_k}$$

$$= \frac{1}{Z_J} 2^{|\Lambda|} \sum_{n=0}^\infty \frac{1}{n!} \sum_{A_1} \cdots \sum_{A_n} 1_{\{A \triangle A_1 \triangle \cdots \triangle A_n = \emptyset\}} \prod_{k=1}^n J_{A_k},$$

which is clearly nonnegative provided the $J_A \geq 0$ for all A. To prove also Griffiths'
second inequality, we write, using (3.12),

$$\frac{\partial^2}{\partial J_A \partial J_B} \log Z_J = \frac{1}{Z_J^2} \left(\sum_x x_A x_B e^{\sum_C J_C x_C} \right) \left(\sum_y e^{\sum_C J_C y_C} \right)$$

$$- \frac{1}{Z_J^2} \left(\sum_x x_A e^{\sum_C J_C x_C} \right) \left(\sum_y y_A e^{\sum_C J_C y_C} \right)$$

$$= \frac{1}{Z_J^2} \sum_{x,y} \left(x_A x_B - x_A y_B \right) e^{\sum_C J_C (x_C + y_C)}.$$

Using the facts that $y_B = (x_B)^2 y_B = x_B (xy)_B$ and $x_A x_B = x_{A \triangle B}$, we may rewrite
our formula as

$$\frac{\partial^2}{\partial J_A \partial J_B} \log Z_J = \frac{1}{Z_J^2} \sum_{x,y} x_{A \triangle B} \left(1 + (xy)_B \right) e^{\sum_C J_C x_C \left(1 + (xy)_C \right)}$$

$$= \frac{1}{Z_J^2} \sum_{x,z} x_{A \triangle B} \left(1 + z_B \right) e^{\sum_C J_C x_C \left(1 + z_C \right)}$$

$$= \frac{1}{Z_J^2} \sum_z \left(1 + z_B \right) \sum_x x_{A \triangle B} e^{\sum_C J_C^z x_C},$$

where we have defined $J_C^z := \left(1 + z_C\right) J_C$. Since $|z_C| = 1$, we have $J_C^z \geq 0$ for all z, hence by Griffiths' first inequality

$$\sum_x x_{A\triangle B}\, e^{\sum_C J_C^z x_C} \geq 0$$

for each $z \in \{-1, +1\}^\Lambda$. Summing over x we obtain Griffiths' second inequality. ∎

The monotonicity of the spontaneous magnetization in β follows from Proposition 3.9 and the following simple consequence of Proposition 3.11.

Lemma 3.12 (Monotonicity of magnetization) *For any finite set $\Lambda \subset \mathbb{Z}^d$ and $i \in \Lambda$, one has*

$$\frac{\partial}{\partial \beta} \int \mu_+^{\Lambda,\beta}(\mathrm{d}x)x(i) \geq 0.$$

Proof We claim that $\mu_+^{\Lambda,\beta} = \mu_J$ for a suitable function J. Indeed, up to an irrelevant additive constant, we may rewrite our Hamiltonian as

$$H_+^\Lambda(x) = -\tfrac{1}{2}\sum_{\{i,j\}\in\mathcal{B}_\Lambda} x(i)x(j) - \tfrac{1}{2}\sum_{(i,j)\in\partial\mathcal{B}_\Lambda} x(i).$$

In view of this, our finite volume Gibbs measures are generated by the function J defined by

$$J_{\{i,j\}} := \tfrac{1}{2}\beta$$

if $i, j \in \Lambda$, $|i - j| = 1$,

$$J_{\{i\}} := \tfrac{1}{2}\beta|\{j \in \mathbb{Z}^d\backslash\Lambda : |i - j| = 1\}$$

and $J_A := 0$ in all other cases. It is now clear that increasing β means increasing the function J and hence, by Proposition 3.11, increasing $\int \mu_J(\mathrm{d}x)x(i)$. ∎

The monotonicity of $m^*(\beta, d)$ in d is proved in a similar way. Indeed, if $d \leq d'$, then we may view $\mathbb{Z}^d$ as a subset of $\mathbb{Z}^{d'}$. With positive boundary conditions, if we switch on the interaction between sites inside $\mathbb{Z}^d$ and sites in $\mathbb{Z}^{d'}\backslash\mathbb{Z}^d$, then by Proposition 3.11 this will lead to a higher magnetization in any point in $\mathbb{Z}^d$.
We conclude this section with the following result.

Lemma 3.13 (Right-continuity) *The spontaneous magnetization $m^*(\beta)$ is a right-continuous function of β.*

Proof Let $\overline{\nu}_\beta$ denote the upper invariant law at inverse temperature β and let $\beta_n \downarrow \beta$. Using the compactness of our state space, going to a subsequence if necessary, we may assume that $\overline{\nu}_{\beta_n} \Rightarrow \nu$ for some probability law ν. Just as in Lemma 2.23, we can show that ν is an invariant law for the stochastic Ising model with inverse temperature β. Moreover, since $\beta \mapsto m^*(\beta)$ is nondecreasing, we must have

$$\lim_{\beta_n \downarrow \beta} m^*(\beta_n) = \int \nu(\mathrm{d}x)x(0) \geq m^*(\beta).$$

Since $\overline{\nu}_\beta$ is the largest invariant law w.r.t. the stochastic order, we must have

$$\int \nu(\mathrm{d}x)x(0) \leq \int \overline{\nu}_\beta(\mathrm{d}x)x(0) = m^*(\beta),$$

proving our claim. ∎

3.6 Existence of a phase transition

We conclude this chapter with two of the oldest results in the field, namely, the result by Ising on the nonexistence of a phase transition for his model in dimension $d = 1$, and the result by Peierls on the existence of a phase transition in dimensions $d \geq 2$. We start with Ising's result.

Lemma 3.14 (No phase transition in one dimension) *In dimension $d = 1$, for each $\beta \geq 0$, there exists a unique infinite-volume Gibbs measure μ associated with the formal Hamiltonian (3.9) and inverse temperature β. If $X = (X(i))_{i \in \mathbb{Z}}$ is a random variable with law μ, then X is a stationary Markov chain with transition probabilities*

$$\mathbb{P}\big[X(i+1) \neq X(i) \,\big|\, X(i)\big] = \frac{e^{-\beta}}{e^{-\beta} + 1}. \tag{3.13}$$

Proof Let $\overline{\nu}$ be the upper invariant law of the one-dimensional Ising model with inverse temperature β and let $X = (X(i))_{i \in \mathbb{Z}}$ is a random variable with law $\overline{\nu}$. We claim that X is a Markov chain. By Proposition 3.9 it suffices to prove that for any finite interval $\Lambda_n = \{-n, \dots, n\}$, the finite-volume Gibbs measures $\mu_+^{\Lambda_n, \beta}$ are the laws of a finite Markov chain. Let X^{Λ_n} be a random variable with law $\mu_+^{\Lambda_n, \beta}$. We need to show that for any $-n \leq k \leq n$, the random variables

$$(X^{\Lambda_n}(i))_{-n \leq i < k} \quad \text{and} \quad (X^{\Lambda_n}(i))_{k < i \leq n}$$

are conditionally independent given $X^{\Lambda_n}(k)$. But this follows from Lemma 3.5 and the structure of the finite-volume Gibbs measures $\mu_y^{\Lambda_n \setminus \{k\}, \beta}$ with $y(i) = +1$ for $i \notin \Lambda_n$ and $y(k) = -1$ or $+1$.

Since (by Proposition 3.8) the upper invariant law is invariant under translations and mirror images, the Markov chain $X = (X(i))_{i \in \mathbb{Z}}$ is stationary and reversible. Set

$$p := \mathbb{P}\big[X(i+1) = +1 \,\big|\, X(i) = -1\big] \quad \text{and} \quad q := \mathbb{P}\big[X(i+1) = -1 \,\big|\, X(i) = +1\big].$$

Then

$$\mathbb{P}[X(i) = +1] = \frac{p}{p+q}.$$

From the fact that X is an infinite volume Gibbs measure for the Ising model, by Lemma 3.4, we know that

$$\frac{\mathbb{P}[X(i) = +1 \,|\, X(i-1) = -1 = X(i+1)]}{\mathbb{P}[X(i) = -1 \,|\, X(i-1) = -1 = X(i+1)]} = \frac{e^{-2\beta}}{1}.$$

Since

$$\mathbb{P}[X(i-1) = -1, X(i) = +1, X(i+1) = -1] = \frac{p}{p+q}pq,$$
$$\mathbb{P}[X(i-1) = -1, X(i) = -1, X(i+1) = -1] = \frac{p}{p+q}(1-p)^2,$$

this leads to the equation

$$\frac{pq}{(1-p)^2} = e^{-2\beta}.$$

Likewise, since

$$\frac{\mathbb{P}[X(i) = +1 \,|\, X(i-1) = -1, X(i+1) = +1]}{\mathbb{P}[X(i) = -1 \,|\, X(i-1) = -1, X(i+1) = +1]} = \frac{e^{-\beta}}{e^{-\beta}}$$

and

$$\mathbb{P}[X(i-1) = -1, X(i) = +1, X(i+1) = +1] = \frac{p}{p+q}p(1-q),$$
$$\mathbb{P}[X(i-1) = -1, X(i) = -1, X(i+1) = +1] = \frac{p}{p+q}(1-p)p,$$

we see that

$$\frac{1-q}{1-p} = \frac{e^{-\beta}}{e^{-\beta}},$$

hence $p = q$. By our previous equation this implies

$$\Big(\frac{p}{1-p}\Big)^2 = e^{-2\beta},$$

which in turn implies (3.13). It follows that $\mathbb{E}[X(i)] = p/(p+q) = 1/2$, hence $m^*(\beta, 1) = 0$ for all $\beta \geq 0$. ∎

Since $m^*(\beta, d)$ is nondecreasing in d, in order to prove the existence of a phase transition in dimensions $d \geq 2$, it suffices to treat the case $d = 2$.

Figure 3.3: Peierls argument for Ising model.

Proposition 3.15 (Estimate on critical temperature) *One has $m^*(\beta, 2) > 0$ for all $\beta > \log 3$.*

Proof We will use the original Peierls argument from [Pei36]. Let

$$\Lambda_n := \{-n, \ldots, n\}^2.$$

We may view Λ_n as a graph with edges between nearest neighbors. In this picture, for a given spin configuration $x \in \{-1, = 1\}^{\Lambda_n}$, we may group the -1 spins and $+1$ spins into connected components, each surounded by a closed curve (see Figure 3.3).

There is a one-to-one correspondence between configurations of curves and configurations of spins. In particular, the origin has a $+1$ spin if and only if it is surrounded by an even number of curves. More formally, for each $x \in \{-1, +1\}^{\Lambda_n}$, define $\overline{x} \in \{-1, +1\}^{\Lambda_{n+1}}$ by

$$\overline{x}(i) := \begin{cases} x(i) & \text{if } i \in \Lambda_n, \\ +1 & \text{if } i \in \partial \Lambda_n, \end{cases}$$

let $\mathcal{E}_n$ be the collection of all pairs $\{i, j\}$ with $|i - j| = 1$, $i, j \in \Lambda_{n+1}$, and define $\Gamma(x) \subset \mathcal{E}_n$ by

$$\Gamma(x) := \big\{\{i, j\} : \overline{x}(i) \neq \overline{x}(j)\big\}.$$

Let

$$\mathcal{G}_n := \left\{ \Gamma(x) : x \in \{-1, +1\}^{\Lambda_n} \right\}$$

be the configuration of all 'configurations of curves'. Then the probability of seeing a certain configuration of curves is given by

$$\rho(\{\Gamma\}) = \frac{1}{Z} e^{-\beta |\Gamma|},$$

where $|\Gamma|$ is the total length of the curves in the configuration Γ and

$$Z := \sum_{\Gamma \in \mathcal{G}_n} e^{-\beta |\Gamma|}$$

is a normalization constant. Now let $\gamma \subset \mathcal{E}$ be a collection of nearest-neighbor edges that form a closed curve (not a configuration of curves but just one single curve) surrounding the origin. We can ask what the probability is of seeing a configuration of curves in which this this particular curve is present. This probability is, of course,

$$\frac{1}{Z} \sum_{\Gamma \in \mathcal{G}_n : \, \gamma \subset \Gamma} e^{-\beta |\Gamma|}$$

$$= \frac{\sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|}}{\sum_{\Gamma} e^{-\beta |\Gamma|}}$$

$$\leq \frac{\sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|}}{\sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|} + \sum_{\Gamma: \, \gamma \cap \Gamma = \emptyset} e^{-\beta |\Gamma|}}$$

$$= \frac{\sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|}}{\sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|} + e^{\beta |\gamma|} \sum_{\Gamma: \, \gamma \subset \Gamma} e^{-\beta |\Gamma|}} = \frac{e^{-\beta |\gamma|}}{e^{-\beta |\gamma|} + 1} \leq e^{-\beta |\gamma|}.$$

Here we use that for every configuration of curves in which γ is present, there is another configuration in which γ is completely removed, which is a factor $e^{\beta |\gamma|}$ more likely than the configuration in which γ is present. Since there are at most $k3^k$ different curves γ of length k surrounding the origin, we find that the expected number of curves surrounding the origin can be estimated from above by

$$\sum_{k=4}^{\infty} k3^k e^{-k\beta}.$$

By choosing β sufficiently small, we can make this number as close to zero as we wish; in particular, this proves that (uniformly (!) in n) $\int \mu_+^{\Lambda_n, \beta}(\mathrm{d}x) x(0) > \frac{1}{2}$ for β sufficiently large.

Unfortunately, this does not quite give the explicit bound we are after. If $\beta > \log 3$, then we see that the expected number of curves surrounding the origin is finite (where, again, our estimate is uniform in n), but this is not enough to conclude that $\int \mu_+^{\Lambda_n,\beta}(\mathrm{d}x)x(0) > \frac{1}{2}$, hence $m^*(\beta) > 0$.

To fix this problem, we use a trick. We fix some $m \leq n$ and look at the proportion of probabilities

$$\frac{\int \mu_+^{\Lambda_n,\beta}(\mathrm{d}x)\mathbb{1}_{\{x(i)=-1\ \forall i\in\Lambda_m\}}}{\int \mu_+^{\Lambda_n,\beta}(\mathrm{d}x)\mathbb{1}_{\{x(i)=+1\ \forall i\in\Lambda_m\}}}.$$

We note that the event $\{x(i) = -1\ \forall i \in \Lambda_m\}$ occurs if and only if there are no contours inside Λ_m and there is an odd number of contours surroundig Λ_m. Likewise, the event $\{x(i) = +1\ \forall i \in \Lambda_m\}$ occurs if and only if there are no contours inside Λ_m and there is an even number of contours surroundig Λ_m. We can estimate the proportion of the probabilities of these events by estimating the expected number of contours surrounding Λ_m, conditional on the event that there are no contours inside Λ_m. By the same arguments as above, this expectation can be estimated by

$$\sum_{k=4m}^{\infty} k3^k e^{-k\beta},$$

which in case $\beta > \log 3$ can be made arbitrarily small by choosing m sufficiently large. Now, letting $\Lambda_n \uparrow \infty$ while keeping m fixed, using Proposition 3.9, we see that the upper invariant measure $\overline{\nu}$ satisfies

$$\frac{\int \overline{\nu}(\mathrm{d}x)\mathbb{1}_{\{x(i)=-1\ \forall i\in\Lambda_m\}}}{\int \overline{\nu}(\mathrm{d}x)\mathbb{1}_{\{x(i)=+1\ \forall i\in\Lambda_m\}}} < 1$$

for some m. In particular, this shows that $\overline{\nu}$ is not symmetric with respect to a simultaneous flip of all spins, hence $\overline{\nu} \neq \underline{\nu}$. As we have already seen, this implies that $m^*(\beta) > 0$. ∎

3.7 Other topics

For the Ising model on $d = 2$, Onsager has shown [Ons44] that

$$\beta_{\mathrm{c}} = \log(1 + \sqrt{2})$$

and

$$m^*(\beta, 2) = \left(1 - \sinh(\beta)^{-4}\right)^{1/8} \qquad (\beta \geq \beta_{\mathrm{c}}),$$

where

$$\sinh(\beta) = \tfrac{1}{2}\big(e^{\beta} - e^{-\beta}\big)$$

is the sinus hyperbolicus. Note that in light of this, the estimate $\beta_{\mathrm{c}} \leq \log 3$ arising from Proposition 3.15 is not so bad! Onsager's solution also implies that

$$m^*(\beta, 2) \sim (\beta - \beta_{\mathrm{c}})^{1/8} \quad \text{as } \beta \downarrow \beta_{\mathrm{c}},$$

which shows that the critical exponent associated with the spontaneous magnetrization is $1/8$ in dimension $d = 2$. It is supposed that

$$m^*(\beta, 3) \sim (\beta - \beta_{\mathrm{c}})^{0.308} \quad \text{as } \beta \downarrow \beta_{\mathrm{c}},$$

but there is no mathematical theory to explain this. (There is -nonrigorous- renormalization group theory that sort of 'explains' this and even allows one to calculate the critical exponent with some precision.) This critical exponent can actually be measured and has been experimentally observed for varous magnetic systems and gasses near the critical point. Obviously, these physical systems are locally not very similar to the (nearest-neighbor) Ising model, but it is believed (and up to some level understood by renormalization group theory) that this critical exponent is universal and shared by a large number of different models.

Similar to what we know for the contact process, one can prove that for the Ising model, all spatially homogeneous infinite volume Gibbs measures are convex combinations of $\overline{\nu}$ and $\underline{\nu}$. In dimension 2, these are in fact all infinite volume Gibbs measures, but, contrary to what we saw for the contact process, in dimensions $d \geq 3$ there exist infinite volume Gibbs measures for the ising model that are not translation invariant.

Generalizing from the Ising model, one may look at models where spins can take $q = 2, 3, \dots$ values, described by Gibbs measures with a Hamiltonian of the form (3.9). These models are called *Potts models*. An interesting feature of these models is that while the spontaneous magnetization $m^*(\beta)$ is (supposed to be) a continuous of β for the Ising model, it is known that the same is not always true for Potts models. Ising and Potts models can be studied in a nice uniform framework using the random cluster model.

Appendix A

K-dependence

By definition, for $k \geq 0$, one says that a collection $(X_i)_{i \in \mathbb{Z}^d}$ of random variables, indexed by the integer square lattice, is *k-dependent* if for any $A, B \subset \mathbb{Z}^d$ with

$$\inf\{|i - j| : i \in A, \ j \in B\} > k,$$

the collections of random variables $(X_i)_{i \in A}$ and $(X_j)_{j \in B}$ are independent of each other. Note that in particular, 0-dependence means independence.

The most important property associated with k-dependence is that a collection of k-dependent Bernoulli random variables with success probability p can be stochastically estimated from below by a collection of independent Bernoulli random variables with a success probability $\tilde{p}$ that has the property that $\tilde{p} \to 1$ as $p \to 1$. It is a bit unfortunate that the term 'k-dependence' as it is standardly used explicity (and only) refers to random variables on $\mathbb{Z}^d$, while in fact, as the next theorem shows, for the property just mentioned the precise spatial structure is not very important. The next theorem is taken from [Lig99, Thm B26], who in turn cites [LSS97].

Theorem A.1 (K-dependence) *Let Λ be a countable set and let $p \in (0, 1)$, $K < \infty$. Assume that $(\chi_i)_{i \in \Lambda}$ are Bernoulli random variables with $P[\chi_i = 1] \geq p$ $(i \in \Lambda)$, such that for each $i \in \Lambda$ there exists a $\Delta_i \subset \Lambda$ with $i \in \Delta_i$ and $|\Delta_i| \leq K$, such that*

$$\chi_i \text{ is independent of } (\chi_j)_{j \in \Lambda \setminus \Delta_i}.$$

Then it is possible to couple $(\chi_i)_{i \in \Lambda}$ to a collection of independent Bernoulli random variables $(\tilde{\chi}_i)_{i \in \Lambda}$ with

$$P[\tilde{\chi}_i = 1] = \tilde{p} := \left(1 - (1 - p)^{1/K}\right)^2, \tag{A.1}$$

in such a way that $\tilde{\chi}_i \leq \chi_i$ for all $i \in \Lambda$.

Proof of Theorem A.1 Since we can always choose some arbitrary denumeration of Λ, we may assume that $\Lambda = \mathbb{N}$. Our strategy will be as follows. We will choose $\{0,1\}$-valued random variables $(\psi_i)_{i \in \Lambda}$ with $P[\psi_i = 1] = r$, independent of each other and of the $(\chi_i)_{i \in \mathbb{N}}$, and put

$$\chi'_i := \psi_i \chi_i \qquad (i \in \mathbb{N}). \tag{A.2}$$

Note that the $(\chi'_i)_{i \in \mathbb{N}}$ are a 'thinned out' version of the $(\chi_i)_{i \in \mathbb{N}}$. In particular, $\chi'_i \le \chi_i$ $(i \in \mathbb{N})$. We will show that for an appropriate choice of r,

$$P[\chi'_n = 1 \mid \chi'_0, \dots, \chi'_{n-1}] \ge \tilde{p} \tag{A.3}$$

for all $n \ge 0$, and we will show that this implies that the $(\chi'_i)_{i \in \mathbb{N}}$ can be coupled to independent $(\tilde{\chi}_i)_{i \in \Lambda}$ as in (A.1) in such a way that $\tilde{\chi}_i \le \chi'_i \le \chi_i$ $(i \in \mathbb{N})$.
We start with the latter claim. Imagine that (A.3) holds. Set

$$p'_n(\varepsilon_0, \dots, \varepsilon_{n-1}) := P[\chi'_n = 1 \mid \chi'_0 = \varepsilon_0, \dots, \chi'_{n-1} = \varepsilon_{n-1}] \tag{A.4}$$

whenever $P[\chi'_0 = \varepsilon_0, \dots, \chi'_{n-1} = \varepsilon_{n-1}] > 0$. Let $(U_n)_{n \in \mathbb{N}}$ be independent, uniformly distributed $[0,1]$-valued random variables. Set

$$\tilde{\chi}_n := 1_{\{U_n < \tilde{p}\}} \qquad (n \in \mathbb{N}) \tag{A.5}$$

and define inductively

$$\chi'_n := 1_{\{U_n < p'_n(\chi'_0, \dots, \chi'_{n-1})\}} \qquad (i \in \mathbb{N}). \tag{A.6}$$

Then

$$P[\chi'_n = \varepsilon_n, \dots, \chi'_0 = \varepsilon_0] = p_n(\varepsilon_0, \dots, \varepsilon_{n-1}) \cdots p_0. \tag{A.7}$$

This shows that these new χ'_n's have the same distribution as the old ones, and they are coupled to $\tilde{\chi}_i$'s as in (A.1) in such a way that $\tilde{\chi}_i \le \chi'_i$.
What makes life complicated is that (A.3) does not always hold for the original $(\chi_i)_{i \in \mathbb{N}}$, which is why we have to work with the thinned variables $(\chi'_i)_{i \in \mathbb{N}}$.[1] We observe that

$$P[\chi'_n = 1 \mid \chi'_0 = \varepsilon_0, \dots, \chi'_{n-1} = \varepsilon_{n-1}] = rP[\chi_n = 1 \mid \chi'_0 = \varepsilon_0, \dots, \chi'_{n-1} = \varepsilon_{n-1}]. \tag{A.8}$$

---

[1]Indeed, let $(\phi_n)_{n \ge 0}$ be independent $\{0,1\}$-valued random variables with $P[\phi_n = 1] = \sqrt{p}$ for some $p < 1$, and put $\chi_n := \phi_n \phi_{n+1}$. Then the $(\chi_n)_{n \ge 0}$ are 1-dependent with $P[\chi_n = 1] = p$, but $P[\chi_n = 1 | \chi_{n-1} = 0, \chi_{n-2} = 1] = 0$.

We will prove by induction that for an appropriate choice of r,

$$P[\chi_n = 0 \,|\, \chi'_0 = \varepsilon_0, \ldots, \chi'_{n-1} = \varepsilon_{n-1}] \leq 1 - r. \tag{A.9}$$

Note that this is true for $n = 0$ provided that $r \leq p$. Let us put

$$\begin{aligned}
E_0 &:= \{i \in \Delta_i : 0 \leq i \leq n-1, \ \varepsilon_i = 0\}, \\
E_1 &:= \{i \in \Delta_i : 0 \leq i \leq n-1, \ \varepsilon_i = 1\}, \\
F &:= \{i \notin \Delta_i : 0 \leq i \leq n-1\}.
\end{aligned} \tag{A.10}$$

Then

$$\begin{aligned}
&P[\chi_n = 0 \,|\, \chi'_0 = \varepsilon_0, \ldots, \chi'_{n-1} = \varepsilon_{n-1}] \\
&\quad = P\big[\chi_n = 0 \,\big|\, \chi'_i = 0 \ \forall i \in E_0, \ \chi_i = 1 = \psi_i \ \forall i \in E_1, \ \chi'_i = \varepsilon_i \ \forall i \in F\big] \\
&\quad = P\big[\chi_n = 0 \,\big|\, \chi'_i = 0 \ \forall i \in E_0, \ \chi_i = 1 \ \forall i \in E_1, \ \chi'_i = \varepsilon_i \ \forall i \in F\big] \\
&\quad = \frac{P\big[\chi_n = 0, \ \chi'_i = 0 \ \forall i \in E_0, \ \chi_i = 1 \ \forall i \in E_1, \ \chi'_i = \varepsilon_i \ \forall i \in F\big]}{P\big[\chi'_i = 0 \ \forall i \in E_0, \ \chi_i = 1 \ \forall i \in E_1, \ \chi'_i = \varepsilon_i \ \forall i \in F\big]} \\
&\quad \leq \frac{P\big[\chi_n = 0, \ \chi'_i = \varepsilon_i \ \forall i \in F\big]}{P\big[\psi_i = 0 \ \forall i \in E_0, \ \chi_i = 1 \ \forall i \in E_1, \ \chi'_i = \varepsilon_i \ \forall i \in F\big]} \\
&\quad = \frac{P\big[\chi_n = 0 \,\big|\, \chi'_i = \varepsilon_i \ \forall i \in F\big]}{P\big[\psi_i = 0 \ \forall i \in E_0, \ \chi_i = 1 \ \forall i \in E_1 \,\big|\, \chi'_i = \varepsilon_i \ \forall i \in F\big]} \\
&\quad \leq \frac{1 - p}{(1 - r)^{|E_0|} P\big[\chi_i = 1 \ \forall i \in E_1 \,\big|\, \chi'_i = \varepsilon_i \ \forall i \in F\big]} \leq \frac{1 - p}{(1 - r)^{|E_0|} \, r^{|E_1|}},
\end{aligned} \tag{A.11}$$

where in the last step we have used K-dependence and the fact that

$$P\big[\chi_i = 1 \ \forall i \in E_1 \,\big|\, \chi'_i = \varepsilon_i \ \forall i \in F\big] \geq r^{|E_1|}. \tag{A.12}$$

We claim that (A.12) is a consequence of the induction hypothesis (A.9). Indeed, we may assume that the induction hypothesis (A.9) holds regardless of the ordering of the first n elements, so without loss of generality we may assume that $E_1 = \{n-1, \ldots, m\}$ and $F = \{m-1, \ldots, 0\}$, for some m. Then the left-hand side of (A.12) may be written as

$$\begin{aligned}
&\prod_{k=m}^{n-1} P\big[\chi_k = 1 \,\big|\, \chi_i = 1 \ \forall m \leq i < k, \ \chi'_i = \varepsilon_i \ \forall 0 \leq i < m\big] \\
&\quad = \prod_{k=m}^{n-1} P\big[\chi_k = 1 \,\big|\, \chi'_i = 1 \ \forall m \leq i < k, \ \chi'_i = \varepsilon_i \ \forall 0 \leq i < m\big] \geq r^{n-m}.
\end{aligned} \tag{A.13}$$

If we assume moreover that $r \geq \frac{1}{2}$, then $r^{|E_1|} \geq (1 - r)^{|E_1|}$ and therefore the right-hand side of (A.11) can be further estimated as

$$\frac{1 - p}{(1 - r)^{|E_0|} \, r^{|E_1|}} \leq \frac{1 - p}{(1 - r)^{|\Delta_n \cap \{0, \ldots, n-1\}|}} \leq \frac{1 - p}{(1 - r)^{K-1}}. \tag{A.14}$$

We see that in order for our proof to work, we need $\frac{1}{2} \leq r \leq p$ and

$$\frac{1 - p}{(1 - r)^{K-1}} \leq 1 - r. \tag{A.15}$$

In particular, choosing $r = 1 - (1 - p)^{1/K}$ yields equality in (A.15). Having proved (A.9), we see by (A.8) that (A.3) holds provided that we put $\tilde{p} := r^2$. ∎

Bibliography

[BFM78]   R.C. Brower, M.A. Furman and M. Moshe. Critical exponents for the reggeon quantum spin model. *Phys. Lett. B* 76, 213–219, 1978.

[BG90]   C. Bezuidenhout and G. Grimmett. The critical contact process dies out. *Ann. Probab.* 18(4), 1462–1482, 1990.

[BG91]   C. Bezuidenhout and G. Grimmett. Exponential decay for subcritical contact and percolation processes. *Ann. Probab.* 19(3), 984–1009, 1991.

[Dob71]   R.L. Dobrushin. Markov processes with a large number of locally interacting components: existence of a limit process and its ergodicity. *Problems Inform. Transmission* 7, 149–164, 1971.

[Dur88]   R. Durrett. *Lecture notes on particle systems and percolation.* Wadsworth & Brooks/Cole, Pacific Grove, 1988.

[Gla63]   R.J. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.* 4, 294–307, 1963.

[Har74]   T.E. Harris. Contact interactions on a lattice. *Ann. Probab.* 2, 969–988, 1974.

[HS05]   R. van der Hofstad and A. Sakai. Critical points for spread-out self-avoiding walk, percolation and the contact process above the upper critical dimensions. *Probab. Theory Relat. Fields* 132(3), 438–470, 2005.

[Isi25]   E. Ising. Beitrag zur Teorie des Ferromagnetismus. *Zeitschrift für Physik* 31, 253–258, 1925.

[Lig85]   T.M. Liggett. *Interacting Particle Systems.* Springer-Verlag, New York, 1985.

[Lig92]   T.M. Liggett. The survival of one-dimensional contact processes in random environments. *Ann. Probab.* 20, 696-723, 1992.

[Lig95]   T.M. Liggett. Improved upper bounds for the contact process critical value. *Ann. Probab.* 23, 697–723, 1995.

[Lig99]   T.M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes.* Springer-Verlag, Berlin, 1999.

[LSS97]   T.M. Liggett, R.H. Schonmann and A.M. Stacey. Domination by product measures. *Ann. Probab.* 25(1), 71–95, 1997.

[Loe63]   M. Loève. *Probability Theory, 3rd edition.* Van Nostrand, Princeton, 1963.

[Ons44]   L. Onsager. Crystal statistics, I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* 65, 117–149, 1944.

[Pei36]   R. Peierls. On Ising's model of ferromagnetism. *Proc. Camebridge Philos. Soc.* 36, 477-481, 1936.

[Rem08]   D. Remenik. The contact process in a dynamic random environment. *Ann. Appl. Probab.* 18(6), 2392-2420, 2008.

[ZG85]   H. Ziezold and C. Grillenberger. On the critical infection rate of the one-dimensional contact process: numerical results. ???

Index