THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Case studies in omniparametric simulation

FREDRIK LUNDIN

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
AND GÖTEBORG UNIVERSITY
SE-412 96 Göteborg, Sweden

Göteborg, Sweden 2006

Case studies in omniparametric simulation
Fredrik Lundin

ISBN 10: 91-628-6747-4
ISBN 13: 978-91-628-6747-8
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Göteborg, Sweden 2006

Abstract

In the field of particle systems and growths models simulation is an important tool. When explicit calculations are too complex or impossible to perform we may use simulations instead. We adapt a new technique here denoted omniparametric simulation, to the two-type Richardson, Ising and Potts models. Omniparametric means simulating for all parameter values at the same time giving us something else than ordinary samples, but by fixingthe parameter value we can always retrieve an ordinary sample. We use only one dimensional parameters, so for the random cluster and Potts models we fix q at some value and consider it known. For the two-type Richardson model we use symmetry and rescale time to eliminate one of the two parameters.

We study We study asymmetric simultaneous survival for the two-type Richardson model using omniparametric simulations. The belief is that if both types are equally strong the can survive for all times but if one type is stronger than the other this can not happen. We do not find any indication of the existence of so called exceptional values < 1 where simultaneous survival may be possible.. We develop a simple test procedure to see how strong the indications against exceptional values are and also which exceptional values tests may rule out, and also consider how large subsets of $\mathbb{Z}^2$ we must use.

For the Ising and Potts models we use omniparametric simulations to find smooth estimates of functions for model characteristics such as connection probabilities and susceptbility. The characteristics are then used for parameter estimation, we construct both point estimate and confidence intervals. Based on partial observations we develop three methods, two using asymptotic theory, and on non-asymptotic. The method for constructing point estimate are the same for all three approaches, the difference lies in ho we capture the variance of the statistic. We perform extensive testing of the methods and elaborate some on the difference between the model used in simulations and the experienced from data.

Keywords: growth model, Ising model, Markov chain, omnithermal simulation, omniparametric simulation, percolation, Potts model, parameter estimation, partially observations, random cluster model, Richardson model, simulation driven parameter estimation, twp-type Richardson model

MSC 2000 subject classifications : 60C35, 60F05, 60J25, 62F10, 62F12, 62F25, 62F40, 62M05, 62M30, 65K35, 82B20, 82B27, 82B43, 82C22

Preface

The thesis

My studies as a graduate student are mainly focused on simulations and the technique for doing simulations for all parameter values of a model at the same time, so called omniparametric simulation. This thesis presents work from two projects. Project one concerns the omniparametric simulation for the two-type Richardson model and was introduced to me around Christmas 2000. The second project, concerning parameter estimation in the Ising and Potts models was presented to me just before defending my licentiate thesis in late spring 2003. Both projects originate from my supervisor, professor Olle Häggström.

Acknowledgements

Phd. studies are not something you do on your own, it does not matter if you work in large research groups or on your own, as I have. The professional trick is to find people who answer your questions, simple as that. You may spend some time interpreting answers, but without the possibility to ask, what do you do? Outside the department you just have to find people tolerating an absent minded mathematician, a slightly more complicated task. I am fortunate, both has been easy for my and as a result I have come to rely on many people over the years for support, help and encouragement, both professionaly and privately.

The thesis you hold in your hand is, as I would like to think, the refined result of many hours of work. Let me introduce to you some of the people I have come to rely on, perhaps you are one of them ...

My supervisor Olle Häggström has, apart from helping me get the phd. position in 2000, answered countless of more or less spontaneous questions and also provided unvaluable help regarding many difficult theoretical issues. I can not stress enough the importance of having

someone to ask questions when you are a graduate student. I am grateful.

My co-supervisor Johan Jonasson has contributed with valueable comments and suggestions.

Ida Looser has proof-read some of the chapters. In my efforts of becoming a better english writer she provided valuable help in finding some of the annoying grammatical errors popping up here and there.

Torgny Lindvall and Patrik Albin introduced me to mathematical statistics and probability theory when I was a undergraduate student. They have been inspiring teachers over the years.

The lunch bunch, the coffee table bunch, the pub crowd, cycling colleagues, teachers, administrators, and all other at the department of mathematical sciences. There are many ways to make a life easier, and they tend to vary over the years, especially as a graduate student since old students leave and new arrive, continuously. Many people help in making the department a nice and pleasant place to be active in, I thank you all for contributing.

All my friends in Göteborg and surroundings, what would I do without you. Maggis, Tassen and Tiger, Michael, Carin, Bamse, Bella och Millie, Maria and Bosse, Anders, Benita, Isak and Anton, Malin and her cats Milton and Frasse, Josephine, and many more. For those I have forgotten to mention, I apologise . . .

My family has provided undivided support over the years, mostly by just being there and believing in me.

My beloved Emma, since you came into my life it has been so much more fun. With such a warm heart you should have a book dedicated to you, so let it be this one . . .

Fredrik Lundin,
Göteborg, 16th of January 2006.

Contents

Part I

Omniparametric simulation

CHAPTER 1

Introduction

Stochastic models may be simple to formulate, and natural questions may come easily to mind. Strikingly often these questions are, when formulated into theorems, as hard to give an exact answer to as they are easily formulated. Models in the field of statistical physics are a particular case in point, as are stochastic growth models. When analysing these models we have at our disposal a number of theorems, often giving us information about the asymptotic case, when time, or space in terms of finite regions, tends to infinity. In real world applications however, we deal with finite models and the asymptotic results have to be complemented by such things as convergence rates and estimation of approximation errors. When no results are available we may turn to simulations trying to extract some behaviour from the output of our computer programs.

When using simulation we traditionally fix all parameter values and run the simulation. In most cases this is no problem, but if we are studying phenomenons like phase transitions or probabilistic behaviour at a certain critical parameter value we may be in trouble. Perhaps the finite set of possible parameter values we can use in a simulation does not cover the interesting parameter values of a certain model. A rather new technique, omniparametric simulation, enables us to simulate a model for all parameter values at the same time, even if the parameter set is uncountable.

Before introducing necessary terminology and notation, let us for a while focus on the treatment of parameters and also on the more general aspects of omniparametric simulation.

1.1 Fixed parameter simulation

A mathematical model has a number of parameters, for example temperature or infection rate, or both. Traditionally the simulation is done for fixed values of these parameters, and inference done based on the simulation results. Results for such an analysis are valid for the chosen parameters values only, and rough approximations be means of interpolations between the chosen parameter values may or may not be valid depending on the circumstances. In some cases the general behaviour of the model is the same for all parameters values, in other cases not. With this approach models with a continuous parameter space can only be simulated for a finite number of parameter values. Hopefully these parameter values are spread throughout the parameter space, or at least over interesting regions.

The approach poses no problem if the model has a nice behaviour over the represented parameter space, and if any interesting behaviour can be detected by using parameter values in that set.

Let us for the moment focus on the simulation of particle systems like independent bond percolation or the Ising and Potts models. These models changes their behaviour if the parameter values are changed, in some cases quite drastically. This dramatic change in behaviour is called a phase transition, and the different regimes for the models are called phases. Different phases have different probabilistic behaviour, sometimes expressed by the existence of multiple probability measures.

The parameter range for different phases however small they are in the parameter space, may consist of a continuous set of values, and can then be detected by a proper choice of parameter values. However if the model has a different phase for a single parameter value this can not easily be detected. Let us call such values of the parameter exceptional values. If this set of exceptional values is unknown it may or may not be possible to detect by means of ordinary fixed parameter simulation. For known exceptional values the problem is reduced to the problem of representing these parameter values in the computer.

The problem with exceptional values occurs when they are not known, and we want to use computer simulations to find them or indications of their existence. We can not by increasing the set of simulated parameter values hope to find exceptional behaviour, since any continuous

parameter can only be simulated for a finite number of different values and we will, with probability one, miss the interesting value. The solution to this problem is to simulate over the entire parameter space at the same time, thus covering the behaviour of the model over the entire parameter space in one single sample.

1.2 Omniparametric simulation

The basic idea behind omniparametric simulation is a coupling of the model over all parameter values in the parameter space. We start with the concept of coupling, elaborate some on the representation of omniparametric models, and finish with some historic remarks.

Coupling of processes over the parameter space

This is a brief presentation of the subject, for a general description of the coupling technique see [Lin92] .

Consider two stochastic processes, X and Y defined on probability spaces $(E_X, \mathcal{E}_X, \mathbb{P}_X)$ and $(E_Y, \mathcal{E}_Y, \mathbb{P}_Y)$ respectively. A coupling of X and Y is a probability measure μ on the measurable space $(E_X \times E_Y, \mathcal{E}_X \times \mathcal{E}_Y)$ such that

$$(i) \quad \mathbb{P}_Y = \mu \circ \pi_X^{-1}$$

$$(ii) \quad \mathbb{P}_Y = \mu \circ \pi_Y^{-1}$$

where $\pi_X : E_X \times E_Y \to E_X$ and $\pi_Y : E_X \times E_Y \to E_Y$ are projections. This makes the coupling a probability measure on a product space with prescribed marginals. This definition is given for a pair, but there is nothing keeping us from performing a coupling of several, sometimes even an uncountable number of processes simultaneously. Next we consider a small generic example, stressing the importance of an effective representation.

Consider a set of $0/1$-valued random variables parametrised by a parameter $\theta \in [0, 1]$ such that

$$X_\theta \stackrel{\mathcal{D}}{=} \mu_\theta$$

for all θ. A coupling of all these variables is a measure μ on the product space $\{0, 1\}^{[0,1]}$ such that $\mu_\theta = \mu \circ \pi_\theta^{-1}$, for $\theta \in [0, 1]$, where π_θ is a projection from $\{0, 1\}^{[0,1]}$ to $\{0, 1\}$. When using the coupling for omniparametric simulation we simulate the probability measure μ on the

space $\{0,1\}^{[0,1]}$, and represent a configuration in this space by a partition (A_0, A_1) of $[0,1]$ such that

$$X_\theta = \begin{cases} 1, & \theta \in A_1 \\ 0, & \theta \in A_0 \end{cases}$$

holds. If we manage to do the coupling in a nice way we can represent the partition with a single number, a threshold θ_t , such that $A_0 = [0, \theta_t]$ and $A_1 = (\theta_t, 1]$.

The complexity of the representation is especially important when using an iterative simulation scheme such as the Gibbs sampler. If we do not keep the complexity of the representation at a constant level the needed amount of computer memory will exceed all realistic limits, making the algorithm impossible to use.

1.2.1 Representation levels

Given a class of random variables $\mathbb{X} = \{X_\theta : \theta \in [0,1]\}$ each having distribution μ_θ and the omniparametric variable X and its measure μ there are different ways of representing X in a simulation.

- *Omniparametric measure* Given an explicit expression for μ we can calculate the probability for different values of X directly, and generate samples (at least in theory) by $\mu^{-1}(u)$ where $u \stackrel{\mathcal{D}}{=} U[0,1]$.
- *Omniparametric random variable* Without an expression for μ we may nevertheless have an algorithm, such as a Gibbs sampler, to generate X according to the correct distribution. An example is the omniparametric random cluster model in Section 8.1.
- *Indirect representation* We may not be able to represent the omniparametric variable directly, instead we have some representation of it, making it possibility to calculate the fixed parameter variable for any parameter value. An example is the omniparametric Potts model presented in Section 8.2.

1.2.2 A brief history

The central tool for constructing omniparametric random variables and simulation algorithms is the coupling. The concept of coupling was first introduced by Doeblin [Doe38] just before the second world war. Strassen was the first to introduce probability measures with prescribed

marginals in 1965 [Str65] , an important result regarding couplings. Strassen's result however involves only two measures which does not suffice to construct omniparametric variables over infinite parameter spaces.

To our knowledge the first construction resembling the omniparametric coupling was done by Stepanov in 1970 [Ste70] to study connectedness in Erdös-Renyi random graphs by coupling them over all paremeter values. In 1991 Higuchi [Hig91] presents what he calls a level set representation of the ferromagnetic Ising model. It is a coupling of the Ising ferromagnet over all values of the external field, while the interaction parameter is kept constant. Higuchi built his construction on a coupling deviced by Holley [Hol74] a few years after Stepanov's article. The purpose of Higuchi's coupling is the study of the percolation probability as a function of external field. In 1995 Grimmet [Gri95] used the ideas of Higuchi to introduce a coupling of random cluster processes for all values of the edge probability, while the parameter relating to cluster size, q, is fixed above 1. Grimmett also makes two notes about extending the model. The first is to couple the processes for all parameter values of the two-dimensional parameter making it fully omniparametric. The other one is concerned with the fixed parameter q, letting it take values below 1.

To our knowledge the level set representations of stochastic particle systems was used for theoretical purposes only until 1996, when Propp and Wilson published their work [PW96] on perfect simulation using coupling from the past (CFTP). They propose an omniparametric simulation algorithm for the random cluster model with q held fixed at some value ≥ 1 and p as the parameter over which domain simulations are generated.

1.3 Basic notation and terminology

We treat models living on graphs $\mathbb{L}^d$ with vertex set $\mathbb{Z}^d$, and edge set

$$\mathbb{E}^d = \{\langle x, y\rangle : x, y \in \mathbb{Z}^d,\ |x - y| = 1\}$$

where

$$|x| = \sum_{k=1}^{d} |x_k|,\ x = (x_1, ..., x_d) \in \mathbb{Z}^d$$

is the usual graph metric on the d-dimensional cubic lattice. When studying parts of the lattice we will do this in finite boxes, emphasized

by denoting boxes

$$\mathbb{B}_n^d = (\mathbb{Z}_n^d, \mathbb{E}_n^d)$$

where

$$\mathbb{Z}_n^d = \{x \in \mathbb{Z}^d : |x| \leq n\}$$

and

$$\mathbb{E}_n^d = \{\langle x, y\rangle : x, y \in \mathbb{Z}^d,\ |x - y| = 1\}$$

is the vertex and edge set respectively. When the number of dimensions are understood we will omit d from notation and simply write

$$\mathbb{B}_n = (\mathbb{Z}_n, \mathbb{E}_n)$$

instead. The boundary of a box $\mathbb{B}_n$ is defined as $\partial\mathbb{B}_n = \mathbb{B}_{n+1} - \mathbb{B}_n$.

Apart from this basic notation we introduce new symbols as we proceed. All used notation is listed in Appendix A starting on page 205.

1.4 Outline of the thesis

In Chapter 2 we present a small percolation exampel illustrating some advatages of the omniparametric simulation technique when studying connection probabilities. The rest of the thesis is divided into two parts.

Part II, chapters 3-5, treat the a model for spatial growth and competition known as two-type Richardson model. We introduce the necessary models and construct a coupling for the omniparametric model. The purpose of this part is to study simultaneous survival of both types in the two-type Richardson model and the tool is simulations.

Part III, chapter 6-12 which the reader will note makes up the main bulk of this thesis, treat applications of omniparametric simulation to parameter estimation in two of the most well-known examples arising in Gibbsian statistical mechanics: the Ising and Potts models.

For a more detailed description of each part see chapters 3 and 6 on pages 21 and 57 respectively.

CHAPTER 2

Percolation example

As a first concrete example, and a kind of warm-up for our more involved simulation algorithms in later chapters, we here consider omniparametric simulation of a very simple stochastic system, so called bond percolation on a special graph.

The idea is the following. We let each edge in the graph be open with probability p and closed with probability $1-p$ and study the probability that there exists a path of open edges between two vertices.

Consider independent bond percolation on the two-dimensional lattice, $\mathbb{L}^2 = (\mathbb{Z}^2, \mathbb{E}^2)$,

$$\mathbb{E}^2 = \{\langle x, y\rangle : x, y \in \mathbb{Z}^2,\ |x - y| = 1\}$$

with parameter p. When the number of dimensions d is understood we omit that from notation, writing $\mathbb{E}$ instead of $\mathbb{E}^d$. Every edge is "on" with probability p and "off" with probability $1-p$, independently of each other. Let $\mathbb{B}_n = (\mathbb{Z}_n, \mathbb{E}_n)$ where

$$\mathbb{Z}_n = \mathbb{Z}^2 \cap [-n, n]^2 \quad , \quad \mathbb{E}_n = \{\langle x, y\rangle \in \mathbb{E} : x, y \in \mathbb{Z}_n\}$$

be a finite subgraph of $\mathbb{L}^2$, a so called box or box graph. Also let $\partial\mathbb{B}_n = \mathbb{B}_{n+1} \setminus \mathbb{B}_n$ be the boundary of $\mathbb{B}_n$ (see Figure 2.1).

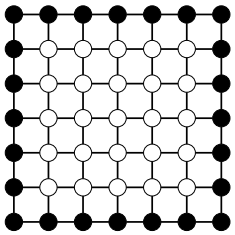

Figure 2.1: *The box B_2 (white) with its boundary $\partial\mathbb{B}_2$ (black).*

Generating a sample $\sigma_{\sigma_{\text{fix}}}$ is normally carried out using the following simple scheme.

1. Fix the parameter $p \in [0, 1]$.
2. Generate independently for each edge e a random number u_e uniformly distributed over $[0, 1]$.
3. Assign edge e the value 1 ("on") if $u_e \leq p$ and 0 ("off") otherwise. Do this for all edges in $\mathbb{E}_n$.

Implicit in this traditional approach is a more powerful idea: couple the process for all $p \in [0, 1]$.

1. Generate independently for each edge a random number u_e, uniformly distributed over $[0, 1]$, and assign this values to the edge.

This simple assignment give us the omniparametric sample $\sigma_{\text{omni}} \in [0, 1]^{\mathbb{E}_n}$. To generate an ordinary configuration with 0's and 1's for a fixed p we perform step 3 above for all edges.

Corresponding to each σ_{omni} is a subgraph of $\mathbb{B}_n$ having $\mathbb{Z}_n$ as vertex set and $\{e \in E_n : \sigma_{\text{fix}}(e) = 1\}$, the set of open edges as edge set. The problem we study below is the connection properties of this subgraph $\mathbb{B}_n$ for larger and larger n. Hopefully we get an estimate of the probability that two vertices are connected as on $\mathbb{Z}_n$ as $n \to \infty$. We shall also examine the difference between the two simulation approaches, both in results and simulation execution time.

2.1 Connection probabilities

Given two vertices $x, y \in \mathbb{Z}^2$ we let $\{x \leftrightarrow y\}$ denote the event that there is a path of open edges between the two vertices. In graph theoretic

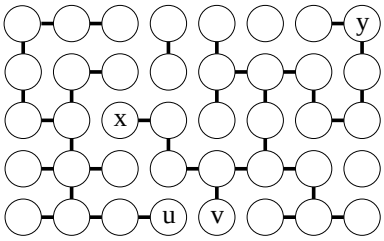

Figure 2.2: *A small rectangle of $\mathbb{Z}^2$, where there is a path between vertices x and y but not between u and v. When the rectangle is enlarged however there could be an open path between u and v.*

terms they are in the same connected component after we have removed all "off" edges from the original graph. The connection probabilities for different pairs of vertices are often difficult or impossible to treat by ordinary analysis so simulation is the tool to use. Let

$$f_{x,y}(p) = \mathbb{P}_p(x \leftrightarrow y)$$

be the connection probability for vertices x and y as a function of p. We will try to estimate it using both omniparametric and fixed parameter simulation, and see if we gain anything by using the omniparametric approach. Note that the function is monotone so any estimate $\hat{f}_{x,y}(p)$ should also be monotone.

As seen in Figure 2.2 the connection probability for a vertex pair could depend on a large, but finite, number of edges. To see why this is the case we use the following argument, given any vertices x and y.

If $\{x \leftrightarrow y\}$ happens then there is some finite path between these vertices, on which the event $\{x \leftrightarrow y\}$ depend, thus there exists $M < \infty$ such that $\{x \leftrightarrow y\}$ depends only on edges in B_M. If $\{x \leftrightarrow y\}$ does not happen then x and y are located in two disjoint connected components, of which only one can be infinite since the infinite connected component is unique (for a proof see [Gri99]). The smaller one is finite and contained in some finite box B_M making the event $\{x \leftrightarrow y\}$ depend only on edges in B_M. Note that the number M is a random variable depending of x, y and σ_{fix}. From the argument above it follows that M is unbounded but almost surely finite. During simulation we simulate larger and larger boxes, and stop when reaching box B_M. This process will then terminate with probability one. Weather or not the available computer re-

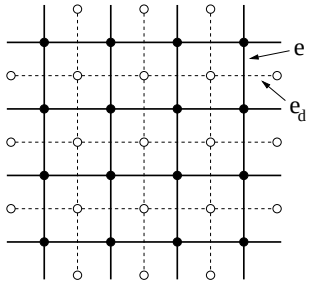

Figure 2.3: *A subset of the grid $\mathbb{L}^2$ (solid) and its dual $\mathbb{L}^2_d$ (dashed). For every edge e in the lattice there is one unique corresponding (and crossing) edge e_d in the dual. Whenever e is "off" in $\mathbb{L}^2$ its companion e_d is "off" if $\mathbb{L}^2_d$.*

sources is enough to simulate the model on B_M is a different question, important though in any application.

A special case

For most pairs we do not know how to calculate these connection probabilities exactly, but there is one exception. For nearest neighbours x and y we can, by using the dual $\mathbb{L}^2_d$ (see Figure 2.3) of $\mathbb{L}^2$, calculate $\mathbb{P}(x \leftrightarrow y)$ in the special case when $p = 1/2$. For any two vertices $x, y \in \mathbb{Z}^2$ let $\rho(x, y)$ denote a path between x and y not containing the edge (x, y), and let $\{\rho(x, y) = 1\}$ and $\{\rho(x, y) = 0\}$ denote the events that this path is open or closed respectively. Consider two nearest neighbour vertices x and y, see Figure 2.4. If $\{x \leftrightarrow y\}$ happens the edge (x, y) is either open or if (x, y) is closed then some path $\rho(x, y)$ is open. If $\{x \not\leftrightarrow y\}$ happens then the edge (x, y) is closed and there exists some closed path $\rho(u, v)$ in the dual, keeping x and y apart (see Figure 2.4). We express this as follows.

$$\begin{array}{ll} (i) & \mathbb{P}_p(x \leftrightarrow y) = \mathbb{P}_p(e = 1) + \mathbb{P}_p(e = 0)\mathbb{P}_p(\ \exists \rho(x, y) : \rho(x, y) = 1\ ) \\ (ii) & \mathbb{P}_p(x \not\leftrightarrow y) = \mathbb{P}_p(e_d = 0)\mathbb{P}_p(\ \exists \rho(x, y) : \rho(u, v) = 1\ ) \end{array}$$

For $p = 1/2$ we have

$$\mathbb{P}_{1/2}(\ \exists \rho(x, y) : \rho(x, y) = 1\ ) = \mathbb{P}_{1/2}(\ \exists \rho(u, v) : \rho(u, v) = 0\ )$$

giving $\mathbb{P}_{1/2}(x \leftrightarrow y) = 3/4$. The argument is well known but we have not been able to trace its roots, and obviously it breaks down for $p \neq 1/2$.

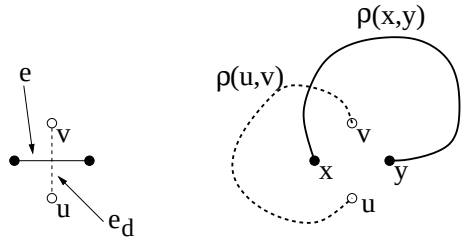

Figure 2.4: *To the left the edge e is on and x and y are connected. In the right diagram u and v are connected by a closed path not containing e_d. Both e and therefore e_d is "off" and if the path $\rho(u, v)$ in the dual is closed so there can be no open path in $\mathbb{L}^2$ connecting x and y.*

2.2 A fixed parameter simulation scheme

Let x and y be arbitrary vertices in $\mathbb{Z}^2$. We start the simulation with a rather small box $B_n = [-n, n]^2$ and see if $\{x \leftrightarrow y\}$ happens or not. If we can not decide (see Figure 2.5 for an explanation) we increase the box size and continues. For $z \in \mathbb{Z}^2$ let $C(z)$ be the cluster containing z. The algorithm continues until $C(x) = C(y)$ happens, or if $C(x) \neq C(y)$ until

$$C(x) \subseteq B_{n-1} \text{ or } C(y) \subseteq B_{n-1} \text{ , if } C(x) \neq C(y)$$

happens.

Termination is a consequence of uniqueness of the infinite cluster in the supercritical phase.

2.3 Omniparametric algorithm

The omniparametric simulation algorithm returns, instead of just a "1" or "0" for every edge, a threshold p_t corresponding to $\sigma_{\text{omni}}(e) = 1$ if $p \geq p_t$ and $\sigma_{\text{omni}}(e) = 0$ otherwise.

The threshold for any path ρ between two vertices x and y has a threshold denoted by $p_t(\rho)$, equal to the largest edge threshold among its edges such that the path is open if $p_t(\rho) \leq p$ and closed otherwise. Let $\Delta^n_{x,y}$ be the set of all paths between x and y contained within the finite box B_n and let

$$p_t(x, y) = \inf \left(\bigcup_{n \in \mathbb{N}} \{p_t(\rho) : \rho \in \Delta^n_{x,y} \} \right)$$

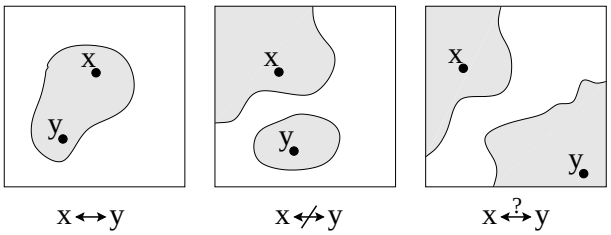

Figure 2.5: *On a finite box one out of three events will happen. In the first both x and y are in the same cluster. In the middle x and y are in different connected components of which at least one is fully contained in the box. To the right x and y are contained in different components in the box, but we can not decide weather or not these components will connect on a larger box, to do this we have to increase the box size sufficiently.*

be the smallest value p for which the event $\{x \leftrightarrow y\}$ occur. For any finite subset $A \subset \mathbb{Z}^2$ we let

$$p_t(x, A) = \inf \bigcup_{y \in A} p_t(x, y)$$

be the smallest p for which there is an open path between x and any of the vertices in A.

We can now formulate the omniparametric simulation algorithm as follows. Given $x, y \in \mathbb{Z}^2$ we start with a small box B_n and compute $p_t(x, y)$, $p_t(x, \partial B_n)$ and $p_t(y, \partial B_n)$. If

$$p_t(x, y) \leq \max(p_t(x, \partial B_n), p_t(y, \partial B_n))$$

we stop and return $p_t(x, y)$, since for any p x and y are either connected or at least one of them is disconnected from the boundary ∂B_n. If however

$$p_t(x, y) > \max(p_t(x, \partial B_n), p_t(y, \partial B_n))$$

we increase the box size with Δn, recompute $p_t(x, y)$, $p_t(x, \partial B_{n+\Delta n})$ and $p_t(y, \partial B_{n+\Delta n})$, and check again. This process continues until

$$p_t(x, y) \leq \max(p_t(x, \partial B_{n+\Delta n}), p_t(y, \partial B_{n+\Delta n})).$$

For fixed parameter simulation, termination is a consequence of uniqueness of the infinite cluster in the supercritical phase. For the

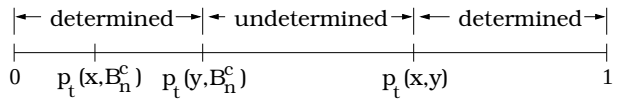

Figure 2.6: *The event $\{x \leftrightarrow y\}$ is determined for $p < p_t(y, \partial B_{n+1})$ since there is no path between x and y, and at lest on of them is not connected to the boundary by an open path. For $p \geq p_t(x, y)$ there is an open path between x and y. For $p \in [p_t(x, \partial B_{n+1}), p_t(x, y))$ there there are open paths between both x and y and the boundary, but no path between them within the box. For p in this interval we have to look at a larger box to determine the situation.*

omniparametric algorithm we need something stronger, namely uniqueness of the infinite cluster simultaneous for all values of p at the same time, established by Alexander [Ale95] in 1995.

2.4 Simulation results

We study the probabilities for two connection events, $\{(0,0) \leftrightarrow (1,0)\}$ and $\{(0,0) \leftrightarrow (5,0)\}$. In Figure 2.7 we see the results, the estimated connection probabilities for both methods.

As expected the curves originating from omniparametric simulations are monotone, while the others are not. Using ordinary simulation there is nothing in the algorithm suggesting that the result should be monotone, each batch of simulations (one for each p) is carried out independently of all others.

For a certain set of values for p say $0, 0.01, 0.02, ..., 0.99, 1.00$ one could perhaps achieve this by running more simulations, but the problem itself remain. No matter how many simulations we do, there will always be possible to find an estimate which is not monotone by dividing the interval for p into finer and finer intervals.

The omniparametric simulation algorithm however has the monotonicity built in. No matter how small the number of simulations is the estimated function originating from omniparametric samples will always be monotone.

There is a large difference in speed. The two simulation techniques are used to generate 1000 samples for each p. The fixed parameter algorithm does 1000 simulations for every value of p, while the omniparametric algorithm does a total of 1000 simulation, independent of the number of values for p.

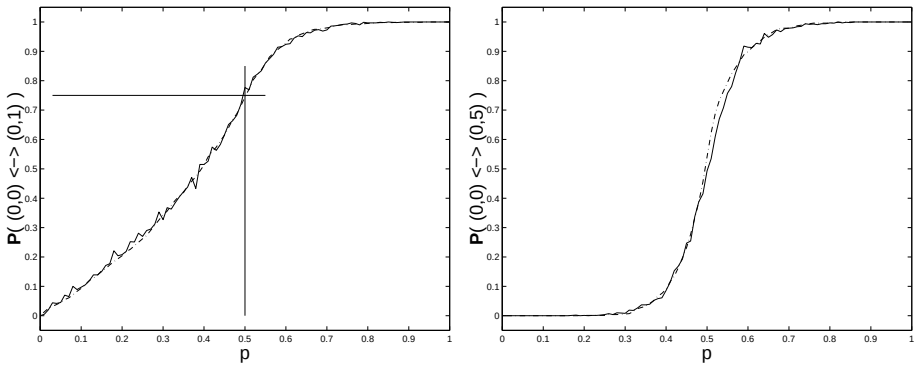

Figure 2.7: *Simulated connection probabilities for events* $\{(0,0) \leftrightarrow (0,1)\}$ *and* $\{(0,0) \leftrightarrow (0,5)\}$*. The solid curves are the probabilities estimated by ordinary simulation (100000 simulations, 1000 for each* p*) and the dashed curves are estimates from omniparametric simulations (1000 simulations). In the left diagram we see both curves closely approximate the probability* $\mathbb{P}_{1/2}(x \leftrightarrow y) = 3/4$*.*

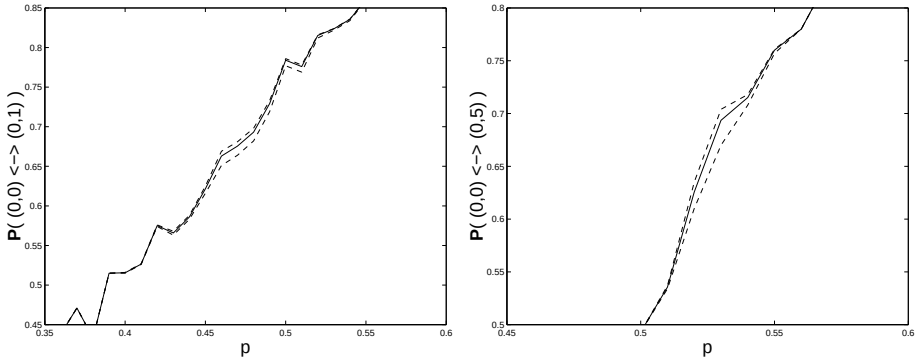

Figure 2.8: *The fixed parameter simulations both suffer from the upper limit of the box. In some interval around* $p = 1/2$ *the curves are only approximated. The solid line is the approximation when disregarding the aborted simulations. The upper and lower dashed lines are estimates using all simulations, including the aborted ones.*

There is a large difference in the total number of simulations we need to run in order to generate our estimates. The fixed parameter scheme used a total of 100000 simulations while we used only 1000 for the omniparametric scheme.

The extra computation for the fixed parameter algorithm becomes critical when p is in a small interval around its critical value, $p_c = 1/2$. In this interval we expect large clusters, making both algorithms slow. Though there is no theoretical difference between the two approaches there is a practical one. While the omniparametric algorithm may encounter large clusters in every simulation, with a certain probability, the ordinary algorithm has the same probability of suffering from a slowdown in $1000c$ simulations, where c is the number of p-values in a small interval around p_c.

Since there is a small probability for the simulations to use larger boxes than the computer can handled we use a fixed largest box size, in this case 500. If the event $I_{\{x \leftrightarrow y\}}$ is not yet determined when reaching this upper limit the simulation is aborted and marked accordingly. These aborted simulations are later used for getting upper and lower estimates of the connection probability. None of the omniparametric simulations did encounter any problems with this limit, while some of the fixed parameter simulations did. In Figure 2.8 we see how this limit affects the simulations.

In later chapters, when trying to estimate connection probabilities in the Ising and Potts models, we will encounter this problem again (see Section 10.1, page 131).

Part II

The two type Richardson model

CHAPTER 3

Introduction

In this part of the thesis we will take into consideration a certain growth model, the two-type Richardson model, and address the question regarding simultaneous survival as time tends to infinity. This question has been partially answered so far, and today it is known that simultaneous survival can not happen when the infections are not equally strong, for almost all parameter values. There is however a strong belief that "almost all" can be replaced by "all" in this statement. What remains is to rule out simultaneous survival for a countable set of parameter values. We use omniparametric simulation and see how far towards an answer the simulations bring us.

The rest of this part is organised as follows. In Section 3.1 and 3.2 we briefly describe the growth models in use, the Richardson model and its extension, the two-type Richardson model.

In Chapter 4 we properly introduce the two-type Richardson model and give some results. We also introduces the omniparametric two-type Richardson model and describe its relation to the ordinary fixed parameter model.

Chapter 5 describe simulations. First a theoretical base is established, then we do simulations, and finally perform the analysis.

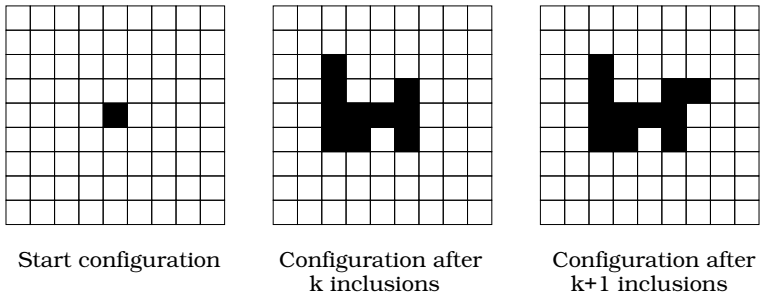

Figure 3.1: *The Richardson model on the square tessellation*

3.1 The Richardson model

The Richardson model [Ric73] is a stochastic growth model in d-dimensions, where d is any finite number. We can think of the volume on which the models grows as a volume divided into cells. From the beginning all the cells have the same state, all but one which has a certain interesting property, for example an infection. When time passes this infection will spread, as infected cells affect uninfected ones.

In mathematical terms the Richardson model is a Markovian growth process in continuous time. The infected area grows as infected cells infect uninfected ones, one by one. The rate by which uninfected cells are infected is proportional to the number of infected neighbours an uninfected cell have. Let $\{X_t\}_{t\geq 0}$ denote the state of the infection at time t and let $N_R(t)$ denote the number of infected neighbour cells of the cell R. For a certain cell R the probability for being infected during a small time interval $[t, t+h]$ is given by the following

$$\mathbb{P}(\ R \text{ infected at time } t+h \mid R \text{ not infected at time } t\ ) = h(1+o(1))N_R(t)$$

as $h \downarrow 0$ In Figure 3.1 we see a simple example of the Richardson growth process on the square tessellation. The only difference between the middle diagram and the right one is an extra infected cell. Instead of studying the model in continuous space, one could use a graph, replacing cells with nodes and each neighbour relation between two cells with an edge.

In his article from 1973 Richardson studied and stated a theorem regarding the shape of the infected region. The result states that as

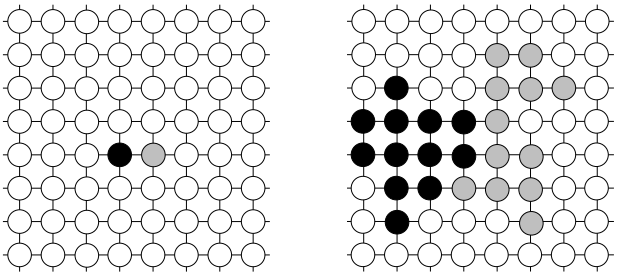

Figure 3.2: *The two-type Richardson model on a small box in $\mathbb{Z}^2$. Initial configuration (left) and configuration after some time (right).*

$t \to \infty$ the shape of the infected area tends to some non-random shape. The question regarding the exact shape remains unanswered.

3.2 The two-type Richardson model

The two-type Richardson model extends the Richardson model by adding another type of infection. It was introduced by Häggström and Pemantle [HP98] in 1998 and extended by the same authors [HP00] in 2000.

The process starts with every vertex in $\mathbb{Z}^d$ having type zero (white), except for vertices in a region of type one (grey) and vertices in a region of type two (black). For theoretical purposes it is enough to let the initial configuration be the simple one in Figure 3.2 (left) and then study how that configuration evolves through time.

As time passes each coloured (black or grey) vertex colour its uncoloured (white) neighbours in a Poisson process with a colour dependent intensity. After being coloured a vertex keeps its colour for all times. See Section 4.1 for a more precise description of the model.

CHAPTER 4

The two-type Richardson model

4.1 Definition

The two-type Richardson model is a growth model on the lattice $\mathbb{Z}^d$ introduced by Häggstrom and Pemantle [HP98] . In this extension of the ordinary Richardson model [Ric73] two infections are competing for space and each vertex is infected or not, if infected it has one of two types.

The process starts at $t = 0$ with a certain initial configuration, that is, two disjoint subsets ξ_1 and ξ_2 of $\mathbb{Z}^d$ where all vertices of ξ_1 have type one, and all vertices of ξ_2 have type two. The rest of the lattice have type zero. The model evolves through time as 0's changes to 1's or 2's at rates depending on the nearest neighbour configuration, while the 1's and 2's do not change at all. Let type one and two have infection rates λ_1 and λ_2 respectively. Due to time scaling and symmetry we only need to study the case when $\lambda_1 = 1$, and $\lambda_2 = \lambda \in [0, 1]$. Let us now state this in a precise manner.

We construct a graph with vertex set $\mathbb{Z}^d$ and edge set

$$\mathbb{E} = \{\langle x, y\rangle : \ x, y \in \mathbb{Z}^d, \ |x - y| = 1\}$$

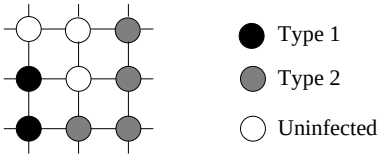

Figure 4.1: *The uninfected vertex in the middle changes to type 1 at rate λ_1, and to type 2 at rate $2\lambda_2$*

where

$$|x| = \sum_{k=1}^{d} |x_k|,\ x = (x_1, ...x_d) \in \mathbb{Z}^d$$

is the metric in $\mathbb{Z}^d$. An element $\{0,1,2\}^{\mathbb{Z}^d}$ is called a configuration and is an assignment of state 0,1 or 2 to each vertex in $\mathbb{Z}^d$, we denote it by $\Xi^\lambda(t)$. If $\Xi^\lambda(t)$ is the configuration at time t with parameter λ we denote the state of a vertex $u \in \mathbb{Z}^d$ by $\Xi^\lambda_u(t)$. Also let $\eta^\lambda_1(t)$ and $\eta^\lambda_2(t)$ denote the sets of edges having type one and two respectively. Let the initial configuration ξ_1, ξ_2 be defined by

$$\xi_i = \{v \in \mathbb{Z}^d : \Xi^\lambda_v(0) = i\},\ i = 1, 2.$$

The pair (ξ_1, ξ_2) can be any pair of disjoint finite subsets of $\mathbb{Z}^d$, but for simplicity we will assume that both are connected subsets of the graph.

The evolution of the model is defined by the infection process. If a vertex has type one or two it infects all its uninfected neighbours at a certain type dependent rate, and the probability that such a neighbour will be infected in a short time interval is proportional to the length of the interval. If a type zero vertex has k_1 type one neighbours and k_2 type two neighbours it changes to type one at rate $k_1\lambda_1$, and to type two at rate $k_2\lambda_2$. Since the parameters λ_1 and λ_2 do not change over time we get two homogeneous Poisson processes of infection events for each edge, one for each direction. We think of these as processes active for all non-negative times, but events are only interesting when the source vertex of the edge is infected and the target vertex is not.

To reduce the number of parameters in the model we use time scaling. For infection rates λ_1, λ_2 suppose $\lambda_1 \geq \lambda_2$ and rescale by using rates

$$\lambda'_1 = 1 \text{ and } \lambda'_2 = \frac{\lambda_2}{\lambda_1}$$

instead.

Before proceeding with behaviour and results we present a representation of the model making it easier to simulate. First we introduce thinning of Poisson processes.

Consider a homogeneous Poisson process, with intensity λ. If we let each event occur with probability p, independently of each other, it is a straightforward calculation to show that the process of remaining events also is a Poisson process, but with intensity λp. The remaining process is called the thinned Poisson process.

Assign now independently to all edges two unit rate Poisson processes, one for each direction of an infection event, and two sequences of independent uniformly distributed $[0,1]$-random variables. The evolution of the model is now determined by the sequences of Poisson events along with the random numbers. At each event one vertex tries to infect another along the common edge. Assume that such an event takes place at time t, and let t^+ be the time just after this event. Also let v be the infecting vertex and w its victim. If v is infected and w is not the following happens.

$$\Xi_w(t^+) = \begin{cases} 1, & \Xi_v(t) = 1 \\ 2, & \Xi_v(t) = 2,\ u \leq \lambda \\ 0, & \Xi_v(t) = 2,\ u > \lambda \end{cases}$$

where u is the random number associated with this event. If w is infected or v is not, nothing happens. Let $\mathbb{P}^{\lambda}_{\xi^1,\xi^2}$ denote the probability measure for the described process.

4.2 Behaviour and results

The remaining part of this chapter is devoted to the study of simultaneous survival, that is, the event that both infection types continues to infect vertices for all times. First we consider the asymptotic behaviour, then analyse the situation for finite boxes.

Asymptotic properties

Start the process with initial configuration, $\xi_1 = \mathbf{0} = \{(0,0,..,0)\}$ and $\xi_2 = \mathbf{1} = \{\ (1,0,\ldots,0)\ \}$, and let it evolve for all times. There will always be infected vertices having uninfected ones as neighbours, so the infected area will get larger and larger and as time tends to infinity the infection will spread to all vertices of $\mathbb{Z}^d$, leaving no vertex uninfected.

We study $\Xi(t)$ as $t \to \infty$. Let G_1 denote the event that type one survives as $t \to \infty$, and define G_2 accordingly. In the sequel we will consider the model for a fixed but arbitrary λ. For simplicity we suppress λ in notation. There are three possible scenarios.

1. Vertices of type one surrounds all type two vertices, and only type one grows to infinity. Type one is said to strangle type two, denote this event $G_1 \setminus G_2$.

2. Type two strangles type one, denoted $G_2 \setminus G_1$.

3. Neither type strangles the other, both types continues to infect vertices as $t \to \infty$, denoted $G_1 \cap G_2$.

There is always a possibility for type one to strangle type two since it is the stronger type, and before the region infected by type one has become too large there is always a possibility for type two to strangle type one. So both scenario one and two has positive probability as long as $\lambda > 0$. The third scenario is more complicated. If both types survive to infinity, this reflects some kind of power balance. For the case $\lambda = 1$ and $d = 2$ Häggstrom and Pemantle [HP98] showed that $\mathbb{P}^{1,\lambda}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) > 0$. During 2005 both Hoffman [Hof05] and Garet and Marchand [GM05] have extended this result to $d \geq 3$ The main theorem in [HP98] states (Theorem 4.1 below) that for almost all λ we can not have simultaneous survival of both types in the limit.

Theorem 4.1 ***Häggstrom-Pemantle***
Consider the two-type Richardson model on $\mathbb{Z}^d,\ d \geq 2$. Then

$$\mathbb{P}^{\lambda,1}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) = 0$$

for all but at most countably many choices of λ.

The result is valid for all finite initial configurations where no type is strangled by the other, see [HP00] . But it does not say anything about the situation for any particular value of λ. For a fixed λ we can not use this theorem and the same situation arises in traditional simulation. We can not use fixed parameter simulation to rule out the existence of some $\lambda_e \in (0, 1)$ such that

$$\mathbb{P}^{\lambda_e,1}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) > 0.$$

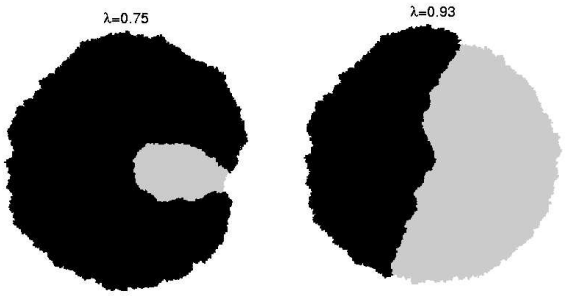

Figure 4.2: *A realisation of the model for two different values of λ. In the right diagram both types seem to have survived in some kind of balance, but in the left type 1 (black) is about to surround the weaker type 2. If conjecture 4.1 is correct, then the apparent balance on the right is doomed to break down in the long run.*

It should be easier to achieve simultaneous survival the closer to 1 λ is implying that $\mathbb{P}_{\mathbf{0},\mathbf{1}}^{\lambda_e,1}(G_1 \cap G_2) > 0$ is monotone in λ.Thus Theorem 4.1 strongly suggests

$$\mathbb{P}_{\mathbf{0},\mathbf{1}}^{\lambda,1}(G_1 \cap G_2) = 0$$

for all $\lambda \neq 1$. If such monotonicity actually hold the result follows for all λ. There are however more general graphs where the monotonicity does not hold, see Deijen and Häggstrom [DH05] . We now present the conjecture from Häggstrom and Pemantle [HP00] , which is also the starting point and main reason for this part of the thesis.

Conjecture 4.1
For the two-type Richardson model on $\mathbb{Z}^d,\ d \geq 2$, we have

$$\mathbb{P}_{\mathbf{0},\mathbf{1}}^{\lambda_1,\lambda_2}(G_1 \cap G_2) = 0$$

whenever $\lambda_1 \neq \lambda_2$.

Another open question is the asymptotic shape of the infected area. Already in his 1973 article [Ric73] Richardson stated the existence of such an asymptotic shape. Some further results concerning it has been published over the years, see [DL81] , but the question regarding exact shape remains unanswered.

Finite boxes

Let us now consider the model on finite boxes. No matter how slowly the weaker type infects its neighbours there is always a positive probability for it to strangle the stronger type. This event might happen at any time, but is most likely to happen in the beginning before the area infected by the stronger type has become too large. This also follows from Proposition 4.1 in [HP00] , where it is proven that a small advantage for the stronger type is asymptotically enough to ensure that it will strangle the weaker type. This proposition is a crucial part of the proof of Theorem 4.1, we state it below, adapted to our notation.

We need a definition of size of an infected area. Although the asymptotic shape B is unknown we can use it to define this measure of size, $|\cdot|$. On $\mathbb{Z}^2$ B is of infinite size, but we can view B as a finite set on $\mathbb{R}^2$ by letting the distance between each vertex (starting with distance 1) shrink to zero as the finite boxes tend to infinity. For any set $A \subseteq \mathbb{Z}^d$ let $|A| = \inf\{t : A \subseteq tB\}$. For any two $u, v \in \mathbb{R}$ we define the set $S(u, v)$ of pairs of configurations as follows

$$S(u,v) = \{(\xi^1, \xi^2) : |\xi^1| \leq u,\ |\xi^2| \geq v\}$$

We will use this for $u < v$ to denote subsets of pairs of configurations, (ξ^1, ξ^2), where ξ^1 always will be a little bit smaller than ξ^2.

Proposition 4.1
Fix $\lambda \in (0,1)$ and let $1 < a < b$. Then

$$\lim_{t\to\infty} \sup_{(\xi^1,\xi^2)\in S(ta,tb)} \mathbb{P}^{\lambda,1}_{\xi^1,\xi^2}(G_1) = 0$$

Let $G^\lambda_{i,n}, i = 1, 2$ be the event that type i survives as long as the set of infected vertices are contained within B_n and define a function $p_n(\lambda)$ by the following.

$$p_n(\lambda) = \mathbb{P}(G^\lambda_{1,n} \cap G^\lambda_{2,n})$$

We need to formulate and prove two properties of the sequence $\{p_n\}_{n=1}^\infty$.

Lemma 4.1
For each n, $p_n(\lambda)$ is continuous.

We defer the proof of Lemma 4.1 until all necessary notation has been introduced, see page 39.

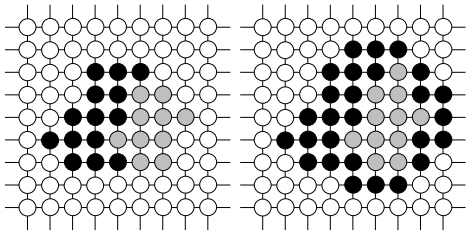

Figure 4.3: *The configuration ω in boxes B_3 and B_4.*

Lemma 4.2
For any $\lambda \in (0, 1]$ and any n, we have $p_{n+1}(\lambda) < p_n(\lambda)$

Proof :
If $G^{\lambda}_{1,n+1} \cap G^{\lambda}_{2,n+1}$ happens then both types has survived on B_n which is enough to ensure $G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}$. Let $\omega \in \Omega$ be the outcome in fig. 4.3. The set of type one, $\eta_1(t)$ is almost surrounded in the smaller box (B_3), and if vertex v is infected by the type two infection, then $\eta_2(t)$ surround $\eta_1(t)$ in B_4. So $\omega \in (G^{\lambda}_{1,2} \cap G^{\lambda}_{2,2}) \setminus (G^{\lambda}_{1,3} \cap G^{\lambda}_{2,3})$. This can happen for each finite n. Thus

$$(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) \setminus (G^{\lambda}_{1,n+1} \cap G^{\lambda}_{2,n+1}) \neq \emptyset$$

for all n and the result follows.
□

4.3 Simulation, a simple scheme

Given the proposed model with intensities 1 and λ for types one and two respectively, a simple simulation algorithm is proposed in Figure 4.4.

We think of the infection processes as unit rate Poisson processes, because of this we can choose an infecting vertex u at random. The next step is to choose a neighbour v to infect, also uniformly at random. If u has type one and v type zero, vertex v gets infected with type one. If u has type two and v has type zero we use thinning. With probability λ we infect v with type two. If u has type zero or v does not have type zero no infection takes place. The function computing the stop criteria can be any function as long as it returns TRUE in finite time.

```
graph G
graph G.setInitialConfig()
repeat
  u = G.getRandomVertex()
  v = G.getRandomNeighbour(u)
  if u.type1 ≡ TRUE AND v.uninfected ≡ TRUE then
    v.type = 1
  end if
  if u.type2 ≡ TRUE AND v.uninfected ≡ TRUE then
    if rnd ≤ λ then
      v.type = 2
    end if
  end if
until G.computeStopCriteria() ≡ TRUE
```

Figure 4.4: A simple simulation scheme for the two-type Richardson model.

4.4 The omniparametric two-type Richardson model

The omniparametric model is based on the ordinary two-type Richardson model. If we looked upon the ordinary model as a model for spreading infections, the omniparametric model is about spreading information. The information tell us for which parameter values each vertex is uninfected or has one of two infection types.

We use the same graph, $(\mathbb{Z}^d, \mathbb{E})$, and the same set of independent Poisson processes for all edges in $\mathbb{E}$, each equipped with a sequence of i.i.d. uniformly distributed $[0,1]$-random numbers. Let a configuration Θ be an element of $\{[0,1] \times [0,1]\}^{\mathbb{Z}^d}$, and let

$$\Theta(t) = \{(a_{v,t}, b_{v,t}) \in [0,1] \times [0,1] : \ a_{v,t} \leq b_{v,t}, \ v \in \mathbb{Z}^d\}$$

denote the configuration at time $t \geq 0$. The configuration for a vertex v at time t is denoted $\Theta_v(t)$, and we let

$$\Omega_O = \{(a,b) : a, b \in [0,1], \ a \leq b\}^{\mathbb{Z}^d}$$

denote the sample space.

Given $\Theta(t)$ we can obtain the fixed parameter configuration $\Xi(t)$ for any $\lambda \in [0,1]$ in the following manner. We assign each vertex v type one if $\lambda \leq a_{v,t}$, type two if $\lambda > b_{v,t}$ and type zero if $\lambda \in (a_{v,t}, b_{v,t}]$ (see Figure

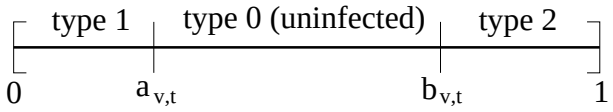

Figure 4.5: *Meaning of threshold representation in the omniparametric two-type Richardson model*

4.5). For this assignment to work we must have $a_{v,t} \leq b_{v,t}$ for all $t \geq 0$ since a vertex can have one infection type only. A initial configuration is an assignment of threshold values to all vertices such that for each $\lambda \in (0, 1]$ we have a valid initial configuration in the ordinary model.

The evolution of the omniparametric model is more complicated than for the fixed parameter model. Instead of propagating infection types we propagate threshold values. Any vertex having type one or two for some fixed but arbitrary λ can propagate this to any neighbour w if $\lambda \in (a_{w,t}, b_{w,t}]$. The idea is the following. We consider the sequence of Poisson events just as in the ordinary model. Assume that, at time t, a vertex v tries to propagate its information to vertex w, and let t^+ be the time just after this event. Let $\Theta_w(t) = (a_{w,t}, b_{w,t})$ be the state of vertex w at time. This vertex is already infected for $\lambda \in [0, a_{w,t}] \cup (b_{w,t}, 1]$ so only for $\lambda \in (a_{w,t}, b_{w,t}]$ an infection can take place. For vertex v we have three cases.

(i) For $\lambda \in [0, a_{v,t}]$ v infects w if $\lambda \in (a_{w,t}, b_{w,t}]$.

(ii) For $\lambda \in (b_{v,t}, 1]$ v infects w if $\lambda \in (a_{w,t}, b_{w,t}]$.

(iii) For $\lambda \in (a_{v,t}, b_{v,t}]$ no infection occur.

Infections of type one are spread each time, so at time t^+ we let a_{w,t^+} be equal to $\max(\ a_{w,t},\ \min(a_{v,t}, b_{w,t})\ )$. For an illustration see Figure 4.6. Infections of type two spread with probability λ, accomplished by using the random number u assigned to the Poisson event, and let the type two infection spread whenever $u \leq \lambda$. When changing the limit $b_{w,t}$ we do this depending on $\max(u, b_{v,t})$ instead of just $b_{v,t}$, thus $b_{w,t+} = \min(\ b_{w,t},\ \max(\gamma, a_{w,t})\ )$, where $\gamma = \max(u, b_{v,t})$.

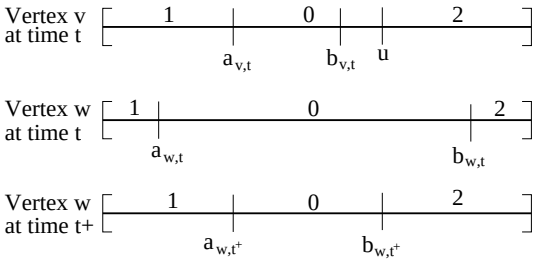

Figure 4.6: *An example of how information propagates from vertex v to vertex w.*

Summary: If at time t vertex v tries to propagate information to vertex w, having configurations $\Theta_v(t) = (a_{v,t}, b_{v,t})$, $\Theta_w(t) = (a_{w,t}, b_{w,t})$ respectively, then

$$\begin{aligned} a_{w,t^+} &= \max(\ a_{w,t},\ \min(a_{v,t}, b_{w,t})\ ) \\ b_{w,t^+} &= \min(\ b_{w,t},\ \max(\gamma, a_{w,t})\ ) \end{aligned}$$

is the updated configuration at w. Given $\xi_1, \xi_2 \subseteq \mathbb{Z}^d$ we define the initial configuration as follows.

Definition 4.1 ***Initial configuration***
For $\xi_1, \xi_2 \in \mathbb{Z}^d$. Define $\Theta(0)$ by

$$\Theta_v(0) = \begin{cases} (1,1), & v \in \xi_1 \\ (0,0), & v \in \xi_2 \\ (0,1), & \textit{otherwise} \end{cases} , \quad v \in \mathbb{Z}^d$$

where ξ_1, ξ_2 are two finite disjoint and connected subsets of $\mathbb{Z}^d$.

Assume that a Poisson event happens at time t. Vertex v tries to affect vertex w along the common edge. We now define the evolution operator **S**.

Definition 4.2 ***The evolution operator***
Given an infection event at time t where vertex v tries to infect vertex w, let u be a $U[0,1]$ random number and define the evolution operator $\mathbf{S}$ *by the following.*

$$\mathbf{S}(v, w, u, \Theta(t)) = \Theta(t^+)$$

where $\Theta_z(t^+) = \Theta_z(t)$ for any vertex $z \neq w$ and

$$\Theta_w(t^+) = (\ \max(\ a_{w,t},\ \min(a_{v,t}, b_{w,t})\ )\ ,\ \ \min(\ b_{w,t},\ \max(u,\ b_{v,t},\ a_{w,t})\ )\ )$$

When there is a need to emphasise the evolution at a certain vertex without focusing on the origin of the infection we denote this by $\mathbf{S}(\Theta_v(t))$. Next step is to make sure that the threshold order is preserved during evolution, otherwise the state of a vertex may not be unique for all $\lambda \in [0,1]$.

Lemma 4.3 ***The evolution operator preserves threshold order***
If for any vertex $w \in \mathbb{Z}^d$ we have $\Theta_w(t) = (a_{w,t}, b_{w,t})$ such that $a_{w,t} \leq b_{w,t}$ then $\mathbf{S}(\Theta_w(t)) = (a_{w,t^+}, b_{w,t^+})$ will satisfy $a_{w,t^+} \leq b_{w,t^+}$.

Proof :
We must establish that the order of the thresholds at w does not change when applying the evolution operator. Let a Poisson event take place at time t and suppose some vertex v is trying to infect w. Assume $a_{v,t} \leq b_{v,t}$ and $a_{w,t} \leq b_{w,t}$. There are five different cases to analyse.

Case 1: Assume that $a_{v,t} \leq b_{v,t} \leq a_{w,t} \leq b_{w,t}$.
$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) = a_{w,t}$
$b_{w,t^+} = \min(b_{w,t}, \max(u, b_{v,t}, a_{w,t})) \geq a_{w,t}$
$a_{w,t^+} \leq b_{w,t^+}$

Case 2: Assume that $a_{v,t} \leq a_{w,t} \leq b_{v,t} \leq b_{w,t}$.
$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) = a_{w,t}$
$b_{w,t^+} = \min(b_{w,t}, \max(u, b_{v,t}, a_{w,t})) \geq b_{v,t}$
$a_{w,t^+} = a_{v,t} \leq b_{v,t} \leq b_{w,t^+}$

Case 3: Assume that $a_{w,t} \leq a_{v,t} \leq b_{v,t} \leq b_{w,t}$.
$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) = a_{v,t}$
$b_{w,t^+} = \min(b_{w,t}, \max(u, b_{v,t}, a_{w,t})) \geq b_{v,t}$
$a_{w,t^+} = a_{v,t} \leq b_{v,t} \leq b_{w,t^+}$

Case 4: Assume that $a_{w,t} \leq a_{v,t} \leq b_{w,t} \leq b_{v,t}$.
$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) = a_{w,t}$
$b_{w,t^+} = \min(b_{w,t}, \max(u, b_{v,t}, a_{w,t})) = b_{w,t}$
$a_{w,t^+} = a_{w,t} \leq b_{w,t} = b_{w,t^+}$

Case 5: Assume that $a_{w,t} \leq b_{w,t} \leq a_{v,t} \leq b_{v,t}$.
$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) = b_{w,t}$
$b_{w,t^+} = \min(b_{w,t}, \max(u, b_{v,t}, a_{w,t})) = b_{w,t}$
$a_{w,t^+} \leq b_{w,t^+}$

Since we have $a_{w,t^+} \leq b_{w,t^+}$ in all five cases, the result follows.

□

We are now ready to state the relation between the omniparametric and the fixed parameter model.

4.5 Relation to the fixed parameter model

Given a configuration $\Theta(t)$ and some $\lambda \in [0,1]$ we may want to compute the fixed parameter configuration $\Xi^\lambda(t)$. We do this with a projection mapping π_λ, defined as follows.

Definition 4.3 ***Projection mapping***
Let $\pi_\lambda : \Omega_O \to \Omega$ for $\lambda \in [0,1]$ be the projection operator defined by

$$\pi_\lambda(\Theta_v(t)) = \begin{cases} 1, & \lambda \in [0, a_{v,t}] \\ 0, & \lambda \in (a_{v,t}, b_{v,t}] \\ 2, & \lambda \in (b_{v,t}, 1] \end{cases}$$

for each $v \in \mathbb{Z}^d$.

We also need to establish

$$\pi_\lambda(\Theta(t)) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2} \ , \quad t \geq 0$$

which is done in Theorem 4.2. The proof consists of proving two properties of the evolution operator. It must preserve the order of threshold values (see Lemma 4.3), and mimic the dynamics of the two-type Richardson model for any fixed $\lambda \in [0,1]$, that is, the operator has to be consistent with the behaviour of the two-type Richardson model. We establish the second property next.

Lemma 4.4 ***The evolution operator*** S ***is consistent***

Assume that a Poisson event takes place at time t*. Let* $\Theta(t)$ *be a valid omniparametric configuration for all times prior to* t*, that is,*

$$\pi_\lambda(\Theta(s)) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2}$$

for all $s \in [0,t]$ *and all* $\lambda \in [0,1]$*. Then*

$$\pi_\lambda(\ \mathbf{S}(\ \Theta(t)\ )\ ) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2}$$

holds for all $\lambda \in [0,1]$

Proof :
Let a Poisson event take place at time t and let $\Theta(t^+) = \mathbf{S}(\Theta(t))$ be the configuration just after this event. We need to prove that the evolution operator has the correct behaviour in every situation. Let $\lambda \in [0,1]$ be arbitrary but fixed. There are nine cases to analyse.

Case 1: $\pi_\lambda(\Theta_v(t)) = \pi_\lambda(\Theta_w(t)) = 0$, implies $a_{v,t} < \lambda \le b_{v,t}$, $a_{w,t} < \lambda \le b_{w,t}$ and

$$a_{w,t^+} = \max(a_{w,t}, \underbrace{\min(a_{v,t}, b_{w,t})}_{=a_{v,t}}) < \lambda$$

$$b_{w,t^+} = \min(b_{w,t}, \underbrace{\max(\gamma, a_{w,t})}_{\ge\lambda}) \ge \lambda$$

where $\gamma = \max(u, b_{v,t}) \ge \lambda$.
Thus $a_{w,t^+} < \lambda \le b_{w,t^+}$ implying $\pi_\lambda(\mathbf{S}(\Theta_w(t))) = 0$

Case 2: $\pi_\lambda(\Theta_v(t)) = 0,\ \pi_\lambda(\Theta_w(t)) = 1$, implies $a_{v,t} < \lambda \le b_{v,t}$, $\lambda \le a_{w,t}$ and

$$a_{w,t^+} = \max(a_{w,t}, \underbrace{\min(a_{v,t}, b_{w,t})}_{=a_{v,t}}) = a_{w,t} \ge \lambda$$

Thus $a_{w,t^+} \ge \lambda$ gives $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 1$

Case 3: $\pi_\lambda(\Theta_v(t)) = 0,\ \pi_\lambda(\Theta_w(t)) = 2$, implies $a_{v,t} < \lambda \le b_{v,t}$, $b_{w,t} < \lambda$ and

$$b_{w,t^+} = \min(\underbrace{b_{w,t}}_{<\lambda}, \max(\gamma, a_{w,t})) < \lambda$$

Thus $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 2$

Case 4: $\pi_\lambda(\Theta_v(t)) = 1,\ \pi_\lambda(\Theta_w(t)) = 0$, implies $\lambda \le a_{v,t},\ a_{w,t} < \lambda \le b_{w,t}$ and

$$a_{w,t^+} = \max(a_{w,t}, \underbrace{\min(a_{v,t}, b_{w,t})}_{\ge\lambda}) \ge \lambda$$

Thus $a_{w,t^+} \ge \lambda$ gives $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 1$

Case 5: $\pi_\lambda(\Theta_v(t)) = 1,\ \pi_\lambda(\Theta_w(t)) = 1$, implying $a_{v,t} \ge \lambda$ and $a_{w,t} \ge \lambda$.

$$a_{w,t^+} = \max(a_{w,t}, \min(a_{v,t}, b_{w,t})) \ge \lambda$$

Since $a_{w,t^+} \ge \lambda$ we have $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 1$.

Case 6: $\pi_\lambda(\Theta_v(t)) = 1,\ \pi_\lambda(\Theta_w(t)) = 2$, implies $a_{v,t} \ge \lambda,\ b_{w,t} < \lambda$ and

$$b_{w,t^+} = \min(\underbrace{b_{w,t}}_{<\lambda}, \max(\gamma, a_{w,t})) < \lambda$$

Since $b_{w,t^+} < \lambda$ we have $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 2$.

Case 7: $\pi_\lambda(\Theta_v(t)) = 2,\ \pi_\lambda(\Theta_w(t)) = 0$, implies $b_{v,t} < \lambda,\ a_{w,t} < \lambda \le a_{w,t}$ and

$$a_{w,t^+} = \min(a_{w,t}, \underbrace{\min(a_{v,t}, b_{w,t})}_{<\lambda}) < \lambda$$

$$b_{w,t^+} = \min(\underbrace{b_{w,t}}_{\ge\lambda}, \max(u, \underbrace{b_{v,t}}_{<\lambda}, \underbrace{a_{w,t}}_{<\lambda})) \begin{cases} < \lambda, & \text{if } u < \lambda \\ \ge \lambda, & \text{if } u \ge \lambda \end{cases}$$

Thus $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 2$ with probability λ, and $\pi_\lambda(\mathbf{S}(\Theta_w(t^+))) = 0$ with probability $1 - \lambda$.

Case 8: $\pi_\lambda(\Theta_v(t)) = 2,\ \pi_\lambda(\Theta_w(t)) = 1$, implies $b_{v,t} < \lambda,\ \lambda \le a_{w,t}$ and

$$a_{w,t^+} = \max(a_{w,t}, \underbrace{\min(a_{v,t}, b_{w,t})}_{<\lambda}) \ge \lambda$$

Thus vertex w is left unchanged.

Case 9: $\pi_\lambda(\Theta_v(t)) = 2,\ \pi_\lambda(\Theta_w(t)) = 2$, implies $b_{v,t} < \lambda,\ b_{w,t} < \lambda$ and

$$b_{w,t^+} = \min(\underbrace{b_{w,t}}_{<\lambda}, \max(\gamma, a_{w,t})) < \lambda$$

With $b_{w,t^+} < \lambda$ vertex w is left unchanged.

This can be done for any $\lambda \in [0,1]$. Thus the evolution operator is consistent.
□

By applying this evolution operator repeatedly at every Poisson event the model evolves as time passes.

Theorem 4.2
Given an omniparametric sample $\Theta(t)$ we can calculate an ordinary sample by using the projection operator

$$\pi_\lambda(\Theta(t)) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2}$$

for $\lambda \in [0,1]$ and any $t \geq 0$.

Proof:
Let $\Theta(0)$ be any initial configuration and $\lambda \in [0,1]$ fixed. We apply the projection operator on this configuration vertex by vertex.

$$\pi_\lambda(\Theta_v(0)) = \begin{cases} 1, & \lambda \in [0, a_{v,t}] \\ 0, & \lambda \in (a_{v,t}, b_{v,t}] \\ 2, & \lambda \in (b_{v,t}, 1] \end{cases} \quad , \quad v \in \mathbb{Z}^d$$

Since each vertex $v \in \mathbb{Z}^d$ has $\Theta_v(0) \in \{(1,1)\ (0,0)\ (0,1)\}$ according to the definition of the initial configuration it follows that the projection mapping applied to the initial configuration gives a proper configuration in the fixed parameter model.

Even though time is continuous the model evolves only at each Poisson event. Let $\{t_n\}_{n=1}^\infty$ be the sequence of times for all the Poisson events. Assume that $\pi_\lambda(\Theta(t_n)) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2}$. Let t_n^+ be the time just after the next event. Since the evolution operator is consistent according to Lemma 4.4 we automatically have $\pi_\lambda(\Theta(t_n^+)) \stackrel{\mathcal{D}}{=} \mathbb{P}^\lambda_{\xi^1,\xi^2}$. The result now follows by induction.
□

We are finally ready to prove Lemma 4.1 stated on page 30.

Proof of Lemma 4.1:
Fix some arbitrary $n \in \mathbb{Z}^+$ Let $G^\lambda_{i,n}$ be the event that type i survives until reaching ∂B_n, $i = 1,2$. If $\mathbb{P}(G^\lambda_{1,n})$ and $\mathbb{P}(G^\lambda_{2,n})$ are continuous continuity follows for p_n since ...

$$p_n(\lambda) = \mathbb{P}(G^\lambda_{1,n} \cap G^\lambda_{2,n}) = \mathbb{P}(G^\lambda_{1,n}) - \mathbb{P}(G^\lambda_{1,n} \setminus G^\lambda_{2,n}) = \mathbb{P}(G^\lambda_{1,n}) - \mathbb{P}((G^\lambda_{2,n})^c)$$

Fix $\lambda \in (0,1)$ and a small $\varepsilon > 0$.
Is there a $\delta > 0$ such that $|\mathbb{P}(G^{\lambda}_{1,n}) - \mathbb{P}(G^{\lambda+\delta}_{1,n})| < \varepsilon$?

$$\begin{aligned} |\mathbb{P}(G^{\lambda}_{1,n}) - \mathbb{P}(G^{\lambda+\delta}_{1,n})| &= \mathbb{P}(G^{\lambda}_{1,n} \setminus G^{\lambda+\delta}_{1,n}) \\ &\leq \mathbb{P}(\{a_v \in [\lambda, \lambda+\delta) \text{ for some } v \in \partial B_n\}) \end{aligned}$$

where $(a_v, b_v) = \Theta_v(t)$. Since ∂B_n is a finite set this shrinks to zero as $\delta \downarrow 0$. Hence, there is some small enough δ satisfying $|\mathbb{P}(G^{\lambda}_{1,n}) - \mathbb{P}(G^{\lambda+\delta}_{1,n})| < \varepsilon$. The same type of argument using the other threshold value, b_v, yields continuity for $\mathbb{P}(G^{\lambda}_{2,n})$.

□

CHAPTER 5

Simulation experiments

By using the proposed simulation algorithm to generate samples we can study the behaviour of the two-type Richardson model on finite boxes in $\mathbb{Z}^d$. Although this is possible for any finite d the case studied here is $d = 2$. The purpose of this chapter is to study the probability for simultaneous survival on B_n as a function of λ.

According to Theorem 4.1 we have $\mathbb{P}^{1,\lambda}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) = 0$ for all but at most countably many choices of $\lambda \in (0, 1]$. Let $E \subseteq (0, 1)$ be this set of exceptional values of λ. We have excluded 1 from E since it is already known that $\mathbb{P}^{1,1}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) > 0$. We can formulate our knowledge of the probability for simultaneous survival as follows

$$\forall \lambda \in (0, 1) \setminus E : \ \mathbb{P}^{1,\lambda}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2) = 0$$

Using simulations we will try to find some indications of the existence of any value $\lambda_e \in E$, contradicting the general belief that it is empty.

5.1 Simultaneous unbounded growth in simulations

If $E \neq \emptyset$ this should manifest itself if we simulate on large enough boxes. Let $p_n(\lambda) = \mathbb{P}^{1,\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n})$ and let $p(\lambda)$ denote the probability for simultaneous survival on $\mathbb{Z}^2$. Then $p_n(\lambda) \to p(\lambda)$ pointwise as $n \to \infty$.

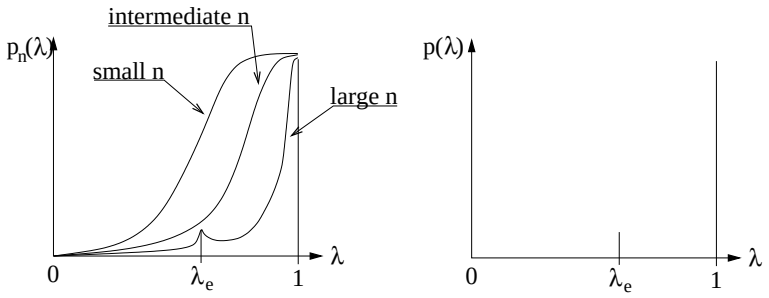

Figure 5.1: *The functions $p_n(\lambda)$ and $p(\lambda)$ when there is an exceptional value λ_e.*

In the absence of exceptional values $p(\lambda)$ will equal

$$p(\lambda) = \begin{cases} c, & \lambda = 1 \\ 0, & \lambda < 1 \end{cases} \quad , \quad \lambda \in (0,1)$$

for some $c > 0$. Assume for the moment that E contains at least on element, and pick $\lambda_e \in E$. Whenever $\lambda = \lambda_e$ there is a positive probability for simultaneous survival and the function $p(\lambda)$ will not be increasing any more. Since the p_n's converge pointwise to p they will not be increasing either if n is large enough. We know from Theorem 4.1 that the E is at most countable so an exceptional value λ_e will have non-exceptional values to the right of it, and manifest itself as in Figure 5.1.

We use simulations to estimate functions p_n, and do this for different values of n. If we find that an estimate $\hat{p}_n$ which is non-increasing it indicates non-increasing p_n, and also existence of exceptional values. If p_n is increasing for all large n then so is $p(\lambda)$ which rules out the existence of exceptional values. The question is now: When is n large? Since we do not know simulations have been carried out for n as large as possible. Later we will use the simulations to estimate how large boxes we need to detect exceptional probabilities bounded from below by some small constant.

5.2 The simulation algorithm

The simulation algorithm follows the omniparametric two-type Richardson model as described in Section 4.4, see Figure 5.2.

```
1: graph G
2: G.setInitialConfig()
3: repeat
4:    u = G.randomVertex()
5:    v = G.randomNeighbour(u)
6:    u.propagateTo(v)
7: until G.calcStopCriteria() ≡ TRUE
```

Figure 5.2: A simulation scheme for the omniparametric two-type Richardson model.

Initial configuration

The smallest possible initial configuration is used. Vertices $\mathbf{0} = (0, 0)$ and $\mathbf{1} = (1, 0)$ are assigned thresholds $(a_{\mathbf{0},0}, b_{\mathbf{0},0}) = (1, 1)$ and $(a_{\mathbf{1},0}, b_{\mathbf{1},0}) = (0, 0)$ respectively. All other vertices $x \in B_n$ are assigned thresholds $(a_{x,0}, b_{x,0}) = (0, 1)$. For every $\lambda \in (0, 1]$ this gives the origin type one, its left neighbour type two and the rest of B_n type zero.

Stop criteria

The algorithm is executed until the first vertex on the boundary ∂B_n is affected. The model is only defined on the box B_n so we have to abort at this stage. If we continue after this the model will not evolve as the two-type Richardson model prescribe.

Simulation setup

We perform simulations for n in

$$n \in \{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 750, 1250, 2500\}$$

and for all but the largest box 1000 simulations are done. Since large boxes take a long time to simulate, we were forced to limit the number of simulations for $n = 2500$ to 100.

After running all simulations the functions p_n are estimated pointwise in the following manner. Each simulation generates threshold values for every vertex. Let T_i be the set of threshold values for a simulation i, and assume we have done m simulations. Then $\mathbf{T} = \cup_{i=1}^{m} T_i$ is the set of values of λ where some vertex changes in the model. Order these

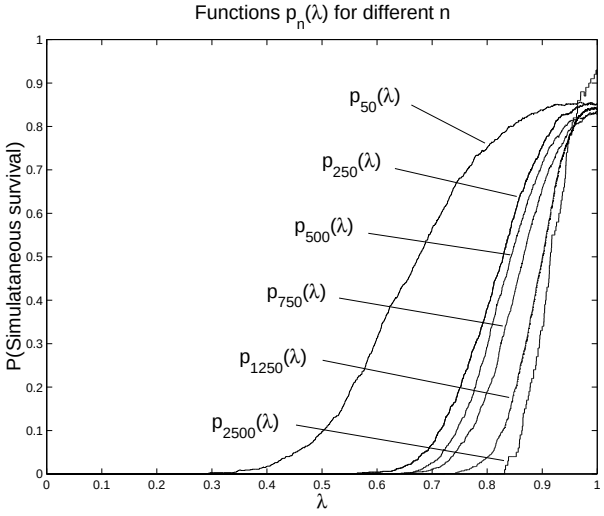

Figure 5.3: *The estimates of the functions $p_n(\lambda)$ for some values of n.*

vales according size.

$$\mathbf{T} = \{\lambda_{(1)}, \lambda_{(2)}, ..., \lambda_{(|T|)}\}$$

For each $\lambda_{(i)} \in \mathbf{T}$ we estimate $p_n(\lambda_{(i)})$ in the natural way.

5.3 Results

In Figure 5.3 we see the estimated functions $\hat{p_n}(\lambda)$ for some of the simulated boxes.

As expected the probability for simultaneous survival (for any fixed λ) is decreasing as the box size increases, even though function $\hat{p}_{2500}(\lambda)$ breaks this monotone pattern. The more erratic behaviour of $\hat{p}_{2500}$ is most likely due to the smaller number of samples. There is no indication of non-monotonicity like in Figure 5.1 in any of the estimated functions. An open question is if the simulated boxes are large enough. If there exists any exceptional value $\lambda_e \approx 1$ with $\mathbb{P}^{1,\lambda_e}_{\mathbf{0},\mathbf{1}}(G^{\lambda_e}_{1,n} \cap G^{\lambda_e}_{2,n}) \approx 0$ we would probably need extremely large boxes to see this. No matter how large boxes we simulate there will however always be possible to have

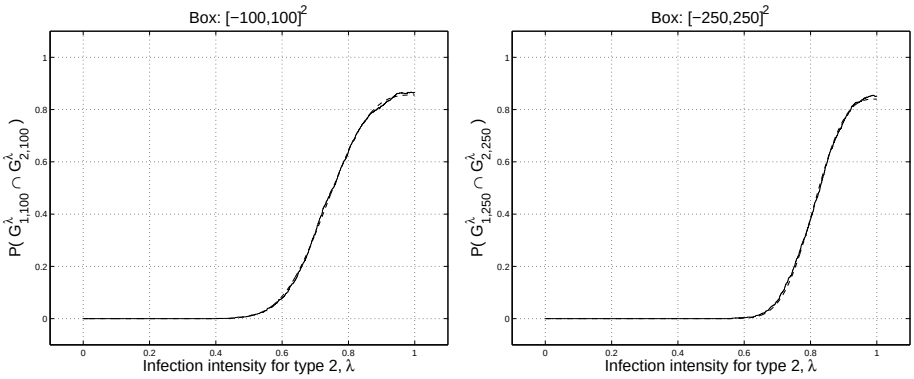

Figure 5.4: *Estimated and fitted function for boxes B_{100} and B_{250}.*

a sufficiently small exceptional probability at an exceptional value close enough to 1.

5.4 Curve fitting

Let us fit a parametric curve to the estimated functions $\hat{p}_n(\lambda)$, and see if we can relate the parameters to box size. There are several reasons for this analysis. First we can use the fitted curves and see how they behave when $n \to \infty$. As we shall see we get indications on expected behaviour, that is, no exceptional values. We may also extract information on how large boxes we need to detect λ_e with small positive exception probability, giving us a clue which exceptional values we may rule out and which we can not say anything about.

When looking at the estimated curves it seems appropriate to fit some s-shaped curve to data. We try with the following.

$$f_n(\lambda) = \alpha_n e^{-\beta_n(1-\lambda)^3}$$

It is tempting to replace 3 in the exponent by a third parameter γ_n to get a closer fit. This does not, however, decrease the difference between f_n and $\hat{p}_n$ enough to motivate such an extension. We now have two parameters α_n and β_n to estimate from the function $\hat{p}_n$. The first parameter α_n is the probability for simultaneous survival when the infections are equally strong , the second one is a shape parameter determining how

steep the curve is.

There is no hope of course for this curve to be correct. Theoretically we have $p_n(0) = \mathbb{P}^{1,0}_{\mathbf{0},\mathbf{1}}(G^0_{1,n} \cap G^0_{2,n}) = 0$ since type two does not infect at all with zero infection intensity, but $\alpha_n e^{-\beta_n} > 0$ for all plausible values of the estimated parameters. Even if there is no theoretical support for this model it is interesting to see how closely we can approximate the estimated functions. Curve fitting can also illustrate how $p_n(\lambda)$ behave for $\lambda \approx 1$ as $n \to \infty$, which is important when studying how fast the stronger type strangle the weaker. With these remarks in mind we proceed. In figures 5.4 and 5.5, we see the result of the procedure.

Note that f_n does not fit the estimated functions well whenever λ is small enough or close to 1. But there is a region, I^1_n (see Figure 5.6), when $\hat{p}_n(\lambda)$ increases the most where $f_n(\lambda)$ fits quite well and perhaps Perhaps $f_n(\lambda)$ captures the behaviour of $\hat{p}_n$ in this region? Above this interval (I^2_n in Figure 5.6) simulations indicate $\hat{p}_n(\lambda) \leq f_n(\lambda)$.

On all boxes the probability for simultaneous survival does not change much when λ is close to 1, but decreases quite rapidly (at least for larger boxes) when λ is below some threshold value $\lambda_{t,n}$. In Figure 5.6 we have $\lambda_{t,n}$ somewhere in the middle of the interval I^2_n. In the simulated cases this threshold $\lambda_{t,n}$ approaches 1 as n gets larger. Perhaps this behaviour is the same for all finite boxes and $\lambda_{t,n} \to 1$ as $n \to \infty$. In that case the behaviour is consistent with the belief that the exceptional set E is empty.

Dependence between parameter and box size

In trying to relate the parameters to box size we plot these as functions of n. In Figure 5.7 there appears to be no strong correlation between n and α_n. Theory indicates that α_n should be decreasing as n gets larger, approaching the limit value $\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G_1 \cap G_2)$ as $n \to \infty$. It's hard to say anything about how α_n approaches its limit values, the simulated data is too spread out and the limit value unknown. Considering Figure 5.7 it seems reasonable however to conclude that $\lim_{n\to\infty} \alpha_n \in [0.80, 0.82]$. The dependence of α_n on n should not be linear since we can not have negative values for α_n. Perhaps some exponential function is more appropriate to model the decay.

The situation for parameter β is different. This parameter models the shape of the function. In Figure 5.7 we see that there is quite strong linear dependence between n and β_n, suggesting $\beta_n \approx 0.40n$. As the finite boxes grows to infinity so does β_n, resulting in the following limit

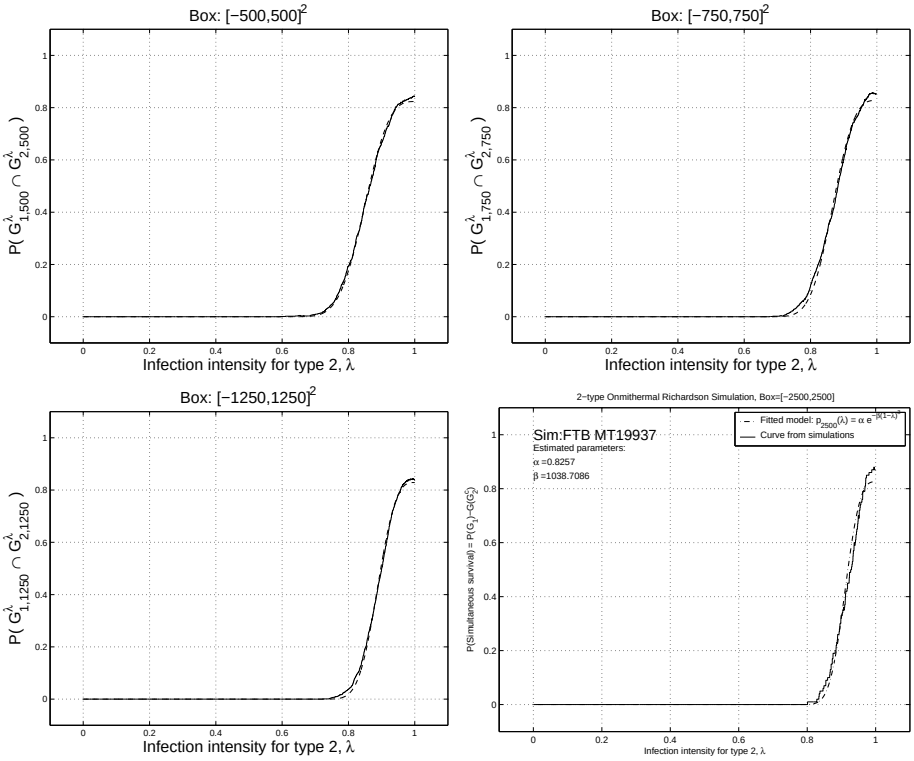

Figure 5.5: *Estimated and fitted function for boxes $\mathbb{B}_{500}$ (upper left), $\mathbb{B}_{750}$ (upper right), $\mathbb{B}_{1250}$ (lower left) and $\mathbb{B}_{2500}$ (lower right). Simulated functions are drawn with solid lines and fitted functions using dashed lines.*

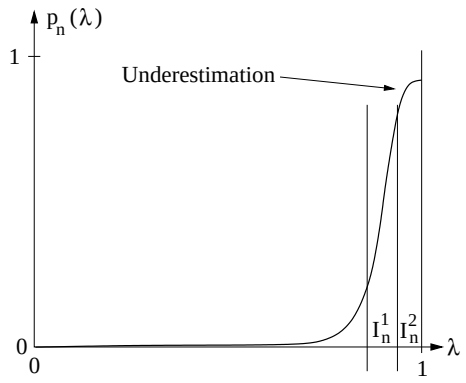

Figure 5.6: *In the interval I_n^1 where $f_n(\lambda)$ closely approximates $\hat{p}_n(\lambda)$. In interval I_n^2 we have $p_n(\lambda) \leq f_n(\lambda)$. We can use the underestimation of the curve in I_n^2 to get a upper bound for the rate of convergence at which $p_n \to p$.*

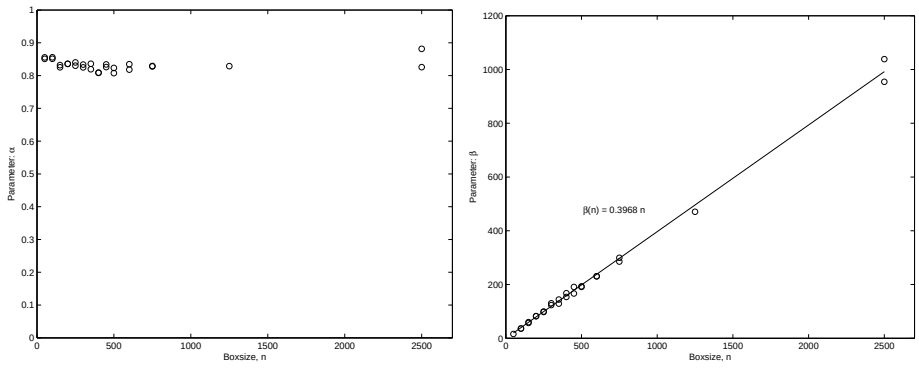

Figure 5.7: *Parameters α_n and β_n for different values of the box size n. Remember the larger uncertainty for the points corresponding to simulations on the largest box, B_{2500}, where we used 100 simulations instead om 1000.*

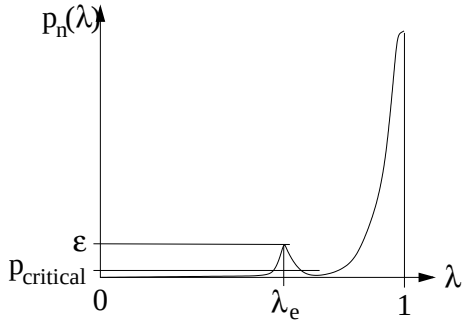

Figure 5.8: *Assuming the existence of* $\lambda_e \in (0, a]$ *such that* $\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda_e}_{1,n} \cap G^{\lambda_e}_{2,n}) \geq \varepsilon$ *we reject the null hypotheses in favour of the alternative if* $\max\{\hat{p}_n(\lambda) : \lambda \in (0, a]\} \leq p_{critical}$.

function.

$$p_\infty(\lambda) = \lim_{n\to\infty} \alpha_n e^{-cn(1-\lambda)^3} = \begin{cases} \alpha, & \lambda = 1 \\ 0, & \lambda < 1 \end{cases}, \quad c = 0.40$$

where $\alpha = \lim_{n\to\infty} \alpha_n$. This is also consistent with the belief that E is empty. If this captures the dependence between n and β_n we can see how quickly the functions f_n converges to $\delta(1-\lambda)$, and get an indication of how the probability of simultaneous survival behave on finite boxes.

5.5 Statistical analysis

A test: Are there exceptional values in $(0, 0.7]$ **?**

All simulation results indicate that the set E of exceptional values for λ is empty, but how strong is the evidence against $E \neq \emptyset$. We fix $\varepsilon > 0$ and test the hypothesis that there exists some $\lambda_e \in (0, a],\ a < 1$, such that $\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) \geq \varepsilon$ against the alternative that there is no such exceptional values.

We perform the test for the largest box where we have $m = 1000$ simulations, that is B_{1250}. Figure 5.5 indicates that $E \cap (0, 0.7] = \emptyset$, so we use $a = 0.7$. (Here we make the cardinal sin of looking at data before deciding on the null hypothesis. We feel that this does not have very severe consequences in this case, but the critical reader is invited to try

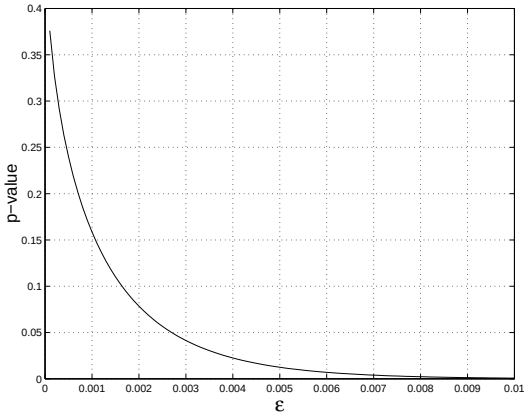

Figure 5.9: *p-values for the test expressed as a function of $p(\lambda_e) = \varepsilon$.*

other values of a.) We now state the hypotheses

$$\begin{aligned} H_0: &\quad \exists \lambda_e \in [0,a],\ \mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda_e}_{1,n} \cap G^{\lambda_e}_{2,n}) \geq \varepsilon \\ H_1: &\quad \forall \lambda \in [0,a],\ \mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) < \varepsilon \end{aligned}$$

Under the null hypothesis the Binomial distribution and the central limit theorem give us a statistic

$$Z = \frac{\hat{p}(\lambda_e) - p(\lambda_e)}{\sqrt{p(\lambda_e)(1-p(\lambda_e))/m}} \overset{\mathcal{D}}{\approx} N(0,1)$$

provided that we have a candidate for λ_e (which we do not). Suppose $p_n(\lambda_e) = \varepsilon$. This suggests that we reject H_0 in favour of H_1 at significance level α if

$$\hat{p}_n(\lambda_e) \leq p_{\text{critical}} = \varepsilon - z_\alpha \sqrt{\frac{\varepsilon(1-\varepsilon)}{m}}$$

We do not however know the value λ_e so we choose a conservative approach and reject H_0 whenever $p_{\max} \leq p_{\text{critical}}$, where

$$p_{\max} = \max\{\hat{p}_n(\lambda) : \lambda \in (0,a]\}$$

As a consequence of the choice of a we have $p_{\max} = 0$, and we can easily calculate the p-values of the test as a function of ε, see Figure 5.9. With

$\alpha = 0.05$ we can reject for $\varepsilon \geq 0.0027$ and if we want $\alpha = 0.01$ we can reject whenever $\varepsilon \geq 0.0054$.

Given simulations on larger boxes we can test the null hypothesis for smaller and smaller values for ε and also do this for larger and larger subsets of $[0, 1]$.

The case $\mathbb{P}(\mathbf{G}^{\lambda}_{\mathbf{1},\mathbf{n}} \cap \mathbf{G}^{\lambda}_{\mathbf{2},\mathbf{n}}) \geq \mathbf{1/2}$

We consider the case when the probability of having at simultaneous survival is at least $1/2$ and see how large λ must be if we let the box size vary. From Figure 5.7 it seems reasonable to assume that

$$\lim_{n\to\infty} \alpha_n \in [0.8, 0.82]$$

We fix $\alpha_n = 0.8$ for all n consider λ such that $\mathbb{P}^{\lambda}_{\xi^1,\xi^2}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) = 1/2$ and let $\lambda_{1/2}(n)$ denote this value, then

$$p_n(\lambda_{1/2}(n)) = \alpha_n e^{-cn(1-\lambda_{1/2}(n))^3} = \frac{1}{2} \quad \Leftrightarrow \quad \lambda_{1/2}(n) = 1 - \left(\frac{\ln(2\alpha_n)}{cn}\right)^{1/3}$$

where $c = 0.40$ (as before). If the probability for simultaneous survival is monotone this is a threshold value for λ above which there is at least probability $1/2$ for simultaneous survival, see Figure 5.10. The theoretical limit of $\lambda_{1/2}$ as $n \to \infty$ is of course 1. It is interesting to see how slowly $\lambda_{1/2}$ converges to its limiting value.

This is an indication of a small region for λ close to 1, where infections are almost equally strong, and in which it takes a very long time for the stronger infection to strangle the weaker.

A conjecture: Lower bound for $\mathbb{P}(\mathbf{G}^{\lambda}_{\mathbf{1},\mathbf{n}} \cap \mathbf{G}^{\lambda}_{\mathbf{2},\mathbf{n}})$ when $\lambda \approx 1$

We will now see what conclusions can be drawn from simulations regarding the behaviour of $\mathbb{P}^{\lambda}_{\xi^1,\xi^2}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n})$ when $\lambda \approx 1$. As claimed before there is a small interval $[1-\delta, 1]$ where it seems as if $\hat{p}_n(\lambda) \geq f_n(\lambda)$, see figures 5.4 and 5.5. Let us formulate this as a conjecture.

Conjecture 5.1
There exists a $\delta \in (0, 1)$ such that

$$\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) \geq \alpha e^{-cn(1-\lambda)^3} \ , \quad c = 0.40 \ , \quad \alpha = \lim_{n\to\infty} \alpha_n$$

whenever $\lambda \in [1 - \delta, 1]$.

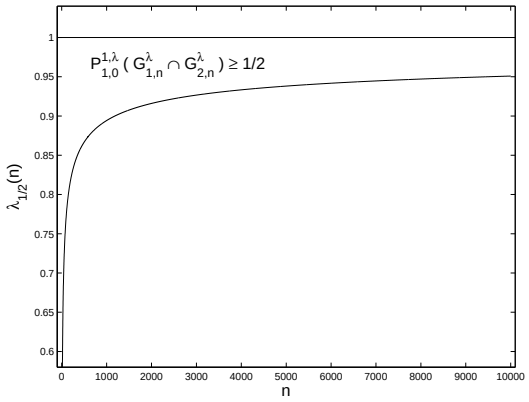

Figure 5.10: *The threshold value $\lambda_{1/2}(n)$ as a function n, illustrating how the region for λ in $[0,1]$ where we have a reasonable probability for simultaneous survival decreases as n grows.*

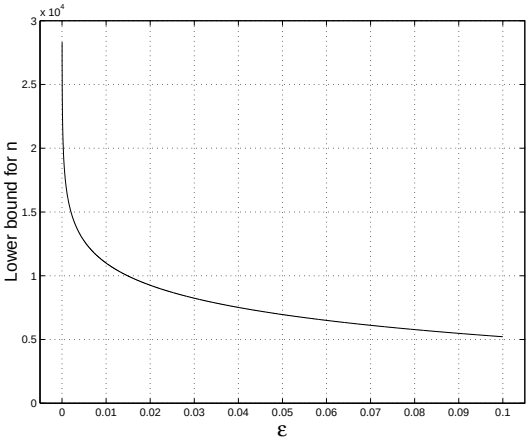

Figure 5.11: *Given a lower bound on the exceptional probability ε for a certain λ_e we can use the conjecture to get a lower bound on the box size we have to use in simulations. In this case $\lambda_e \approx 0.999$ while ε varies over $[0, 0.1]$.*

If true, the conjecture sets an upper bound on the rate at which

$$\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n}) \to \mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1} \cap G^{\lambda}_{2})$$

and also give us an opportunity to determine how large boxes we must use to detect $\lambda_e \in E \cap [1-\delta, 1]$ such that $\mathbb{P}^{\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1} \cap G^{\lambda}_{2}) \geq \varepsilon$ for any $\varepsilon > 0$. For example to detect $\lambda_e \approx 0.999$ with exceptional probability $\varepsilon \leq 0.1$, we rewrite

$$n \geq \frac{\ln\ \alpha - \ln \mathbb{P}^{1,\lambda}_{\mathbf{0},\mathbf{1}}(G^{\lambda}_{1,n} \cap G^{\lambda}_{2,n})}{n(1-\lambda)^3}$$

and see how the lower bound on n behaves as a function of ε, see Figure 5.11.

Part III

Estimation for the Ising and Potts models

CHAPTER 6

Introduction

We now turn our focus to parameter estimation in the Ising and Potts models. The situation is the following. Suppose we have, not necessarily complete, data from some finite region of $\mathbb{Z}^2$. We make the fundamental assumption that data follows either the Ising or q-state Potts model without any external field. The remaining, unknown, parameter in both cases is the inverse temperature β. We explore omniparametric simulation as a tool for make sensible estimates of β. The rest of Part III is divided into seven chapters.

Chapter 7 introduces necessary notation and results for the models we use, i.e. the random cluster, Ising and Potts models. We finish the chapter with some historical notes.

Chapter 8 introduces omniparametric versions of the random cluster and Potts models. As a special case of the Potts model the Ising model is not treated. We elaborate some on the definition of the omniparametric random cluster model since it is the base for our procedures.

Chapter 9 contains further definition of models used to model partially observed data. By introducing a new process on the lattice, independent of the observed data it is possible to model partially observed as well as degraded data. We use a simple degradation process to model partially observed data, and in turn manage to establish central limit theorems, which to our knowledge is a new result.

Chapter 10 focuses on using simulations to explore introduced models. In particular we study characteristics (such as susceptibility) we need later, trying to estimate quantities for which there are no known closed form. In the parameter estimation procedure we will regard these quantities as known at their estimated values.

Chapters 11 and 12 treat parameter estimation. In Chapter 11 we develop procedures, prove necessary results and present simulations for estimating the inverse temperature in the Ising model. Chapter 12 repeat the agenda of Chapter 11 for the q-state Potts model. We estimate the inverse temperature in the case of zero external field and assume q to be known in advance.

CHAPTER 7

Three models in statistical mechanics

In this chapter we present some models from the field of statistical mechanics. Most of them can be defined on arbitrary graphs, but here we will restrict ourselves to the d-dimensional lattice in Euclidean space. We let $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ be the usual graph on $\mathbb{Z}^d$ with edges between vertices at unit distance. We are primarily interested in these models living on finite boxes in $\mathbb{Z}^d$, but also treat the infinite case.

Let $\mathbb{B}_n^d$ be a finite subgraph of $\mathbb{L}^d$ with vertex set $\mathbb{Z}_n^d = \mathbb{Z}^d \cap [0, n-1]^d$ and edge set $\mathbb{E}_n^d = \{\langle x, y\rangle \in \mathbb{E}^d : x, y \in \mathbb{Z}_n^d\}$. The boundary of $\mathbb{B}_n$ is denoted $\partial\mathbb{B}_n^d$ and defined as $\partial\mathbb{B}_n^d = \mathbb{B}_{n+1}^d \setminus \mathbb{B}_n^d$. When the number of dimensions are understood we suppress d in the notation, writing $\mathbb{B}_n,\ \partial\mathbb{B}_n,\ \mathbb{Z}_n$ and $\mathbb{E}_n$.

The rest of the chapter is outlined as follows. We start by introducing general concepts relating to stochastic domination on finite boxes. Here we assume that the model lives on the vertex set but we could as easily present the material as living on the edge set. We then present the random cluster, Ising and Potts models on finite boxes and when required also on the entire infinite lattice. For a more thorough treatment of the subject see [GHM01] and references therein.

7.1 General properties

Let $n \geq 1$ and $d \geq 2$ be arbitrary but fixed and let S be a linearly ordered finite state space. A configuration is an assignment of values in S to each vertex in the d-dimensional box $\mathbb{B}_n$. Given a configuration $X \in S^{\mathbb{B}_n}$ we let $X^{\{v\}}$ denote the corresponding configuration in $S^{\mathbb{B}_n \setminus \{v\}}$.

Definition 7.1 ***Irreducibility***
Fix a probability measure ν on $S^{\mathbb{B}_n}$. Let $S^{\mathbb{B}_n}_{pos}$ be the set of elements in $S^{\mathbb{B}_n}$ having positive ν-probability. We say that ν is irreducible if given any two $X, X' \in S^{\mathbb{B}_n}_{pos}$ we can make vertexwise transformations to X, changing it one vertex at a time into X', such that all intermediate elements are in $S^{\mathbb{B}_n}_{pos}$.

Definition 7.2 ***A partial order***
We define a partial order $\preceq$ on the set $S^{\mathbb{B}_n}$ by the following. For $X, X' \in S^{\mathbb{B}_n}$ we say that X' is larger than X and write $X \preceq X'$ if

$$X(v) \leq X'(v)$$

holds for every vertex $v \in \mathbb{B}_n$.

Definition 7.3 ***Monotonicity***
We say that ν is monotone, if given any $v \in \mathbb{B}_n,\ s \in S$ and $X, Y \in S^{\mathbb{B}_n}$ such that $X^{\{v\}} \preceq Y^{\{v\}}$, we have

$$\nu(X(v) \geq s | X^{\{v\}}) \leq \nu(X(v) \geq s | Y^{\{v\}})$$

whenever $X^{\{v\}}$ and $Y^{\{v\}}$ both have positive probability.

Monotonicity means increased probability for an event when we increase its surroundings according to the partial order $\preceq$.

Let $B_I(S^{\mathbb{B}_n}, \mathbb{R})$ be the set of all increasing (with respect to $\preceq$) real function on the space $S^{\mathbb{B}_n}$.

Definition 7.4 ***Positive correlations***
We say that ν have positive correlations if

$$\forall f, g \in B_I(S^{\mathbb{B}_n}, \mathbb{R}) : \int fg \, d\nu \geq \int f \, d\nu \int g \, d\nu$$

Definition 7.5 ***Stochastic domination***
Let ν_1 and ν_2 be two probability measures on $\mathbb{R}$. We say that ν_2 stochastically dominates ν_1, writing $\nu_1 \preceq_{\mathcal{D}} \nu_2$ whenever

$$\forall f \in B_I(S^{\mathbb{B}_n}, \mathbb{R}) : \int f \, d\nu_1 \leq \int f \, d\nu_2$$

holds.

We can use the following theorem to conclude when we have positive correlations by checking if the measure under investigation in monotone.

Theorem 7.1 ***The FKG inequality***
Let ν be irreducible measure assigning positive probability to the maximal element of $S^{\mathbb{B}_n}$ according to $\preceq$. If ν is monotone it also has positive correlations.

This was proved in [FKG71] , see [GHM01] for an introductory treatment. A FKG system is a system of random variables satisfying the FKG inequality [FKG71] . Many models in statistical mechanics have this property, the Ising and random cluster (for certain regimes of the parameter space) models are among them.

We also present a theorem by Holley [Hol74] stating sufficient conditions for stochastic domination. For a proof see the original article [Hol74] or [GHM01] .

Theorem 7.2 ***Holley's theorem***
Consider probability measures ν_1 and ν_2 defined on state space $S^{\mathbb{B}_n}$. Assume that ν_2 is irreducible and assigns positive probability to the maximal element (according to $\preceq$) of $S^{\mathbb{B}_n}$. Suppose further that

$$\nu_1(X(v) \geq a | X = \xi \text{ off } v) \leq \nu_2(X(v) \geq a | X = \eta \text{ off } v)$$

for any $v \in \mathbb{Z}_n,\ a \in S$, and $\xi, \eta \in S^{\mathbb{B}_n}$ such that $\xi \preceq \eta,\ \nu_1(X = \xi \text{ off } v) > 0$ and $\nu_2(X = \eta \text{ off } v) > 0$. Then $\nu_1 \preceq_{\mathcal{D}} \nu_2$.

Finally we introduce translation invariance for measures on $S^{\mathbb{Z}^d}$.

Definition 7.6 ***Translation invariance***
Let ν be a measure on the infinite state space $S^{\mathbb{Z}^d}$. We say that ν is translation invariant if

$$\nu(\ (X_{v_1}, ..., X_{v_n}) \in A\ ) = \nu(\ (X_{v_1+u}, ..., X_{v_n+u}) \in A\ )$$

holds for any $n \geq 1$, any vertices $v_1, ..., v_n \in \mathbb{Z}^d$, any $u \in \mathbb{Z}^d$ and for all $A \subseteq S^{\{1,...,n\}}$.

Translation invariance means that we can "move around" (but not rotate) events without changing their probability. One problem where this property is important is the study of infinite clusters. If the measure under observation is translation invariant every vertex will have the same probability of being in an infinite cluster (as in Chapter 2), so studying this probability for the origin is enough.

Long range order or not, a note on phase transition

On an infinite graph we have phenomenons not present on finite graphs, for example phase transition. A phase transition manifests itself in different ways, for some models the transition occur by the development of an infinite component and for some models we have non-uniqueness of measures. Regardless of the system under study a change of phase does change the probabilistic behaviour of the system.

All models treated in this chapter are connected to the random cluster model where phase transition manifests itself by the creation of an infinite connected component. In the subcritical phase all connected components are finite and therefore connection probabilities decay towards zero as the distance between vertices increase to infinity. The supercritical phase is characterized by a connection probability bounded away from zero regardless of the distance between vertices due to the existence of an infinite component. We some times call the subcritical phase the unordered phase, and the supercritical phase the ordered phase to emphasize the long range order introduced by infinite components.

For the Ising and Potts models phase transition is characterized by the existence of multiple measures in the supercritical phase, while the subcritical phase is governed by a unique probability measure.

7.2 The random cluster model

The random cluster model is a two parameter family of models, living on the edge set of the graph $\mathbb{L}^d$. We first define the random cluster distribution on finite boxes $\mathbb{B}_n \subset \mathbb{L}^d$ and then proceed with random cluster measures on the whole $\mathbb{L}^d$. We make a distinction between the finite and infinite cases, writing probability *distributions* on $\{0,1\}^{\mathbb{B}_n}$ and probability *measures* on $\{0,1\}^{\mathbb{L}^d}$. This section is mostly based on [GHM01] .

There are two different approaches to the construction of random cluster measures and we treat both since they give rise to different sets of measures. The careful treatment allow us see what kind of measures we might encounter on infinite graphs and if they give rise to unexpected phenomenons on finite boxes.

We treat the random cluster model in some detail since it is the main tool for analysing the correlation structure of the Ising and Potts models. Establishing uniqueness of measure and determining a subset of the unit interval for p where we have exponential decay of the cluster size become especially important tasks.

Definition 7.7 ***The random cluster distribution***
Given $\mathbb{B}_n$ the random cluster distribution on $\mathbb{B}_n$ is a two parameter family of distributions $\phi^{(n)}_{p,q},\ p \in [0,1],\ q \in [0,\infty)$, on the space $\{0,1\}^{\mathbb{E}_n}$. An element $\gamma \in \{0,1\}^{\mathbb{E}_n}$ is called a configuration, and we denote its value at a single edge e by $\gamma(e)$. The subgraph $\mathbb{B}_n(\gamma) = (\mathbb{Z}_n, \mathbb{E}_n(\gamma))$,

$$\mathbb{E}_n(\gamma) = \{e \in \mathbb{E}_n : \gamma(e) = 1\}$$

induced by a configuration γ is called the random cluster graph. A configuration $\gamma \in \{0,1\}^{\mathbb{E}_n}$ is assigned probability

$$\phi^{(n)}_{p,q}(\gamma) = \frac{q^{\kappa(\gamma)}}{Z_{p,q}} \prod_{e \in \mathbb{E}_n} p^{\gamma(e)}(1-p)^{1-\gamma(e)}$$

where $Z_{p,q}$ is a normalizing constant, and $\kappa(\gamma)$ is the number of connected components in $\mathbb{B}_n(\gamma)$.

Next we establish the single-edge conditional probabilities. The proof follows immediately from definitions.

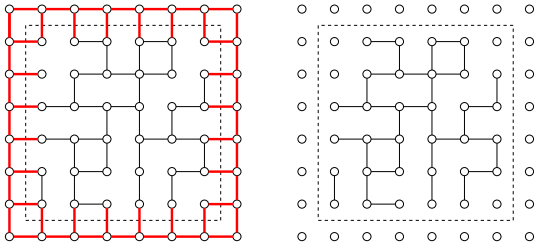

Figure 7.1: *An example on boundary conditions for a small box (within dashed lines) in the random cluster model. The right diagram contains all edges outside the box, the so called wired boundary condition. To the left we see the other extreme containing no edges outside of the box, the so called free boundary condition.*

Lemma 7.1 ***Single-edge conditional probabilities***
Let X be distributed according to the random cluster distribution on the graph $\mathbb{B}_n$ with parameters p and q. Let e be an arbitrary edge and let γ be an arbitrary configuration on $\mathbb{E}_n \setminus \{e\}$ having positive probability. Then the following holds.

$$\phi_{p,q}^{(n)}(X(e) = 1 | X^{\{e\}} = \gamma) = \begin{cases} p, & \text{if } \{x \leftrightarrow y \text{ in } X^{\{e\}}\} \\ \dfrac{p}{p + q(1-p)}, & \text{otherwise} \end{cases}$$

When defining the random cluster distribution on finite boxes in $\mathbb{Z}^d$ some boundary condition is necessary. Here we used of the free boundary condition, meaning that no edges are present outside the box $\mathbb{B}_n$, in contrast to the wired boundary condition where all edges outside $\mathbb{B}_n$ are present (see Figure 7.1).

The parameter p can be interpreted as the probability of assigning an edge the value 1. This is only true whenever $q = 1$, when the random cluster model is equivalent to the independent edge percolation model, any other value of q introduces dependence between edges. Whenever $q < 1$ the distribution favors configurations with a small number of clusters, while $q > 1$ makes it favor configurations with a large number of clusters. We will restrict ourselves to the case $q \geq 1$, since that is precisely the regime where the FKG inequality holds, and we will be

able to simulate the random cluster model efficiently for all $p \in [0,1]$ simultaneously.

Next we establish two facts about the random cluster model, first monotonicity and then positive correlations .

Theorem 7.3 ***Monotonicity***
The random cluster distribution $\phi_{p,q}^{(n)}$ is monotone.

Proof:
Let $e \in E$ be any edge and let σ_1, σ_2 be two configurations on the graph $G^{\{e\}} = (V, E \setminus \{e\})$ such that $\sigma_1 \preceq \sigma_2$. We have $\phi_{p,q}^{(n)}(X(e) \geq 0 | X^{\{e\}} = \cdot) = 1$ regardless of what we condition on, so we only need to consider the case $\phi_{p,q}^{(n)}(X(e) = 1 | X^{\{e\}} = \cdot)$. For $e = \langle x, y \rangle$ let $\{x \leftrightarrow y \text{ in } \sigma\}$ denote the event that there is a path of open edges between x and y in configuration σ, and let $\{x \not\leftrightarrow y \text{ in } \sigma\}$ denote its complement. There are three cases,

$$(i) \quad \{x \not\leftrightarrow y | \sigma_1\} \cap \{x \not\leftrightarrow y | \sigma_2\}$$

$$(ii) \quad \{x \not\leftrightarrow y | \sigma_1\} \cap \{x \leftrightarrow y | \sigma_2\}$$

$$(iii) \quad \{x \leftrightarrow y | \sigma_1\} \cap \{x \leftrightarrow y | \sigma_2\}$$

and in all three the result is a consequence of Lemma 7.1.
□

It is now an easy task to establish positive correlations.

Theorem 7.4 ***Positive correlations***
The random cluster distribution $\phi_{p,q}^{(n)}$ on $\mathbb{B}_n$ has positive correlations for any $q \geq 1$ and any $p \in [0,1]$.

Proof:
If $\phi_{p,q}^{(n)}$ is irreducible and monotone the result follows from Theorem 7.1. Since the underlying graph is finite all random cluster configurations has positive probability and the corresponding measure is irreducible. Monotonicity is established in Theorem 7.3 and we are done.
□

Next we establish stochastic ordering for the random cluster model. The result is partly taken from [GHM01, Corollary 6.7] . For a proof of statement $(i), (ii)$ and (iii) see [GHM01] . Statement (iv) is a consequence of Holley's theorem (page 61) and Lemma 7.1.

Theorem 7.5 ***Stochastic domination***
For the graph $\mathbb{B}_n$, $p, p' \in [0,1]$, $q, q' \geq 1$ *such that* $p \leq p'$ *and* $q \leq q'$ *we have the following for the random cluster distribution.*

(i) $\quad \phi^{(n)}_{p,q} \preceq_{\mathcal{D}} \phi^{(n)}_{p,1}$

(ii) $\quad \phi^{(n)}_{p,q} \succeq_{\mathcal{D}} \phi^{(n)}_{p,p(p+(1-p)q)^{-1}}$

(iii) $\quad \phi^{(n)}_{p,q} \preceq_{\mathcal{D}} \phi^{(n)}_{p',q}$

(iv) $\quad \phi^{(n)}_{p,q'} \preceq_{\mathcal{D}} \phi^{(n)}_{p',q}$

Infinite volume measures

On $\mathbb{L}^d$ there are two different methods for constructing the random cluster measure, both use the definition of random cluster distributions defined on finite boxes.

The DLR method, after Dobrushin, Lanford and Ruelle [Dob68, LR69] defines a measure as of random cluster type if it follows the random cluster distribution on finite regions Λ given the configuration on $\mathbb{Z}^d \setminus \Lambda$, i.e., if its conditional distribution on finite sets agrees with those arising in the finite graph random cluster model. Let $\mathcal{R}_{p,q}$ be the set of all such measures for parameters p and q.

The weak limit construction uses a sequence of finite regions $\{\Lambda_n\}_{n=1}^{\infty}$ such that $\Lambda_n \to \mathbb{Z}^d$ as $n \to \infty$. Without loss of generality we can think of the Λ_n's as finite boxes. The random cluster measure is defined as the weak limit of random cluster distributions on Λ_n, as $n \to \infty$. Let $\mathcal{W}_{p,q}$ be the set of all such measures for parameters p and q.

The two approaches to infinite volume measures does not in general produce the same set of measures. For p and q let $\mathcal{P}_{p,q} = \mathcal{R}_{p,q} \cup \mathcal{W}_{p,q}$ denote the set of random cluster measures. The question regarding uniqueness, that is, if $|\mathcal{P}_{p,q}| = 1$ or not, is to a large extent still open. The general result is the following, quoted from [Gri03] .

Theorem 7.6
For $q \geq 1$ *there exists a countable subset* $D_q \subset [0,1]$ *such that* $|\mathcal{P}_{p,q}| = 1$ *whenever* $p \neq D_q$*.*

It has been conjectured that D_q is empty for small q and a singleton

(consisting of the critical value) when q is large. In two dimensions it is known that D_q is either the empty set or $\{\sqrt{q}(1+\sqrt{q})^{-1}\}$. More information on the sets $\mathcal{R}_{p,q}$ and $\mathcal{W}_{p,q}$ can be found in [Gri95] and [GHM01] . The translation invariant measures $\phi^0_{p,q}$ and $\phi^1_{p,q}$ defined as weak limits of random cluster distributions with free and wired boundary conditions (see Figure 7.1, page 64) respectively, are in both sets, and are extremal in the stochastic ordering sense,

$$\forall \phi \in \mathcal{R}_{p,q} \cup \mathcal{W}_{p,q} : \phi^0_{p,q} \preceq_{\mathcal{D}} \phi \preceq_{\mathcal{D}} \phi^1_{p,q}.$$

Whenever we have non-uniqueness of measures we do have access to these extremal measures. In subsequent chapters the parameter estimation procedure is simulation driven. Suppose we have a data set X measured at some locations in a finite region $\Lambda \subset \mathbb{L}^d$. For some n we will have $\Lambda \subseteq \mathbb{B}_n$ so we can regard the data as having distribution $\phi^{(n)}_{p,q}$. This distribution arises as a conditional distribution from some infinite volume measure, there exists ϕ such that

$$\phi^{(n)}_{p,q}(\cdot) = \phi(\cdot|\mathcal{T}_{\mathbb{L}^d \setminus \mathbb{B}^d_n})$$

where $\mathcal{T}_{\mathbb{Z}^d \setminus \mathbb{B}^d_n}$ is the σ-field of events defined on vertices outside $\mathbb{B}^d_n$. In simulations we assume that $\phi = \phi^0_{p,q}$. For almost all parameter values, $p \in [0,1] \setminus D_q$, this is a correct approach, but on D_q it is not. For a further discussion of non-uniqueness and its implications see page 202.

Phase transition

Phase transition for the random cluster model manifests itself as the creation of an infinite connected component. Given a random cluster random variable X distributed according to $\phi_{p,q},\ q \geq 1,\ p \in [0,1]$ we let $\mathbb{L}^d(X)$ denote the random cluster graph induced by X. The question is:

Does $\mathbb{L}^d(X)$ have a infinite connected component?

To answer this question we let $C(x) = \{y \in \mathbb{Z}^d : x \leftrightarrow y\}$ denote the connected component containing vertex x. Suppose $\phi_{p,q}$ is translation invariant and as a consequence the size of $C(x)$ is independent of x. In the sequel let $\mathbf{0} \in \mathbb{Z}^d$ be the origin, also let

$$C_0 = \{y \in \mathbb{Z}^d : \mathbf{0} \leftrightarrow y\}$$

be the connected component at the origin. We answer the above question by analyzing the following probability.

$$\phi_{p,q}(|C_0| = \infty)$$

The event $\{|C_0| = \infty\}$ does not depend on the state of any finite collection of edges, so by Kolmogorov's 0-1 law in the $q = 1$ case and by ergodicity for $q \neq 1$ it is either 0 or 1. By using stochastic domination and the fact that $\{|C_0| = \infty\}$ is an increasing event we can show the existence of a p_c such that

$$\phi_{p,q}(|C_0| = \infty) = \begin{cases} 0, & p < p_c \\ 1, & p > p_c \end{cases}$$

with p_c depending on q and d. The subcritical phase ($p < p_c$) means that all connected components are finite, while the supercritical phase ($p > p_c$) implies the a.s. existence of an infinite connected component. To which phase p_c belong is generally an open question.

On the square lattice $\mathbb{L}^2$ the situation is more complete. We know for example that

$$p_c(q) = \frac{\sqrt{q}}{1 + \sqrt{q}} \ , \quad q \in \{1, 2\} \cup [25.72, \infty)$$

holds [Gri03] . For $q = 1$ (percolation) we have the famous result by Kesten [Kes80] , $p_c(1) = 1/2$, and for the Ising model we have the corresponding critical inverse temperature

$$\beta_c(2) = \frac{1}{2} \ \ln(1 + \sqrt{2}).$$

The value of the critical temperature for the Ising model was computed by Onsager in 1944 [Ons44] , much earlier than the theoretical foundations for Gibbs measures which was established in the late 1960's. The connection between the computed value by Onsager and the proper theoretical foundations was made in 1973 by Abraham and Martin-Löf [AML73] .

The result for $q \geq 25.72$ was established in [LMR86, LMMS$^+$91] . The general statement

$$p_c(q) = \frac{\sqrt{q}}{1 + \sqrt{q}} \ , \quad q \geq 1$$

remains to our knowledge open.

Connection probabilities

Given a configuration X according to the random cluster measure the connectivity function

$$f_{p,q}(x, y) = \phi_{p,q}(x \leftrightarrow y) \ , \quad x, y \in \mathbb{Z}^d$$

describes the probability that two vertices are in the same connected component in $\mathbb{L}^d(X)$. Our goal is to find a set $R_{\text{mix}} \subseteq [0,1]$ such that if $p \in R_{\text{mix}}$ there exists $C, \alpha \in (0,\infty)$ such that

$$f_{p,q}(x,y) \leq Ce^{-\alpha|x-y|}$$

This implies strong mixing for the random cluster model so let us call R_{mix} the *strong mixing regime*. The strategy is the following: We define an alternative critical point $p_g(q)$, in such a way that for $p < p_g(q)$, we have exponential decay of the cluster size at the origin, and exponential decay of the connection function follows.

For general $d \geq 2,\ q \geq 1$ little is known about the decay of the connectivity function as $|x-y| \to \infty$. In the subcritical regime we do not have infinite clusters so

$$\lim_{|x-y|\to\infty} f_{p,q}(x,y) = 0$$

but we can not say much about the convergence rate. Grimmett and Piza [GP97] have shown that for sufficiently fast polynomial decay we automatically have exponential decay of $f_{p,q}$. For this purpose we define an alternative critical point.

Definition 7.8 ***Alternative critical point***
Let

$$Y(p,q) = \limsup_{n\to\infty} n^{d-1}\phi_{p,q}(\mathbf{0} \leftrightarrow \partial\mathbb{B}_n)$$

for any $p \in [0,1]$ and $q \geq 1$. The alternative critical point $p_g(q)$ is defined as $p_g(q) = \sup\{p :\ Y(p,q) < \infty\}$.

We have $p_g(q) \leq p_c(q)$ and it is believed that equality holds for $q \geq 1$, see [GP97] . For the special case $q = 2$ equality is known to hold (according to Grimmett and Piza, [GP97, page 2]). Below $p_q(q)$ we have sufficiently fast polynomial decay to imply exponential decay, and consequently the central limit theorems are applicable. We do not, however, know $p_g(q)$.

Theorem 7.7 ***Grimmett-Piza***
Let $q \geq 1,\ d \geq 2, p < p_g(q)$. Then there exists $\gamma > 0$ depending on p and q such that

$$\phi_{p,q}(\mathbf{0} \leftrightarrow \partial\mathbb{B}_n) \leq e^{-\gamma n}$$

for all large n.

We continue with a lemma stating that even though the $p_g(q)$'s are unknown for general q they are monotone. It is a simple consequence of monotonicity (Theorem 7.3) for the random cluster distributions.

Lemma 7.2 ***Monotone alternative critical point***
Let $q_1, q_2 \in \mathbb{R}$ be such that $q_1 \leq q_2$, and let $p_g(q)$ be the alternative critical point in the random cluster model with parameter q. Then $p_g(q_1) \leq p_g(q_2)$ holds.

Proof :
Let q_1 and q_2 be as in the statement. Consider the event $\{\mathbf{0} \leftrightarrow \partial\mathbb{B}_n\}$ that there is a path of open edges between the origin and the boundary of the box $\mathbb{B}_n$. Since $\{\mathbf{0} \leftrightarrow \partial\mathbb{B}_n\}$ is an increasing event we have

$$\phi_{p,q_2}(\mathbf{0} \leftrightarrow \partial\mathbb{B}_n) \leq \phi_{p,q_1}(\mathbf{0} \leftrightarrow \partial\mathbb{B}_n)$$

for any $p \in [0,1]$ due to Theorem 7.5. Suppose $p \leq p_g(q_1)$, then Theorem 7.7 and the inequality above allow us to conclude

$$\phi_{p,q_2}(\mathbf{0} \leftrightarrow \partial\mathbb{B}_n) \leq Ce^{-\gamma n}$$

for some $C,\ \gamma \in (0,\infty)$. This implies $p \leq p_g(q_2)$ and we are done.
□

We finish with treating cluster size and prove that the distribution of $d(C_0)$ has exponentially decaying tail if $p < p_g(q)$, a simple consequence of Theorem 7.7.

Definition 7.9 ***Cluster diameter***
Let C be a connected component in a random cluster sample. We denote its diameter by $d(C)$ and define it as

$$d(C) = \sup_{x,y \in C} |x - y|$$

where $|\cdot|$ is the metric on $\mathbb{Z}^d$.

Corollary 7.1 ***Exponentially decaying cluster size***
Let $\phi_{p,q}$ be the random cluster measure for $q \geq 1$ and $p < p_g(q)$. Then the distribution of $d(C_0)$ has exponentially decaying tail.

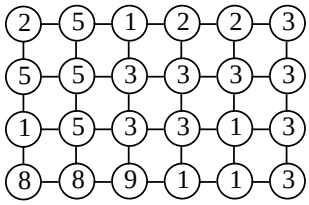

Figure 7.2: *An example of a small Potts configuration for some unknown inverse temperature β and some $q \geq 9$.*

Proof:
If $d(C_0) > n$ there exists at least one $y \in \partial\mathbb{B}_n$ such that $\mathbf{0} \leftrightarrow y$, thus $\{d(C_0) > n\} \subseteq \{\mathbf{0} \leftrightarrow \mathbb{B}_n\}$.The result now follows from Theorem 7.7 since

$$\phi_{p,q}(d(C_0) > n) \leq \phi_{p,q}(\mathbf{0} \leftrightarrow \mathbb{B}_n) \leq Ce^{-\gamma n/2}$$

for some $C, \gamma > 0$.
□

7.3 The Potts model

The Potts model is a two parameter family of models living on the vertex set of a graph $G = (V, E)$. We start by treating finite boxes $\mathbb{B}_n^d \subset \mathbb{Z}^d$ and expand to the infinite case when needed. We have two parameters, the inverse temperature β and q, the number of states a vertex can have.

Definition 7.10 ***The Potts distribution***
Given the graph $\mathbb{L}_n = (\mathbb{Z}_n, \mathbb{E}_n)$ the Potts distribution is a two parameter family of distributions $\mu_{q,\beta}^{(n)},\ \beta \in [0, \infty),\ q \in \{2, 3, ...\}$, on the space $\{1, ..., q\}^{\mathbb{Z}_n}$. An element $\sigma \in \{1, ..., q\}^{\mathbb{Z}_n}$ is called a configuration, and we denote the configuration of a single vertex $x \in \mathbb{Z}_n^d$ by $\sigma(x)$. A configuration $\sigma \in \{1, ..., q\}^{\mathbb{Z}_n}$ is assigned probability

$$\mu_{q,\beta}(\sigma) = \frac{1}{Z_{\beta,q}} \exp\left(-2\beta \sum_{e=\langle x,y\rangle \in \mathbb{E}_n} I_{\{\sigma(x)\neq\sigma(y)\}}(e)\right)$$

where $Z_{\beta,q}$ is a normalising constant. For $q = 2$ we get the famous Ising model (see Section 7.4).

Relation to the random cluster model

There is an interesting relation between the random cluster and Potts models. For the box $\mathbb{B}_n = (\mathbb{Z}_n, \mathbb{E}_n)$, consider the following measure living on the set $\{0,1\}^{\mathbb{E}_n} \times \{1,...,q\}^{\mathbb{Z}_n}$,

$$\rho_{p,q}((\gamma,\omega)) \;\propto \prod_{e=\langle e_x,e_y\rangle \in E} \Big(\; p^{\gamma(e)}(1-p)^{1-\gamma(e)} \; I_{\{(\omega(e_x)-\omega(e_y))\gamma(e)=0\}} \; \Big)$$

for any $(\gamma,\omega) \in \{0,1\}^{\mathbb{E}_n} \times \{1,...,q\}^{\mathbb{Z}_n}$. This is the Edwards-Sokal measure, introduced by Swendsen and Wang [SW87] and made more explicit by Edwards and Sokal [ES88] .

Theorem 7.8
Let $\rho_{p,q}^{vertex}$ and $\rho_{p,q}^{edge}$ be the probability measures obtained by projecting $\rho_{p,q}$ on $\{1,...,q\}^V$ and $\{0,1\}^E$ respectively. Then

$$\rho_{p,q}^{vertex} = \mu_{q,\beta} \;, \quad \rho_{p,q}^{edge} = \phi_{p,q}$$

where $\beta = \dfrac{1}{2}\ln(1-p)$.

A consequence of this result is Corollary 7.2, giving us a method of generating Potts samples from random cluster samples. The method requires that the spins $1,...,q$ are considered equivalent, so when assigning spins we must assign them uniformly at random from $\{1,...,q\}$.

Definition 7.11 ***Random cluster to Potts mapping***
Suppose $\lambda_{p,q} \in \{0,1\}^{\mathbb{E}_n}$ is an arbitrary random cluster configuration. Let $\mathbf{H} : \{0,1\}^{\mathbb{E}_n} \to \{1,...,q\}^{\mathbb{Z}_n}$ be the mapping transforming a random cluster sample into a Potts sample in the following manner. For each connected component C in the graph $G(\lambda_{p,q})$ pick a spin uniformly at random from the set $1,...,q$ and assign it to every vertex in C. Do this independently for each connected component. Return the given spin configuration $\omega_{\beta,q}$, where p and β are related as $p = 1 - e^{-2\beta}$.

Note that the mapping $\mathbf{H}$ defined above is random. For a random cluster configuration γ there are no less than $q^{\kappa(\gamma)}$ corresponding Potts configurations. The mapping chooses one uniformly at random from this set.

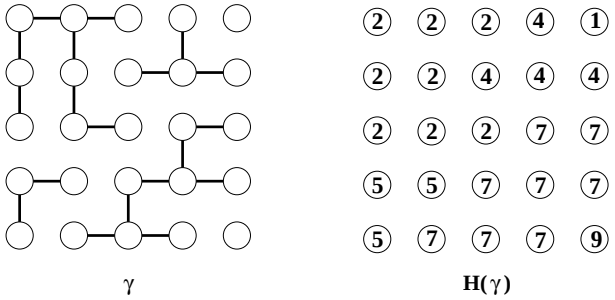

Figure 7.3: *An example of the result of the mapping from the random cluster model to the Potts model. In this case* $q = 9$.

Corollary 7.2
Let $p = 1 - e^{-2\beta}$, *and pick a random spin configuration* $\Omega \in \{1, ..., q\}^{\mathbb{Z}_n}$ *as follows*

1. *Pick* $\Gamma \in \{0, 1\}^{\mathbb{E}^d}$ *according to the random cluster measure* $\phi_{p,g}$.
2. *Let* $\Omega = \mathbf{H}(\Gamma)$

Then Ω *is distributed according to the Potts measure,* $\Omega \stackrel{\mathcal{D}}{=} \mu_{q,\beta}$.

We omit the proofs of Theorem 7.8 and Corollary 7.2, they can both be found in [GHM01] .

Infinite volume limits and phase transition

Let $\{\mathbb{B}_n\}_{n=1}^{\infty}$ be a sequence of boxes such that $\mathbb{B}_n \to \mathbb{L}^d$, and let ξ be some fixed but arbitrary configuration in $\{1, ..., q\}^{\mathbb{Z}^d}$ with positive probability. On each box we define the Potts distributions with parameters q and β as follows.

$$\mu_{q,\beta}^{(n),\xi}(X) = \mu_{q,\beta}^{\xi}(X | X = \xi \text{ on } \mathbb{Z}^d \setminus \mathbb{B}_n^d) \ , \quad X \in \{1, ..., q\}^{\mathbb{Z}^d}$$

Taking the limit as $n \to \infty$ we get

$$\mu_{q,\beta}^{\xi} = \lim_{n \to \infty} \mu_{q,\beta}^{(n),\xi}$$

where $\mu^{\xi}_{q,\beta}$ is called the Potts measure on $\mathbb{L}^d$ with parameters q, β and boundary conditions ξ. The Potts model exhibits a phase transition, and there exists a critical value β_c, possibly infinite, such that for $\beta < \beta_c$ (subcritical phase) we have one unique measure, and for $\beta > \beta_c$ (supercritical phase) there exists multiple measures. The multiplicity of measures in the supercritical phase means that there exists q distinct measures, originating from limit construction having boundary conditions $\xi \equiv k, k \in \{1, ..., q\}$, we denote these measures $\mu^{k}_{q,\beta}$. The phase transition is closely related to the creation of an infinite connected component in the corresponding random cluster measure $\phi_{1-e^{-2\beta},q}$. The infinite connected components introduces long range dependence between vertices, making the boundary affect the configuration at the origin. Since each $\mu^{k}_{q,\beta}$ is translation invariant the effect is the same for every vertex.

On finite boxes we let

$$\mu^{(n),k}_{q,\beta}(X) = \mu^{k}_{q,\beta}(X|X \equiv k \text{ on } \mathbb{Z}^d \setminus \mathbb{B}^d_n) \ , \quad X \in \{1, ..., q\}^{\mathbb{Z}^d}$$

for $k \in \{1, ..., q\}$. On $\mathbb{B}_n$ we experience a finite version of the phase transition, it is not sharp as in the infinite case but "smeared out" over a small interval $[\alpha_1, \alpha_2]$ around β_c. The length of the interval depends in n, and the larger n we take the smaller the interval becomes.

Uniqueness of measures is known to hold at the critical value on $\mathbb{Z}^d$ whenever $d = 2$ or $d \geq 4$, while the three-dimensional case remains open.

Correlations

Let l_1, l_2, l_3, l_4 be vertices in $\mathbb{L}^d$, also let σ be an Potts configuration chosen according to the Potts distribution $\mu^{(n)}_{q,\beta}$. Let ω be the corresponding random cluster configuration according to $\phi^{(n)}_{p,q}$. The distributions are related through p and β as $p = 1 - e^{-2\beta}$.

We start with an important characteristic and continue with correlations.

Definition 7.12 ***Susceptibility***
Let X be distributed according to the Potts model on $\mathbb{Z}^d$ with parameters β and q. The susceptibility for X, denoted by $\sigma^2_X(\beta)$ (or simply σ^2_X) at inverse temperature β is defined as

$$\sigma^2_X(\beta) = \sum_{v \in \mathbb{Z}^d} Cov_{\beta}[X(\mathbf{0}), X(v)].$$

We now proceed with the basic theorem giving us expected values, variances and correlations.

Lemma 7.3
Let X be distributed according to the Potts distribution $\mu_{q,\beta}^{(n)}$, and let l and l' be two distinct vertices in $\mathbb{Z}^d$. Let $\phi_{p,q}^{(n)}$ be the corresponding random cluster distribution with $p = 1 - e^{-2\beta}$, then the following holds.

$$(i) \quad \mathbb{E}[X(l)] = \frac{q+1}{2}$$

$$(ii) \quad \mathit{Var}[X(l)] = \frac{q^2-1}{12}$$

$$(iii) \quad \mathit{Cov}[X(l), X(l')] = \frac{q^2-1}{12}\ \phi_{p,q}^{(n)}(l \leftrightarrow l')$$

$$(iv) \quad \mathit{Corr}[X(l), X(l')] = \phi_{p,q}^{(n)}(l \leftrightarrow l')$$

$$(v) \quad \mathbb{E}[X(l)X(l')] = \left(\frac{q+1}{2}\right)^2 + \phi_{p,q}^{(n)}(l \leftrightarrow l')\ \frac{q^2-1}{12}$$

Proof :
Straightforward calculations and some assembly gives the result. First the expected value

$$\mathbb{E}[X(l)] = \frac{1}{q}\ \sum_{k=1}^{q} k = \frac{q+1}{2}$$

the second moment

$$\mathbb{E}[X(l)^2] = \frac{1}{q}\ \sum_{k=1}^{q} k^2 = \frac{(q+1)(2q+1)}{6}$$

and the variance follows by straightforward calculations.

$$\mathrm{Var}[X(l)] = \frac{(q+1)(2q+1)}{6} - \left(\frac{q+1}{2}\right)^2 = \frac{q^2-1}{12}.$$

Then also the covariance

$$\mathrm{Cov}[X(l), X(l')] \quad = \quad \mathbb{E}[X(l)X(l')] - \mathbb{E}[X(l)]\mathbb{E}[X(l')] = \phi_{p,q}^{(n)}(l \leftrightarrow l')\ \frac{q^2-1}{12}$$

and the correlation

$$\begin{aligned}\mathrm{Corr}[X(l),X(l')] &= \frac{\mathrm{Cov}[X(l),X(l')]}{\sqrt{\mathrm{Var}[X(l)]}\sqrt{\mathrm{Var}[X(l')]}} \\ &= \left(\phi_{p,q}^{(n)}(l\leftrightarrow l')\ \frac{q^2-1}{12}\right)\left(\frac{q^2-1}{12}\right)^{-1/2}\left(\frac{q^2-1}{12}\right)^{-1/2} \\ &= \phi_{p,q}^{(n)}(l\leftrightarrow l')\end{aligned}$$

follows since

$$\begin{aligned}\mathbb{E}[X(l)X(l')] &= \mathbb{E}[X(l)X(l')|l\leftrightarrow l']\mathbb{P}(l\leftrightarrow l') + \mathbb{E}[X(l)X(l')|l\not\leftrightarrow l']\mathbb{P}(l\not\leftrightarrow l') \\ &= \phi_{p,q}(l\leftrightarrow l')\ \frac{1}{q}\ \sum_{k=1}^{q} k^2 + \phi_{p,q}(l\not\leftrightarrow l')\left(\frac{q+1}{2}\right)^2 \\ &= \phi_{p,q}(l\leftrightarrow l')\ \frac{(q+1)(2q+1)}{6} + \phi_{p,q}(l\not\leftrightarrow l')\left(\frac{q+1}{2}\right)^2 \\ &= \left(\frac{q+1}{2}\right)^2 - \phi_{p,q}(l\leftrightarrow l')\ \frac{q^2-1}{12}\end{aligned}$$

and we are done.
□

In the subcritical phase the absence of an infinite cluster implies

$$\mathrm{Cov}[X(u),X(v)] = \left(\frac{q^2-1}{12}\right)\phi_{p,q}(u\leftrightarrow v)\to 0,$$

as the distance between u and v increases. Note that we need not explicitly know the decay rate, exponential decay suffice in order to make the susceptibility

$$\sigma_X^2 = \sum_{v\in\mathbb{Z}^d}\mathrm{Cov}[X(u),X(v)]$$

finite.

7.4 The Ising model

The Ising model is one of the oldest models in statistical mechanics, it dates back to the 1920's. In this section we define the Ising distribution on finite boxes, $\mathbb{B}_n$, and the Ising measure on $\mathbb{Z}^d$. The Ising model is a special case of the Potts model for $q = 2$ where the states are denoted -1 and $+1$ instead of 1 and 2.

Definition 7.13 ***The Ising distribution***
Given the graph $\mathbb{L}_n = (\mathbb{Z}_n, \mathbb{E}_n)$ in on or more dimensions, the Ising distribution is a one parameter family of distributions $\mu_\beta^{(n)}$, $\beta \in [0, \infty)$, on the space $\{-1, 1\}^{\mathbb{Z}_n}$. An element $\sigma \in \{-1, 1\}^{\mathbb{Z}_n}$ is called a configuration, and we denote the configuration of a single vertex $x \in \mathbb{Z}_n^d$ by $\sigma(x)$. A configuration $\sigma \in \{-1, 1\}^{\mathbb{Z}_n}$ is assigned probability

$$\mu_\beta^{(n)}(\sigma) = \frac{1}{Z_\beta} \exp\left(-2\beta \sum_{e=\langle x,y\rangle \in \mathbb{E}_n^d} I_{\{\sigma(x)\neq\sigma(y)\}}(e)\right)$$

where Z_β is a normalizing constant.

The definition is for an arbitrary number d of dimensions, but in the sequel we will restrict ourself to study the cases $d = 2, 3$. Due to certain properties of planar graphs, such as $\mathbb{L}^2$, the theory for the two-dimensional case is better understood than the three-dimensional case.

Infinite volume limits and phase transition

Constructing the Ising measure μ_β as a limit of Ising distributions

$$\mu_\beta^{(n),\xi}(X) = \mu_\beta^\xi(X | X = \xi \text{ on } \mathbb{Z}^d \setminus \mathbb{B}_n^d), \quad X \in \{-1, +1\}^{\mathbb{Z}^d}$$

with suitable boundary conditions ξ, follow the same lines as for the Potts measure (see page 73).

Since the Ising model is a special case of the Potts model (just set $q = 2$) the phase transition is of the same type, but better understood. As for the Potts model we know that in the subcritical phase, $\beta < \beta_c$ we have a unique measure and that the supercritical phase, $\beta > \beta_c$, we have exactly two measures originating from the two boundary conditions (all $+/-$ outside the $\mathbb{B}_n$'s). These are the so called pure phases consisting of a vast ocean of $-/+$ spins (see Figure 7.4 below), with small islands of the other spin type.

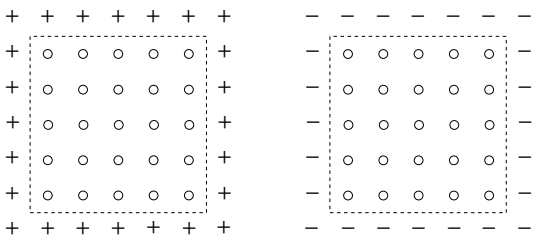

Figure 7.4: *An example on boundary conditions for a small box (within dashed lines) in the Ising model. In the right diagram we have all $+$-spins outside the box, and on the left we see the $-$ boundary condition.*

Correlations

As a preparation for point estimation we prove some results regarding correlation, variance and expectation of the configuration at single vertices. Let l_1, l_2, l_3, l_4 be vertices in $\mathbb{L}_n$, also let σ be an Ising configuration chosen according to the Ising distribution $\mu_\beta^{(n)}$ on box $\mathbb{B}_n$. Let ω be the corresponding random cluster configuration according to $\phi_{p,q}^{(n)}$. The distributions are related through p and β as $p = 1 - e^{-2\beta}$.

Lemma 7.4 ***Two-point correlation***
Let X be distributed according to the Ising measure $\mu_\beta^{(n)}$, and let l and l' be two distinct locations in $\mathbb{Z}_n$. Let $\phi_{p,2}^{(n)}$ be the corresponding random cluster measure with $p = 1 - e^{-2\beta}$. Then the following holds.

$$(i) \quad \mathbb{E}[X(l)] = 0$$

$$(ii) \quad \mathit{Var}[X(l)] = 1$$

$$(iii) \quad \mathit{Corr}[X(l), X(l')] = \phi_{p,2}^{(n)}(l \leftrightarrow l')$$

$$(iv) \quad \mathbb{E}[X(l)X(l')] = \phi_{p,2}^{(n)}(l \leftrightarrow l')$$

$$(v) \quad \mathit{Var}[X(l)X(l')] = 1 - \phi_{p,2}^{(n)}(l \leftrightarrow l')^2$$

Proof :
The result follows using the same type of calculations as in Lemma 7.3.
□

Lemma 7.5 ***Four-point correlation***
Let X be distributed according to the Ising measure $\mu_\beta^{(n)}$, and let l_1, l_2 and l_3, l_4 be four distinct locations in $\mathbb{Z}_n$. Then $X(l_1)X(l_2)$ and $X(l_3)X(l_4)$ are positively correlated with covariance

$$\begin{aligned} \mathrm{Cov}[X(l_1)X(l_2), X(l_3)X(l_4)] &= \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2\} \cap \{l_3 \leftrightarrow l_4\}) \\ &\quad -\phi_{p,2}^{(n)}(l_1 \leftrightarrow l_2)\phi_{p,2}^{(n)}(l_3 \leftrightarrow l_4) \\ &\quad +\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_3 \not\leftrightarrow l_2 \leftrightarrow l_4\}) \\ &\quad +\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_4 \not\leftrightarrow l_2 \leftrightarrow l_3\}). \end{aligned}$$

Proof :
Let l_1, l_2, l_3, l_4 be chosen as in the statement, also let X be a configuration chosen according to the Ising measure $\mu_\beta^{(n)}$.

$$\begin{aligned} \mathrm{Cov}[X(l_1)X(l_2), X(l_3)X(l_4)] &= \mathbb{E}[X(l_1)X(l_2)X(l_3)X(l_4)] \\ &\quad -\mathbb{E}[X(l_1)X(l_2)]\mathbb{E}[X(l_3)X(l_4)] \end{aligned}$$

The second term is given by Lemma 7.4, while for the first term we make the following observation. Given the event I that vertices $l_1,\ l_2,\ l_3$ and l_4 are located in distinct connected components, in the underlying random cluster configuration, their spin is independent of each other and

$$\mathbb{E}[X(l_1)X(l_2)X(l_3)X(l_4)|I] = \prod_{k=1}^{4} \mathbb{E}[X(l_k)|I] = 0.$$

The same holds as long as at least one of the vertices is located in a component different from where the other three are. Whenever the vertices are located pairwise in the same connected components the expectation becomes 1. We define events for these cases.

$\{l_1 \leftrightarrow l_2 \not\leftrightarrow l_3 \leftrightarrow l_4\}$ Vertices l_1, l_2 is located in one connected components and that l_3, l_4 are located in another connected component.

$\{l_1 \leftrightarrow l_2 \leftrightarrow l_3 \leftrightarrow l_4\}$ All four vertices are located in the same connected component.

$\{isolated\}$ At least one of the four vertices is located in a connected components separate from the others.

Just for notation let $X = X(l_1)X(l_2)X(l_3)X(l_4)$.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|\{isolated\}]\phi_{p,2}^{(n)}(\{isolated\}) \\
&\quad +\mathbb{E}[X|\{l_1 \leftrightarrow l_2 \not\leftrightarrow l_3 \leftrightarrow l_4\}]\ \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2 \not\leftrightarrow l_3 \leftrightarrow l_4\}) \\
&\quad +\mathbb{E}[X|\{l_1 \leftrightarrow l_3 \not\leftrightarrow l_2 \leftrightarrow l_4\}]\ \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_3 \not\leftrightarrow l_2 \leftrightarrow l_4\}) \\
&\quad +\mathbb{E}[X|\{l_1 \leftrightarrow l_4 \not\leftrightarrow l_2 \leftrightarrow l_3\}]\ \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_4 \not\leftrightarrow l_2 \leftrightarrow l_3\}) \\
&\quad +\mathbb{E}[X|\{l_1 \leftrightarrow l_2 \leftrightarrow l_3 \leftrightarrow l_4\}]\ \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2 \leftrightarrow l_3 \leftrightarrow l_4\}) \\
&= \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2\} \cap \{l_3 \leftrightarrow l_4\}) + \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_3 \not\leftrightarrow l_2 \leftrightarrow l_4\}) \\
&\quad +\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_4 \not\leftrightarrow l_2 \leftrightarrow l_3\})
\end{aligned}$$

We proceed with the covariance.

$$\begin{aligned}
\mathrm{Cov}_{\mu_\beta^{(n)}}[X(l_1)X(l_2), X(l_3)X(l_4)] &= \phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2\} \cap \{l_3 \leftrightarrow l_4\}) \\
&\quad -\phi_{p,2}^{(n)}(l_1 \leftrightarrow l_2)\phi_{p,2}^{(n)}(l_3 \leftrightarrow l_4) \\
&\quad +\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_3 \not\leftrightarrow l_2 \leftrightarrow l_4\}) \\
&\quad +\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_4 \not\leftrightarrow l_2 \leftrightarrow l_3\})
\end{aligned}$$

The random cluster measure $\phi_{p,2}$ is monotone and therefore has positive correlations, see [GHM01] , implying that

$$\phi_{p,2}^{(n)}(\{l_1 \leftrightarrow l_2\} \cap \{l_3 \leftrightarrow l_4\}) \geq \phi_{p,2}^{(n)}(l_1 \leftrightarrow l_2)\phi_{p,2}^{(n)}(l_3 \leftrightarrow l_4)$$

since $\{l_1 \leftrightarrow l_2\}$ and $\{l_3 \leftrightarrow l_4\}$ both are increasing events, and we are done.
□

7.5 Some natural extensions

The are many ways to extend the presented models, we present a few of the here.

For the Ising and Potts models one way is to introduce random interaction parameters. Instead of letting all interaction have the same strength we introduce a class of random variables

$$\{J(\langle x, y\rangle)\}_{\langle x,y\rangle\in\mathbb{E}_n}$$

representing the strength of the interaction between neighbouring vertices. Different distributions of the interactions lead to different characteristics, and the ordinary mode is a special where the distribution assigns all probability to a single point in the underlying space. We could also remove certain bonds or vertices and study the model on the remaining graph, this method is called (bond or site) dilution and has been studied intensively over the years for both the Ising and Potts models. There is also a version of the q-state Potts model, called the Potts lattice gas, having $q+1$ possible states (possible states are $\{0, 1, ..., q\}$) where state 0 represent as vacant site. There are also the fuzzy models, such as the fuzzy Potts model introduced by Maes and Vande Velde in 1995, [MV95] . It is an example of a hidden Markov model where we simple hide some of the spins in a ordinary Potts realisation.

There are also variants of the random cluster model. We have the asymmetric random cluster model introduced by Alexander [Ale01] 2001, a variant suitable for studying the Potts lattice gas. Another example is the partial random cluster model where a fraction of the non singleton clusters are removed from the random cluster model.

7.6 Historical notes

We treat only briefly the history and development of the Ising, Potts and random cluster models. For more complete historical notes see for example [GHM01] , [Gri03] , [Gri95] and references therein.

7.6.1 The Ising model

The Ising model was developed in the 1920's to study phase transition in one dimension. The first articles was published by Lenz [Len20] in 1920 and by Ising [Isi25] in 1925. It quickly became clear that the Ising model on $\mathbb{Z}$ does not exhibit a phase transition, but it was not until 1936 when Peirls [Pei36] introduced his contour argument that the

existence of a phase transition was proved for $d \geq 2$. In 1944 Onsager [Ons44] computed the value of the critical inverse temperature on $\mathbb{Z}^2$, see comments on page 68.

The critical value divides the domain for the inverse temperature in two phases, the subcritical ($\beta < \beta_c$) and the supercritical ($\beta > \beta_c$). The nature of the phases we partially established in 1972 when Gallavotti and Miracle-Sole [GMS72] proved that the Ising model with periodic boundary on $\mathbb{Z}^d$ with d arbitrary have exactly two pure phases for large enough β. The full result was established in 1975 by Messager and Miracle-Sole [MMS75] proving the corresponding result in two dimensions for $\beta > \beta_c$. The concept of pure phases is emphasized in three dimensions where there, in addition to two pure phases, also exists measures expressing the coexistence of pure phases [Dob72] . The coexistence is such that one half of the space have a majority of $-$ spins and the other half a majority of $+$ spins. The two halves are separated by a sharp interface.

The concept of dilution is together with the introduction of random interactions natural extensions of the Ising model and has as such been intensively studied over the years. For some results see for example [ACCN87] , [Boc83] for the Ising model and [Wu80] , [Wu81] for the Potts model.

In statistical applications of the Ising model we regard the inverse temperature as the unknown parameter and use observations to compute point estimates and confidence intervals. we could construct confidence intervals by direct methods if we know the distribution of some appropriate statistic, or we could use asymptotic theory and the normal distribution through some central limit theorem. There are in the field many central limit theorems regarding different characteristics of the Ising model but here we are only interested in those directly relating the inverse temperature to some entity computed from observed data, fully observed or not.

In 1976 Pickard states a central limit theorem [Pic76] regarding the sample correlation between nearest neighbours in the Ising model. Based on an exact characterisation of the partition function (by Kaufmann [Kau49] in 1949) he proves the result for the fully observed Ising model on a torus in the subcritical regime. In later articles [Pic77] and [Pic79] he extends the results allowing different vertical and horizontal interactions, non-zero magnetic field and supercritical inverse temperature, still assuming a fully observed model.

In 1982 Bolthusen gives a central limit theorem for stationary mixing random fields regarding the mean over bounded regions. The result

is valid for models fulfilling the requirements for strong or ρ-mixing. Bolthausen's result is rather general, allowing us to introduce new models based on the Ising and Potts models and still have a central limit theorem for the asymptotic distribution of relevant statistics.

In 1992 and 1993 Comets and Gidas [CG92, Gid93] present results regarding parameter estimation for the Ising and Potts models for fully and partially observed data. The article treating fully observed data [Gid93] states consistency of ML estimators and also asymptotic normality under suitable conditions. The treatment of partially observed data in [CG92] uses data from Gibbs models not directly observed, but observed through another independent lattice process. This extra process can be used to add noise or remove observations at some sites. Consistency for ML estimators are stated and proved, there is however no result regarding asymptotic normality.

7.6.2 The Potts model

As a natural extension of the Ising model the q-state Potts model was introduced by R.B. Potts in 1952 [Pot52] , an extension of the four state Ashkin-Teller model from 1943 [AT43] . Since the 1970's much effort has been made to explore the model, its properties and relation to other models. Among mathematicians and physicists the phase transition which is more complex than that of the Ising model has been extensively studied. Today many problems are solved, for example the structure of phase transition on the cubic d-dimensional lattice, due to Aizenman, Chayes, Chayes and Newman in 1988 [ACCN88] . Another important result by Kotečky and Schlosman in 1982 [KS82] is that there are multiple measures also at the critical value, which is not the case for the Ising model.

Other questions remain, to our knowledge, open, for example the value for the critical inverse temperature. It in known through Onsager [Ons44] that in the Ising case ($q = 2$) the critical value is

$$\beta_c = \frac{1}{2} \log(1 + \sqrt{q})$$

and it is believed to hold for general q in $\mathbb{Z}^2$, but has only been proved for q sufficiently large [LMMS$^+$91] .

7.6.3 The random cluster model

After observing similarities in electrical networks, percolation and the Ising/Potts models Kastelyn and Fortuin introduced the random cluster in a series of articles [FK72] , [For72a] and [For72b] in the early 1970's. As a two parameter model on the edge set of a graph it can be applied in connection with many spin-systems. For the Potts model correlations transfers to connections in the random cluster model, making is possible to study dependence through stochastic geometry. The random cluster model also experiences a phase transition enabling us to study phase transition in a wider class of spin-models simultaneously. We mention a few of the important articles published over the years.

Aizenman, Chayes, Chayes and Newman studied magnetization in one-dimensional $1/|x-y|^2$ Ising and Potts models [ACCN88] in 1988, the purpose of the article in two-fold and they demonstrate how to use the random cluster model when relating the phase structure of the Ising, Potts and percolation models.

Another important contribution is the coupling between the Potts models and the random cluster model, the Edwards-Sokal coupling [ES88] after the inventors. The coupling makes it possible to generate Potts and Ising samples from random cluster data, and vice versa. An important application is the Edwards-Sokal-Swendsen-Wang algorithm making it possible to simulate the Potts model efficiently for each inverse temperature. Also the omniparametric algorithm extensively used in this thesis is, via the important article by Propp and Wilson [PW96] , a consequence of this coupling.

In 1998 Häggström [Häg98] presented the random cluster model as an important tool for studying phase transitions in percolation, Ising and Potts models, see also [GHM01] by Georgii, Häggström and Maes.

Other examples of applications of the random cluster model are Bouabci and Carneiro's article [BC00] where a representation of the Blume-Capel model is presented, and [MM04] where Mossel and Steel study phase transition on a phylogenetic tree. We also have Pfister and Velenik's article [PV97] on the random cluster representation of the Ashkin-Teller model and many more . . .

CHAPTER 8

Omniparametric models

In this chapter we present omniparametric versions of the random cluster and Potts models. The Ising model is covered as a special case of the Potts model. We present omniparametric versions and projection operators for retrieving the ordinary fixed parameter models. The rest of the chapter is organized as follows. We begin by describing how to couple two random cluster variables in a special way and proceed with the definition of the omniparametric random cluster model and a projection mapping. We repeat the process for the Potts model. For simplicity we present the material for a finite graph $G = (V, E)$.

8.1 The random-cluster case

Let $\{\Gamma_{p,q}\}_{p\in[0,1],q\geq 1}$ be a class of random cluster variables in $\{0,1\}^E$. For each fixed $q_0 \geq 1$ we have a subclass $\{\Gamma_{p,q_0}\}_{p\in[0,1]}$ which we will represent by a single random variable, the omniparametric random cluster variable Γ_{q_0} taking values in $[0,1]^E$. The omniparametric variable is a coupling of the $\Gamma_{p,q}$'s over all $p \in [0,1]$ for a fixed q. We couple them in such a way that $\Gamma_{p_1,q} \leq \Gamma_{p_2,q}$ a.s. whenever $p_1 \leq p_2$. This makes it possible to represent the behaviour of the whole class at a single edge e by a threshold value in $[0,1]$. Saving threshold values for all edges give us a simple description of the behaviour of $\{\Gamma_{p,q}\}_{p\in[0,1]}$ and we denote the collection of thresholds by Γ_q, the omniparametric random cluster

variable.

We start by describing the coupling and a certain Markov chain. We define the omniparametric random cluster distribution as the stationary distribution of this chain. Some effort is made to ensure existence of, and convergence to, the stationary distribution and we will adopt the techniques of Propp and Wilson [PW96] . Finally we define a projection operator from $[0,1]^E$ to $\{0,1\}^E$ and prove that the variable in $\{0,1\}^E$ emerging by projection follows the fixed parameter random cluster model.

The coupling

For fixed parameter simulation we use the Gibbs sampler and marginal probabilities described in Section 7.2. We define the coupling of the variables $\{\Gamma_{p,q}\}_{p\in[0,1]}$ by letting the Gibbs sampler run Markov chains $\{\gamma^{(n)}_{p_1,q}\}_{n=0}^{\infty}$ and $\{\gamma^{(n)}_{p_2,q}\}_{n=0}^{\infty}$ on $\{0,1\}^E$ for fixed arbitrary $p_1, p_2 \in [0,1]$ such that $p_1 \le p_2$. Let $\gamma^{(0)}_{p_1,q}$ and $\gamma^{(0)}_{p_2,q}$ be two arbitrary elements in $\{0,1\}^E$ such that $\gamma^{(0)}_{p_1,q} \preceq \gamma^{(0)}_{p_2,q}$. Given configurations $\gamma^{(n)}_{p_1,q}$ and $\gamma^{(n)}_{p_2,q}$ we update them to $\gamma^{(n+1)}_{p_1,q}$ and $\gamma^{(n+1)}_{p_2,q}$ in the following manner, using the same random number for both variables.

1. Choose an edge $e \in E$ uniformly at random.
2. Let $u \stackrel{\mathcal{D}}{=} U[0,1]$.
3. Update the configurations at e.

$$\gamma^{(n+1)}_{p_1,q}(e) = \begin{cases} 1, & u \le \phi_{p_1,q}(\gamma^{(n)}_{p_1,q}(e) = 1 |\ \gamma^{(n),\{e\}}_{p_1,q}) \\ 0, & \text{otherwise} \end{cases}$$

$$\gamma^{(n+1)}_{p_2,q}(e) = \begin{cases} 1, & u \le \phi_{p_2,q}(\gamma_{p_2,q}(e) = 1 |\ \gamma^{(n),\{e\}}_{p_2,q}) \\ 0, & \text{otherwise} \end{cases}$$

4. Let $\gamma^{(n+1)}_{p_1,q}(f) = \gamma^{(n)}_{p_1,q}(f)$ and $\gamma^{(n+1)}_{p_2,q}(f) = \gamma^{(n)}_{p_2,q}(f)$ for $f \neq e$.

Since $\gamma^{(n)}_{p_1,q} \preceq \gamma^{(n)}_{p_2,q}$ and the random cluster measure is monotone we have

$$\phi_{p_1,q}(\gamma^{(n)}_{p_1,q}(e) = 1 |\ \gamma^{(n),\{e\}}_{p_1,q}) \le \phi_{p_2,q}(\gamma^{(n)}_{p_2,q}(e) = 1 |\ \gamma^{(n),\{e\}}_{p_2,q})$$

ensuring us that $\gamma^{(n+1)}_{p_1,q} \preceq \gamma^{(n+1)}_{p_2,q}$ whenever $q \ge 1$. Asymptotically we have

$$\gamma^{(n)}_{p_1,q} \xrightarrow{\mathcal{D}} \gamma_{p_1,q},\ \gamma^{(n)}_{p_2,q} \xrightarrow{\mathcal{D}} \gamma_{p_2,q}$$

as $n \to \infty$, where $\gamma_{q,1} \stackrel{\mathcal{D}}{=} \phi_{p_1,q}$ and $\gamma_{p_2,q} \stackrel{\mathcal{D}}{=} \phi_{p_2,q}$. Since we do not have an infinite amount of time we use the coupling from the past (CFTP) algorithm by Propp and Wilson [PW96] . We include in this scheme the Markov chains $\{\gamma_{p,q}^{(n)}\}_{n=0}^{\infty}$ for all $p \in [0,1]$ and repeat (at least in theory) the procedure. Due to monotonicity of the coupling there will for every edge e be a unique threshold $p_t(e)$ such that

$$\gamma_{p,q}(e) = \begin{cases} 1, & p \geq p_t(e) \\ 0, & \text{otherwise} \end{cases}$$

We save the threshold values for all edges in a new configuration denoted γ_q, the omniparametric configuration.

A projection operator

We can not treat the omniparametric random cluster model without relating it to the fixed parameter model. For this purpose we define a projection mapping, $\mathbf{P}_p^{\text{RC}}$.

Definition 8.1 ***Random cluster projection mapping***
Given $p \in [0,1]$ we define the projection mapping $\mathbf{P}_p^{RC} : [0,1]^E \to \{0,1\}^E$ edgewise in the following manner

$$\mathbf{P}_p^{RC}(\gamma_o(e)) = \begin{cases} 0, & \textit{if } \gamma_o(e) < p \\ 1, & \textit{otherwise.} \end{cases}$$

Note that the definition itself has nothing to do with the random cluster measure, it is simply a mapping from one space to another. For an illustration see Figure 8.1

A Markov chain

We define a Markov chain, $\{X_n\}_{n=0}^{\infty}$ on $[0,1]^E$ with parameter $q \geq 1$. Our goal is to prove existence of and convergence to a stationary distribution. Let the initial element, $X^{(0)}$, follow any distribution on $[0,1]^E$. The update procedure is the following. Given $X^{(n)}$ we construct $X^{(n+1)}$ by choosing an edge $e = \langle x, y \rangle \in E$ uniformly at random. Let $(X^{(n)})^{\{e\}}$ be the configuration on $G^{\{e\}} = (V, E \setminus \{e\})$ which is equal to $X^{(n)}$ on every edge except e. We let

$$p_t = \min\{p \in [0,1] : \ x \leftrightarrow y \text{ in } \mathbf{P}_p^{\text{RC}}((X^{(n)})^{\{e\}})\}.$$

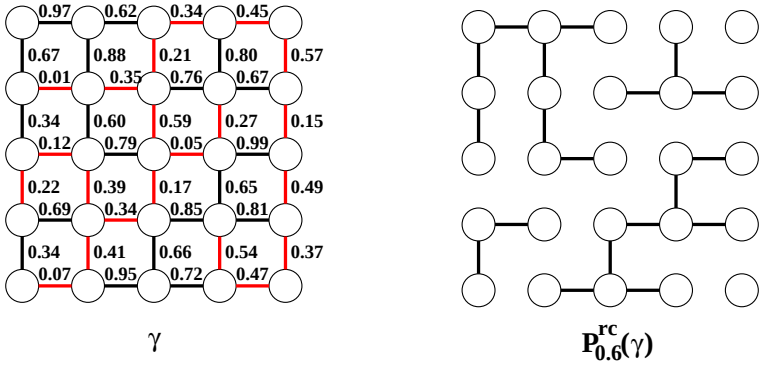

Figure 8.1: *An example of how the projection mapping works.*

be the smallest p such that x and y is connected by an open path in $\mathbf{P}_p^{\text{RC}}((X^{(n)})^{\{e\}})$. Let $f_{p_t,q}(p)$ be the probability of $X^{(n+1)}(e) = 1$ as a function of p. We get

$$f_{p_t,q}(p) = \begin{cases} p(p+q(1-p))^{-1}, & p < p_t \\ p, & \text{otherwise} \end{cases}$$

according to the single-edge conditional probabilities (Lemma 7.1). Sometimes when q is understood we simple write $f_{p_t}(p)$.

It is important that $f(p)$ is monotone, otherwise we loose the nice representation of the omniparametric state. This is why we can not treat the random cluster model for $q < 1$, as seen in Figure 8.2. We also define a version of its inverse.

$$f^{-1}_{p_t,q}(t) = \begin{cases} qt(1-t(q+1))^{-1}, & t < p_t(p_t+q(1-p_t))^{-1} \\ p_t, & p_t(p_t+q(1-p_t))^{-1} \leq t \leq p_t \\ t, & t > p_t \end{cases}$$

Note that if $p_1 \leq p_2$ then $f^{-1}_{p_1,q} \leq f^{-1}_{p_2,q}$ as seen in Figure 8.3.

We define the update function $\phi^q : [0,1] \times E \times [0,1]^E \to [0,1]^E$ edgewise in the following manner.

$$\phi^q(u, e, X^{(n)})(f) = \begin{cases} f^{-1}_{p_t,q}(u), & f = e \\ X^{(n)}(f), & f \neq e \end{cases}$$

for all $f \in E$ and $u \in [0,1]$.

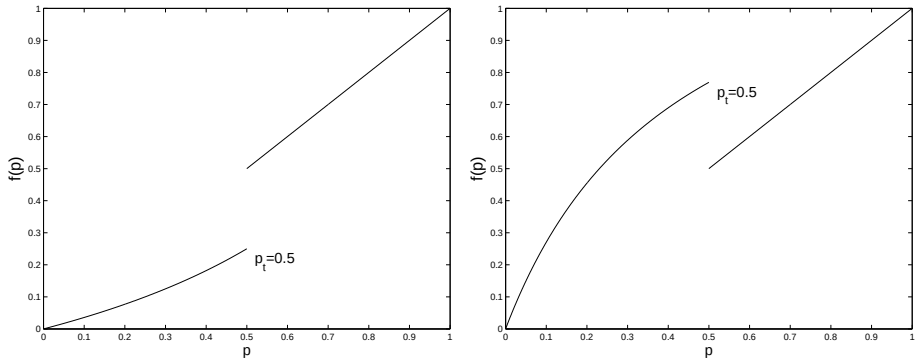

Figure 8.2: *Two examples of marginal probability $f_{p_t,q}(p) = \phi_{p,q}(X(e) = 1|X^{\{e\}})$ for the random cluster model. In both cases we have $p_t = 1/2$. In the left diagram we used $q = 3$ and in the right $q = 0.3$.*

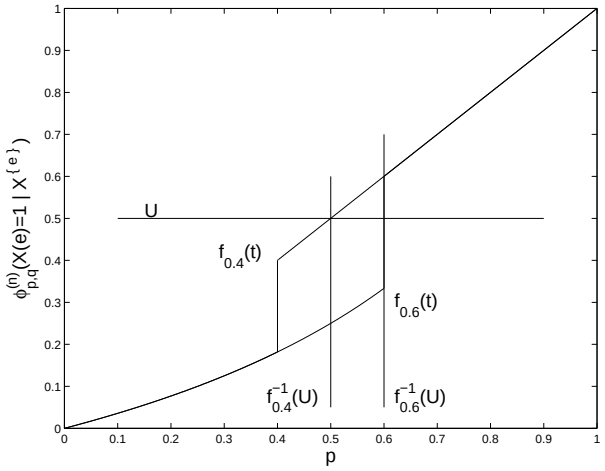

Figure 8.3: *An example of how the update function behaves when we couple the update for two chains by using the same random number. Suppose $X(e) \leq Y(e)$ for all $e \in E$, $X(e) = 0.4$ and $Y(e) = 0.6$. We see that the coupled update preserves that order*

Definition 8.2 ***The omniparametric random cluster Markov chain***
Let $\{e_n\}_{n=0}^{\infty}$ be a random sequence of edges, all selected independently and uniformly at random from the edge set E. Also let $\{U_n\}_{n=0}^{\infty}$ be a sequence of independent $U[0,1]$ random numbers. Let $\{X_n\}_{n=0}^{\infty}$ be a Markov chain on $[0,1]^E$ defined by the following.

(i) $X_0 \stackrel{\mathcal{D}}{=} \rho$
(ii) $X_{n+1} = \phi^q(U_n, e_n, X_n)$

where ρ is an arbitrary distribution on the state space $[0,1]^E$. Then $\{X_n\}_{n=0}^{\infty}$ is called the omniparametric random cluster Markov chain .

In order to adopt the techniques of Propp and Wilson we need a partial order on the state space, see Definition 7.2 on page 60.

We will run two chains, $\{X_n\}_{n=0}^{\infty}$ and $\{Y_n\}_{n=0}^{\infty}$, simultaneously and couple them by using the same random numbers for both chains. At an update we choose an edge $e \in E$ uniformly at random and update $X_n(e)$ and $Y_n(e)$ by using the random number U. We call this procedure the coupled update .

$$X_{n+1} = \phi^q(U, e, X_n) \ , \quad Y_{n+1} = \phi^q(U, e, Y_n)$$

We want $X_n \preceq Y_n$ to imply $X_{n+1} \preceq Y_{n+1}$. It is a simple consequence of of the definition of the update function ϕ^q.

Lemma 8.1 ***Coupled update preserves order***
Let $X, Y \in [0,1]^E$ such that $X \preceq Y$, and let $U \stackrel{\mathcal{D}}{=} U[0,1]$. Then

$$\phi^q(U, e, X) \preceq \phi^q(U, e, Y)$$

for any $e \in E$.

Proof :
Let X and Y be any elements in $[0,1]^E$ such that $X \prec Y$, and let $e \in E$ be arbitrary. Then $X(e) \leq Y(e)$ and as a consequence $f_{X(e)}^{-1} \leq f_{Y(e)}^{-1}$ on $[0,1]$ (see Figure 8.3). The result now follows from the definition of the update function.
□

Suppose that X_0 is the all "1" state, and that Y_0 is the all "0" state, that is,

$$\forall e \in E: \ X(e) = 0, \ Y(e) = 1.$$

We run the two chains forward, generating $\{X_n\}_{n=0}^{\infty}$ and $\{Y_n\}_{n=0}^{\infty}$ by using the coupled update. The question is now: Is there some finite positive M such that $X_n = Y_n$ for $n \geq M$?

To answer this question we begin with the definition of a coupled version of the omniparametric random cluster Markov chain. It is the main tool when proving existence of and convergence to a unique stationary distribution for the omniparametric random cluster Markov chain. The usefulness come from the fact that any other initial configuration X' is squeezed in between the initial configurations of the coupled chains, $\tilde{X}_0 \preceq X' \preceq \tilde{Y}_0$, and the coupled update preserves this order if all three chains are coupled.

Definition 8.3 ***A coupled Markov chain***
Let $\{e_n\}_{n=0}^{\infty}$ is a random sequence of edges, all selected uniformly at random and independently of each other. Also let $\{U_n\}_{n=0}^{\infty}$ be a sequence of independent $U[0,1]$ random numbers. Let $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ be a Markov chain on $[0,1]^E$ defined by the following.

(i) $\quad \forall e \in E: \ \tilde{X}_0(e) = 0$
(ii) $\quad \tilde{X}_{n+1} = \phi^q(U_n, e_n, \tilde{X}_n)$
(i') $\quad \forall e \in E: \ \tilde{Y}_0(e) = 1$
(ii') $\quad \tilde{Y}_{n+1} = \phi^q(U_n, e_n, \tilde{Y}_n)$

Then $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ is called the coupled omniparametric random cluster Markov chain .

Lemma 8.2 ***Coalescence in finite time***
Let $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ be a coupled omniparametric random cluster Markov chain. Then there exists a random time $M < \infty$ such that $\tilde{X}_n = \tilde{Y}_n$ whenever $n \geq M$.

Proof :
We consider finite subchains of $\{\tilde{X}_n, \tilde{Y}_n\}_{n=1}^{2|E|}$ of $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$, such that

$$\mathbb{P}(\tilde{X}_{2|E|} = \tilde{Y}_{2|E|}) > 0.$$

By writing the original chain as a sequence of such subchains we will eventually have coalescence with probability one.

Consider the sequences of random edges $\{e_n\}_{n=0}^{\infty}$ and random numbers $\{U_n\}_{n=0}^{\infty}$. Fix $a, b \in (0,1)$ such that $a < b$. With positive probability there is a subsequence of indices $l, ..., l + 2k - 1, \ k = |E|$, such that

$$\{e_l, ..., e_{l+k-1}\} = \{e_{l+k}, ..., e_{l+2k-1}\} = E,$$

$U_m < a$ for $m = l, ..., l+k-1$, and $U_m > b$ for $m = l+k, ..., l+2k-1$ such that

$$U_{l+k} < U_{l+k+1} < ... < U_{l+2k-1}.$$

Consider the subchain $\{\tilde{X}_n, \tilde{Y}_n\}_{n=l}^{l+k-1}$. Regardless of $(\tilde{X}_l, \tilde{Y}_l)$ we can choose a sufficiently small, making $\tilde{X}_{l+k}(e) < 0.1$ and $\tilde{Y}_{l+k}(e) < 0.1$ for all $e \in E$. Now consider the subchain $\{\tilde{X}_n, \tilde{Y}_n\}_{n=l+k}^{l+2k-1}$. At every update we will use a random number U_n such that $\tilde{X}_n(e) < U_n$ and $\tilde{Y}_n(e) < U_n$ for $n = l+k, ..., l+2k-1$ since the sequence $U_{l+k}, ..., U_{l+2k-1}$ is strictly increasing. The update function will choose the same value for both chains

$$\phi^q(U_n, e, \tilde{X}_n)(e) = \phi^q(U_n, e, \tilde{Y}_n)(e), \ n = l+k, ..., l+2k-1$$

since the thresholds in $\tilde{X}_n$ and $\tilde{Y}_n$ are smaller than U_n regardless of the choice of e. Since $\{e_{l+k}, ..., e_{l+2k-1}\} = E$ we will have $\tilde{X}_{l+2k-1} = \tilde{Y}_{l+2k-1}$ and we say that the chains have coalesced. We can now regard the chain $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ as a sequence of subchains

$$\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{L-1}, \ \{\tilde{X}_n, \tilde{Y}_n\}_{n=L}^{2L-1}, ..., \ \{\tilde{X}_n, \tilde{Y}_n\}_{n=(m-1)L}^{mL-1}, \cdots,$$

all on length $L = 2k$. With probability one some of these subchains will eventually result in coalescence. Suppose this happens for subchain m and let $M = mL$. For $n = ML$ we have $\tilde{X}_n = \tilde{Y}_n$ and since the coupled updates for subsequent pairs of $\tilde{X}_n$'s and $\tilde{Y}_n$'s will result in the same updated value regardless e_n and U_n we have $\tilde{X}_n = \tilde{Y}_n$ for all $n \geq M$ and we are done.
□

We are now ready to prove existence of the stationary distribution for the omniparametric random cluster Markov chain.

Theorem 8.1 ***Existence theorem***
Let $\{X_n\}_{n=0}^{\infty}$ be an omniparametric random cluster Markov chain. Then there exists a distribution μ such that

$$\lim_{n\to\infty} X_n \stackrel{\mathcal{D}}{=} \mu.$$

Proof:
We follow along the same lines as the proof of the main theorem in Propp and Wilsons article [PW96] from 1996. The idea is the following. We involve another two chains and couple them all. The two extra chains has the property that one is a version if the other shifted one time step.

Since they all coalesce eventually one of the two extra chains will follow a distribution having all necessary properties of a stationary distribution. Also as a consequence of coalescence all four chains follow the same distribution at time $t = 0$. Now to the details.

Consider the coupled omniparametric random cluster Markov chain $\{\tilde{X}_n, \tilde{Y}_n\}_{n=-M}^{0}$ started at time $-M$ and stopped at $n = 0$. For this we use the random edge sequence $\{e_n\}_{n=-M}^{0}$ and random number sequence $\{U_n\}_{n=-M}^{0}$. According to Lemma 8.2 we can choose $M < \infty$ large enough such that $\tilde{X}_n = \tilde{Y}_n,\ n \geq M$, with probability 1.
Let us now involve another two chains $\{Z_n\}_{n=-M}^{0}$ and $\{W_n\}_{n=-M}^{0}$ where $W_{-M} = Z_{-M+1} =$. Then

$$W_n \stackrel{\mathcal{D}}{=} Z_{n+1}$$

for any $n \geq -M$. We couple the W_n's and the Z_n's with $\tilde{X}_n$ and $\tilde{Y}_n$ as described in Definition 8.3 now with four chains instead of two. Regardless of the initial states $(W_{-M},\ Z_{-M})$ we will have $W_0 = Z_0$ since $\tilde{X}_0 = \tilde{Y}_0$ and either

$$\tilde{X}_n \preceq Z_n \preceq W_n \preceq \tilde{Y}_n \ \text{ or } \ \tilde{X}_n \preceq W_n \preceq Z_n \preceq \tilde{Y}_n$$

holds. We now let $M \to \infty$ and consider the chains $\{Z_n\}_{n=-\infty}^{0}$ and $\{W_n\}_{n=-\infty}^{0}$. We now have $W_0 \stackrel{\mathcal{D}}{=} Z_0$ since $W_0 = Z_0$ a.s. and since $W_0 \stackrel{\mathcal{D}}{=} Z_{-1}$ we also have $Z_{-1} \stackrel{\mathcal{D}}{=} Z_0$. Remember that Z_0 was constructed from Z_{-1} by executing the update function once so

$$Z_{-1} \stackrel{\mathcal{D}}{=} \phi^q(U_{-1}, e_{-1}, Z_{-1})$$

making the distribution of Z_0 stationary for the $\{Z_n\}_{n=-\infty}^{0}$ chain. The probabilistic behaviour of $\{Z_n\}_{n=-\infty}^{0}$ is the same as for the original chain $\{X_n\}_{n=0}^{\infty}$ making them share the same stationary distribution, and we are done.
□

In order to show convergence towards a stationary distribution we need to measure the distance between distributions. We use the total variation distance.

Definition 8.4 ***Total variation distance***
Let ν_1, ν_2 be two measures on a measurable space $(\Omega, \mathcal{F})$. We define the total variation distance between ν_1 and ν_2 as

$$d_{TV}(\nu_1, \nu_2) = 2 \sup_{A \in \mathcal{F}} |\nu_1(A) - \nu_2(A)|.$$

Theorem 8.2 ***Convergence theorem***
Let $\{X_n\}_{n=0}^{\infty}$ be an omniparametric random cluster Markov chain with stationary distribution μ. Let X_0 follow some arbitrary initial distribution ρ and let $\rho^{(n)}$ denote the distribution of X_n. Then

$$\lim_{n\to\infty} d_{\text{TV}}(\rho^{(n)}, \mu) = 0$$

where $d_{\text{TV}}(\cdot,\cdot)$ is the total variation distance.

Proof:
Let $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ be the coupled omniparametric random cluster Markov chain. Let $X_0 \stackrel{\mathcal{D}}{=} \rho$ and let the chain $\{Y_n\}_{n=0}^{\infty}$ start in the stationary distribution μ. Then $Y_n \stackrel{\mathcal{D}}{=} \mu$ for all $n \geq 0$. We couple all four chains as in Definition 8.3. We bound the total variation distance between the measures of X_n and Y_n by using the coupling inequality [Lin92] in the following way

$$d_{\text{TV}}(X_n, Y_n) \leq 2\mathbb{P}(X_n \neq Y_n) \leq 2\mathbb{P}(\tilde{X}_n \neq \tilde{Y}_n)$$

since either $\tilde{X}_n \preceq X_n \preceq Y_n \preceq \tilde{Y}_n$ or $\tilde{X}_n \preceq Y_n \preceq X_n \preceq \tilde{Y}_n$. Then, due to Lemma 8.2 (page 91) we have

$$\lim_{n\to\infty} d_{\text{TV}}(X_n, Y_n) \leq \lim_{n\to\infty} 2\mathbb{P}(\tilde{X}_n \neq \tilde{Y}_n) = 0$$

and the result follows.
□

A consequence of the convergence theorem is uniqueness of the stationary distribution. If two different chains converge to something, they eventually coalesce according to Lemma 8.2, and therefore must converge to the same distribution.

Theorem 8.3 ***Uniqueness theorem***
The stationary distribution of the omniparametric random cluster Markov chain is unique.

Proof:
Suppose there are two omniparametric random cluster Markov chains

$$\{X_n\}_{n=0}^{\infty}\ ,\quad \{Y_n\}_{n=0}^{\infty}$$

and two distinct stationary distributions μ and μ' such that

$$X_n \xrightarrow{\mathcal{D}} \mu \text{ and } X'_n \xrightarrow{\mathcal{D}} \mu'$$

as $n \to \infty$. Since μ and μ' are distinct there exists $\varepsilon > 0$ such that $d_{\text{TV}}(\mu, \mu') \geq \varepsilon$. Now let $\{\tilde{X}_n, \tilde{Y}_n\}_{n=0}^{\infty}$ by the coupled omniparametric random cluster Markov chain in Definition 8.3. Couple $\{\tilde{X}_n, X_n, Y_n, \tilde{Y}_n\}_{n=0}^{\infty}$ in the same way as in Theorem 8.1. Since either

$$\tilde{X}_n \preceq X_n \preceq Y_n \preceq \tilde{Y}_n \text{ or } \tilde{X}_n \preceq Y_n \preceq X_n \preceq \tilde{Y}_n$$

holds we will eventually have $X_n = Y_n$ according to Lemma 8.2. This is a contradiction since

$$0 < \varepsilon \leq d_{\text{TV}}(\mu, \mu') \leq \lim_{n \to \infty} 2\mathbb{P}(X_n \neq X'_n) = 0$$

and we are done.
□

We are now finally ready for the definition of the omniparametric random cluster model. We say that an element in $\gamma \in [0,1]^E$ is a valid omniparametric configuration if $\mathbf{P}_p^{\text{RC}}(\gamma)$ is a valid random cluster configuration for all $p \in [0,1]$. The random cluster distribution assigns positive probability to any element in $\{0,1\}^E$ (as long as E is finite) and thus any element in $[0,1]^E$ is a valid omniparametric configuration.

Definition 8.5 ***The omniparametric random cluster distribution*** *Let $\{X_n\}_{n=0}^{\infty}$ be an omniparametric random cluster Markov chain according to Definition 8.2 on a graph $G = (V, E)$. We define the omiparametric random cluster distribution, denoted by ϕ_q, as the unique stationary distribution of $\{X_n\}_{n=0}^{\infty}$.*

In order for this definition to make sense the projected variable must have the correct distribution. We need

$$X \overset{\mathcal{D}}{=} \phi_q \;\Rightarrow\; \mathbf{P}_p^{\text{RC}}(X) \overset{\mathcal{D}}{=} \phi_{p,q}.$$

Before stating and prove the necessary theorem we introduce a result giving requirements to ensure convergence towards the stationary distribution for the fixed parameter random cluster Markov chain. The following result, here slightly reformulated, is taken from Norris' book on Markov chains [Nor97, Theorem 1.8.3] .

Lemma 8.3
Let $\{X_n\}_{n=0}^{\infty}$ be an irreducible and aperiodic Markov chain with stationary distribution μ and initial distribution ρ. Then

$$d_{TV}(\rho^{(n)}, \mu) \to 0$$

as $n \to \infty$, where $\rho^{(n)}$ is the distribution of X_n.

Theorem 8.4 ***Projection theorem***

$$X \stackrel{\mathcal{D}}{=} \phi_q \quad \Rightarrow \quad \mathbf{P}_p^{RC}(X) \stackrel{\mathcal{D}}{=} \phi_{p,q}$$

Proof:
Let $\{X_n\}_{n=0}^{\infty}$ be an omniparametric random cluster Markov chain according to Definition 8.2. Fix an arbitrary $q \geq 1$. If for each $p \in [0,1]$ the projected chain

$$\{Y_n\}_{n=0}^{\infty},\ Y_n = \mathbf{P}_p^{\mathrm{RC}}(X_n)$$

has the correct stationary distribution and

$$\lim_{n\to\infty} Y_n \stackrel{\mathcal{D}}{=} \phi_{p,q}$$

holds, we are done.

Fix arbitrary $p \in [0,1]$ and $n \geq 0$. Suppose X_n is such that $\mathbf{P}_p^{\mathrm{RC}}(X_n) \stackrel{\mathcal{D}}{=} \phi_{p,q}$. We update at the randomly selected edge e. We calculate p_t and use a random number generator to retrieve u, uniformly distributed over $[0,1]$, and update the omniparametric configuration

$$X_{n+1}(e) = f_{p_t}^{-1}(u)$$

at e. Note that $p \neq p_t$ a.s. so we only need to consider two cases for Y_{n+1}.

Case $p < p_t$: Let

$$Y_{n+1}(e) = \begin{cases} 1, & u < f_{p_t}(p) \\ 0, & u > f_{p_t}(p) \end{cases}$$

where $u < f_{p_t}(p) \ \Leftrightarrow \ p > f_{p_t}^{-1}(u)$ and $u \geq f_{p_t}(p) \Rightarrow p \leq f_{p_t}^{-1}(u)$.

Case $p > p_t$: Let

$$Y_{n+1}(e) = \begin{cases} 1, & u < f_{p_t}(p) \\ 0, & u > f_{p_t}(p) \end{cases}$$

where $u < f_{p_t}(p) \Rightarrow f_{p_t}^{-1}(u) \leq p$ and $u > f_{p_t}(p) \Rightarrow f_{p_t}^{-1}(u) > p$.

This is exactly the result we get when assigning $f_{p_t}^{-1}(u)$ to $X_{n+1}(e)$ and apply the projection mapping $\mathbf{P}_p^{\text{RC}}$. Thus

$$\mathbf{P}_p^{\text{RC}}(X_{n+1}) \stackrel{\mathcal{D}}{=} \phi_{p,q}$$

implying that $\phi_{p,q}$ is a stationary distribution for the projected chains. If $\{\mathbf{P}_p^{\text{RC}}(X_n)\}_{n=0}^{\infty}$ is irreducible and aperiodic then

$$d_{\text{TV}}(\rho^{(n)}, \phi_{p,q}) \to 0$$

follows from Lemma 8.3.

Let ξ_1 and ξ_2 be two arbitrary elements in $\{0,1\}^E$. Suppose that $\mathbf{P}_p^{\text{RC}}(X_n) = \xi_1$ and let ξ_1^o be one of the corresponding omniparametric states of ξ_1. Then

$$\xi_1(e) = 0 \;\; \Rightarrow \;\; \xi_1^o(e) > p \;\text{ and }\; \xi_1(e) = 1 \;\; \Rightarrow \;\; \xi_1^o(e) \leq p$$

for all edges $e \in E$. With positive probability we can find a sequence of edges $e_1, ..., e_m$ covering E and a sequence of number $u_1, ..., u_m$ such that if

$$X_{n+k} = \phi^q(u_k, e_k, X_{n+k-1}),\; k = 1, ..., m$$

then

$$\xi_2(e) = 0 \Rightarrow X_{n+m}(e) > p \;\text{ and }\; \xi_2(e) = 1 \Rightarrow X_{n+m}(e) \leq p$$

for every edge $e \in E$. Thus the projected chain is irreducible.

Consider $\mathbf{P}_p^{\text{RC}}(X_n)$ and let ξ be one of the corresponding omniparametric states, that is, $X_n = \xi$. When updating X_n to X_{n+1} it happens with positive probability that $U \in [p_t(p_t + q(1-p_t))^{-1}, p_t]$, making ξ and $\mathbf{P}_p^{\text{RC}}(X_n)$ aperiodic states. Aperiodicity for the $\{\mathbf{P}_p^{\text{RC}}(X_n)\}_{n=0}^{\infty}$ chain now follows from irreducibility and we can apply Lemma 8.3.

□

8.2 The Potts case

The Potts model, as we saw, is a two parameter family of models, having measure $\mu_{\beta,q}$, where β is the inverse temperature and q is the number of possible states for each vertex. The omniparametric Potts variable Ω_q is a coupling of the fixed parameter Potts variables, $\{\Omega_{\beta,q}\}_{\beta\in[0,\infty)}$, for all values of the inverse temperature, for fixed q. For the random cluster model we were able to construct the omniparametric random variable directly due to a monotonicity property in the random

cluster model. For the Potts model we will not construct the omniparametric random variable, instead we construct a sequence of fixed parameter variables covering the behavior for all $\beta \in [0, \infty)$, a sort of weak representation.

The idea is the following. We use the connection between the random cluster and Potts models and start with the omniparametric random cluster variable, Γ_q. It is only at these threshold values in Γ_q that something actually changes. Let us denote them $\mathcal{P} = \{p_1, ...p_{|E|}\}$. So for each $k \in \{1, ..., |E| - 1\}$ we have

$$\Gamma_{p,q} = \Gamma_{p',q}$$

for all $p, p' \in [p_k, p_{k+1})$, making the sequence $\{\Gamma_{p_k,q}\}_{k=1}^{|E|}$ completely capture the behaviour of the model for all $p \in [0, 1]$. It is this built in discreteness together with finite graphs that makes it possible to represent the behaviour of an uncountable class with a finite sequence.

We create our Potts random variable by assigning to each connected components in the random cluster graph a uniformly chosen spin, and do this independently for different components. Repeating this procedure for every inverse temperature, $\beta_k = -\log(1 - p_k)/2$, we get random variables with correct distributions. If we fix a vertex v the spin of v could change with every change in inverse temperature, and this is not what we want. We would like the spin of vertices to change only when their connected components are affected by a change in the corresponding random cluster variable.

When constructing the coupling of the Potts model for different temperatures we start at a very high temperature when the vertices are virtually independent. As the temperature decreases and dependence between vertices are introduced (as edges in the random cluster model are added) we merge connected components pairwise into new components. When a new components is formed it takes on the spin of the larger subcomponent. We hope by using this method to eliminate extra fluctuations coming from the extra source of randomness in the **H** mapping, introduced in Definition 7.11 (page 72).

The result of the procedure is a sequence of Potts variables describing the behavior of a system as the inverse temperature runs from zero to infinity.

The coupling

The coupling is rather involved. We need to keep track of two sequences of random variables. One sequence representing the evolution

of the fixed parameter random cluster model as p changes from zero to one, and one sequence for the corresponding Potts variables. The reason for keeping track of both sequences will be clear when we describe the procedure.

Let $G = (V, E)$ be a graph on which we generate an omniparametric random cluster sample γ_q. We start by constructing a sequence of fixed parameter random cluster variables

$$\gamma_{p_k,q} = \mathbf{P}^{rc}_{p_k}(\gamma_q)$$

where $p_k = X(e_k)$ for edges $e_k,\ k \in \{1, ..., |E|\}$. We will now construct a sequence of elements

$$\{(\omega_{\beta_k,q}, \gamma_{p_k,q})\}_{k=1}^{|E|}$$

in $\{1, ..., q\}^V \times \{0, 1\}^E$ where every $\omega_{\beta_k,q}$ is Potts variable corresponding to the random cluster variable $\gamma_{p_k,q}$. For each k we let $\gamma_{p_k,q} = \mathbf{P}^{\text{RC}}_{p_k}(\Gamma_q)$, and construct $\omega_{\beta_{k+1},q}$ recursively using operator $\mathbf{U}$.

Definition 8.6 ***Update operator***
Let $\mathbf{U} : \{1, ..., q\}^V \times \{0, 1\}^E \times \{0, 1\}^E \to \{1, ..., q\}^V \times \{0, 1\}^E$ *be the following mapping*

$$\mathbf{U}(\omega_k, \gamma_k, \gamma_{k+1}) = (\omega_{k+1}, \gamma_{k+1})$$

where the Potts configuration ω_{k+1} *is constructed as follows.*

1. *If* $\kappa(\gamma_{k+1}) < \kappa(\gamma_k)$ *two connected components* C_1, C_2 *has merged and we proceed with step 2. Otherwise let* $\forall v \in V : \omega_{k+1}(v) = \omega_k(v)$*, and stop.*

2. *If* $|C_1| \neq |C_2|$ *let* $C_{max} = \max\{C_1, C_2\}$ *and* $C_{min} = \min\{C_1, C_2\}$*. If* $|C_1| = |C_2|$ *let some predefined method decide which component is largest. Finally let* $a = \omega(u)$ *for some* $u \in C_{max}$*.*

3. *Let*

$$\omega_{k+1}(v) = \begin{cases} a, & v \in C_{min} \\ \omega_k(v), & \textit{otherwise} \end{cases}$$

The method for determining which component is largest in step two has to be predefined to ensure that $\mathbf{U}$ is a deterministic mapping. We are now ready to construct the sequence $\{(\Omega_{\beta_k,q}, \Gamma_{p_k,q})\}_{k=1}^{|E|}$.

Start: $(\Omega_{\beta_1,q},\ \Gamma_{p_1,q}) = (\ \mathbf{H}(\Gamma_{p_1,q}),\ \ \mathbf{P}^{\text{RC}}_{p_1}(\Omega_q)\ )$
Recursive step : $(\Omega_{\beta_{k+1},q},\ \Gamma_{p_{k+1},q}) = \mathbf{U}(\Omega_{\beta_k,q},\ \Gamma_{p_k,q},\ \Gamma_{p_{k+1},q})\ ,\quad k \geq 1$

Theorem 8.5
Let $\{(\Omega_{\beta_k,q}, \Gamma_{p_k,q})\}_{k=1}^{|E|}$ *be the sequence constructed above, and let* $p_k = 1 - e^{-2\beta_k}$. *Then for each* $k = 1, ..., |E|$, *the following holds.*

(i) $\Gamma_{p_k,q} \stackrel{\mathcal{D}}{=} \phi_{p_k,q}$

(ii) $\Omega_{\beta_k,q} \stackrel{\mathcal{D}}{=} \mu_{\beta_k,q}$

Proof:
Claim (i) follows directly from the definition of the omniparametric random cluster model. Claim (ii) follows if

$$\Omega_{\beta_k,q} \stackrel{\mathcal{D}}{=} H(\Gamma_{p_k,q}) \ , \quad k = 1, .., |E|$$

by Corollary 7.2, page 73.

We use induction over k to prove $\Omega_{\beta_k,q} \stackrel{\mathcal{D}}{=} H(\Gamma_{p_k,q})$. For $k = 1$ the statement follows by definition since $\Omega_{\beta_1,q} = \mathbf{H}(\Gamma_{p_1,q})$. For $k \geq 2$ let us consider the Potts random variable $\mathbf{H}(\Gamma_{p_k,q})$, suppose $\Omega_{\beta_k,q} \stackrel{\mathcal{D}}{=} \mu_{\beta_k,q}$. We must show that the spins of the connected components in $\Omega_{\beta_{k+1},q}$ are independent and uniformly distributed over $\{1, ..., q\}$. If the underlying graph $\Omega_{\beta_k,q}$ has the same number of connected components as $\Omega_{\beta_{k+1},q}$ then $\Omega_{\beta_k,q} = \Omega_{\beta_{k+1},q}$ and we are done. If not two components are merged and vertices in the smaller component is (deterministically) assigned the spin of the larger component. The spin of the larger component in $\Omega_{\beta_k,q}$ follows the uniform distribution over the spin set. Thus the distribution of the spin of the newly formed component also has uniform distribution, and we are done.
□

CHAPTER 9

Models for incomplete data

We define versions of the Ising and Potts models related to the parameter estimation procedures used in subsequent chapters. The basic idea is that given Ising or Potts data we estimate the random cluster connection probability between the origin and one of its neighbours. Given an estimate of the connection probability it is easy to calculate an estimate of the inverse temperature.

We need to define properly what we mean with *"data at locations in a finite set"*. One approach is to choose the finite set with some strategy in mind and use all points contained therein, or choose a finite region and then use only a subset of the implicated vertices. We would like a theory allowing us to freely choose a subset of any finite region, compute parameter estimates and construct confidence intervals. We take the first step towards such a theory and use boxes as our finite regions, and within these we choose our subset of vertices uniformly at random. This approach allow us to use available theory to construct both estimates and confidence intervals.

In order to capture the variance we use a central limit theorem by Bolthausen [Bol82] . This requires finite regions $\{\Lambda_n\}_{n=0}^{\infty}$ such that

$$\Lambda_n \subseteq \Lambda_{n+1},\ \Lambda_n \to \mathbb{Z}^d \text{ and } \frac{|\Lambda_n|}{|\partial\Lambda_n|} \to 0$$

as $n \to \infty$, making any vertex $v \in \mathbb{Z}^d$ belong to all but finitely many Λ_n.

We are however free to choose the shape of the Λ_n's as we like, and our sequence of boxes $\{\mathbb{B}_n\}_{n=0}^{\infty}$ fulfills these requirements.

The rest of the chapter is outlined as follows. We begin with a simple procedure for randomly selecting vertices in $\mathbb{Z}^d$ in Section 9.1. In Section 9.2 we focus on simple pairwise interactions between neighbouring vertices, developing theory both for the Ising and Potts models including central limit theorems. We continue with more general interactions in Section 9.3 and develop theory along the same lines. We finish with Section 9.4 with comments regarding strong mixing and central limit theorems.

9.1 Random point selection in $\mathbb{Z}^d$

We present here a simple method of generating sets of marked points. The idea is to generate these sets independently of the lattice process and then use them to define marked versions of the lattice process, in this case marked interaction versions of the Ising and Potts models.

Definition 9.1 ***Marked points***
To all vertices in $\mathbb{Z}^d$ we associate independent and identically distributed $Bernoulli(p_s)$ random variables, $M(\mathbb{Z}^d) = \{M(v),\ v \in \mathbb{Z}^d\}$. We let

$$M_1(\mathbb{Z}^d) = \{v : M(v) = 1\}$$

be the set of marked points. We call p_s the selection probability. When the underlying vertex set is understood we write M for the class of random variables and M_1 for the set of marked points. We let P_M denote the probability distribution for $M(\mathbb{Z}^d)$.

9.2 Pairwise interactions

9.2.1 The Ising case

When estimating the parameter in the Ising model we will base our estimate on pairwise interactions in a sample and for this purpose we define a version of the Ising model, the marked Ising interaction model.

It is relatively easy to compute point estimates, it is just a matter of estimating the expected value of the model at the origin as a function for all values of β and use its inverse. We would like to be able to construct confidence intervals also, but then we also need the distribution and variance of the observable which we don not have. Instead

state and prove a central limit theorem, which is only possible when the underlying Ising process is subcritical.

Definition 9.2 ***The marked Ising interaction model***
Let X be distributed according to the Ising distribution $\mu_\beta^{(n)}$ on $\mathbb{B}_n$ at some inverse temperature β and let $M = \{M(v) : v \in \mathbb{Z}_n\}$ be a set of marked points according to Definition 9.1 with selection probability p_s. Let $R = \{R(v) : v \in \mathbb{Z}^d\}$ be a class of independent identically $U\{1, ..., d\}$ distributed random variables. The marked Ising interaction model Y is defined vertex wise as

$$\forall v \in \mathbb{Z}_n : \; Y(v) = M(v)X(v) \sum_{k=1}^{d} I_{\{R(v)=k\}}(k)X(v + e_k)$$

where e_k is a unit vector with zero in every position except in position k. The distribution for this model is denoted by $\mu_{\beta,p_s}^{mark,(n)}$. We define the marked Ising interaction measure as a weak limit of distributions

$$\mu_{\beta,p_s}^{mark} = \lim_{n\to\infty} \mu_{\beta,p_s}^{mark,(n)}.$$

For questions regarding the limit construction of $\mu_{\beta,p_s}^{\text{mark}}$ see [GHM01] where the subject is treated for the ordinary Ising and Potts models. Similar arguments give the existence of $\mu_{\beta,p_s}^{\text{mark}}$.

The Potts, Ising and random cluster models on $\mathbb{Z}^d$ are all symmetric with respect to rearrangements of directions, meaning that

$$\mathbb{P}((x_1..., x_k, ..., x_l, ..., x_d) \leftrightarrow (x_1, ..., x_k + 1, ..., x_l, ..., x_d))$$

$$= \mathbb{P}((x_1, ..., x_k, ..., x_l, ..., x_d) \leftrightarrow (x_1, ..., x_k, ..., x_l + 1, ..., x_d))$$

for any $k, l \in \{1, ..., d\}$. This directional invariance allow us to use $e_1 = (1, 0, ..., 0)$ as our preferred direction when considering connection probabilities between neighbours.

In an application the natural way to proceed is to first construct a set of marked points $M_1 \subseteq \mathbb{Z}_n$, and a set of directions R. We observe the Ising model at sites in $M_1' = M_1 \cup \{v + e_{R(v)} : v \in M_1\}$ and write $X_M(v) = X(v)$ if $v \in M_1'$ and $X_M(v) = 0$ otherwise. For completeness let $M_0 = \mathbb{Z}_n \setminus M_1'$ denote the set of vertices where we do not observe the Ising configuration.

What remains is to establish a few simple facts about the marked Ising interaction model.

Theorem 9.1
Let Y follow the marked Ising interaction distribution, $\mu^{mark,(n)}_{\beta,p_s}$, and let $l, l' \in \mathbb{B}_n$ be arbitrary but distinct vertices such that $|l - l'| > 1$. Also let $\phi^{(n)}_{1-e^{-2\beta},2}$ be the corresponding random cluster distribution. Then

$$(i) \quad \mathbb{E}[Y(l)] = p_s \phi_{p,2}(l \leftrightarrow l + e_1)$$

$$(ii) \quad \mathit{Var}[Y(l)] = p_s - p_s^2 \phi_{p,2}(l \leftrightarrow l + e_1)^2$$

$$\begin{aligned}(iii) \quad \mathit{Cov}[Y(l), Y(l')] = \frac{p_s^2}{d^2} \sum_{k=1}^{d} \sum_{m=1}^{d} \big(& \phi_{p,2}(\{l \leftrightarrow l + e_k\} \cap \{l' \leftrightarrow l' + e_m\}) \\ & \phi_{p,2}(l \leftrightarrow l' \not\leftrightarrow l + e_k \leftrightarrow l' + e_m) \\ & \phi_{p,2}(l \leftrightarrow l' + e_m \not\leftrightarrow l + e_k \leftrightarrow l') \big) \\ - p_s^2 \, \phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})^2 &\end{aligned}$$

and for l, l' such that $|l - l'| = 1$ we have $l' = l + e_n$ for some $n \in \{1, ..., d\}$ and

$$\begin{aligned}(iv) \quad \mathit{Cov}[Y(l), Y(l')] \;=\; & \frac{p_s^2}{d^2} \sum_{m=1}^{d} \phi_{p,2}(l \leftrightarrow l' + e_m) \\ & + \frac{p_s^2}{d^2} \sum_{k \neq n} \sum_{m=1}^{d} \big(\phi_{p,2}(\{l \leftrightarrow l + e_k\} \cap \{l' \leftrightarrow l' + e_m\}) \\ & \qquad \phi_{p,2}(l \leftrightarrow l' \not\leftrightarrow l + e_k \leftrightarrow l' + e_m) \\ & \qquad \phi_{p,2}(l \leftrightarrow l' + e_m \not\leftrightarrow l + e_k \leftrightarrow l') \big) \\ & - p_s^2 \, \phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})^2\end{aligned}$$

follows.

Proof:
Straightforward calculations give us the result for the expectation

$$\begin{aligned}
\mathbb{E}[Y(l)] &= \mathbb{E}\left[M(v)\sum_{k=1}^{d} I_{\{R(v)=k\}}X(v)X(v+e_k)\right] \\
&= \mathbb{E}[M(l)]\ \sum_{k=1}^{d}\mathbb{E}[I_{\{R(v)=k\}}]\ \mathbb{E}[X(v)X(v+e_k)] \\
&= p_s\phi_{p,2}(l \leftrightarrow l+e_1)
\end{aligned}$$

and the variance.

$$\begin{aligned}
\mathbf{Var}[Y(l)] &= \mathbb{E}\left[M(l)^2\left(\sum_{k=1}^{d} I_{\{R(v)=k\}}X(v)X(v+e_k)\right)^2\right] - \mathbb{E}[Y(l)]^2 \\
&= p_s\ \mathbb{E}\left[\left(\sum_{k=1}^{d} I_{\{R(v)=k\}}X(v)^2X(v+e_k)^2\right)\right] - \mathbb{E}[Y(l)]^2 \\
&= p_s\sum_{k=1}^{d}\mathbb{E}[I_{\{R(v)=k\}}]\ \mathbb{E}[X(v)^2X(v+e_k)^2] - \mathbb{E}[Y(l)]^2 \\
&= p_s\ \frac{1}{d}\ \sum_{k=1}^{d}\ \mathbb{E}[X(v)^2X(v+e_k)^2] - \mathbb{E}[Y(l)]^2 \\
&= p_s - p_s^2\phi_{p,2}(l \leftrightarrow l+e_1)^2
\end{aligned}$$

The covariance

$$\mathbf{Cov}[Y(l),Y(l')] = \frac{p_s^2}{d^2}\sum_{k=1}^{d}\sum_{m=1}^{d}\ \mathbf{Cov}[X(l)X(l+e_k),X(l')X(l'+e_m)]$$

$$
\begin{aligned}
= \quad & \frac{p_s^2}{d^2} \sum_{k=1}^{d} \sum_{m=1}^{d} \Big(\phi_{p,2}^{(n)}(\{l \leftrightarrow l + e_k\} \cap \{l' \leftrightarrow l' + e_m\}) \\
& \qquad + \phi_{p,2}^{(n)}(\{l \leftrightarrow l' \not\leftrightarrow l + e_k \leftrightarrow l' + e_m\}) \\
& \qquad + \phi_{p,2}^{(n)}(\{l \leftrightarrow l' + e_m \not\leftrightarrow l + e_k \leftrightarrow l'\}) \Big) \\
& - p_s^2 \ \phi_{p,2}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})^2
\end{aligned}
$$

follows from Lemma 7.5 and since $\phi_{p,2}^{(n)}(l \leftrightarrow l + e_k) = \phi_{p,2}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})$ for any l and any k. Also if $|l' - l| = 1$, we have $l' = l + e_n$ for some $n \in \{1, ..., d\}$) and

$$
\begin{aligned}
\mathrm{Cov}[Y(l), Y(l')] \quad = \quad & \frac{p_s^2}{d^2} \ \sum_{k=1}^{d} \sum_{m=1}^{d} \mathrm{Cov}[X(l)X(l + e_k), X(l')X(l' + e_m)] \\
= \quad & \frac{p_s^2}{d^2} \left(\sum_{m=1}^{d} \mathrm{Cov}[X(l)X(l + e_n), X(l')X(l' + e_m)] \right. \\
& \left. + \sum_{k \neq n} \sum_{m=1}^{d} \mathrm{Cov}[X(l)X(l + e_k), X(l')X(l' + e_m)] \right) \\
= \quad & \frac{p_s^2}{d^2} \sum_{m=1}^{d} \phi_{p,2}(l \leftrightarrow l' + e_m) \\
& + \frac{p_s^2}{d^2} \ \sum_{k \neq n} \sum_{m=1}^{d} \big(\ \phi_{p,2}(\{l \leftrightarrow l + e_k\} \cap \{l' \leftrightarrow l' + e_m\}) \\
& \qquad \phi_{p,2}(l \leftrightarrow l' \not\leftrightarrow l + e_k \leftrightarrow l' + e_m) \\
& \qquad \phi_{p,2}(l \leftrightarrow l' + e_m \not\leftrightarrow l + e_k \leftrightarrow l') \ \big) \\
& - p_s^2 \ \phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})^2
\end{aligned}
$$

follows, and we are done.
□

9.2.2 The Potts case

We now generalize to the Potts model. Instead of using pairwise products as in the Ising model we use the indicator function for the event that the colors at two sites are the same.

Definition 9.3 ***The marked Potts interaction model***
Let X be distributed according to the Potts distribution $\mu^{(n)}_{\beta,q}$ at some inverse temperature β and $q \geq 1$. Let M be a set of marked points according to Definition 9.1 with selection probability p_s. Let $\{R(v) : v \in \mathbb{Z}^d\}$ be a class of independent identically distributed $U\{1, ..., d\}$ random variables. The marked Potts interaction model Y is defined as

$$\forall v \in \mathbb{Z}_n.\ Y(v) = M(v) \sum_{k=1}^{d} I_{\{R(v)=k\}}(2I_{\{X(v)=X(v+e_k)\}} - 1)$$

where e_k is a unit vector in $\mathbb{Z}^d$ with a zero in every position except in position k. The distribution for the model is denoted by $\mu^{mark,(n)}_{\beta,q,p_s}$. We define the marked Potts interaction measure as a weak limit of distributions

$$\mu^{mark}_{\beta,q,p_s} = \lim_{n\to\infty} \mu^{mark,(n)}_{\beta,q,p_s}.$$

We finish this section by establishing a few simple facts about marked Potts interaction model.

Theorem 9.2
Let Y follow the marked Potts interaction distribution, let X be distributed according to the Potts distribution and let l and l' be two arbitrary but distinct vertices in $\mathbb{B}_n$ such that $|l - l'| > 1$. Also let $\phi^{(n)}_{1-e^{-2\beta},q}$ be the corresponding random cluster distribution. Then

$$(i) \quad \mathbb{E}[Y(l)] = p_s \left(\frac{2}{q} + 2\left(1 - \frac{1}{q}\right) \phi_{p,q}(l \leftrightarrow l + e_1) - 1 \right)$$

$$(ii) \quad \mathit{Var}[Y(l)] = p_s - p_s^2 \left(\frac{2}{q} + 2\left(1 - \frac{1}{q}\right) \phi_{p,q}(l \leftrightarrow l + e_1) - 1 \right)^2$$

$$(iii) \quad \mathbf{Cov}[Y(l), Y(l')] = \frac{4p_s^2}{d^2} \sum_{k=1}^{d} \sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)) - p_s^2 \left(\frac{1}{q} + 2 \left(1 - \frac{1}{q} \right) \phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1 \right)^2$$

and for l, l' such that $|l - l'| = 1$ we have $l' = l + e_n$ for some $n \in \{1, ..., d\}$ and

$$(iv) \quad \mathbf{Cov}[Y(l), Y(l')] = \frac{4p_s^2}{d^2} \left(\sum_{m=1}^{d} \left(\frac{1}{q} + 2 \left(1 - \frac{1}{q} \right) \phi_{p,q}(l \leftrightarrow l' + e_m) - 1 \right) + \sum_{k \neq n} \sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)) \right) - p_s^2 \left(\frac{1}{q} + 2 \left(1 - \frac{1}{q} \right) \phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1 \right)^2$$

where $X(l,k) = \{X(l) = X(l + e_k)\}$ and $X(l,l',k) = \{X(l) = X(l' + e_k)\}$.

Proof:
We start with an observation regarding the probability of equal configuration at two different sites. Remember the connection between the Potts and random cluster models, we may obtain a Potts sample by simply colour connected components in a random cluster sample at random.

$$\begin{aligned} \mathbb{P}(X(v) = X(v + e_k)) &= \phi_{p,q}(v \leftrightarrow v + e_k) + \frac{1}{q} \phi_{p,q}(v \not\leftrightarrow v + e_k) \\ &= \frac{1}{q} + \left(1 - \frac{1}{q} \right) \phi_{p,q}(v \leftrightarrow v + e_k) \end{aligned}$$

Now, the expected value at a location l follows.

$$\begin{aligned}
\mathbb{E}[Y(l)] &= \mathbb{E}\left[M(v)\sum_{k=1}^{d} I_{\{R(v)=k\}}(2I_{\{X(v)=X(v+e_k)\}}-1)\right] \\
&= \mathbb{E}[M(l)]\ \sum_{k=1}^{d}\mathbb{E}[I_{\{R(v)=k\}}]\ (2\mathbb{P}(X(v)=X(v+e_k))-1) \\
&= \frac{p_s}{d}\ \sum_{k=1}^{d}\left(2\left(\frac{1}{q}+\left(1-\frac{1}{q}\right)\ \phi_{p,q}(v\leftrightarrow v+e_k)\right)-1\right) \\
&= p_s\ \left(\frac{2}{q}+2\left(1-\frac{1}{q}\right)\phi_{p,q}(v\leftrightarrow v+e_1)-1\right)
\end{aligned}$$

Preparing for the variance ...

$$\begin{aligned}
\mathbb{E}[Y(l)^2] &= \mathbb{E}\left[\left(M(v)\sum_{k=1}^{d} I_{\{R(v)=k\}}(2I_{\{X(v)=X(v+e_k)\}}-1)\right)^2\right] \\
&= \mathbb{E}[M(v)^2]\ \mathbb{E}\left[\sum_{k=1}^{d} I_{\{R(v)=k\}}\ \underbrace{(2I_{\{X(v)=X(v+e_k)\}}-1)^2}_{=1}\right] \\
&= \mathbb{E}[M(v)] \\
&= p_s
\end{aligned}$$

... calculating the variance ...

$$\mathrm{Var}[Y(l)] = p_s - p_s^2\ \left(\frac{2}{q}+2\left(1-\frac{1}{q}\right)\phi_{p,q}(v\leftrightarrow v+e_1)-1\right)^2$$

... finally the covariance ...

$$\mathrm{Cov}[Y(l),Y(l')] = \frac{p_s^2}{d^2}\ \sum_{k=1}^{d}\sum_{m=1}^{d}\mathrm{Cov}[2I_{X(l,k)}-1, 2I_{X(l',m)}-1]$$

$$
\begin{aligned}
&= \frac{4p_s^2}{d^2} \sum_{k=1}^{d}\sum_{m=1}^{d} \mathrm{Cov}[I_{X(l,k)}, I_{X(l',m)}] \\
&= \frac{4p_s^2}{d^2} \sum_{k=1}^{d}\sum_{m=1}^{d} \Big(\mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)) \\
&\quad - \mu_{p,q}^{(n)}(X(l) = X(l+e_k))\mathbb{P}(X(l') = X(l'+e_m)) \Big) \\
&= \frac{4p_s^2}{d^2} \sum_{k=1}^{d}\sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)) \\
&\quad - p_s^2 \left(\frac{1}{q} + 2\left(1-\frac{1}{q}\right)\phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1\right)^2
\end{aligned}
$$

and if $|l' - l|$ we have $l' = l + e_n$ for some $n \in \{1, ..., d\}$ and

$$
\mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)\}) = \mu_{p,q}^{(n)}(X(l,l',m)) = \frac{1}{q} + \left(1-\frac{1}{q}\right)\phi_{p,q}(l \leftrightarrow l' + e_m)
$$

follows, and as a consequence we also have

$$
\begin{aligned}
\mathrm{Cov}[Y(l), Y(l')] &= \frac{4p_s^2}{d^2} \sum_{k=1}^{d}\sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,k) \cap X(l',m)) \\
&\quad - p_s^2 \left(\frac{1}{q} + 2\left(1-\frac{1}{q}\right)\phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1\right)^2 \\
&= \frac{4p_s^2}{d^2} \left(\sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,l',m)) \right. \\
&\quad \left. + \sum_{k \neq n}\sum_{m=1}^{d} \mu_{p,q}^{(n)}(X(l,n) \cap X(l',m)) \right) \\
&\quad - p_s^2 \left(\frac{1}{q} + 2\left(1-\frac{1}{q}\right)\phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1\right)^2
\end{aligned}
$$

$$= \frac{4p_s^2}{d^2}\left(\sum_{m=1}^{d} p_s^2\left(\frac{1}{q}+2\left(1-\frac{1}{q}\right)\phi_{p,q}(l\leftrightarrow l'+e_m)-1\right)^2 + \sum_{k\neq n}\sum_{m=1}^{d}\mu_{p,q}^{(n)}(X(l,n)\cap X(l',m))\right)$$

$$- p_s^2\left(\frac{1}{q}+2\left(1-\frac{1}{q}\right)\phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{1})-1\right)^2$$

and finally we are done.
□

9.2.3 Central limit theorems

Next we present the central limit theorem by Bolthausen [Bol82] we use for the marked interaction models. First some definitions. Let $\{\Lambda_n\}_{n=1}^{\infty}$ be a sequence of subsets in $\mathbb{Z}^d$ such that

$$\lim_{n\to\infty}\Lambda_n = \mathbb{Z}^d \quad \text{and} \quad \lim_{n\to\infty}\frac{|\partial\Lambda_n|}{|\Lambda_n|} = 0.$$

For $\Lambda \subseteq \mathbb{Z}^d$ let $\mathcal{F}_\Lambda = \sigma(\{X(v) : v \in \Lambda\})$ be the σ-algebra of events on Λ. We defined the distance function between sets as

$$d(A,B) = \inf_{x\in A, y\in B}|x-y|$$

for any two $A, B \in \mathbb{Z}^d$. Also let $L_2(\mathcal{F}_\Lambda)$ be the set of $\mathcal{F}_\Lambda$ measurable random variables with finite second moments. The mixing coefficients, $\alpha_{k,l}(m)$ and $\rho(m)$, used when defining so called strong mixing and ρ-mixing respectively, are defined as follows.

$$\alpha_{k,l}(m) = \sup\left\{ |\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_1)\mathbb{P}(A_2)| \quad : \quad \begin{array}{l} A_i\in\mathcal{F}_{\Lambda_i},\ i=1,2,\\ |\Lambda_1|\le k,\ |\Lambda_2|\le l,\\ d(\Lambda_1,\Lambda_2)\ge m\end{array}\right\}$$

$$\rho(m) = \sup\{|\mathrm{Cov}[Y_1,Y_2]| : Y_i\in L_2(\mathcal{F}_{\{v_i\}}),\ \mathbb{E}[|Y_i|^2]\le 1,\ i=1,2,\ d(v_1,v_2)\ge m\}$$

We proceed with the theorem.

Theorem 9.3 ***Bolthausen's central limit theorem***
Let $\{X(v) : v \in \mathbb{Z}^d\}$ be a real valued stationary random field with shift invariant joint laws such that $\mathbb{E}[|X(v)|^2] < \infty$ for all $v \in \mathbb{Z}^d$. Also let $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequence of subsets of $\mathbb{Z}^d$ defined above. If

$$\sum_{m=1}^{\infty} m^{d-1} \alpha_{k,l}(m) < \infty$$

for $k + l \leq 4$, $\alpha_{1,\infty}(m) = o(m^{-d})$ and if either (a) or (b) as defined below

$$(a) \quad \sum_{m=1}^{\infty} m^{d-1} \rho(m) < \infty$$

$$(b) \quad \exists \delta > 0 \ \text{ such that } \ v \in \mathbb{Z}^d : \ \mathbb{E}[|X(v)|^{2+\delta}] < \infty$$

$$\text{and } \ \sum_{m=1}^{\infty} m^{d-1} \alpha_{1,1}(m)^{\delta/(2+\delta)} < \infty$$

is satisfied then we have

$$\sigma_X^2 = \sum_{v \in \mathbb{Z}^d} Cov[X(0), X(v)] < \infty$$

and

$$\sum_{v \in \Lambda_n} \frac{(X(v) - \mathbb{E}[X(v)])}{\sigma_X |\Lambda_n|^{1/2}} \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$.

Next we use Bolthausen's theorem to prove similar central limit theorems for the marked Ising and marked Potts interaction models.

Theorem 9.4 ***Marked Ising interaction CLT***
Let Y be distributed according to the marked Ising interaction measure with parameter β. Suppose the selection probability is p_s. Let $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequence of finite boxes such that $\Lambda_n \to \mathbb{Z}^d$. If the underlying Ising measure is subcritical then

$$\sum_{v \in \Lambda_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\Lambda_n|^{1/2}} \xrightarrow{\mathcal{D}} Z$$

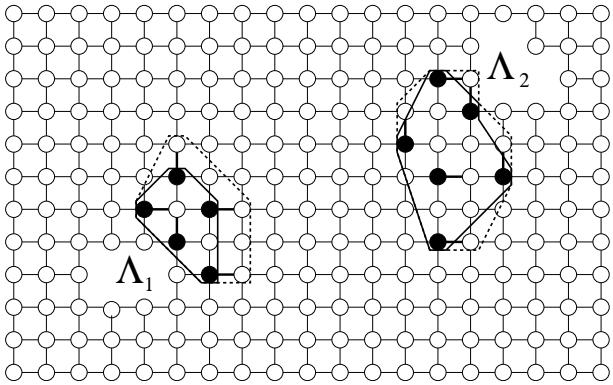

Figure 9.1: *An illustration of vertex sets Λ_1 and Λ_2 in Theorem 9.4.*

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$ *and*

$$\sigma_Y^2 = \sum_{v \in \mathbb{Z}^d} \mathit{Cov}[Y(0), Y(v)]$$

Proof:
Consider the mixing coefficient for strong mixing, $\alpha_{k,l}(m)$ for any finite k and l. Let the Ising configuration X we based Y on be generated using the random cluster measure $\phi_{1-e^{-2\beta},q}$, and let W be the corresponding random cluster sample.

Suppose that X is subcritical, and as a consequence the underlying random cluster variable is also subcritical. Let $\Lambda_1,\ \Lambda_2$ be two arbitrary but fixed finite subsets in $\mathbb{Z}^d$ such that $d(\Lambda_1, \Lambda_2) \geq m$. If Λ_1 has diameter d_1 then event A_1 depends on vertices in a set with diameter at most d_1+2 since we are using pairwise interactions between neighbours. Now let D be the event that there exists a cluster B such that $\Lambda_1 \cap B \neq \emptyset$ and $\Lambda_2 \cap B \neq \emptyset$, then B has diameter at least $d(\Lambda_1, \Lambda_2) - 2$, see Figure 9.1. Note that $\mathbb{P}(D) > 0$ for any finite distance between Λ_1 and Λ_2, and also

$$\mathbb{P}(D) \leq Ce^{-\gamma(m-2)},\ C, \gamma > 0$$

for some $C, \gamma \in (0, \infty)$ according to Grimmet and Piza [GP97, page 2] . Then for any $A_i \in \mathcal{F}_{\Lambda_i},\ i \in \{1, 2\}$, we have

$$\begin{aligned}
&|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| \\
&\quad = |\mathbb{P}(A_1 \cap A_2 \cap D) + \mathbb{P}(A_1 \cap A_2 \cap D^c) \\
&\qquad - (\mathbb{P}(A_1 \cap D) + \mathbb{P}(A_1 \cap D^c))(\mathbb{P}(A_2 \cap D) \\
&\qquad + \mathbb{P}(A_2 \cap D^c))| \\
&\quad = |\mathbb{P}(A_1 \cap A_2 \cap D) + \mathbb{P}(A_1 \cap A_2 \cap D^c) \\
&\qquad - \mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2 \cap D^c) \\
&\qquad - \mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D^c) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2 \cap D)| \\
&\quad \leq |\mathbb{P}(A_1 \cap A_2 \cap D) - \mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D)| \\
&\qquad + |\mathbb{P}(A_1 \cap A_2 \cap D^c) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2 \cap D^c)| \\
&\qquad + |\mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D^c) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2 \cap D)| \\
&\quad \overset{(*)}{=} |\mathbb{P}(A_1 \cap A_2 \cap D) - \mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D)| \\
&\qquad + |\mathbb{P}(A_1 \cap D)\mathbb{P}(A_2 \cap D^c) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2 \cap D)| \\
&\quad = \mathbb{P}(D)\ (\ |\mathbb{P}(A_1 \cap A_2|D) - \mathbb{P}(A_1|D)\mathbb{P}(A_2 \cap D)| \\
&\qquad + |\mathbb{P}(A_1|D)\mathbb{P}(A_2 \cap D^c) - \mathbb{P}(A_1 \cap D^c)\mathbb{P}(A_2|D)|\ ) \\
&\quad \leq 2\mathbb{P}(D) = 2\phi_{p,2}(D) \leq Ce^{-\gamma m}
\end{aligned}$$

for some $C, \gamma > 0$, where the equality at $(*)$ comes from the fact that on D^c the events A_1 and A_2 are independent. Since the bound is independent of the sets Λ_1, Λ_2 it also holds for the supremum over all such pairs, and we have $\alpha_{k,l}(m) \leq Ce^{-\gamma m}$. Let $l \to \infty$ and define $\alpha_{k,\infty}(m)$ as

$$\alpha_{k,\infty}(m) = \lim_{l \to \infty} \alpha_{k,l}(m) \leq Ce^{-\gamma m}.$$

Let $\Lambda_1 = \{v\}$ for arbitrary but fixed $v \in \mathbb{Z}^d$. We use the same argument as for finite l to establish that $\alpha_{k,\infty}(m) \leq Ce^{-\gamma m}$. It now follows that

$$\sum_{m=1}^{\infty} m^{d-1}\alpha_{k,l}(m) \leq C \sum_{m=1}^{\infty} m^{d-1}e^{-\gamma m} < \infty$$

and $\alpha_{1,\infty}(m) = o(m^{-d})$ from $\alpha_{k,\infty}(m) \leq Ce^{-\gamma m}$ for $k < \infty$. For requirement (b) note that for any $\delta > 0$ we have

$$\mathbb{E}[|X(v)|^{2+\delta}] < \infty$$

since $Y(v) \in \{-1, 0, +1\}$ for all $v \in \mathbb{Z}^d$. Also

$$\sum_{m=1}^{\infty} m^{d-1} \alpha_{1,1}(m)^{\delta/(2+\delta)} \leq C \sum_{m=1}^{\infty} m^{d-1} e^{-\gamma\delta m/(2+\delta)} < \infty$$

holds since $\gamma\delta/(2+\delta) > 0$. The result now follows from Bolthausen's theorem and we are done.
□

The central limit theorems only work if correlations in the underlying Ising model decay exponentially. For the Ising models this happens throughout the subcritical phase (according to Grimmett and Piza [GP97]), when

$$\beta < \beta_c = \frac{1}{2} \log(1 + \sqrt{2}),$$

corresponding to a critical p-value

$$p_c = \frac{\sqrt{2}}{1 + \sqrt{2}}$$

in the random cluster model.

For the q-state Potts model we do not know if correlations decay exponentially whenever $\beta < \beta_c$, but we do know that if correlations decay sufficiently fast then exponential decay follows. In the same spirit as Grimmett and Piza [GP97] we define an alternative critial point

$$\beta_g(q) = -\frac{1}{2} \log(1 - p_g(q))$$

for β where $p_g(q)$ is according to Definition 7.8. For $\beta < \beta_g(q)$ we have, due to Theorem 7.7, exponential decay for correlations. Due to the special connection between the Ising/Potts model and the random cluster model we can express Ising and Potts correlations using random cluster connection probabilities. The proof of Theorem 9.4 is based entirely on this connection and the exponential decay of correlations. We therefore state a version of Theorem 9.4 for the Potts model without proof, the extension is tedious but straightforward.

First a note on susceptibility. There is no explicit expression for it, instead we use simulation to make estimates. The same holds for the susceptibility in the marked Potts interaction model. For both characteristics see Definition 7.12 (page 74) and simulation results in Section 10.2 (page 137).

Theorem 9.5 ***Marked Potts interaction CLT***
Let Y be distributed according to the marked Potts interaction measure with parameters β and q. Suppose the selection probability is p_s. Let $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequence of finite boxes such that $\Lambda_n \to \mathbb{Z}^d$. If

$$\beta < \ln \sqrt{1 + \frac{1}{\sqrt{2}}}$$

then

$$\sum_{v \in \Lambda_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\Lambda_n|^{1/2}} \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$ and

$$\sigma_Y^2 = \sum_{v \in \mathbb{Z}^d} \mathrm{Cov}[Y(0), Y(v)]$$

9.3 General interactions

Consider the Ising or Potts model on the lattice $\mathbb{Z}^d$, and let X be distributed according to μ_β or $\mu_{\beta,q}$. Suppose we have a sequence of measurements around certain vertices, $l_k,\ k = 1, ..., n$. We do some computation based on measurements around each of the l_k's and assign the value, denoted $Y_k = Y(l_k)$ to l_k. The computed values, $Y_1, ..., Y_n$, then serves as a base for further analysis.

We start with defining the so called measurement method, and its two parts, the measurement configuration and the measurement function. We proceed with the definitions of the general Ising and Potts interaction models along with central limit theorems. We finish this section with an example of general interactions, so called box interactions, using all available interactions within a small box.

Definition 9.4 ***Measurement configuration***
Let $B_R \subset \mathbb{R}^d$ be a finite convex set containing the origin in its interior such that

$$\sup_{u,v \in B_R} d(u,v) \leq L$$

for some $L \in (0, \infty)$. A measurement configuration B is the set of integer points in B_R, and we say that it has range L.

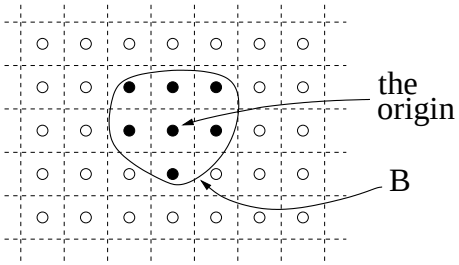

Figure 9.2: *An example of a measurement configuration around the origin. The value at the origin is calculated based on the values at vertices $u_1, ..., u_7$. The measurement configurations around other vertices are just translations of the set B and its vertices $u_1, ..., u_7$.*

Definition 9.5 ***Measurement function***
Let $f : \mathbb{R}^m \to \mathbb{R}$ be a bounded function, and let $\{X(v) : v \in \mathbb{Z}^d\}$ be a random field. We say that f is a measurement function if its domain is $\{X(v) : v \in B\}$ for some measurement configuration B.

When we write "we have a measurement method" we mean that both the measurement configuration and the measurement function are fixed. Let (B, f) denote the measurement method. We finish this section with definitions of the two models.

Definition 9.6 ***The general Ising interaction model***
Let X be distributed according to the Ising measure μ_β at inverse temperature β. Let (B, f) be a range L measurement method on $\mathbb{B}_n$ and suppose that M is a set of marked points according to Definition 9.1 with selection probability p_s. The general Ising interaction model Y is defined as

$$Y(v) = M(v)f(v),\ v \in \mathbb{Z}^d.$$

The distribution for the model is denoted by $\mu^{g,(n)}_{\beta,f,M,p_s}$. We define the general Ising interaction measure as the weak limit of distributions

$$\mu^{g}_{\beta,f,M,p_s} = \lim_{n\to\infty} \mu^{g,(n)}_{\beta,f,M,p_s}.$$

Definition 9.7 ***The general Potts interaction model***
Let X be distributed according to the Potts measure $\mu_{\beta,q}$ at inverse temperature β and some $q \geq 1$. Let (B, f) be a range L measurement configuration on $\mathbb{B}_n$ and suppose that M is a set of marked points according to Definition 9.1 with selection probability p_s. The general Potts interaction model Y is defined as

$$Y(v) = M(v)f(v),\ v \in \mathbb{Z}^d.$$

The distribution for the model is denoted by $\mu^{g,(n)}_{\beta,q,f,L,p_s}$. We define the general Potts interaction measure as the weak limit of distributions

$$\mu^{g}_{\beta,f,M,p_s,=} \lim_{n\to\infty} \mu^{g,(n)}_{\beta,f,M,p_s,.}$$

9.3.1 Central limit theorems

The conclusion in central limit theorem by Bolthausen is based on mixing conditions. By a slight modification of the proof of Theorem 9.4 we are able to adapt the central limit theorems for the pairwise interaction models to the general interaction case. The requirement is that all interactions are of finite range.

Theorem 9.6 ***General Ising interaction CLT***
Let Y be distributed according to the general Ising interaction measure, μ^{g}_{β,f,M,p_s}. Let $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequence of finite boxes such that $\Lambda_n \to \mathbb{Z}^d$. If the underlying Ising measure is subcritical then

$$\sum_{v\in\Lambda_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\Lambda_n|^{1/2}} \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$ and

$$\sigma_Y^2 = \sum_{v\in\mathbb{Z}^d} \mathit{Cov}[Y(0), Y(v)]$$

Proof:
Consider the mixing coefficient for strong mixing, $\alpha_{k,l}(m)$ for any finite k and l. Let the Ising configuration X we based Y on be generated using the random cluster measure $\phi_{1-e^{-2\beta},q}$, and let W be the corresponding random cluster sample.

Suppose that X is subcritical, and as a consequence the underlying random cluster variable is subcritical. Let $\Lambda_1,\ \Lambda_2$ be two arbitrary but fixed finite subsets in $\mathbb{Z}^d$ such that $d(\Lambda_1,\Lambda_2)\geq m$. If Λ_1 has diameter d_1 then event A_1 depends on vertices in a set with diameter at most d_1+M. Now let D be the event that there exists a cluster B such that $\Lambda_1\cap B\neq\emptyset$ and $\Lambda_2\cap B\neq\emptyset$, then B has diameter at least $d(\Lambda_1,\Lambda_2)-2M$. Note that $\mathbb{P}(D)>0$ for any finite distance between Λ_1 and Λ_2, and $\mathbb{P}(D)\leq Ce^{-\gamma(m-2M)},\ C,\gamma>0$ according to Grimmett and Piza [GP97] (page 2). Then for any $A_i\in\mathcal{F}_{\Lambda_i},\ i\in\{1,2\}$, we have the following.

$$
\begin{aligned}
&|\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_1)\mathbb{P}(A_2)| \\
&\qquad = |\mathbb{P}(A_1\cap A_2\cap D)+\mathbb{P}(A_1\cap A_2\cap D^c) \\
&\qquad\quad -(\mathbb{P}(A_1\cap D)+\mathbb{P}(A_1\cap D^c))(\mathbb{P}(A_2\cap D) \\
&\qquad\quad +\mathbb{P}(A_2\cap D^c))| \\
&\qquad = |\mathbb{P}(A_1\cap A_2\cap D)+\mathbb{P}(A_1\cap A_2\cap D^c) \\
&\qquad\quad -\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2\cap D^c) \\
&\qquad\quad -\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D^c)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2\cap D)| \\
&\qquad \leq |\mathbb{P}(A_1\cap A_2\cap D)-\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D)| \\
&\qquad\quad +|\mathbb{P}(A_1\cap A_2\cap D^c)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2\cap D^c)| \\
&\qquad\quad +|\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D^c)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2\cap D)| \\
&\qquad \overset{(*)}{=} |\mathbb{P}(A_1\cap A_2\cap D)-\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D)| \\
&\qquad\quad +|\mathbb{P}(A_1\cap D)\mathbb{P}(A_2\cap D^c)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2\cap D)| \\
&\qquad = \mathbb{P}(D)\ (\ |\mathbb{P}(A_1\cap A_2|D)-\mathbb{P}(A_1|D)\mathbb{P}(A_2\cap D)| \\
&\qquad\quad +|\mathbb{P}(A_1|D)\mathbb{P}(A_2\cap D^c)-\mathbb{P}(A_1\cap D^c)\mathbb{P}(A_2|D)|\ ) \\
&\qquad \leq 2\mathbb{P}(D)=2\phi_{p,2}(D)\leq Ce^{-\gamma(m-2M)}
\end{aligned}
$$

for some $C,\gamma>0$, where the equality at $(*)$ comes from the fact that on D^c the events A_1 and A_2 are independent. Since the bound is independent of the sets Λ_1,Λ_2, it also holds for the supremum over all such pairs, and we have $\alpha_{k,l}(m)\leq Ce^{-\gamma(m-2M)}$.

Let $l\to\infty$ and define $\alpha_{k,\infty}(m)$ as

$$\alpha_{k,\infty}(m)=\lim_{l\to\infty}\alpha_{k,l}(m)\leq Ce^{-\gamma(m-2M)}.$$

Let $\Lambda_1=\{v\}$ for arbitrary but fixed $v\in\mathbb{Z}^d$. Then we can use the same argument as for finite l to establish that $\alpha_{k,\infty}(m)\leq Ce^{-\gamma(m-2M)}$. It now

follows that

$$\sum_{m=1}^{\infty} m^{d-1}\alpha_{k,l}(m) \leq C \sum_{m=1}^{\infty} m^{d-1}e^{-\gamma(m-2M)} < \infty$$

and $\alpha_{1,\infty}(m) = o(m^{-d})$ from $\alpha_{k,\infty}(m) \leq Ce^{-\gamma(m-2M)}$ for $k < \infty$. For the last requirement (b) note that for any $\delta > 0$ we have

$$\mathbb{E}[|X(v)|^{2+\delta}] < \infty$$

since the measurement function is bounded and

$$\sum_{m=1}^{\infty} m^{d-1}\alpha_{1,1}(m)^{\delta/(2+\delta)} \leq C \sum_{m=1}^{\infty} m^{d-1}e^{-\gamma\delta(m-2M)/(2+\delta)} < \infty$$

since $\gamma\delta/(2+\delta) > 0$. The result now follows from Bolthausen's theorem, and we are done.
□

We state corresponding theorem for the Potts case without proof, the extension is as before, tedious but straightforward.

Theorem 9.7 ***General Potts interaction CLT***
Let Y be distributed according to the general Potts interaction measure, μ^g_{q,β,f,M,p_s}. Let $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequence of finite boxes such that $\Lambda_n \to \mathbb{Z}^d$. If

$$\beta < \ln\sqrt{1 + \frac{1}{\sqrt{2}}}$$

then

$$\sum_{v\in\Lambda_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y|\Lambda_n|^{1/2}} \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$ and

$$\sigma_Y^2 = \sum_{v\in\mathbb{Z}^d} \mathit{Cov}[Y(0), Y(v)]$$

9.3.2 An example: box interactions

There are a many possibilities when defining general interaction models. In this thesis we restrict ourselves to study one of them, here

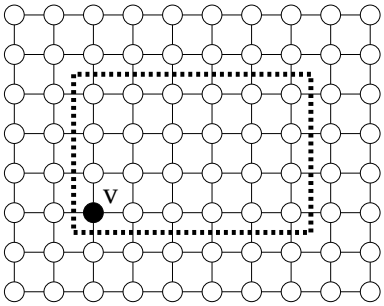

Figure 9.3: *An example of the measurement configuration $B(v)$ (within the dashed line) around a vertex v. Here we use $a = 3$ and $b = 5$.*

denoted box integrations. Instead of using pairwise neighbour interactions we use pairwise interactions between vertices within a box or rectangle at some vertex v. For an example of the measurement configuration see Figure 9.3. The corresponding measurement function is a simple average of pairwise interactions within the $a \times b$-box.

We are now ready to define the models for the Ising and Potts cases respectively, and prove basic results.

The Ising case

First the definition is the same spririt as in previous chapters.

Definition 9.8 ***The marked Ising box interaction model***
Let X be distributed according to the Ising distribution $\mu_\beta^{(n)}$ on $\mathbb{B}_n$ at inverse temperature β. Fix arbitrary but finite $a, b \in \mathbb{Z}^+$. For measurement configuration $B = ([0, a] \times [0, b]) \cap \mathbb{Z}^2$ and $v \in \mathbb{B}_n$ we let $B(v) = \{u+v : u \in B\}$ and defined the measurement function as

$$f(v) = \frac{1}{|E(v)|} \sum_{e \in E(v)} X(e)$$

where $E(v) = \{\langle l, l' \rangle \in B(v)^2 : l \neq l'\}$. We define the marked Ising box interaction model as

$$Y(v) = M(v)f(v) \ , \quad v \in \mathbb{B}_n$$

where M is a set of marked points according to Definition 9.1 with selection probability p_s. The corresponding distribution is denoted by $\mu^{box,(n)}_{\beta,p_s}$.

We define the Ising box interaction measure as a weak limit of distributions

$$\mu^{box}_{\beta,p_s} = \lim_{n\to\infty} \mu^{box,(n)}_{\beta,p_s}$$

as $n \to \infty$.

We state the basic theorem with expected values, variance and covariance.

Theorem 9.8
Let Y be distributed according to the marked Ising box interaction model, and let X be distributed according to the corresponding Ising distribution. Then for distinct vertices $v, w \in \mathbb{B}_n$ such that $|v - w| > \max(a, b)$ we have the following.

$$\mathbb{E}[Y(v)] = \frac{p_s}{|E(v)|} \sum_{\langle l,l'\rangle \in E(v)} \phi_{p,q}(l \leftrightarrow l')$$

$$\mathit{Var}[Y(v)] = \frac{p_s}{|E(v)|^2} \sum_{\langle l,l'\rangle \in E(v)} \sum_{\langle \tilde{l},\tilde{l}'\rangle \in E(v)} (\mathbb{E}[X(l,l')X(\tilde{l},\tilde{l}')] - p_s\, \mathbb{E}[X(l,l')]\mathbb{E}[X(\tilde{l},\tilde{l}')])$$

$$\mathit{Cov}[Y(v), Y(w)] = \frac{p_s^2}{|E(v)|^2} \sum_{\langle l,l'\rangle \in E(v)} \sum_{\langle \tilde{l},\tilde{l}'\rangle \in E(w)} \mathit{Cov}[X(l,l'), X(\tilde{l},\tilde{l}')]$$

where $X(l,l') = X(l)X(l')$ for unordered pairs $\langle l, l'\rangle \in E(v)$.

Proof:
Let X and Y be defined as in the statement, and let $v, w \in \mathbb{B}_n$ be distinct. Then the expected value

$$\mathbb{E}[Y(v)] = \frac{\mathbb{E}[M(v)]}{|E(v)|} \sum_{\langle l,l'\rangle \in E(v)} \mathbb{E}[X(l,l')] = \frac{p_s}{|E(v)|} \sum_{\langle l,l'\rangle \in E(v)} \phi_{p,q}(l \leftrightarrow l')$$

the variance

$$\begin{aligned}
&\mathrm{Var}[Y(v)] \\
&= \mathbb{E}\left[\left(\frac{M(v)}{|E(v)|}\sum_{\langle l,l'\rangle\in E(v)} X(l,l')\right)^2\right] - \mathbb{E}\left[\frac{M(v)}{|E(v)|}\sum_{\langle l,l'\rangle\in E(v)} X(l,l')\right]^2 \\
&= \frac{p_s}{|E(v)|^2}\sum_{\langle l,l'\rangle\in E(v)}\sum_{\langle \tilde{l},\tilde{l}'\rangle\in E(v)} (\ \mathbb{E}[X(l,l')X(\tilde{l},\tilde{l}')] - p_s\ \mathbb{E}[X(l,l')]\mathbb{E}[X(\tilde{l},\tilde{l}')]\ )
\end{aligned}$$

and the covariance

$$\begin{aligned}
&\mathrm{Cov}[Y(v),Y(w)] \\
&\quad= \mathbb{E}\left[\left(\frac{M(v)}{|E(v)|}\sum_{\langle l,l'\rangle\in E(v)} X(l,l')\right)\left(\frac{M(w)}{|E(v)|}\sum_{\langle \tilde{l},\tilde{l}'\rangle\in E(w)} X(\tilde{l},\tilde{l}')\right)\right] \\
&\qquad- \mathbb{E}\left[\frac{M(v)}{|E(v)|}\sum_{\langle l,l'\rangle\in E(v)} X(l,l')\right]\ \mathbb{E}\left[\frac{M(w)}{|E(v)|}\sum_{\langle \tilde{l},\tilde{l}'\rangle\in E(v)} X(\tilde{l},\tilde{l}')\right] \\
&\quad= \frac{p_s^2}{|E(v)|^2}\sum_{\langle l,l'\rangle\in E(v)}\sum_{\langle \tilde{l},\tilde{l}'\rangle\in E(w)} \mathrm{Cov}[X(l,l'),X(\tilde{l},\tilde{l}')]
\end{aligned}$$

follows by straightforward calculations, and we are done.
□

The Potts case

We extend the defintion for the Ising model to general q using indicator function instead of products. A consequence is that the basic theorem will give us much more complicated expressions for the expected value, variance and covariance compared to the Ising case. It will affect the estimation of susceptibility in Chapter 10, where we will use omniparametric Potts instead of omniparametric random cluster data. It will have a surprisingly positive effect on the estimated function.

Definition 9.9 ***The marked Potts box interaction model***
Let X be distributed according to the Potts distribution $\mu^{(n)}_{\beta,q}$ on $\mathbb{B}_n$ for some fixed $q \geq 2$ at inverse temperature β. Fix arbitrary but finite $a, b \in \mathbb{Z}^+$. For measurement configuration $B = ([0,a] \times [0,b]) \cap \mathbb{Z}^2$ and $v \in \mathbb{B}_n$ we let $B(v) = \{u + v : u \in B\}$ and defined the measurement function as

$$f(v) = \frac{1}{|E(v)|} \sum_{\langle l,l' \rangle \in E(v)} (2I_{\{X(l)=X(l')\}} - 1)$$

where $E(v) = \{\langle l, l' \rangle \in B(v)^2 : l \neq l'\}$. The marked Potts box interaction model is defined as

$$Y(v) = M(v)f(v) \ , \quad v \in \mathbb{B}_n$$

where M is a set of marked points according to Definition 9.1 with selection probability p_s. The corresponding distribution is denoted by $\mu^{box,(n)}_{\beta,q,p_s}$.

We define the Potts box interaction measure as a weak limit of distributions

$$\mu^{box}_{\beta,q,p_s} = \lim_{n\to\infty} \mu^{box,(n)}_{\beta,q,p_s}$$

as $n \to \infty$.

Now the basic theorem with its lengthy calculations.

Theorem 9.9
Let Y be distributed according to the marked Potts box interaction model with parameters $q \geq 2$, β and p_s. Let X follow the corresponding Potts distribution on box $\mathbb{B}_n$. For simplicity let e_v denote the edge $\langle v, v' \rangle$. Then for distinct vertices $v, w \in \mathbb{B}_n$ such that $|v - w| > \max(a, b)$ we have

$$\begin{aligned}
\mathbb{E}[Y(v)] &= p_s \left(\frac{2}{q} - 1 + 2 \, \frac{q-1}{q} \, A_{p,q} \right) \\
\mathit{Var}[Y(v)] &= 4p_s(B_{p,q}(\mathbf{0}) - A_{p,q}) - p_s^2 \left(\frac{1}{q} + 2 \, \frac{q-1}{q} \, A_{p,q} \right)^2 + 1 \\
\mathit{Cov}[Y(v), Y(w)] &= 4p_s^2(B_{p,q}(v-w) - A_{p,q}) - p_s^2 \left(\frac{1}{q} + 2 \, \frac{q-1}{q} \, A_{p,q} \right)^2
\end{aligned}$$

where

$$A_{p,q} = \frac{1}{|E(v)|} \sum_{e_l \in E(\mathbf{0})} \phi_{p,q}(l \leftrightarrow l')$$

and

$$B_{p,q}(v) = \frac{1}{|E(v)|^2} \sum_{e_l \in E(\mathbf{0})} \sum_{e_u \in E(v)} \mu^{(n)}_{\beta,q}(\{X(l) = X(l')\} \cap \{X(u) = X(u')\}).$$

Proof:
Let Y and X be distributed as in the statement. For distinct $v, w \in \mathbb{B}_n$ we get the following. For simplicity of notation let $L(e_u)$ denote the event $\{X(u) = X(u')\}$. Also define $A_{p,q}(v)$ and $B_{p,q}(v, w)$ as follows.

$$\begin{aligned} A_{p,q}(v) &= \frac{1}{|E(v)|} \sum_{e_l \in E(v)} \phi_{p,q}(l \leftrightarrow l') \\ B_{p,q}(v,w) &= \frac{1}{|E(v)|^2} \sum_{e_l \in E(v)} \sum_{e_u \in E(w)} \mu^{(n)}_{\beta,q}(L(e_l) \cap L(e_u)) \end{aligned}$$

Translation invariance for the Potts and random cluster measures imply $A_{p,q}(v) = A_{p,q}(\mathbf{0}) = A_{p,q}$ and $B_{p,q}(v, w) = B_{p,q}(\mathbf{0}, w - v) = B_{p,q}(w - v)$ for any pair of vertices $w, v \in \mathbb{B}_n$. Now the result follows for the expectation,

$$\begin{aligned} \mathbb{E}[Y(v)] &= \frac{\mathbb{E}[M(v)]}{|E(v)|} \sum_{e_l \in E(v)} (2I_{L(e_l)} - 1) \\ &= \frac{p_s}{|E(v)|} \sum_{e_l \in E(v)} \left(\frac{2}{q} + 2\, \frac{q-1}{q}\, \phi_{p,q}(l \leftrightarrow l') - 1 \right) \\ &= p_s \left(\frac{2}{q} - 1 + \frac{q-1}{q|E(v)|} \sum_{e_l \in E(v)} \phi_{p,q}(l \leftrightarrow l') \right) \\ &= p_s \left(\frac{2}{q} - 1 + 2\, \frac{q-1}{q}\, A_{p,q} \right) \end{aligned}$$

the variance,

$$\begin{aligned}
\text{Var}[Y(v)] &= \mathbb{E}\left[\left(\frac{M(v)}{|E(v)|}\sum_{e_l\in E(v)}(2I_{L(e_l)}-1)\right)^2\right]-\mathbb{E}[Y(v)]^2 \\
&= \frac{p_s}{|E(v)|^2}\sum_{e_l\in E(v)}\sum_{e_u\in E(v)}4\mu^{(n)}_{\beta,q}(L(e_l)\cap L(e_u)) \\
&\quad -\frac{p_s}{|E(v)|^2}\sum_{e_l\in E(v)}\sum_{e_u\in E(v)}2(\,\mu^{(n)}_{\beta,q}(L(e_l))+\mu^{(n)}_{\beta,q}(L(e_u))\,) \\
&\quad +1-p_s^2\left(\frac{2}{q}-1+2\,\frac{q-1}{q}\,A_{p,q}\right)^2 \\
&= 4p_s\,B_{p,q}(v,v)+1-p_s^2\left(\frac{2}{q}-1+2\,\frac{q-1}{q}\,A_{p,q}\right)^2 \\
&\quad -2\,\frac{p_s}{|E(v)|}\left(\sum_{e_l\in E(v)}\mu^{(n)}_{\beta,q}(L(e_l))+\sum_{e_u\in E(v)}\mu^{(n)}_{\beta,q}(L(e_u))\right) \\
&= 4p_s\,B_{p,q}(v,v)+1-p_s^2\left(\frac{2}{q}-1+2\,\frac{q-1}{q}\,A_{p,q}\right)^2-4p_sA_{p,q}(v) \\
&= 4p_s(B_{p,q}(v)-A_{p,q})-p_s^2\left(\frac{2}{q}-1+2\,\frac{q-1}{q}\,A_{p,q}\right)^2+1
\end{aligned}$$

and finally the covariance

$$\begin{aligned}
&\text{Cov}[Y(v),Y(w)] \\
&=\mathbb{E}\left[\left(\frac{M(v)}{|E(v)|}\sum_{e_l\in E(v)}(2I_{L(e_l)}-1)\right)\left(\frac{M(w)}{|E(v)|}\sum_{e_l\in E(w)}(2I_{L(e_u)}-1)\right)\right] \\
&-\mathbb{E}[Y(v)]\mathbb{E}[Y(w)]
\end{aligned}$$

$$= \frac{p_s}{|E(v)|^2} \sum_{e_l \in E(v)} \sum_{e_u \in E(w)} 4\mu^{(n)}_{\beta,q}(L(e_l) \cap L(e_u)) + 1$$

$$- \frac{p_s}{|E(v)|^2} \sum_{e_l \in E(v)} \sum_{e_u \in E(w)} 2(\ \mu^{(n)}_{\beta,q}(L(e_l)) + \mu^{(n)}_{\beta,q}(L(e_u))\ )$$

$$-p_s^2 \left(\frac{2}{q} - 1 + 2\ \frac{q-1}{q}\ A_{p,q}(v) \right) \left(\frac{2}{q} - 1 + 2\ \frac{q-1}{q}\ A_{p,q}(w) \right)$$

$$= 4p_s\ B_{p,q}(v,w) + 1$$

$$-p_s^2 \left(\frac{2}{q} - 1 + 2\ \frac{q-1}{q}\ A_{p,q}(v) \right) \left(\frac{2}{q} - 1 + 2\ \frac{q-1}{q}\ A_{p,q}(w) \right)$$

$$-2\ \frac{p_s}{|E(v)|} \left(\sum_{e_l \in E(v)} \mu^{(n)}_{\beta,q}(L(e_l)) + \sum_{e_u \in E(w)} \mu^{(n)}_{\beta,q}(L(e_u)) \right)$$

$$= 4p_s B_{p,q}(w-v) + 1 - p_s^2 \left(\frac{2}{q} - 1 + 2\ \frac{q-1}{q}\ A_{p,q} \right)^2 - 4p_s A_{p,q}$$

and we are done.
□

9.4 Strong mixing and central limit theorems

The central limit theorems presented in this chapter are only applicable if correlations are decaying sufficiently fast (see mixing coefficients in and before Theorem 9.3, page 112). Correlations in the Ising and Potts models are closely related to connection probabilities in the corresponding random cluster models. If we have exponentially decaying connection probabilities in the random cluster model exponentially decaying correlations follows in the Ising and Potts models.

As a consequence of monotonicity of the alternative critical point (Definition 7.8 and Lemma 7.2, pages 69 and 70) we have exponential decay of random cluster connection probabilities whenever

$$p < p_g(2) = \frac{\sqrt{2}}{1+\sqrt{2}}$$

and there is a corresponding limit for the inverse temperature β in the

Ising and Potts model. We define the strong mixing region as follows.

Definition 9.10 ***Strong mixing regime***
Let R^p_{mix} and R^β_{mix} defined as

$$R^p_{mix} = \left[0,\ \frac{\sqrt{2}}{1+\sqrt{2}}\right],\ R^\beta_{mix} = \left[0,\ \ln\sqrt{1+\frac{1}{\sqrt{2}}}\right]$$

be the regimes of strong mixing for p and β respectively.

When using a central limit theorem to construct a confidence interval for the inverse temperature β we sometimes do not know for sure that $\beta \in R^\beta_{\text{mix}}$, and if not the central limit theorem does not hold. If however $\beta \in R^\beta_{\text{mix}}$ we say that the confidence interval is reliable. We will elaborate more on this in chapters 11 and 12.

CHAPTER 10

Exploring theory: simulation studies

In this chapter we study two important characteristics, connection probabilities and susceptibility, for the models defined in Chapter 9. Since all parameter estimation is carried out in two dimensions we restrict to this case here also.

10.1 Connection probabilities

In subsequent chapters we need an estimate of the connection probability between nearest neighbours in the random cluster representation. We use omniparametric simulation to estimate

$$\phi_{p,q}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1}) \ , \quad p \in [0,1]$$

for $q = 2$ (Ising case) and $q = 4$ where $\mathbf{0} = (0,0)$ and $\mathbf{1} = (1,0)$. Due to translation and rotation invariance of the random cluster distributions the event $\{\mathbf{0} \leftrightarrow \mathbf{1}\}$ is sufficient.

The procedure is the following. We generate omniparametric random cluster samples $X_1, ..., X_m$. In each sample we calculate the smallest p for which $\mathbf{0}$ and $\mathbf{1}$ is connected by open edges, and denote it by

$$X_k = \inf\{\tilde{p} : \mathbf{0} \leftrightarrow \mathbf{1} \text{ in } \mathbf{P}_{\tilde{p}}^{\text{RC}}(\Gamma_q)\}, \ k = 1, ..., m$$

where Γ_q is the omniparametric random cluster variable and $\mathbf{P}_{\tilde{p}}^{\text{RC}}$ is the projection operator (Definition 8.1, page 87). Then $\phi_{p,q}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})$ is the distribution function of X_k since

$$\phi_{p,q}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1}) = \mathbb{P}(\{\gamma_o : \mathbf{0} \leftrightarrow \mathbf{1} \text{ in } \mathbf{P}_p^{rc}\gamma_o\}) = \mathbb{P}(X_k \leq p)$$

and we write $F^{(q)}(p)$ for $\phi_{p,q}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})$. We estimate $F^{(q)}$ by its empirical counterpart

$$\hat{F}_m^{(q)}(p) = \frac{1}{m} \sum_{k=1}^{m} I_{\{X_k < p\}}(p).$$

The Glivenko-Cantelli theorem [Dur96] gives us uniform convergence of $\hat{F}_m^{(q)}$ to $F^{(q)}$,

$$\sup_{p \in [0,1]} |\hat{F}_m^{(q)}(p) - F^{(q)}(p)| \xrightarrow{a.s.} 0,$$

as $m \to \infty$. Next we introduce the Kolmogorov-Smirnov distribution [Smi48] which help us construct confidence intervals for the remaining discrepancy.

Definition 10.1 ***The Kolmogorov-Smirnov distribution***
The Kolmogorov-Smirnov distribution on the positive real line is the distribution having

$$F_\Delta(x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-k^2x^2}$$

as distribution function.

For an illustration of the density of the Kolmogorov-Smirnov distribution see Figure 10.2. The following theorem (here slightly reformulated) from Feller [Fel48] give the distribution for the difference between a distribution function and its empirical counterpart.

Theorem 10.1
Suppose that $F(x)$ is continuous distribution function and let $\hat{F}_m$ be the empirical distribution. If D_m is defined as follows

$$D_m = \sup_{x \in \mathbb{R}} |F(x) - \hat{F}_m(x)|,$$

then for every fixed $x \geq 0$

$$\mathbb{P}(\sqrt{m}\, D_n \leq x) \to F_\Delta(x)$$

as $n \to \infty$, where $F_\Delta(x)$ is the Kolmogorov-Smirnov distribution function.

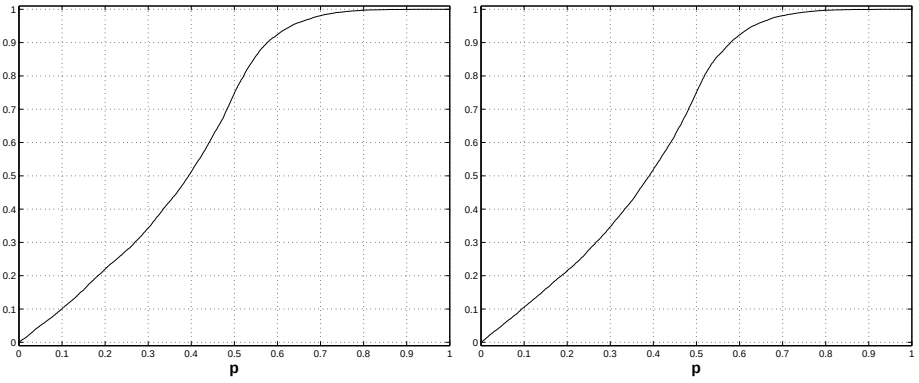

Figure 10.1: *To the left we see the result of estimating $F^{(2)}$, the Ising case, by using $m = 10000$ omniparametric samples. To the right we see the corresponding result for $F^{(4)}$*

With $m = 10000$, as in Figure 10.2, we get for all $p \in [0, 1]$ simultaneously, a level α confidence interval for $F^{(q)}(p)$,

$$F^{(q)}(x) \in [\hat{F}_m^{(q)}(p) - D, \hat{F}_m^{(q)}(p) + D] \quad (\alpha)$$

where D is according to Table 10.1.

α	0.95	0.975	0.99	0.995	0.9990	0.9995	0.9999
D	0.0136	0.0148	0.0165	0.0175	0.0195	0.0204	0.0247

Table 10.1: Percentiles on the Kolmogorov-Smirnov distribution.

In Figure 10.1 we see the results for the Ising case, $q = 2$ (left), and for $q = 4$ (right).

Mixed connection probabilities

Next we study the sum of connection probabilities emerging in the box interaction models, see Section 9.3.2 (page 120).

Let Y be distributed according to the marked Ising box interaction model, using $a, b \geq 1$. Let $B = ([0, a] \times [0, b]) \cap \mathbb{Z}^2$ be a rectangle in $\mathbb{Z}^2$ and let

$$E = \{\langle l, l' \rangle \in B^2 : l \neq l' \}$$

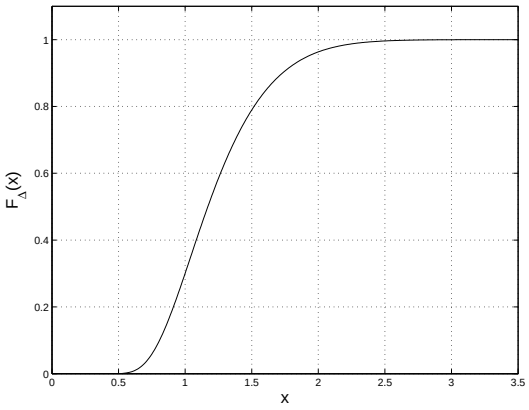

Figure 10.2: *The Kolmogorov-Smirnov distribution, for the scaled deviation* $\sqrt{m}D = 100D$.

be the set of unordered pairs of vertices in B. Suppose we choose a pair $\langle l, l' \rangle$ uniformly at random from E and let

$$Y(l, l') = \inf\{\tilde{p} : l \leftrightarrow l' \text{ in } \mathbf{P}^{\text{RC}}_{\tilde{p}}(\Gamma_q)\}.$$

be the threshold value of that pair, then

$$\begin{aligned}
\mathbb{P}(Y(l, l') \leq p) &= \sum_{\langle \tilde{l}, \tilde{l}' \rangle \in E} \mathbb{P}(Y(\tilde{l}, \tilde{l}') \leq p | \text{ choose } \tilde{l}, \tilde{l}') \mathbb{P}(\text{ choose } \tilde{l}, \tilde{l}') \\
&= \frac{1}{|E|} \sum_{\langle \tilde{l}, \tilde{l}' \rangle \in E} \mathbb{P}(X(\tilde{l}, \tilde{l}') \leq p) \\
&= \frac{1}{|E|} \sum_{\langle \tilde{l}, \tilde{l}' \rangle \in E} \phi^{(n)}_{p,2}(\tilde{l} \leftrightarrow \tilde{l}')
\end{aligned}$$

is a mixture of distribution functions, and we proceed along the same lines as in the last section. In Figure 10.3 we see the results for the Ising model and Figure 10.4 display the results for the 4-state Potts model. Note that if we use larger a and b we get a steeper curve which result in different characteristics of our estimation of β depending on how sharp

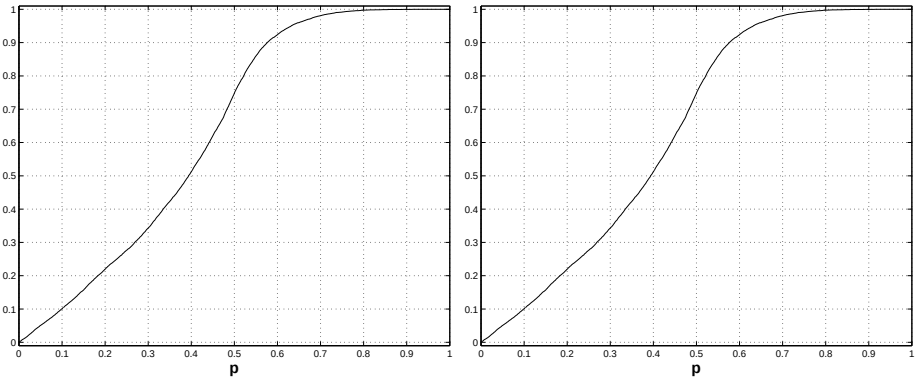

Figure 10.3: *Some examples of the mixed connection probability $f_{avg}^{a,b}(p)$ for the marked Ising box interaction model using $a \times b$-boxes. In this case we have $a = b = 2$ (left) and $a = b = 3$ (right).*

our estimate of the connection probability is. We do not explore this further, instead we use $a = b = 2$ in later chapters.

10.1.1 Estimation of connection probabilities

In Chapter 2 we treated an omniparametric approach for estimating connection probabilities between neighbouring sites in the bond percolation model on $\mathbb{Z}^2$. According to Figure 2.8 there is an interval for p around its critical value p_c where we have difficulties getting a proper answer, some simulations simply could not give an answer weather two sites are connected or not since there could be a connecting path of open edges reaching outside the simulated area. The problem is the same when simulating estimating connection probabilities in the random cluster model.

The simulations presented in this chapter are set up so that we only mark vertices as connected based on data from finite regions. That is, vertices are marked as connected if they are connected by a path of open edges within the simulated box, and since

$$\mathbb{P}(u \leftrightarrow v \text{ on } \mathbb{B}_n) < \mathbb{P}(u \leftrightarrow v \text{ on } \mathbb{Z}^2)$$

we get an under estimation of the connection probabilities. This is a choice we have made in order to get manageable execution times for our

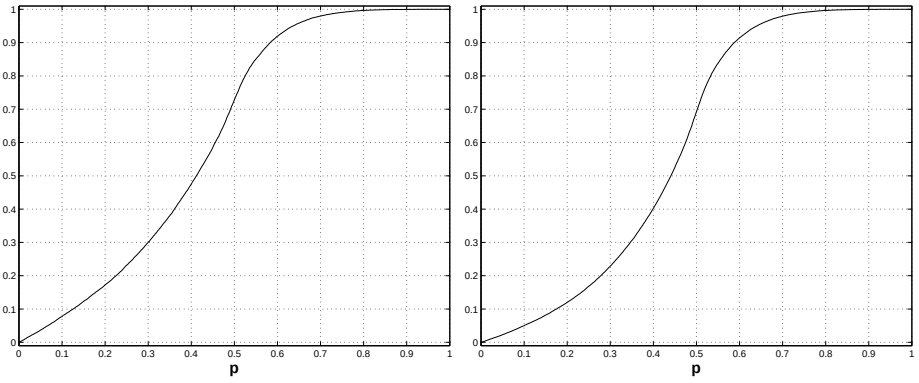

Figure 10.4: *Some examples of the mixed connection probability $f_{avg}^{a,b}(p)$ for the marked Potts box interaction model with $q = 4$, using $a \times b$-boxes. In this case we have $a = b = 2$ (left) and $a = b = 3$ (right).*

simulation programs. To what extent we deviate from the true value is hard to say. We can only use as large boxes as possible and hope for the best.

The problem remains if we try to estimate connection probabilities from omniparametric Ising or Potts data since we use the component structure in the omniparametric random cluster samples for their generation.

10.2 Susceptibility

The marked Ising interaction model

Let Y follow the marked Ising interaction measure on $\mathbb{Z}^2$. In this section we use simulation to estimate the susceptibility

$$\sigma_Y^2 \quad = \quad \sum_{v \in \mathbb{Z}^2} \mathrm{Cov}[Y(0), Y(v)].$$

From Theorem 9.1 (page 104) we get an expression for the covariance $\mathrm{Cov}[Y(0), Y(v)]$ when $|v| > 1$. For the four points in $\{w : |w| = 1\}$ the covariance is slightly more complicated and we do not treat it explicitly.

In subsequent chapters we will use the central limit theorems (Theorems 9.4, 9.5, 9.6 and 9.7) to capture the asymptotic variation of certain statistics. A vital part of these theorems is the susceptibility, therefore a careful study is motivated.

We cannot study the proper susceptibility so we have to be content with the truncated version, the box susceptibility

$$\sigma^2_{Y,n,\beta} = \sum_{v\in\mathbb{B}_n} \mathrm{Cov}[Y(0), Y(v)].$$

We study $\sigma^2_{Y,n,\beta}$ for $\beta \in [0, \beta_c)$ since otherwise the susceptibility is infinite and no central limit theorems are available. The idea is to simulate the model on sufficiently large boxes $\mathbb{B}_n$ and hopefully get a reasonable estimate of $\sigma^2_{Y,n,\beta}$. Due to limited computer resources we limit ourself to the box $\mathbb{B}_{50}$.

The simulation algorithm is based on estimation of connection probabilities in the random cluster model according to Lemma 7.5 (page 79). We believe it is more accurate since we do not use the extra randomness involved in generating omniparametric Ising samples. For simulation results see Figure 10.5.

As expected we see the truncated susceptibility grow larger and larger for $\beta \approx \beta_c$. Somewhat unexpected is that the susceptibility decays for $\beta > \beta_c$ a plausible explanation is the following. Fix $v \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}$, then the covariance

$$\begin{aligned}\mathrm{Cov}[Y(\mathbf{0}), Y(v)] \quad = \quad & \frac{p_s^2}{4}\sum_{k=1}^{2}\sum_{m=1}^{2} \quad (\ \phi_{p,2}(\{\mathbf{0}\leftrightarrow e_k\}\cap\{v\leftrightarrow v+e_m\}) \\ & \qquad +\phi_{p,2}(\mathbf{0}\leftrightarrow v \not\leftrightarrow e_k \leftrightarrow v+e_m) \\ & \qquad +\phi_{p,2}(\mathbf{0}\leftrightarrow v+e_m \not\leftrightarrow e_k \leftrightarrow v) \\ & \qquad -\ \phi_{p,2}(\mathbf{0}\leftrightarrow \mathbf{1})^2 \ )\end{aligned}$$

may approach p_s^2 as $p \to 1$ ($\beta \to \infty$) since the probabilities

$$\phi_{p,2}(\{\mathbf{0}\leftrightarrow e_k\}\cap\{v\leftrightarrow v+e_m\}) \ \text{ and } \ \phi_{p,2}(\mathbf{0}\leftrightarrow\mathbf{1})^2$$

both approach 1 while the events

$$\{\mathbf{0}\leftrightarrow v \not\leftrightarrow e_k \leftrightarrow v+e_m\}$$

and

$$\{\mathbf{0}\leftrightarrow v+e_m \not\leftrightarrow e_k \leftrightarrow v\}$$

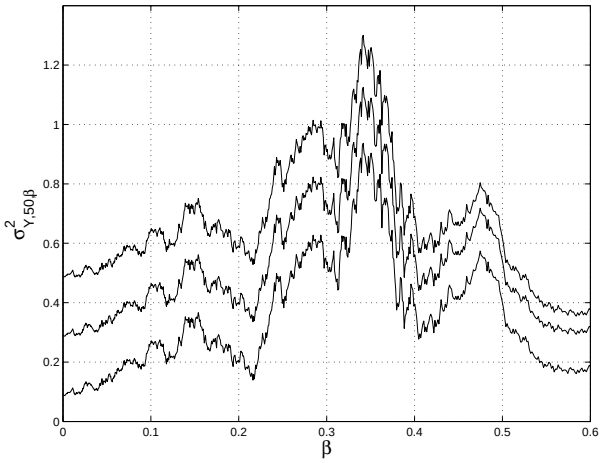

Figure 10.5: *An estimation of the susceptibility $\sigma^2_{Y,50,p}$ in the marked Ising interaction model on the box $\mathbb{B}_{50}$ for $p_s = 0.1$ (lower curve), $p_s = 0.3$ (middle) and $p_s = 0.5$ (upper curve), using 100 000 omniparametric samples.*

become more and more unlikely. The last two events depend on existing open paths between vertices at some distance and the non-existence of open paths between neighbouring vertices. We end up with a situation where covariances between the state at different vertices decrease due to less variation while the dependence increase.

We would expect something smoother from an estimation procedure using so many simulations but instead we get something rather crude. The most likely explanation is the estimation of probabilities for unlikely events in the expression for $\mathrm{Var}[Y(\mathbf{0}), Y(v)]$ when the origin and v are far apart. For example, the event $\{\mathbf{0} \leftrightarrow v \not\leftrightarrow e_k \leftrightarrow v + e_m\}$ where the origin is connected to v and e_k is connected to $v + e_m$, but neither $\mathbf{0}$ nor v are connected to their neighbours e_k and $v + e_m$ respectively. This is highly unlikely and require massive simulation to generate a smooth curve for its probability.

The marked Potts interaction model

We repeat the estimation procedure for the marked Potts interaction model using $q = 4$. The procedure is completely analogous to the

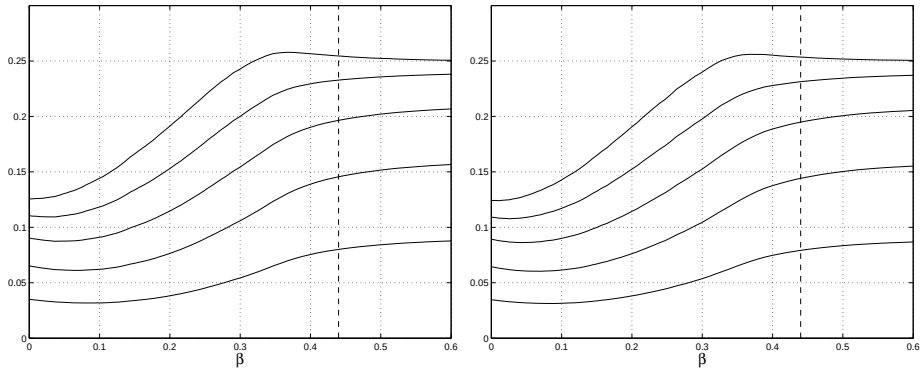

Figure 10.6: *An estimation of the susceptibility $\sigma^2_{Y,n,\beta}$ in the marked Potts interaction model with $q = 4$ on box $\mathbb{B}_{10}$ (right) and $\mathbb{B}_{20}$ (left) for $p_s = 0.1$ (lower curve), $p_s = 0.2,\ 0.3,\ 0.4$ and $p_s = 0.5$ (upper curve). The difference between the two are small, $|\sigma^2_{Y,20,\beta} - \sigma^2_{Y,10,\beta}| \leq 0.0015$ regardless of β.*

previous section, and we present the results in Figure 10.6.

Instead of estimate the susceptibility through omniparametric random cluster samples we use omniparametric Potts samples, generated according to Section 8.2 (page 97). The reason is that the expression for the marked Potts interaction covariance is too complex, it is much simpler to estimate it directly from the omniparametric Potts samples. As we see in Figure 10.7 the curves are much smoother than the corresponding ones for the Ising model. The approach has a negative side, computer time. It is much more demanding than just simulating the random cluster model, as a consequence we have to limit the simulated area further. Instead of using box $\mathbb{B}_{50}$ as for the Ising model we use $\mathbb{B}_{20}$. When using the estimate later we will, as before, treat the susceptibility as known at its estimated value. We will also generate Potts samples on larger boxes than $\mathbb{B}_{20}$ and hope that the difference between $\sigma^2_{Y,20,\beta}$ and $\sigma^2_{Y,50,\beta}$ is negligible for β small enough.

The marked Ising box interaction model

We generate omniparametric Ising samples and calculate the truncated susceptibility $\sigma^2_{n,m,k}$ for the marked Ising box interaction model.

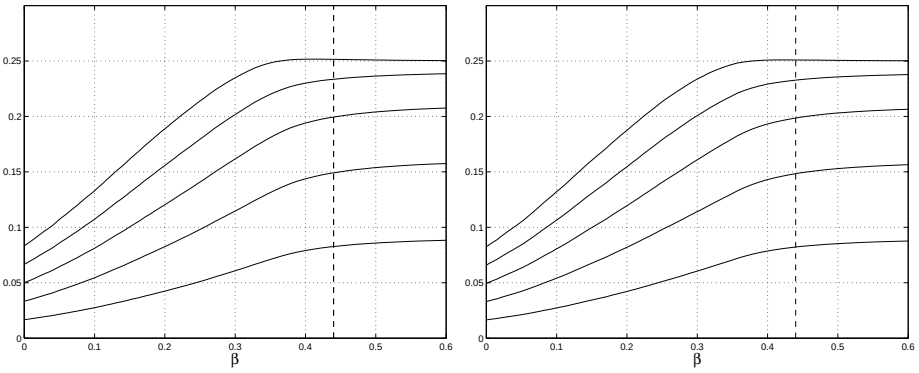

Figure 10.7: *An estimation of the susceptibility in the marked Ising box interaction model,* $\sigma^2_{2,2,10,\beta}$ *(left) and* $\sigma^2_{2,2,20,\beta}$ *(right), on boxes* $\mathbb{B}_{10}$ *and* $\mathbb{B}_{20}$ *respectively, for* $p_s \in \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$*. The curves are in increasing order regarding selection probability. Again the difference between the two are small,* $|\sigma^2_{2,2,20,\beta} - \sigma^2_{2,2,10,\beta}| \leq 0.000375$ *regardless of* β*.*

Due to limited computer resources and limited time we manage boxes $\mathbb{B}_{10}$ and $\mathbb{B}_{20}$, using $a = b = 2$. We see the result in Figure 10.7. As expected the susceptibility becomes larger around the critical value.

The marked Potts box interaction model

By using omniparametric Potts samples we estimate the truncated susceptibility for the marked Potts box interaction model. We see the results in Figure 10.8

The results are similar to the corresponding results for the Ising model, and there is also a rather small difference between $\sigma^2_{2,2,10,\beta}$ and $\sigma^2_{2,2,20,\beta}$ indicating that the sum of covariance terms for $v \in \mathbb{B}_{20} \setminus \mathbb{B}_{10}$ is negligible.

Final remarks

As we have seen the difference between the estimated susceptibility based on data from different box sizes are small, perhaps negligible. In Table 10.2 we see the respective differences. The marked Ising interaction model is excluded since we used another estimation technique.

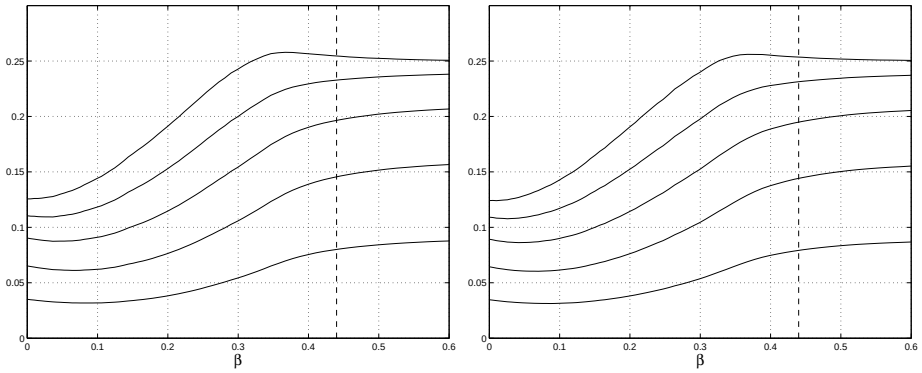

Figure 10.8: *An estimation of the susceptibility in the marked Potts box interaction model with $q = 4$, $\sigma^2_{2,2,10,\beta}$ (left) and $\sigma^2_{2,2,20,\beta}$ (right), on boxes $\mathbb{B}_{10}$ and $\mathbb{B}_{20}$ respectively, for $p_s \in \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$. The curves are in increasing order regarding selection probability, $p_s = 0.1$ in the bottom and $p_s = 0.5$ on top. Again the difference between the two are small, $|\sigma^2_{2,2,20,\beta} - \sigma^2_{2,2,10,\beta}| \leq 0.00030$ regardless of β.*

Since the differences are so small we will try to use the estimated susceptibilities on box sizes larger than 20 for all three models. The validation procedure will show that this give us reasonable results, see chapters 11 and 12.

Model	Absolute difference $\times 10^{-4}$
marked Potts interaction	15
marked Ising box interaction	3.7441
marked Potts box interaction	2.9816

Table 10.2: *We see how much the estimated susceptibilities deviate between data from $\mathbb{B}_{10}$ and $\mathbb{B}_{20}$.*

CHAPTER 11

Parameter estimation in the Ising model

11.1 Introduction

We present three different approaches to parameter estimation in the Ising model, based on incomplete data. All methods are carried out on the two-dimensional lattice.

The first method is based on pairwise interactions and asymptotic theory. We estimate the probability of an open connection between two neighbouring vertices in the random cluster model from Ising data. From omniparametric random cluster samples we estimate the connection probability as a function of $p = 1 - e^{-2\beta}$. The estimated function together with our statistic result in a point estimate of p corresponding to a point estimate of inverse temperature β. We capture the variance by using a central limit theorem and establish confidence intervals.

The second method is a generalization of method one, based on general interactions instead of pairwise. By using more general interactions we may extract more information out of a given data set and we hope to get sharper estimates. Apart from the extension to general interactions method one and two are identical.

The third approach is more direct and not based on asymptotic theory. Given data X we calculate a statistic $T(X)$. We now view $T(X)$ as

a function of the parameter β, and estimate it using omniparametric simulations. The simulation step results in a mean value curve for the function $T(X)(\beta)$ and also empirical percentiles from order statistics. Using these characteristics we can easily compute point estimates and confidence intervals for β.

The rest of the chapter is outlined as follows. Next we introduce necessary notation and present methods. Sections 11.3 and 11.4 treat method one. Sections 11.5 and 11.6 present the corresponding material for method two and in sections 11.7 and 11.8 we treat the third, non-asymptotic, method.

11.2 Analysing simulations, method and comments

For all three methods we perform validation and analyse precision and reliability. The analysis is carried out for data on $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$ for selection probabilities $p_s = 0.10,\ 0.20$, also for $p_s = 0.4,\ 0.6,\ 0.8$ and $p_s = 1.0$ on $\mathbb{B}_{20}$.

- In the validation procedure we estimate the confidence level from data, and compare it with the desired one.
- By precision we mean the width of the confidence interval. For a fixed confidence level we analyse how much data is required to achieve a certain precision.
- For the asymptotic methods a confidence interval is valid if the central limit theorem is applicable, and that happens when $\beta < \beta_c$ for the Ising model. We can of course not know if that is the case from the start. We assume $\beta < \beta_g(q)$, compute the confidence interval $[\beta_{\text{low}}, \beta_{\text{up}}]$, and if

  $$[\beta_{\text{low}}, \beta_{\text{up}}] \subset R^{\beta}_{\text{mix}}$$

  (Definition 9.10, page 128) then we say that the interval is reliable, otherwise we mark it as unreliable. The approach is conservative in the sense that for some β close to β_c we will make an error and mark the confidence intervals as unreliable. The non-asymptotic method does not have this limitation.
- By phase determination we mean the probability of being able to determine in which phase the system is in. For values of the inverse temperature far from the critical value there are no problems, but for $\beta \approx \beta_c$ we run into difficulties. We look at the region where

we have a reasonably high probability of being able to separate the subcritical phase from the supercritical. For the non-asymptotic this approach works for all β, while for the asymptotic methods we can only make a decision in the subcritical phase.

We generate n omniparametric samples, $X_1, ..., X_n$, and calculate the statistic $S(X_1), ..., S(X_n)$, each a function of $p \in [0, 1]$. Based on these functions we calculate confidence intervals, estimate confidence levels for validation, average width, the probability that the confidence interval does not contain β_c (reliability), and analyse phase determination.

We now introduce necessary notation. Let the two-dimensional random variable

$$(B_{\text{low}}^{n,\alpha,p_s}(\omega), B_{\text{up}}^{n,\alpha,p_s}(\omega))$$

represent the bounds of a level α confidence interval based on data ω within box $\mathbb{B}_n$ using selection probability p_s. Let

$$W_{n,\alpha,p_s}(\omega) = B_{\text{up}}^{n,\alpha,p_s}(\omega) - B_{\text{low}}^{n,\alpha,p_s}(\omega)$$

be the length of the interval, and let

$$V = \{\omega : \beta_c \notin [B_{\text{low}}^{n,\alpha,p_s}(\omega), B_{\text{up}}^{n,\alpha,p_s}(\omega)]\}$$

be the event that the confidence interval does not cover the critical value. We estimate the confidence level by

$$\hat{\alpha} = \frac{1}{n} \sum_{k=1}^{n} I_{[B_{\text{low}}^k, B_{\text{up}}^k]}(\beta)$$

for $\beta \in (0, \infty)$.

There is always a chance of getting a confidence interval of infinite length, a one-sided interval. Any confidence interval for p close enough to 1 or including 1 results in an interval for β approaching infinity. When analysing the interval width we condition on finite confidence intervals for β whenever there are infinite ones.

11.3 Pairwise interactions

Our first approach is to use a simple measurement configuration, $B = \{(0, 0), (1, 0)\}$, and a measurement function $f(a, b) = ab$. This method makes measurements follow the marked Ising interaction model if we choose our the vertices where we make measurements uniformly at random. Our goal is to estimate the inverse temperature and our main tool

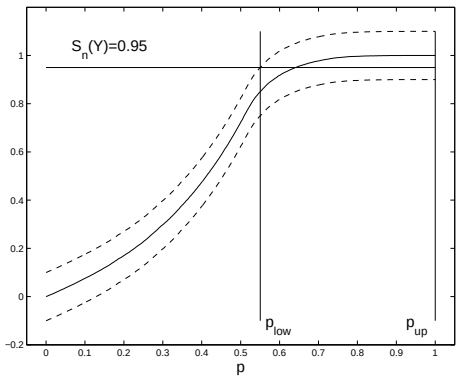

Figure 11.1: *An illustration of how the construction of confidence intervals may produce infinite or one-sided intervals. A confidence interval for p containing 1 results in a one-sided interval for β. A consequence is that we can not say much about the precision of our estimate of β.*

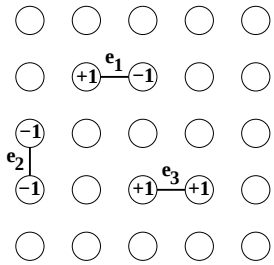

Figure 11.2: *We start with data collected in pairs. The idea is to estimate the probability that there is an edge between the two vertices in the underlying random cluster sample. In this case the data consists of the three values for the edges e_1, e_2 and e_3, $X(\mathcal{L}) = (-1, +1, +1)$.*

the random cluster distribution. We start with point estimation and continue with the more complicated interval estimation.

Given a randomly chosen set of locations in the form of edges

$$\mathcal{L} = \{e_k : k = 1, ..., n\} \ , \quad e_k = (l_k, l'_k)$$

and measurements X at each end vertex, see Figure 11.2. Suppose X follow the Ising model at some inverse temperature β_0. We calculate $Y(l_k),\ k = 1, ..., n$ with the measurement function,

$$Y(l_k) = f(\ X(l_k),\ X(l'_k)\ ) = X(l_k)X(l'_k).$$

Next we introduce the statistic, establish unbiasedness and under subcriticality also a central limit theorem and consistency.

Definition 11.1 ***The statistic***
Let $\mathbb{B}_n$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Ising interaction distribution on $\mathbb{B}_n$. We define the statistic $S_n(Y)$ as

$$S_n(Y) = \frac{1}{p_s|\mathbb{B}_n|} \sum_{v \in \mathbb{B}_n} Y(v).$$

It might appear as if

$$S'_n(Y) = \frac{1}{N} \sum_{v \in \mathbb{B}_n} Y(v) \ , \quad N = |\{v \in \mathbb{B}_n : Y(v) \neq 0\}|$$

is a more natural choice of statistic, but since $N = 0$ with positive probability $S'_n(Y)$ has infinite mean. Choosing $S'_n(Y)$ instead of $S_n(Y)$ will also complicate the weak convergence towards a normal distribution in the central limit theorem stated below.

Theorem 11.1
Let Y be distributed according to the marked Ising interaction model, and let $\phi_{p,2}$ be the corresponding random cluster measure. Let $S_n(Y)$ be defined as in Definition 11.1. Then $S_n(Y)$ is unbiased and consistent as an estimator of $\phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})$. If the underlying Ising process is subcritical $S_n(Y)$ is consistent and we have

$$\frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} \ (\ S_n(Y) - \phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})\ ) \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0, 1)$.

Proof:
Let $S_n(Y)$ be as defined above. It is indeed unbiased,

$$\mathbb{E}[S_n(Y)] = \mathbb{E}\left[\frac{1}{p_s|\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} Y(v)\right] = \frac{1}{p_s|\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n}\mathbb{E}[Y(v)] = \phi_{p,2}(\mathbf{0}\leftrightarrow\mathbf{1})$$

and consistence follows from unbiasedness since

$$\mathrm{Var}[S_n(Y)] = \frac{1}{p_s^2|\mathbb{B}_n|^2}\sum_{v\in\mathbb{B}_n}(p_s - p_s^2\phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{1})^2) = \frac{1-p_s\phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{1})^2}{p_s|\mathbb{B}_n|}\to 0$$

as $n\to\infty$. From Bolthausen's central limit theorem we have

$$\sum_{v\in\mathbb{B}_n}\frac{(Y(v)-\mathbb{E}[Y(v)])}{\sigma_Y|\mathbb{B}_n|^{1/2}} \xrightarrow{\mathcal{D}} Z_0$$

where Z_0 follows the standard normal distribution. Considering this expression we make the following rearrangements.

$$\begin{aligned}\sum_{v\in\mathbb{B}_n}\frac{(Y(v)-\mathbb{E}[Y(v)])}{\sigma_Y|\mathbb{B}_n|^{1/2}} &= \frac{1}{\sigma_Y|\mathbb{B}_n|^{1/2}}\left(\sum_{v\in\mathbb{B}_n}Y(v) - p_s|\mathbb{B}_n|\phi_{p,2}^{(n)}(\mathbf{0}\leftrightarrow\mathbf{1})\right)\\ &= \frac{(p_s|\mathbb{B}_n|)^{-1}\sum_{v\in\mathbb{B}_n}Y(v) - \phi_{p,2}^{(n)}(\mathbf{0}\leftrightarrow\mathbf{1})}{\sigma_Y\ p_s^{-1}\ |\mathbb{B}_n|^{-1/2}}\\ &= \frac{S_n(Y) - \phi_{p,2}^{(n)}(\mathbf{0}\leftrightarrow\mathbf{1})}{\sigma_Y\ p_s^{-1}\ |\mathbb{B}_n|^{-1/2}}\end{aligned}$$

It follows that

$$\frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}}\ (\ S_n(Y) - \phi_{p,2}^{(n)}(\mathbf{0}\leftrightarrow\mathbf{1})\ ) \xrightarrow{\mathcal{D}} Z$$

where $Z\stackrel{\mathcal{D}}{=}N(0,1)$. Now

$$|S_n(Y) - \phi_{p,2}^{(n)}(\mathbf{0}\leftrightarrow\mathbf{1})| \xrightarrow{P} 0,$$

and as a consequence also consistency follows from the above together with unbiasedness, and we are done.
□

Note that consistency only follows if we expand the region from which we take our data, increasing p_s in order to use more data within a given region is not enough.

The point estimator

From data we calculate our statistic $S_n(Y)$. Through simulation we estimate

$$F(p) = \phi_{p,2}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})$$

by the empirical distribution function $\hat{F}_m(p)$ (Section 10.1, page 129). Our estimate of p is $\hat{p} = \hat{F}_m^{-1}(\ S_n(Y)\ )$. The relation between p and β give us an estimate of β.

Summing up: Given the statistic $S_n(Y)$ our estimate of β is

$$\hat{\beta} = -\ \frac{1}{2}\ \log(\ 1 - \hat{F}_m^{-1}(\ S_n(Y)\ )\ )$$

The interval estimator

Under the assumption of subcriticality we compute a level α confidence interval $[\beta_{\text{low}}, \beta_{\text{up}}]$ for β. We later verify if this assumption is supported by data or not. The variance of $\hat{\beta}$ has two sources. The statistic $S_n(Y)$ may vary and the estimated function $\hat{F}_m$ also has some variation, and we have to make two choices regarding confidence levels.

- α_S, for $\phi_{p,2}^{(n)}(\mathbf{0} \leftrightarrow \mathbf{1})$ based on $S_n(Y)$.
- α_F, for F_m based on $\hat{F}_m$.

We construct level α_S and level α_F confidence intervals respectively, making the confidence interval for β have level $\alpha = \alpha_S\alpha_F$. For simplicity we let $\alpha_S = \alpha_F = \sqrt{\alpha}$.

In any application we may of course estimate $F(p)$ to arbitrary precision, assume $F(p) = \hat{F}_m(p)$ and set $\varepsilon = 0$ (see Figure 11.3). This not only removes the extra variation from estimating $F(p)$ but also allow us to use a slightly smaller percentile, z_α instead of $z_{\sqrt{\alpha}}$, in the computations below.

When computing the confidence interval the selection parameter and the measurement region are both given. We have data X from some finite region $\Lambda = \mathbb{B}_n$ in the Ising model, and compute marked Ising interaction data Y. The statistic $S_n(Y)$ is our estimate of $F(p)$ for some unknown p and

$$F(p) = S_n(Y) \pm z_{(1+\sqrt{\alpha})/2}\ \frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}}$$

is a level $\sqrt{\alpha}$ confidence interval for $F(p)$, where $z_{(1+\alpha)/2}$ is the two-sided α-quantile of the standard normal distribution.

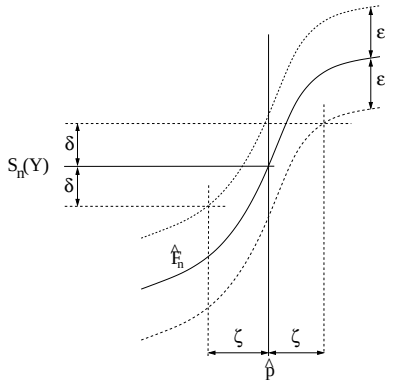

Figure 11.3: *An illustration of how a confidence interval of p around $\hat{p}$ is related to the corresponding confidence interval of $S_n(Y)$ and $\hat{F}_n(p)$. We can calculate ζ through the relationship $\hat{F}_n(p-\zeta)+\varepsilon = S_n(Y)-\delta$, where $S_n(Y)$, $\varepsilon = m^{-1/2}D_{\sqrt{\alpha}}$, $\delta = z_{(1+\sqrt{\alpha})/2}\ \sigma_Y\ p_s^{-1}\ |\mathbb{B}_n|^{-1/2}$ and $\hat{F}_n$ are known through data and simulations.*

We use the estimate of $F(p)$ and the maximum deviation

$$\sup_{p\in[0,1]} |\hat{F}_m(p) - F(p)| \leq \frac{D_{\sqrt{\alpha}}}{\sqrt{m}}$$

according to the Kolmogorov-Smirnov distribution, where D_α is the α-quantile and m is the number of omniparametric samples used to construct $\hat{F}_m(p)$.

For simplicity let $\varepsilon = D_{\sqrt{\alpha}}/\sqrt{m}$. For an illustration of how the confidence interval and confidence band affect the variation of $\hat{p}$ see Figure 11.3. Let $p_1 = p-\zeta$ and $p_2 = p+\zeta$. We then find p_1, p_2 by solving the following two equations.

$$F_n(p_1)+\varepsilon = S_n(Y)-\delta\ ,\quad F_n(p_2)-\varepsilon = S_n(Y)+\delta$$

The estimates

$$\hat{p}_1 = \hat{F}_n^{-1}(\ S_n(Y)-\delta-\varepsilon\ )\ \text{ and }\ \hat{p}_2 = \hat{F}_n^{-1}(\ S_n(Y)+\delta+\varepsilon\ )$$

follows. The interval for p depends heavily on the variation of $S_n(Y)$. If the variation is too large the confidence interval will not give us any

information at all, so keeping the variation of $S_n(Y)$ small becomes important. Finally we transform the confidence interval for p to an interval for β by letting

$$\beta_{\text{low}} = -\frac{1}{2}\,\log(1-\hat{p_1}) \;\text{ and }\; \beta_{\text{up}} = -\frac{1}{2}\,\log(1-\hat{p_2}).$$

We summarize the above in a theorem.

Theorem 11.2 ***Confidence interval for the inverse temperature*** *Let Y be distributed according to the marked Ising interaction model on a finite box $\mathbb{B}_n$ at inverse temperature β and let β_c be the critical temperature. Let $S_n(Y)$ and $\hat{F}_m$ be defined as above. Then $[\beta_{low}, \beta_{up}]$ is a level α confidence interval for the inverse temperature β. The bounds are given by*

$$\beta_{low} = -\frac{1}{2}\,\log\left(1-\hat{F}_m^{-1}\left(\,S_n(Y) - z_{(\sqrt{\alpha}+1)/2}\,\frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} - \frac{D_{\sqrt{\alpha}}}{\sqrt{m}}\,\right)\right)$$

and

$$\beta_{up} = -\frac{1}{2}\,\log\left(1-\hat{F}_m^{-1}\left(\,S_n(Y) + z_{(\sqrt{\alpha}+1)/2}\,\frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} + \frac{D_{\sqrt{\alpha}}}{\sqrt{m}}\,\right)\right)$$

where z_α and D_α are quantiles from the standard normal distribution and the Kolmogorov-Smirnov distribution respectively. The confidence interval is valid if $\beta_c \notin [\beta_{low}, \beta_{up}]$.

Note that not only the location but also the width of a confidence interval is a random variable. The width depends on $S_n(Y)$ through $\hat{F}_m(p)$, see Figure 11.3.

11.4 Simulation study: pairwise interactions

We simulate the statistic S_n for boxes $\mathbb{B}_{10},\ \mathbb{B}_{20},\ \mathbb{B}_{30},\ \mathbb{B}_{40}$ and $\mathbb{B}_{50}$ using selection probabilities, $p_s = 0.1, 0.2$. Extra simulations are generated on $\mathbb{B}_{20}$ for $p_s \in \{0.40,\ 0.60,\ 0.80,\ 1.00\}$. For each combination $(\mathbb{B}_n, p_s)$ we generate 1000 simulations.

Validation

In Figure 11.4 and B.1 (pages 150, 210) we see the result of the procedure for $p_s = 0.10$ and $p_s = 0.20$ for $\alpha = 0.95$ and all box sizes. The

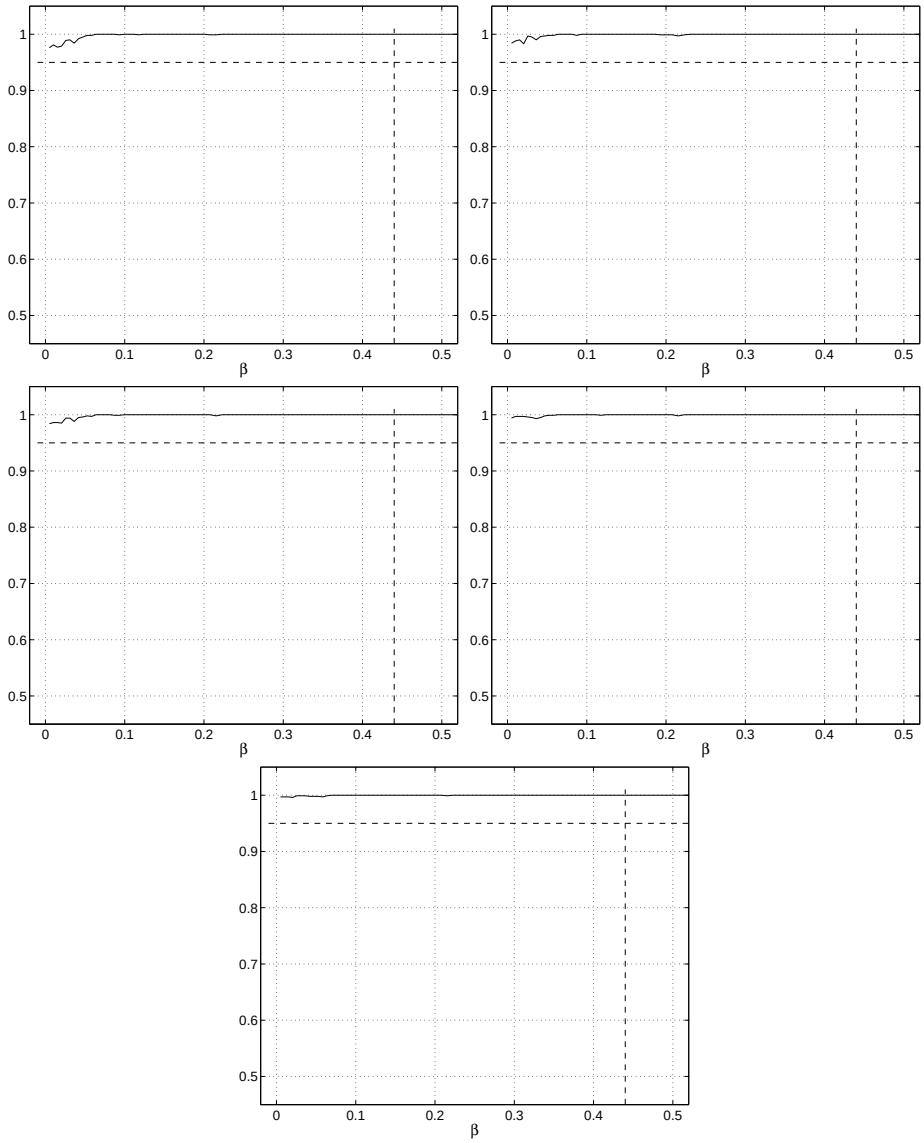

Figure 11.4: *Estimated confidence levels on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). The desired confidence level is $\alpha = 0.95$. The results are given for $p_s = 0.10$. For corresponding results with $p_s = 0.20$ see Figure B.1 (page 210).*

validity of the confidence intervals are good on all five box sizes. We get an estimated confidence level of 1 or almost 1 despite the used level $\alpha = 0.95$, indicating an overestimation of the variance of $S_n(Y)$.

In Figure 11.5 we see a deviation of the estimated confidence level for large values of p_s. The estimated level never reaches below 0.95 but the deviation is systematic and grows larger as p_s increases. A plausible explanation for the deviation is an error in the estimation of the susceptibility $\sigma^2_{Y,\beta}$ by its truncated counterpart $\sigma^2_{Y,n,\beta}$ (see Section 10.2). For values of β close to the critical the sum of covariance terms for all $v \in \mathbb{Z}^2 \setminus \mathbb{B}_n$ may not be negligible. Another plausible explanation is the following. Since the statistic is an estimate of the connection probability we expect to have

$$\mathbb{E}[S_n(Y)] \approx \phi_{p,2}(\mathbf{0} \leftrightarrow \mathbf{1})$$

with only a small random deviation. In Figure 11.6 we see that the deviation (between sample mean and expected) is almost zero except for $\beta \geq 0.3$. With such smooth estimates we would expect no or almost no deviation at all between the two curves. In the case $p_s = 0.1$ the deviation is negligible relative to the expected deviation (according to Theorem 11.2), but for $p_s = 0.8$ the deviation is substantial, resulting in a systematically lower estimated confidence level. According to Section 10.2 this deviation may depend on the way we implemented the validation procedure and not on the estimation itself, see Figure 11.7 (page 153) where we have simulated the statistic on $\mathbb{B}_{20}$ using a much wider margin than before.

Precision

The question is how much data is needed to attain a certain precision? We study this question for $\alpha = 0.95$.

Consider the width W_{α,n,p_s} of a confidence interval. We estimate $\mathbb{E}[W_{n,\alpha,p_s}]$ and the truncated expected width, $\mathbb{E}[W_{n,\alpha,p_s} | W_{n,\alpha,p_s} < \infty]$. See Figure 11.8 for results. Using the smallest box size, $\mathbb{B}_{10}$ and $p_s = 0.1$, we use 44 data points and we achieve confidence intervals having width between 0.2 and 0.6 depending on the true value of β. We even suffer a risk of only getting a one-sided interval if $\beta > 0.05$ (see Figure 11.8, upper left diagram). By using twice as many data points within the same area we get a slightly more narrow interval, but there is still a possibility of getting a one-sided interval now for $\beta \geq 0.14$.

For $\beta \lesssim 0.23$ on $\mathbb{B}_{20}$ using the same selection probability there is a large chance of getting two-sided intervals while for $\beta \geq 0.23$ we get one-sided intervals with high probability.

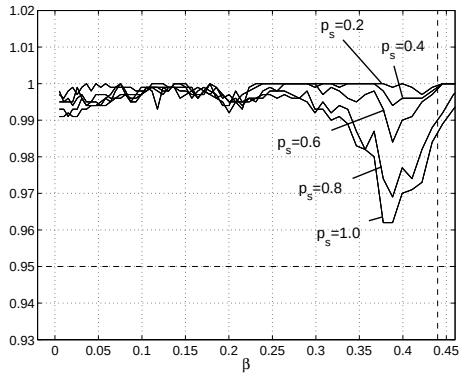

Figure 11.5: *Estimated confidence levels $\hat{\alpha}$ for confidence intervals based on data within box $\mathbb{B}_{20}$, for different choices of selections probabilities, $p_s = 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0$.*

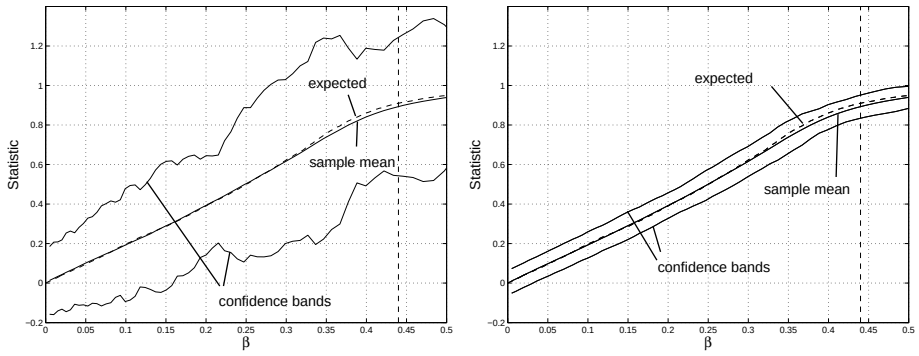

Figure 11.6: *In the left diagram we see the mean value of the statistic S (solid line) and confidence bands for the true value using the central limit theorem (Theorem 9.6) and confidence level $\alpha = 0.95$ for $p_s = 0.1$ on $\mathbb{B}_{20}$. The curves for $\mathbb{E}[S_n(Y)]$ as a function of β are dashed. The left diagram shows the corresponding information for $p_s = 0.8$, also on $\mathbb{B}_{20}$.*

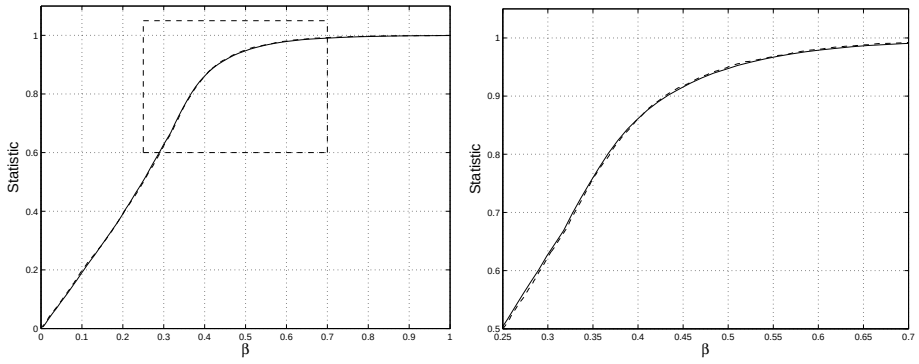

Figure 11.7: *Running simulation on box $\mathbb{B}_{20}$ with larger margins, using only a $\mathbb{B}_{20}$ box on $\mathbb{B}_{70}$ gives another result when comparing the mean sample value for the statistic with its expected value. The systematic deviation seen in Figure 11.6 is gone and instead we see a more random deviation. The full diagram is on the left and a magnified version on the right.*

If we want to reduce the risk of getting one-sided intervals we have to use complete data on box $\mathbb{B}_{20}$ or at least $p_s = 0.2$ on box $\mathbb{B}_{50}$ corresponding to 2040 data points. To achieve an expected precision of ≈ 0.1 or less we have to use selection probability $p_s = 0.2$ on boxes $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$, or $p_s \geq 0.6$ on $\mathbb{B}_{20}$, equivalent to around 1100 data points, or more (see Table 11.1).

Our sharpest estimate comes from simulation on box $\mathbb{B}_{20}$ with complete data ($p_s = 1$), and there we have used 1681 data points. Suppose we spread out these observations over a larger area, assuming we get the same precision, we can then use $p_s \approx 0.46$ on box $\mathbb{B}_{30}$, $p_s \approx 0.26$ on box $\mathbb{B}_{40}$ or $p_s \approx 0.17$ on box $\mathbb{B}_{50}$.

Reliability

We estimate $P(V)$ from data on boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$, see Figure 11.9. By using large enough boxes (for $p_s = 0.2$) we can almost remove the probability for unreliable confidence intervals for $\beta < 0.35$, using on average 2040 data points. The shape of the curves in Figure 11.9 suggests that the closer the actual value of β is to β_c the amount of extra data required, to get a reliable confidence interval, is substantial.

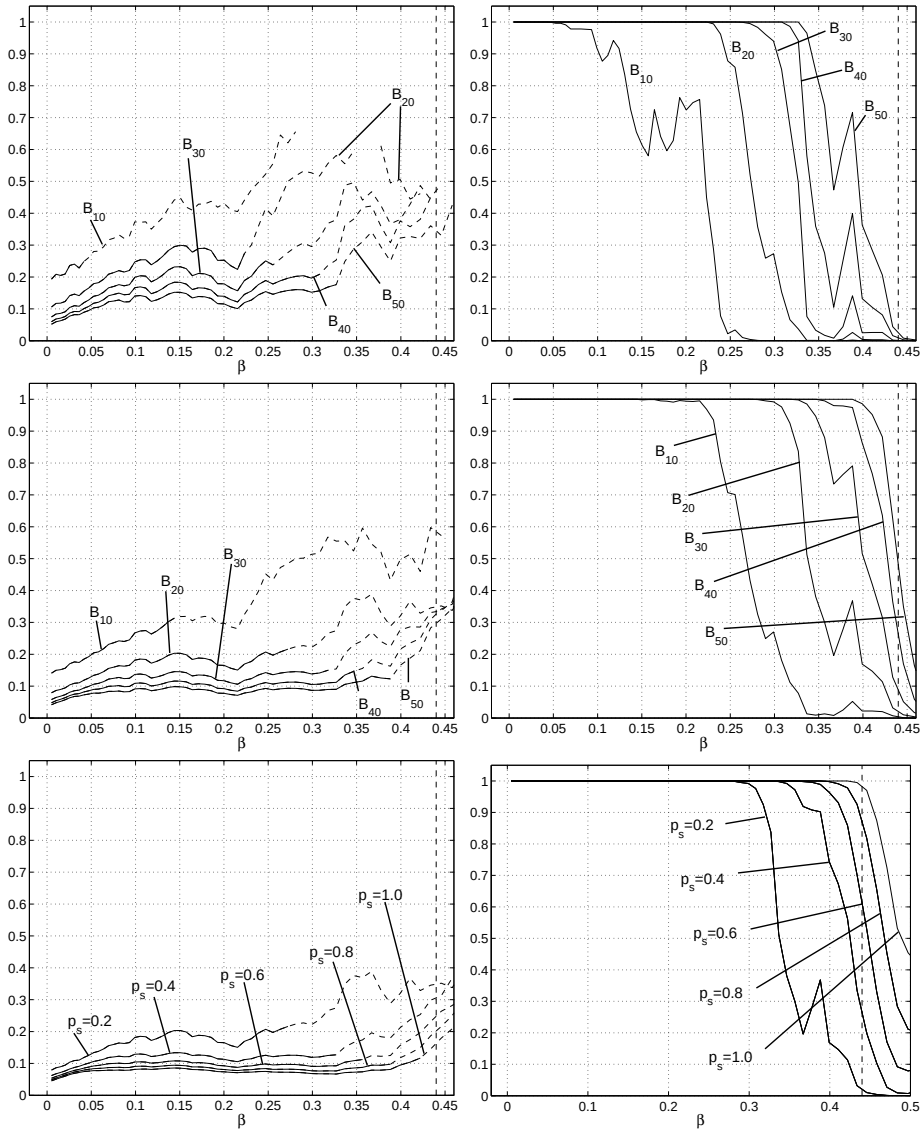

Figure 11.8: *The width of level* 0.95 *confidence intervals for different boxes, using* $p_s = 0.1$ *(upper left),* $p_s = 0.2$ *(middle left) for boxes* $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ *and* $\mathbb{B}_{50}$. *Below (also left) we see the corresponding curves on box* $\mathbb{B}_{20}$ *for* $p_s = 0.2$, $p_s = 0.4$, $p_s = 0.6$, $p_s = 0.8$ *and* $p_s = 1.0$. *Solid lines for* $\mathbb{E}[W_{n,\alpha,p_s}]$ *and dashed lines for* $\mathbb{E}[W_{n,\alpha,p_s} | W_{n,\alpha,p_s} < \infty]$. *The right diagrams show us* $\mathbb{P}(W_{n,\alpha,p_s} < \infty)$ *for the corresponding data sets.*

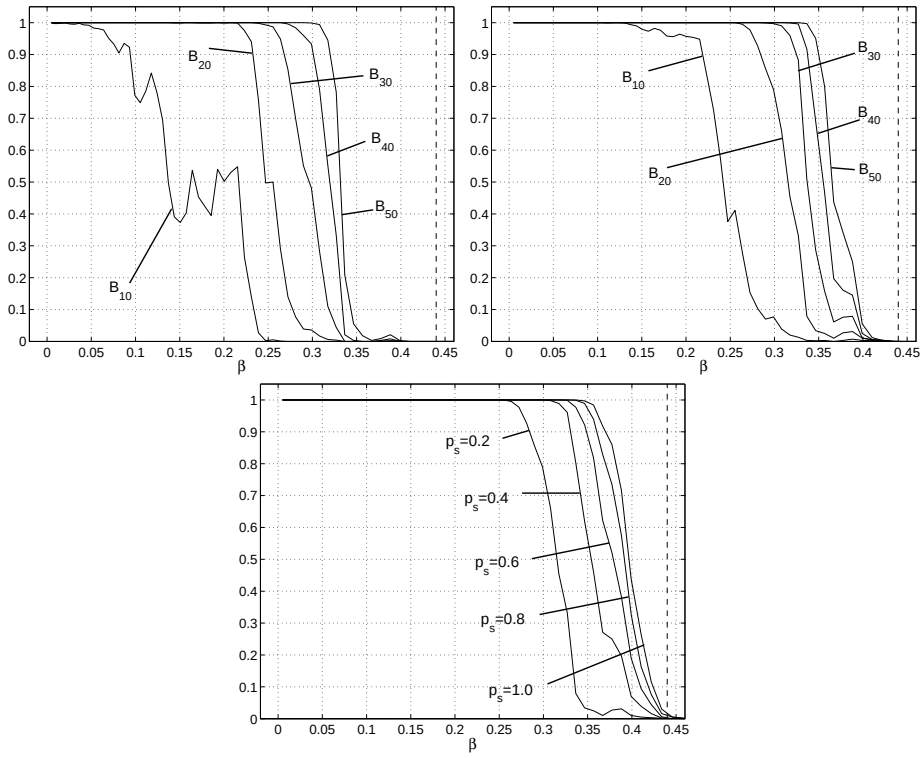

Figure 11.9: *Estimations of* $\mathbb{P}(V)$ *on all boxes using* $p_s = 0.1$ *(left),* $p_s = 0.2$ *(right) and for* $p_s = 0.2, 0.4, 0.6, 0.8$ *and* 1.0 *on* $\mathbb{B}_{20}$ *(lower). Desired level is* $\alpha = 0.95$*.*

p_s	$\mathbb{B}_{10}$	$\mathbb{B}_{20}$	$\mathbb{B}_{30}$	$\mathbb{B}_{40}$	$\mathbb{B}_{50}$
0.1	44	168	372	656	1020
0.2	88	336	744	1312	2040
0.4	–	672	–	–	–
0.6	–	1086	–	–	–
0.8	–	1344	–	–	–
0.2	–	1681	–	–	–

Table 11.1: *The expected amount of data points, $p_s|\mathbb{B}_n| = p_s(2n+1)^2$, used on average on boxes of different sizes for the considered selection probabilities. Note that each data point may require two measurements*

11.5 Box interactions

We now consider a more general form of interactions. Let the measurement region $B(v)$ be an $a \times b$-box in $\mathbb{Z}^2$ having its lower left corner at vertex v, and let $E(v)$ be the set of unordered pairs of vertices in B. We define the measurement function

$$f_{a,b}(v) = \frac{1}{|E(v)|} \sum_{\langle l,l' \rangle \in E(v)} X(l,l')$$

where $X(l,l') = X(l)X(l')$ and X follows the Ising distribution. Let M be a set of marked points according to Section 9.1. By letting

$$Y(v) = M(v) f_{a,b}(v)$$

the data $\{Y(v) : v \in \mathbb{Z}^2\}$ follows the marked Ising box interaction distribution. We repeat the procedure from the last two sections, using an average of connection probabilities instead of the pairwise connection probability.

We begin with the necessary definitions, the statistic and average connection probabilities.

Definition 11.2 ***The box statistic***
Let $\mathbb{B}_n$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Ising box interaction distribution on $\mathbb{B}_n$. We define the statistic $S_n(Y)$ as

$$S_n(Y) = \frac{1}{p_s|\mathbb{B}_n|} \sum_{v \in \mathbb{B}_n} Y(v).$$

Instead of the random cluster connection probability between neighbours we use an average of connection probabilities over all pairs of vertices in $E(v)$. The estimation procedure was presented in Section 10.1.

Definition 11.3 ***Average connection probability***
Let $\mathbb{B}$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Ising box interaction distribution on $\mathbb{B}_n$. We define the average connection probability $f_{avg}^{a,b}$ as

$$f_{avg}^{a,b}(p) = \frac{1}{|E(\mathbf{0})|} \sum_{\langle l,l' \rangle \in B(\mathbf{0})^2} \phi_{p,2}^{(n)}(l \leftrightarrow l')$$

We are now ready to state and prove unbiasedness (straightforward calculation) and conditional consistency (Theorem 9.6, page 118) for $S_n(Y)$.

Theorem 11.3
Let Y be distributed according to the marked Ising box interaction model, and let $\phi_{1-e^{-2\beta},2}$ be the corresponding random cluster measure. Let $S_n(Y)$ be defined as in Definition 11.2. Then $S_n(Y)$ is an unbiased estimator of $f_{avg}^{a,b}(\mathbf{0})$. If the underlying Ising process is subcritical then

$$\frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} \left(S_n(Y) - f_{avg}^{a,b}(\mathbf{0}) \right) \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$, and $S_n(Y)$ is consistent.

Proof:
Let $S_n(Y)$ be as defined above. It is indeed unbiased.

$$\begin{aligned}
\mathbb{E}[S_n(Y)] &= \mathbb{E}\left[\frac{1}{p_s|\mathbb{B}_n|} \sum_{v \in \mathbb{B}_n} \frac{1}{|E(v)|} \sum_{\langle l,l' \rangle \in B(v)^2} Y(l,l') \right] \\
&= \frac{1}{p_s|\mathbb{B}_n||E(v)|} \sum_{v \in \mathbb{B}_n} \sum_{\langle l,l' \rangle \in B(v)^2} \mathbb{E}[Y(l,l')] \\
&= \frac{|\mathbb{B}_n|}{p_s|E(v)||\mathbb{B}_n|} \sum_{\langle l,l' \rangle \in B(v)^2} p_s \phi_{p,2}^{(n)}(l \leftrightarrow l') \\
&= f_{\text{avg}}^{a,b}(p)
\end{aligned}$$

From Bolthausen's central limit theorem we have

$$\sum_{v\in\mathbb{B}_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\mathbb{B}_n|^{1/2}} \xrightarrow{\mathcal{D}} Z_0$$

where Z_0 follows the standard normal distribution. We make the following rearrangements,

$$\begin{aligned}
\sum_{v\in\mathbb{B}_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\mathbb{B}_n|^{1/2}} &= \frac{1}{\sigma_Y |\mathbb{B}_n|^{1/2}} \left(\sum_{v\in\mathbb{B}_n} Y(v) - p_s |\mathbb{B}_n| f^{a,b}_{\text{avg}}(\mathbf{0}) \right) \\
&= \frac{(p_s|\mathbb{B}_n|)^{-1} \sum_{v\in\mathbb{B}_n} Y(v) - f^{a,b}_{\text{avg}}(\mathbf{0})}{\sigma_Y \; p_s^{-1} \; |\mathbb{B}_n|^{-1/2}} \\
&= \frac{S_n(Y) - f^{a,b}_{\text{avg}}(\mathbf{0})}{\sigma_Y \; p_s^{-1} \; |\mathbb{B}_n|^{-1/2}}
\end{aligned}$$

and

$$\frac{\sigma_Y}{p_s |\mathbb{B}_n|^{1/2}} \; (\; S_n(Y) - f^{a,b}_{\text{avg}}(\mathbf{0}) \;) \xrightarrow{\mathcal{D}} Z \qquad (*)$$

where $Z \overset{\mathcal{D}}{=} N(0,1)$. Consistency now follows since

$$|S_n(Y) - f^{a,b}_{\text{avg}}(\mathbf{0})| \xrightarrow{P} 0$$

is implied by unbiasedness and $(*)$.
□

We estimate $f^{a,b}_{\text{avg}}$ with $\hat{f}^{a,b}_{\text{avg}}$ by using $m = 10000$ simulations and use it to produce a point and interval estimations for β. Apart from using another function the method in this section is identical to the one used for the marked Ising interaction model in Section 11.3. We restate slightly reformulated versions here.

Point and interval estimators

Given the statistic $S_n(Y)$ our estimate of β is

$$\hat{\beta} = -\frac{1}{2} \; \log(\; 1 - (\hat{f}^{a,b}_{\text{avg}})^{-1} (\; S_n(Y) \;) \;)$$

The confidence interval in given by the following result.

Theorem 11.4 ***Confidence interval for the inverse temperature*** *Let Y be distributed according to the marked Ising box interaction model on a finite box $\mathbb{B}_n$ at inverse temperature β and let β_c be the critical temperature. Let $S_n(Y)$ and $\hat{f}_{avg}^{a,b}$ be defined as above. Then $[\beta_{low}, \beta_{up}]$ is a level α confidence interval for the inverse temperature β. The bounds are given by*

$$\beta_{low} = -\frac{1}{2}\ \log\left(1 - (f_{avg}^{a,b})^{-1}\left(\ S_n(Y) - z_{(\sqrt{\alpha}+1)/2}\ \frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} - \frac{D_{\sqrt{\alpha}}}{\sqrt{m}}\ \right)\right)$$

and

$$\beta_{upp} = -\frac{1}{2}\ \log\left(1 - (f_{avg}^{a,b})^{-1}\left(\ S_n(Y) + z_{(\sqrt{\alpha}+1)/2}\ \frac{\sigma_Y}{p_s|\mathbb{B}_n|^{1/2}} + \frac{D_{\sqrt{\alpha}}}{\sqrt{m}}\ \right)\right)$$

where z_α and D_α are quantiles from the standard normal distribution and the Kolmogorov-Smirnov distribution respectively. The confidence interval is valid if $\beta_c \notin [\beta_{low}, \beta_{up}]$.

As for a proof, see Section 11.3 and the discussion before Theorem 11.2.

11.6 Simulation study: box interactions

We simulate the statistic S for boxes $\mathbb{B}_{10}$ and $\mathbb{B}_{20}$ using $a = b = 2$ and selection probabilities, $p_s = 0.1, 0.2$. We generate 100 omniparametric samples Y_k for each box size, calculate statistics

$$S_n(Y_1), ..., S_n(Y_{100})$$

and then proceed as before with validation, precision and reliability, see Section 11.4 for a detailed description of the procedure.

Validation

In Figure 11.11 and 11.10 we see the simulation results regarding estimated confidence level.

As in Section 11.4 the confidence levels are met for $p_s = 0.1, 0.2$ on all box sizes, but there is a difference for more closely spaced observations on $\mathbb{B}_{20}$, when we use $p_s > 0.2$. In Figure 11.10 we see a drop in estimated confidence level from expected 0.95 to ≈ 0.7. This deviation is consistent with the behaviour of the sample mean related to expected sample mean in Figure 11.12. We see that for relatively sparse data

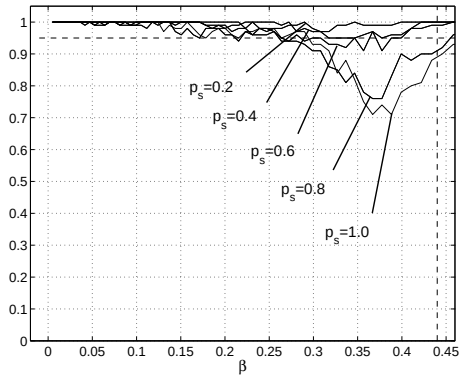

Figure 11.10: *Estimated confidence levels $\hat{\alpha}$ for confidence intervals based on data within box $\mathbb{B}_{20}$, for different choices of selections probabilities, $p_s = 0.2, 0.4, 0.6, 0.8$ and $p_s = 1.0$.*

($p_s = 0.1$) the deviation is small compared to the width of the confidence interval, while for $p_s = 0.8$ the deviation is of the same magnitude as the allowed deviation according to Theorem 11.4.

Precision

We study both expected width of the confidence intervals, $\mathbb{E}[W_{n,\alpha,p_s}]$ and its truncated counterpart, $\mathbb{E}[W_{n,\alpha,p_s}|W_{n,\alpha,p_s} < \infty]$, see Figure 11.13 for simulation results. As expected we achieve more accurate estimates on box $\mathbb{B}_{20}$ than on $\mathbb{B}_{10}$ regardless of selection probability.

In order to achieve expected precision $\mathbb{E}[W_{n,\alpha,p_s}] \leq 0.1$ we need at least selection probability $p_s = 0.2$ on $\mathbb{B}_{20}$ which is equivalent to using 336 data points on average. Note that a data point in this case is based on no less than four measurements, so we actually need more than 1300 measurements for the wanted precision.

There is also a risk for unbounded, or one-sided intervals. On the smaller box we have that risk for $\beta \geq 0.25$ using $p_s = 0.1$ and for $\beta \geq 0.30$ using $p_s = 0.2$. On the larger box that the risk is limited to β larger than ≈ 0.32. In Figure 11.13 (right column) we see how the probability for bounded interval width (two-sided intervals) varies with the true value of β.

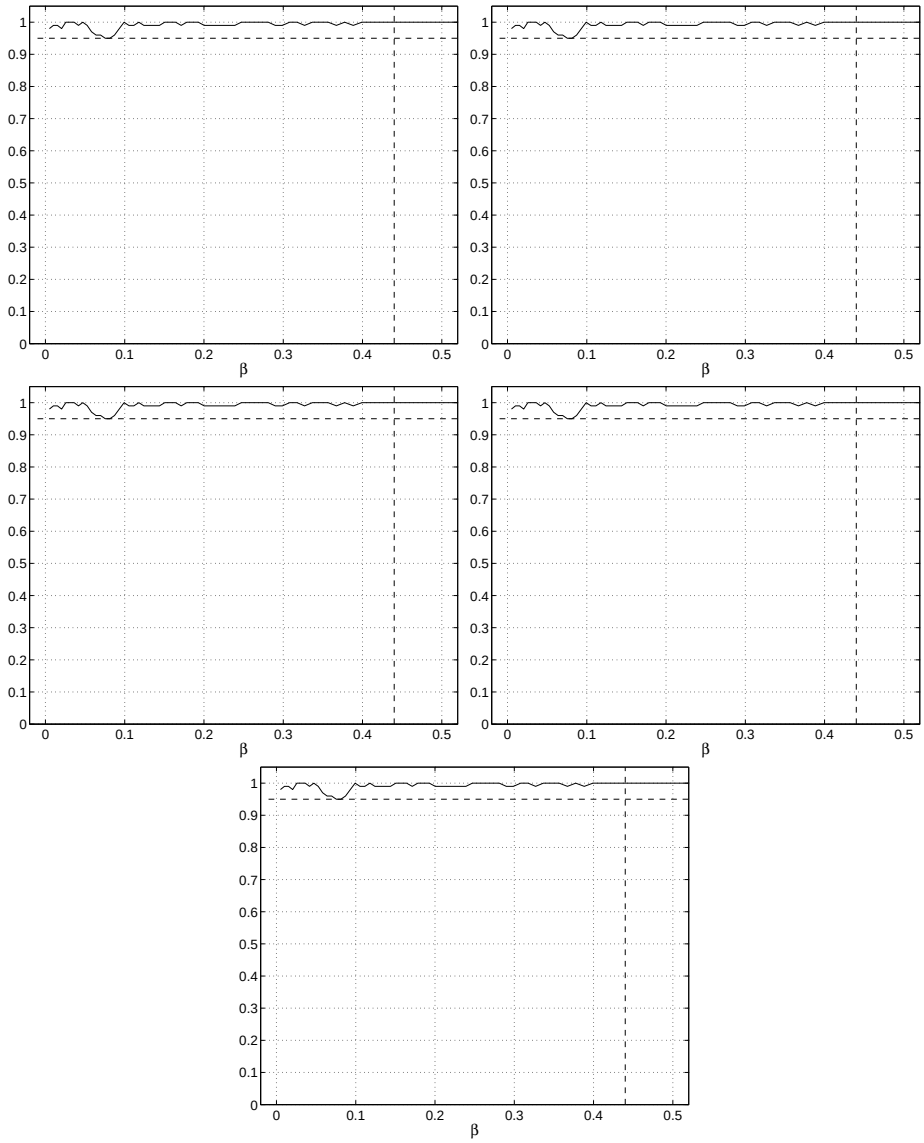

Figure 11.11: *Estimated confidence levels for the confidence interval for β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). Desired confidence level is $\alpha = 0.95$. The results are given for $p_s = 0.10$. For corresponding results using $p_s = 0.20$ see Figure B.2 on page 211.*

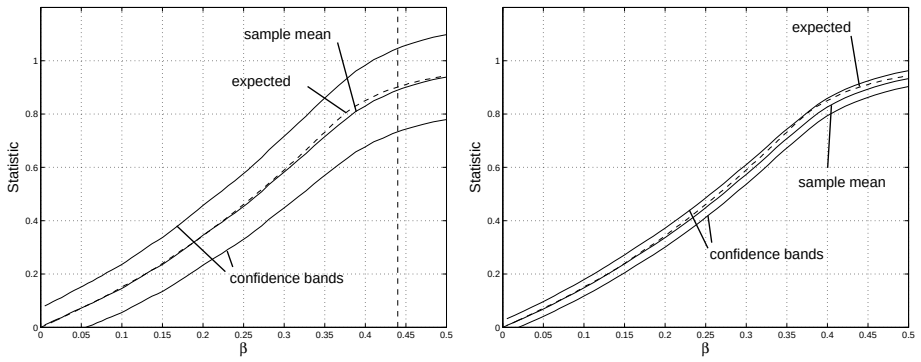

Figure 11.12: *In the left diagram we see the mean value of the statistic $S_n(Y)$ (solid line) and confidence bands for the true value using the central limit theorem (Theorem 9.6) and confidence level $\alpha = 0.95$ for $p_s = 0.1$ on $\mathbb{B}_{20}$. The curves for $\mathbb{E}[S_n(Y)]$ as a functions of β are dashed. The left diagram shows the corresponding information for $p_s = 0.8$, also on $\mathbb{B}_{20}$.*

Reliability

We naturally want the probability of a non-valid confidence interval to be as small as possible, hopefully close to zero, but it is only possible for sufficiently small β, see Figure 11.14 for results. More data give narrower confidence intervals, and as a consequence we get reliable results for larger values of β. On the smaller box we can, by using $p_s = 0.2$, get reliable results with large probability for $\beta \leq 0.25$, and on the larger box the same holds for $\beta \leq 0.33$.

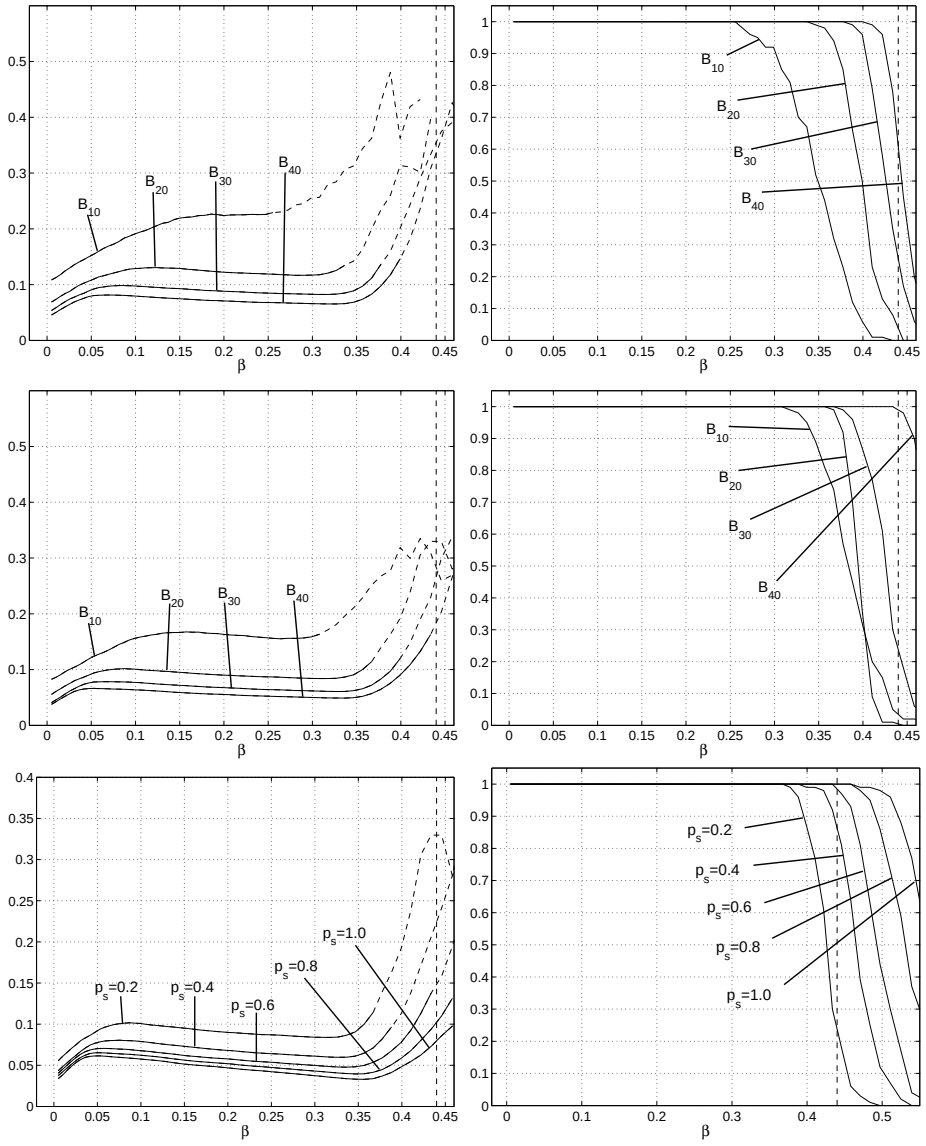

Figure 11.13: *Width of level* 0.95 *confidence intervals for different boxes, using* $p_s = 0.1$ *(upper left) and* $p_s = 0.2$ *(middle left) on* $\mathbb{B}_{10}$ *and* $\mathbb{B}_{20}$*. In the lower left diagram we see the corresponding information for* $p_s = 0.2, = 0.4, 0.6, 0.8$ *and* $p_s = 1.0$ *on* $\mathbb{B}_{50}$*. Solid lines represent* $\mathbb{E}[W_{n,\alpha,p_s}]$*, and dashed lines* $\mathbb{E}[W_{n,\alpha,p_s}|W_{n,\alpha,p_s} < \infty]$*. The right column shows* $\mathbb{P}(W_{n,\alpha,p_s} < \infty)$ *in the corresponding situations.*

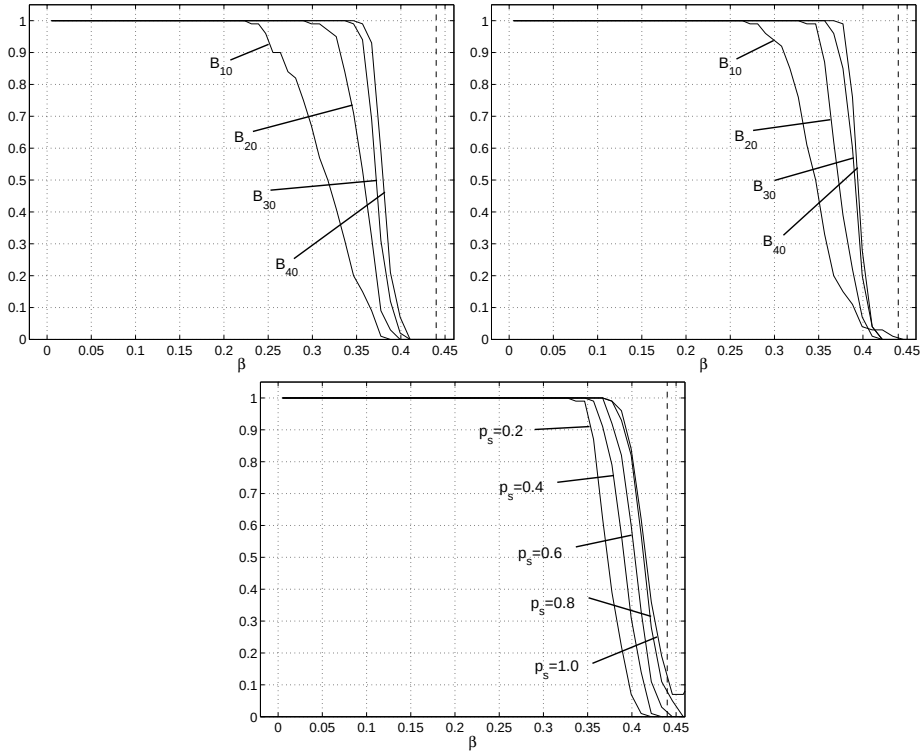

Figure 11.14: *Estimations of $\mathbb{P}(V^c)$ as functions of β for boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$ using $p_s = 0.1$ (left) and $p_s = 0.2$ (right), both for $\alpha = 0.95$.*

11.7 A non-asymptotic method

Instead of using central limit theorems to capture the variance of a statistic we use simulations. The idea is the following: Given a statistic T we generate omniparametric random cluster samples $X_1, ..., X_n$ and calculate T as a function

$$f_k^T(p) = T(\ \mathbf{P}_p^{\text{Ising}}(X_k)\ )\ , \quad k = 1, ..., n$$

of $p = 1 - e^{-2\beta}$. We make inference for p since it varies over the unit interval and is easier to handle than β which varies over the positive real line. This approach works for all values of p, regardless of phase. Given point and interval estimates of p it is easy to convert them into corresponding estimates of β. We are free to choose the statistic T, as long as

$$f(p) = \mathbb{E}_p[T(X)]$$

is reasonable as a function of p and of course as long as it is independent of data. We will use pairwise interactions as our T since it is a convenient way to compare this direct approach to parameter estimation to methods using asymptotic theory.

From simulations we get a sequence of functions,

$$f_1^T, f_2^T, ..., f_n^T,$$

and for each $p \in [0, 1]$ we order the values $f_1^T(p), ..., f_n^T(p)$ in increasing order, denoting the ordered sequence by

$$f_{(1)}^T(p), ..., f_{(n)}^T(p).$$

Suppose we want a level α confidence interval. We use the empirical $1 - \alpha/2$- and $(1 + \alpha)/2$-quantiles,

$$f^T_{(\lfloor n(1-\alpha/2)\rfloor)}(p) \ \text{ and } \ f^T_{(\lceil n(1+\alpha)/2\rceil)}(p)$$

as bounds for the confidence interval, giving us a confidence band with upper and lower limits

$$f_{\alpha,\text{low}}(p) = f^T_{(\lfloor n(1-\alpha/2)\rfloor)}(p), \text{and } f_{\alpha,\text{up}}(p) = f^T_{(\lceil n(1+\alpha)/2)\rceil)}(p)$$

respectively. We also construct a mean value curve

$$f_{\text{avg}}(p) = \frac{1}{n}\ \sum_{k=1}^{n} f_k^T(p)\ , \quad p \in [0, 1]$$

for calculating point estimates.

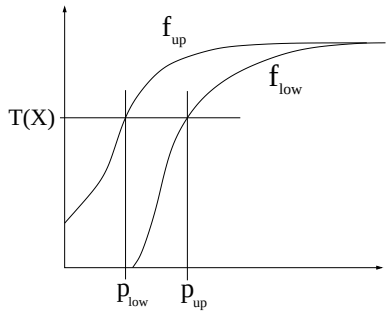

Figure 11.15: *An example of how method three works. Note that if the curves f_{low} and f_{up} defines a level α confidence band for $f(p) = \mathbb{E}_p[T(X)]$ the interval $[p_{low},\ p_{up}]$ is a level α confidence interval for p.*

11.7.1 The point and interval estimators

Given data X and the calculated statistic $T(X)$ we let

$$\hat{\beta} = -\ \frac{1}{2}\ \log(\ 1 - f_{\text{avg}}^{-1}(\ T(X)\ )\ )$$

be our point estimate of β. A level α conservative confidence interval for β is $[\beta_{\text{low}}, \beta_{\text{up}}]$ as described below.

Theorem 11.5 ***Confidence interval for the inverse temperature***
Let Y be distributed according to the Ising model on a finite box $\mathbb{B}_n$ at inverse temperature β and let β_c be the critical temperature. Let T be any statistic with finite first and second moments. Generate omniparametric Ising samples $X_1, ..., X_n$ and calculate quantile functions $f_{1-\alpha,low}$ and $f_{1-\alpha,up}$ as described above. Then $[\beta_{low}, \beta_{up}]$ is a level α confidence interval for the inverse temperature β, with bounds as follows.

$$\beta_{low} = -\ \frac{1}{2}\ \log(1 - p_{low}), \quad p_{low} = \min\{p : f_{\alpha,up}(p) = T(X)\}$$

$$\beta_{up} = -\ \frac{1}{2}\ \log(1 - p_{up}), \quad p_{up} = \max\{p : f_{\alpha,low}(p) = T(X)\}$$

Proof:
Suppose we have a confidence interval $[a, b]$ for β constructed according to the statement with intended coverage probability α. Consider the

corresponding interval $[p_a, p_b]$ for p where $p_a = 1 - e^{-2a},\ p_b = 1 - e^{-2b}$. Let f_{up} and f_{low} be the percentile functions based on n omniparametric Ising samples. If the samples are distributed according to the omniparametric Ising distribution then

$$\mathbb{P}\left(\ \forall p \in [0,1] : \mathbb{E}_p[T(X)] \in [f_{\text{low}}(p), f_{\text{up}}(p)]\ \right) \approx \alpha$$

where $\mathbb{E}_p[\cdot]$ is the expectation under the assumption that p is the correct parameter value. Then

$$\{p : T(X) \in [f_{\text{low}}(p), f_{\text{up}}(p)]\}$$

is a level α confidence set for p (see Figure 11.15). This set may or may not be an interval, depending on how smooth the quantile estimate are. So by making a conservative choice and letting

$$p_{\text{low}} = \min\{p : T(X) \in [f_{\text{low}}(p), f_{\text{up}}(p)]\}$$

and

$$p_{\text{up}} = \max\{p : T(X) \in [f_{\text{low}}(p), f_{\text{up}}(p)]\}$$

be boundaries of the set we get a confidence interval. These are the exact boundaries we get by using the expressions in the statement, and we are done.
□

The interval is conservative since functions $f_{\alpha,\text{low}},\ f_{\alpha,\text{up}}$ may not be monotone. This happens especially when we base the statistic on a small number of observations.

11.7.2 Generating percentile functions

Given a fixed measurement configuration we generate omniparametric Ising samples $X_1, ..., X_n$ and calculate the percentile functions. The question is how large we should make n? By choosing a larger n our percentile estimates become better and better, but it will not reduce the band width,

$$f_{1-\alpha,\text{up}}(\beta) - f_{1-\alpha,\text{low}}(\beta),$$

in order to do so, we have to use more data, that is, extend the measurement region, increase the selection probability, or both.

Unfortunately there are, to our knowledge, no available asymptotic theory for order statistics with discrete parent distribution. Without

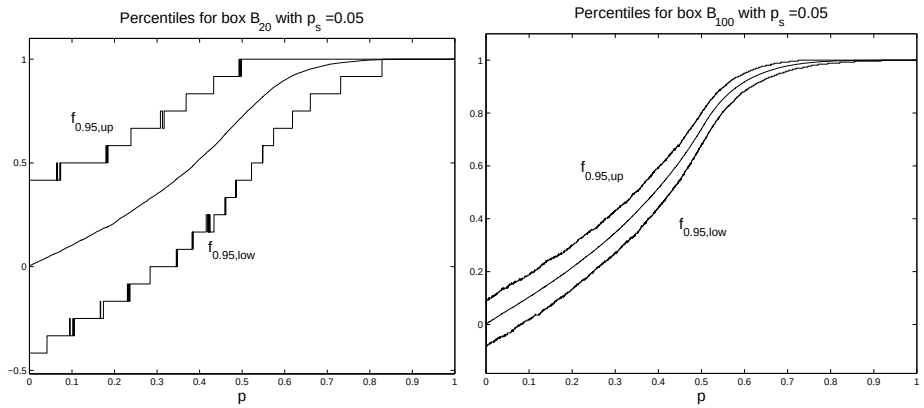

Figure 11.16: *We see the percentile functions $f_{\alpha,low}$ and $f_{\alpha,up}$ for boxes $\mathbb{B}_{10}$ and $\mathbb{B}_{50}$ both using selection probability 0.05. The extra data used on box $\mathbb{B}_{50}$ comprises of measurements at another 505 locations, compared to the 24 locations used on $\mathbb{B}_{10}$.*

any estimate of the precision of the sample percentiles we use 1000 omniparametric simulations and for each inverse temperature we use

$$\begin{aligned} \beta_{\text{low}} &= X_{(\lfloor n(1-\alpha)/2 \rfloor)} \\ \beta_{\text{up}} &= X_{(\lfloor n(1+\alpha)/2 \rfloor)]} \end{aligned}$$

as bounds for our level α confidence interval, see Figure 11.16.

11.8 Simulation study: the non-asymptotic method

We generate for each box size and each selection probability 100 omniparametric samples and study validation, precision and phase determination. We test the procedure on the previously used pairwise interactions. There is an important difference however. In this setting we view the measurement function and the randomly selected locations together as a statistic. When generating the 100 samples we use the locations as fixed and only the underlying values of the random field are varied. By generating the locations at random as before we could view them as a part of the randomization and for each simulation not

just generate the underlying data but also the locations. Our approach however give us the possibility to choose the locations more freely.

Validation

The simulation results are on display in figures 11.17 and B.3 (page 170,212). The estimated confidence levels are fulfilled on all box sizes and all selection probabilities. We can also use complete or almost complete data and still get confidence intervals with expected confidence level, see fig B.4 (page 213).

Precision

We study precision for $p_s = 0.1, 0.2$ on boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$, and also for $p_s = 0.4, 0.6, ..., 1.0$ on $\mathbb{B}_{20}$. The results are shown in Figure 11.18.

Suppose we want a confidence interval of width 0.1 or less. We see that on box B_{30} using at least $p_s = 0.1$ we almost have an average width of less than 0.1. For that alternative we use on average 360 data points. The combinations $\mathbb{B}_{40}, p_s = 0.05$ and $\mathbb{B}_{50}, p_s = 0.05$ also has average width around 0.1, using 320 and 500 data points respectively. If we are content with an interval width around 0.2 we only need approximately 80 data points, using combinations $\mathbb{B}_{10},\ p_s = 0.20$ and $\mathbb{B}_{20},\ p_s = 0.05$.

Phase determination

We consider data on the usual boxes, results are given in Figure 11.19 (page 172).

Given enough data we can, for certain intervals of the inverse temperature, determine which phase the system is in. On box $\mathbb{B}_{10}$ we can only hope to establish subcriticality, for $\beta < 0.2$ if we use $p_s \geq 0.2$. On box $\mathbb{B}_{20}$ we can for $p_s \geq 0.10$ determine both phases if β are sufficiently small or sufficiently large. By using box size $\mathbb{B}_{50}$ there is only a small interval around β_c where we can not determine the phase of the underlying system. Using selection probability $p_s \geq 0.2$ we can now determine the phase for $\beta \leq 0.4$ or $\beta \geq 0.51$.

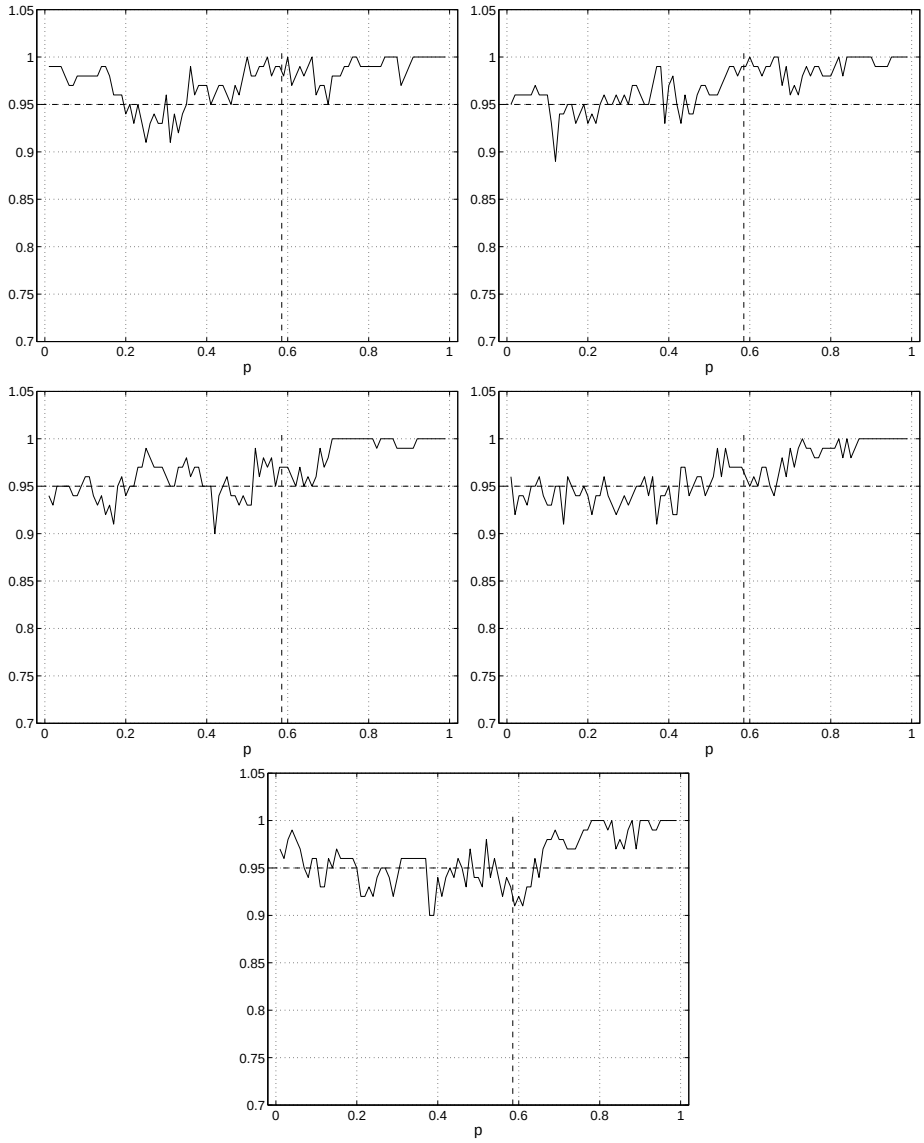

Figure 11.17: *Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). The desired confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.10$. For corresponding results with $p_s = 0.20$ see figures B.3 and B.4 on pages 212 and 213 in Appendix B.*

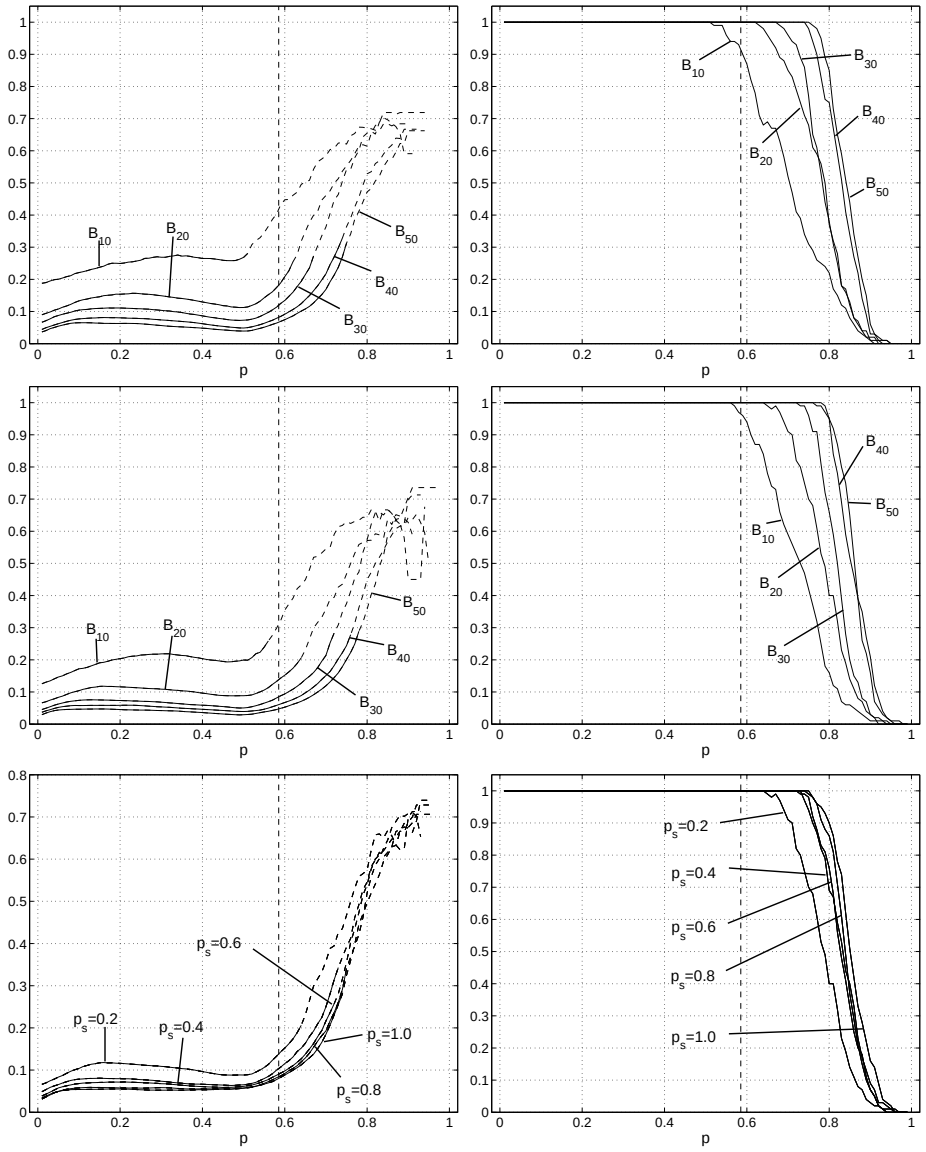

Figure 11.18: *Width of level* 0.95 *confidence intervals for different boxes, using* $p_s = 0.1$ *(upper left) and* $p_s = 0.2$ *(middle left) for boxes* $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ *and* $\mathbb{B}_{50}$ *and for* $p_s \in \{0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0\}$ *on* $\mathbb{B}_{20}$ *(lower left). Solid lines represent* $\mathbb{E}[W_{n,\alpha,p_s}]$*, and dashed lines* $\mathbb{E}[W_{n,\alpha,p_s}|W_{n,\alpha,p_s} < \infty]$*. In the right column we see* $\mathbb{P}(W_{n,\alpha,p_s} < \infty)$ *for the corresponding data sets.*

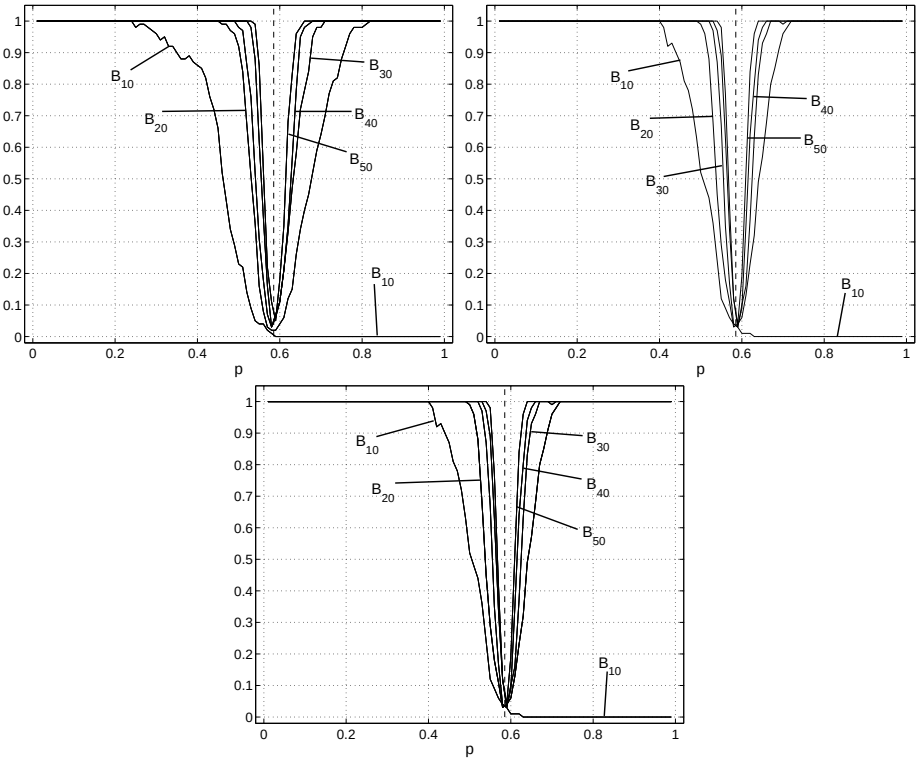

Figure 11.19: *We see how* $\mathbb{P}(\beta_c \notin [\beta_{low}, \beta_{up}])$ *vary in the different situations for* $p = 1 - e^{-2\beta}$*. The upper two diagrams show us the situation on boxes* $\mathbb{B}_{10}$*,* $\mathbb{B}_{20}$*,* $\mathbb{B}_{30}$*,* $\mathbb{B}_{40}$ *and* $\mathbb{B}_{50}$ *for* $p_s = 0.1$ *(left) and* $p_s = 0.2$ *(right). The lower diagram shows us the the situation on box* $\mathbb{B}_{20}$ *for different choices of* p_s*. The confidence level is* $\alpha = 0.95$*.*

CHAPTER 12

Parameter estimation in the Potts model

12.1 Introduction

We now turn our focus to the Potts model. The Potts model has two parameters, the inverse temperature β, and the number of types q a vertex can assume. Throughout the chapter we will assume that q is known and fixed at some value. We adjust and apply the methods from Chapter 11 regarding the Ising model, two asymptotic methods and one non-asymptotic. The presentation is kept short, only material essentially different from the chapter regarding the Ising model will be presented. For a full description of all details see Chapter 11.

When using central limit theorems we must ensure that the correlations are decaying exponentially and we can only do that if $\beta < \beta_c^I$ where β_c^I is the critical inverse temperature for the Ising model. See Definition 7.8, Theorem 7.7 and Lemma 7.2 on pages 69 and 70.

There is one exception from Chapter 11. The section regarding phase determination for the non-asymptotic method is left out due to the simple fact that $\beta_c(q)$ is unknown for $q = 4$ and that β_c^I has no meaning in this case.

12.2 Pairwise interactions

We define the appropriate statistic $S_n(Y)$ for data observed within finite boxes, establish unbiasedness and consistency. With Theorem 9.2 in mind we make the following definition

Definition 12.1 ***The statistic***
Let $\mathbb{B}_n$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Potts interaction distribution on $\mathbb{B}_n$. We define the statistic $S_n(Y)$ as

$$S_n(Y) = \left(\frac{q}{q-1}\right)\left(\frac{1}{2\ p_s\ |\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} Y(v) + \frac{1}{2} - \frac{1}{q}\right).$$

Now to the Potts version of Theorem 11.1.

Theorem 12.1
Let Y be distributed according to the marked Potts interaction model, and let $\phi_{p,q}$ be the corresponding random cluster measure. Let the statistic $S_n(Y)$ be as defined in Definition 12.1. Then $S_n(Y)$ is unbiased as an estimator of $\phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{1})$. If $\beta<\beta_c^I$ then $S_n(Y)$ is consistent and we have

$$\left(\frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)}\right)^{-1}\ (\ S_n(Y) - \phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{1})\ ) \xrightarrow{\mathcal{D}} Z$$

where $Z \stackrel{\mathcal{D}}{=} N(0,1)$.

Proof:
Let $S_n(Y)$ be as defined above and let for simplicity $A = q(q-1)^{-1}$. Then it is indeed unbiased.

$$\begin{aligned}\mathbb{E}[S_n(Y)] &= A\ \left(\frac{1}{2\ p_s\ |\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n}\mathbb{E}[Y(v)] + \frac{1}{2} - \frac{1}{q}\right)\\ &= A\ \left(\frac{1}{2\ p_s\ |\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} p_s\left(\frac{2}{q} + \frac{2}{A}\ \phi_{p,q}(\mathbf{0}\leftrightarrow\mathbf{e}_1) - 1\right) + \frac{1}{2} - \frac{1}{q}\right)\end{aligned}$$

$$= A \left(\left(\frac{1}{q} + \frac{1}{A} \phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{e}_1) - \frac{1}{2} \right) + \frac{1}{2} - \frac{1}{q} \right)$$

$$= \phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{e}_1)$$

From Bolthausen's central limit theorem we have

$$\sum_{v \in \mathbb{B}_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\mathbb{B}_n|^{1/2}} \xrightarrow{\mathcal{D}} Z_0$$

where Z_0 follows the standard normal distribution. We make the following rearrangements.

$$\sum_{v \in \mathbb{B}_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y |\mathbb{B}_n|^{1/2}}$$

$$= \frac{1}{\sigma_Y |\mathbb{B}_n|^{1/2}} \left(\sum_{v \in \mathbb{B}_n} Y(v) - |\mathbb{B}_n| p_s \left(\frac{2}{q} + \frac{2}{A} \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) - 1 \right) \right)$$

$$= \frac{2p_s |\mathbb{B}_n|^{1/2}}{\sigma_Y} \left(\frac{1}{2p_s |\mathbb{B}_n|} \sum_{v \in \mathbb{B}_n} Y(v) - \frac{1}{q} - \frac{1}{A} \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) + \frac{1}{2} \right)$$

$$= \frac{1}{A} \frac{2p_s |\mathbb{B}_n|^{1/2}}{\sigma_Y} \left(A \left(\frac{1}{2p_s |\mathbb{B}_n|} \sum_{v \in \mathbb{B}_n} Y(v) + \frac{1}{2} - \frac{1}{q} \right) - \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) \right)$$

$$= \frac{1}{A} \frac{2p_s |\mathbb{B}_n|^{1/2}}{\sigma_Y} \left(S_n(Y) - \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) \right)$$

It follows that

$$\left(\frac{q-1}{q} \right) \frac{2p_s |\mathbb{B}_n|^{1/2}}{\sigma_Y} \left(S_n(Y) - \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1}) \right) \xrightarrow{\mathcal{D}} Z$$

where $Z \overset{\mathcal{D}}{=} N(0,1)$. Now

$$|S_n(Y) - \phi^{(n)}_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1})| \xrightarrow{P} 0,$$

and as a consequence also consistency follows from the above together with unbiasedness, and we are done.
□

Next we establish point and interval estimators.

The point estimator

The statistic $S_n(Y)$ is our estimate of the connection probability, and our point estimate for the inverse temperature becomes

$$\hat{\beta} = -\frac{1}{2}\ \log(\ 1 - \hat{F}_m^{-1}(\ S_n(Y)\ )$$

The interval estimator

Constructing the confidence interval is done in the same manner as in Section 11.3 with one difference, we use the standard deviation of Theorem 12.1.

Theorem 12.2 ***Confidence interval for the inverse temperature*** *Consider the Potts model on a finite box $\mathbb{B}_n$ at inverse temperature β for some $q \geq 2$ and let β_c^I be the critical temperature for the Ising model. Let $S_n(Y)$ and $\hat{F}_m$ be defined as above. Then $[\beta_{low}, \beta_{up}]$ is a level α^2 confidence interval for the inverse temperature β. The bounds are given by*

$$\beta_{lower} = -\frac{1}{2}\ \log\left(1 - \hat{F}_m^{-1}\left(\ S_n(Y) - z_\alpha\ \frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)} + \frac{D_\alpha}{\sqrt{m}}\ \right)\right)$$

and

$$\beta_{upper} = -\frac{1}{2}\ \log\left(1 - \hat{F}_m^{-1}\left(\ S_n(Y) + z_\alpha\ \frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)} - \frac{D_\alpha}{\sqrt{m}}\ \right)\right)$$

where z_α and D_α are quantiles from the standard normal distribution and the Kolmogorov-Smirnov distribution respectively. The confidence interval is valid if $\beta_c^I \notin [\beta_{low}, \beta_{up}]$.

12.3 Simulation study: pairwise interactions

We simulate the statistic S_n for boxes $\mathbb{B}_{10},\ \mathbb{B}_{20},\ \mathbb{B}_{30},\ \mathbb{B}_{40}$ and $\mathbb{B}_{50}$ using selection probabilities, $p_s = 0.1, 0.2$. Extra simulations are generated on $\mathbb{B}_{20}$ for $p_s \in \{0.40,\ 0.60,\ ...,\ 1.00\}$. For each combination $(\mathbb{B}_n, p_s)$ we generate 1000 simulations.

Validation

In Figure 12.1 and B.5 (pages 178, 214) we see the result of the procedure for $p_s = 0.10$ and $p_s = 0.20$ for $\alpha = 0.95$ and all box sizes.

The validity of the confidence intervals are good. As for the Ising model we get an estimated confidence level of ≈ 1, again indicating an overestimation of the variance of $S_n(Y)$. In Figure 12.2 we see a deviation of the estimated confidence level for large values of p_s. The estimated level becomes as low as ≈ 0.8. An explanation is the failure of $S_n(Y)$ as an unbiased estimator of $\phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1})$ for some values of β. In Figure 12.3 we see that the mean of $S_n(Y)$ deviate slightly from its expected value when using $p_s = 0.1$ on $\mathbb{B}_{20}$ and more when we use $p_s = 0.8$. Compared to the width of the confidence band the deviation in the $p_s = 0.8$ case is substantial. A possible reason for this deviation is the same as in the Ising case, see Section 11.4.

Precision

We study how much data we need to attain a certain precision of our estimates for confidence level $\alpha = 0.95$. For simulation results see Figure 12.4.

Using the smallest box size, $\mathbb{B}_{10}$ and $p_s = 0.1$, we use 44 data points and we achieve confidence intervals having width between 0.2 and ≈ 0.37 depending on the true value of β. We also suffer a risk of getting a one-sided interval for $\beta > 0.23$ (see Figure 12.4, upper left diagram). By using twice as many data points within the same area we get a slightly more narrow interval ($0.1 \leq W \leq 0.35$), but there is still a possibility of getting a one-sided interval, now for $\beta \geq 0.29$. If we want to reduce the risk of getting one-sided intervals we have to use complete or almost complete data ($p_s = 0.8, 1.0$) on box $\mathbb{B}_{20}$ or at least $p_s = 0.2$ on box $\mathbb{B}_{40}$ corresponding to 1312 data points (see Table 11.1 on page 156). To achieve an expected precision of ≈ 0.1 or less we have to use selection probability $p_s = 0.2$ on boxes $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$, or $p_s \geq 0.4$ on $\mathbb{B}_{20}$. The expected amount of data, $p_s|\mathbb{B}_n|$, we have used is given in Table 11.1.

Our sharpest estimate comes from simulation on box $\mathbb{B}_{50}$ using $p_s = 0.2$ resulting in estimated expected width between 0.04 and 0.06 for $\beta \leq 0.4$, and for this we have used 2040 data points. Suppose we spread out these observations over a larger area, assuming we get the same precision, we can then use $p_s \approx 0.46$ on box $\mathbb{B}_{30}$, $p_s \approx 0.26$ on box $\mathbb{B}_{40}$ or $p_s \approx 0.17$ on box $\mathbb{B}_{50}$.

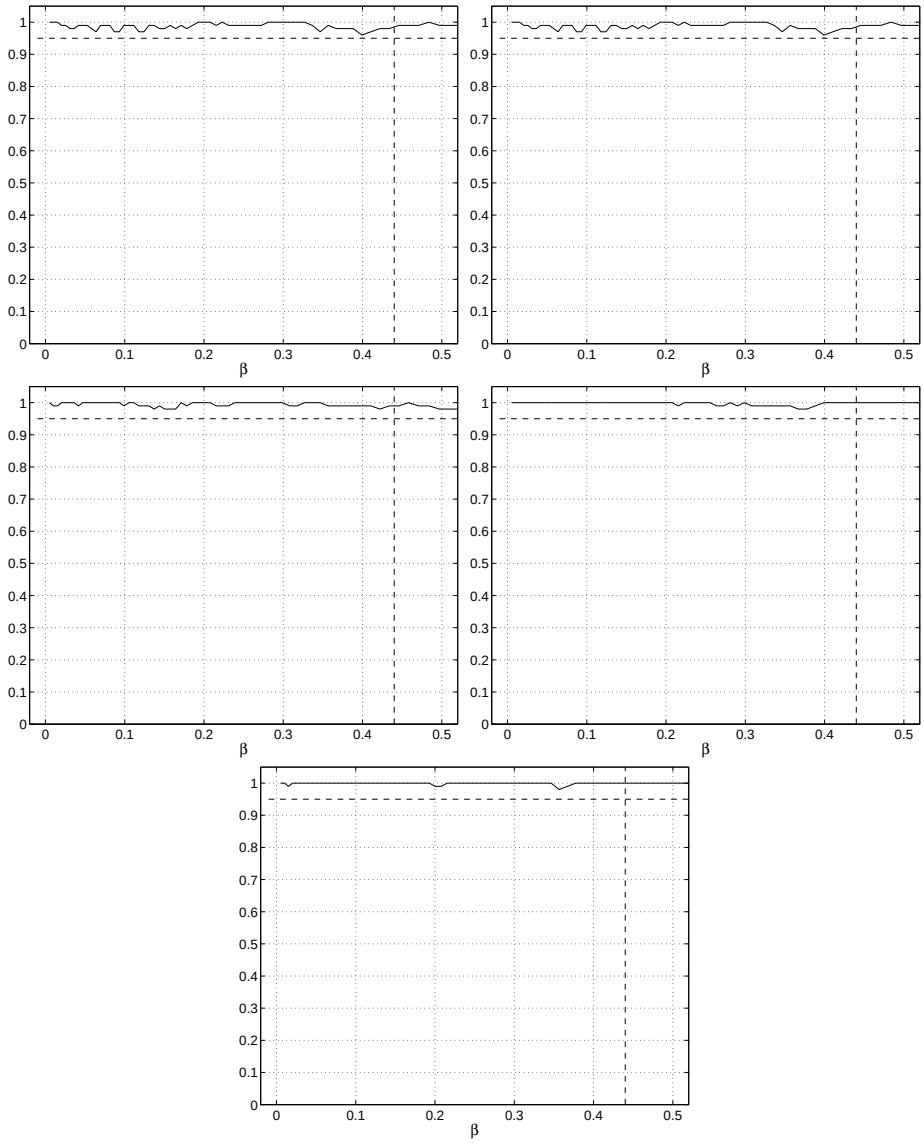

Figure 12.1: *Estimated confidence level on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). The desired confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.10$. For corresponding results with $p_s = 0.20$ see Figure B.5 (page 214).*

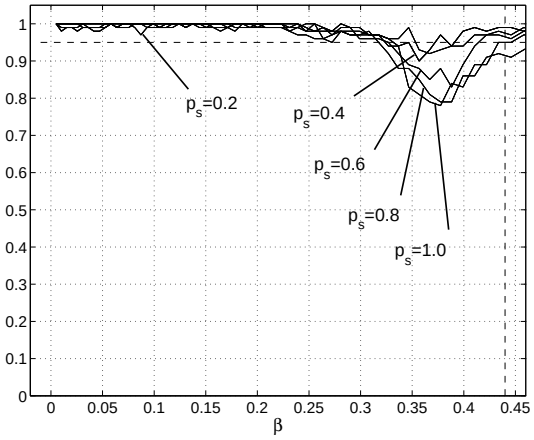

Figure 12.2: *Estimated confidence levels $\hat{\alpha}$ for confidence intervals based on data within box $\mathbb{B}_{20}$, for different choices of selections probabilities, $p_s = 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.0$.*

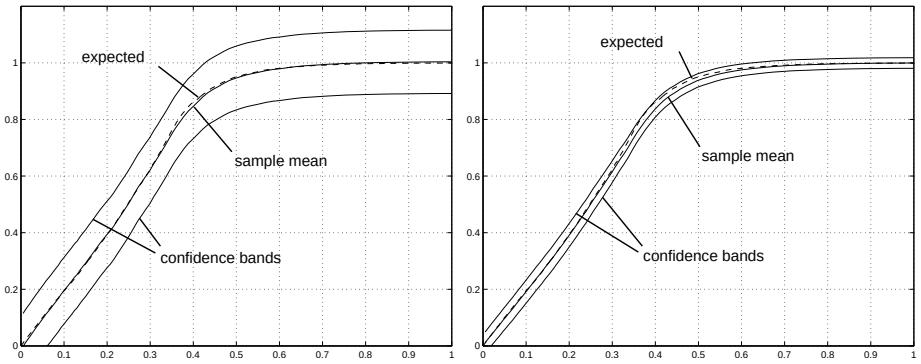

Figure 12.3: *In the left diagram we see the mean value of the statistic S (solid line) and confidence bands for the true value using the central limit theorem (Theorem 9.6) and confidence level $\alpha = 0.95$ for $p_s = 0.1$ on $\mathbb{B}_{20}$. The right diagram shows the corresponding information for $p_s = 0.8$, also on $\mathbb{B}_{20}$.*

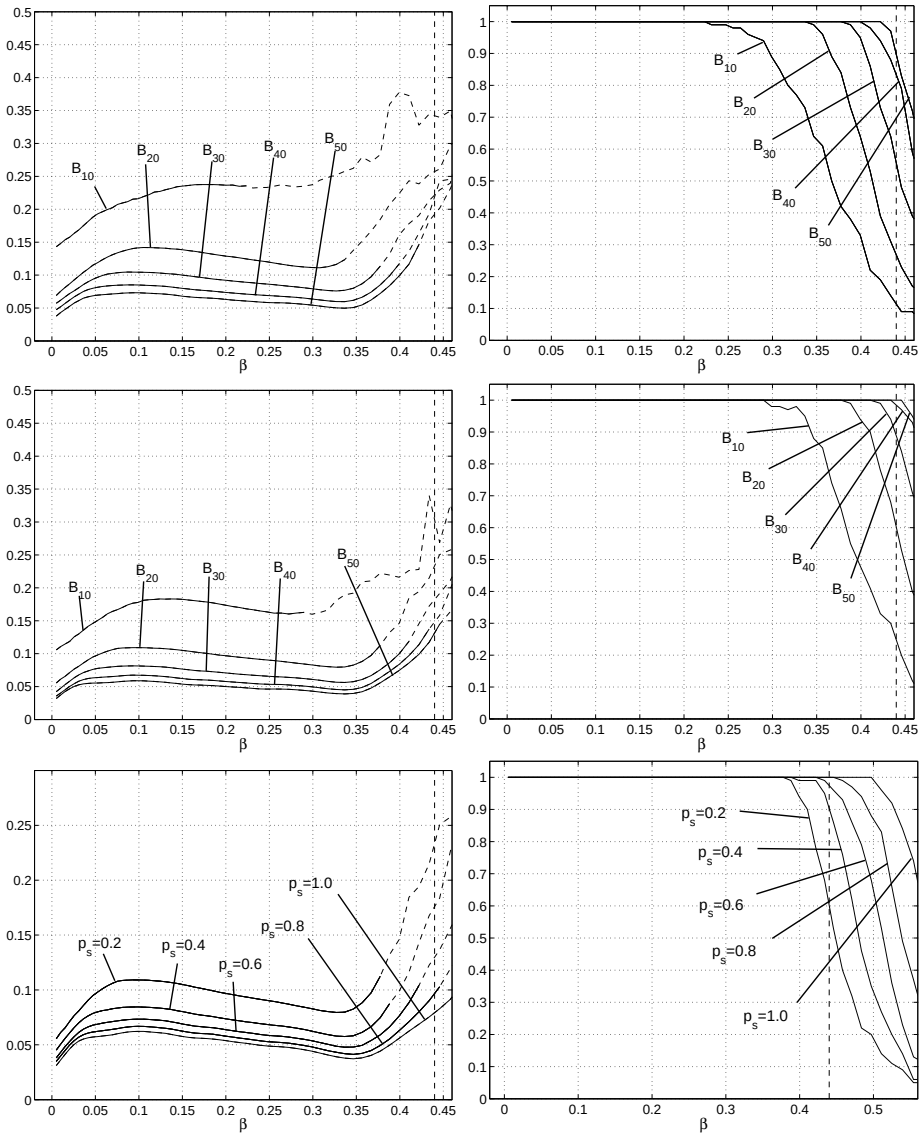

Figure 12.4: *The width of level* 0.95 *confidence intervals for different boxes, using* $p_s = 0.1$ *(upper left),* $p_s = 0.2$ *(middle left) on boxes* $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ *and* $\mathbb{B}_{50}$. *Below (left) we see the corresponding curves for box* $\mathbb{B}_{20}$ *for* $p_s = 0.2$, 0.4, 0.6, 0.8 *and* $p_s = 1.0$. *Solid lines represent* $\mathbb{E}[W_{n,\alpha,p_s}]$ *and dashed lines* $\mathbb{E}[W_{n,\alpha,p_s} | W_{n,\alpha,p_s} < \infty]$. *The diagrams in the right column show* $\mathbb{P}(W_{n,\alpha,p_s} < \infty)$ *for the corresponding data sets.*

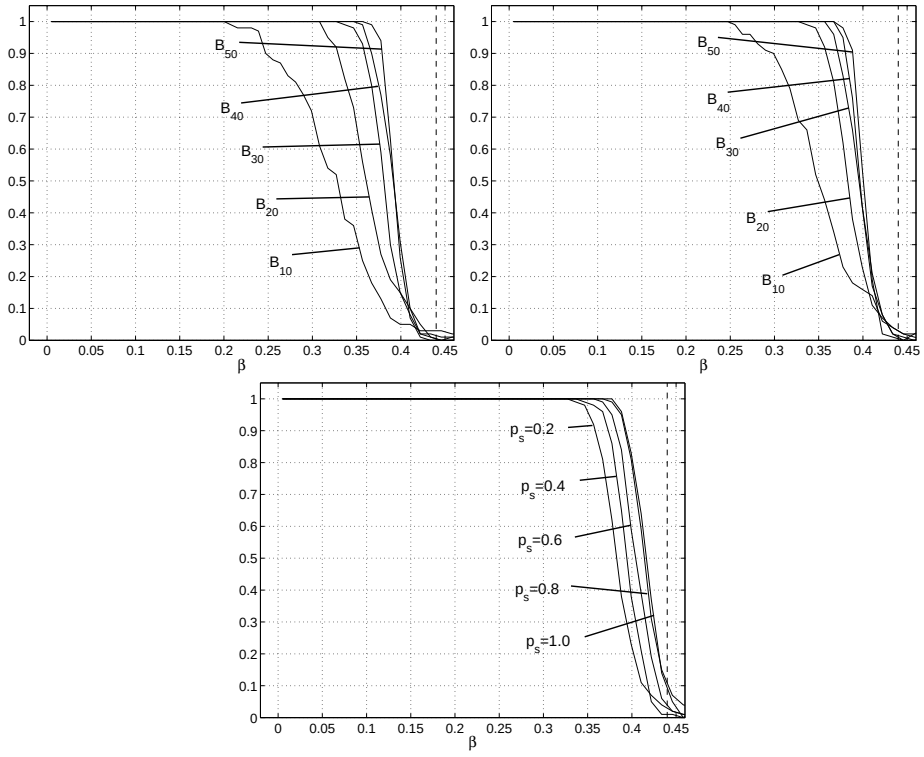

Figure 12.5: *Estimation of $\mathbb{P}(V)$, the probability of getting a valid confidence interval, as functions of β for boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$ using selection probabilities $p_s = 0.1$ (upper left) and $p_s = 0.2$ (upper right) and confidence level $\alpha = 0.95$. Below we see the corresponding probability for $p_s = 0.2, 0.4, 0.6, 0.8, 1.0$.*

Reliability

We estimate $P(V)$ from data on boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ and $\mathbb{B}_{50}$, for results see Figure 12.5. By using large enough boxes (for $p_s = 0.2$) we can almost remove the probability for faulty confidence intervals for $\beta < 0.35$, this approach requires on average 2040 data points (see Table 11.1). The shape of the curves in Figure 12.5 suggests that the closer the actual value of β is to β_c the amount of extra data required to be almost sure to get a reliable confidence interval, is substantial.

12.4 Box interactions

We now proceed with box interactions for the Potts model. We follow along the same lines as in section 11.5 for the Ising model. We define the statistic, $S_n(Y)$, average connection probability and establish unbiasedness and consistency for $S_n(Y)$.

Definition 12.2 ***The box statistic***
Let $\mathbb{B}_n$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Potts box interaction distribution on $\mathbb{B}_n$ for some $q \geq 2$. We define the statistic $S_n(Y)$ as

$$S_n(Y) = \left(\frac{q}{q-1}\right)\left(\frac{1}{2\ p_s\ |\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} Y(v) + \frac{1}{2} - \frac{1}{q}\right).$$

Now to the average connection probability, the box interaction version of pairwise connection probabilities in Section 12.2.

Definition 12.3 ***Average connection probability***
Let $\mathbb{B}$ be a box in $\mathbb{Z}^2$ and suppose we have data Y distributed according to the marked Potts box interaction distribution on $\mathbb{B}_n$ for some $q \geq 2$. We define the average connection probability statistic $f^{a,b}_{avg}(p)$ as

$$f^{a,b}_{avg}(p) = \frac{1}{|E(\mathbf{0})|}\sum_{\langle l,l'\rangle\in E(\mathbf{0})} \phi^{(n)}_{p,q}(l \leftrightarrow l')$$

We also state the box interaction version of Theorem 12.1.

Theorem 12.3
Let Y be distributed according to the marked Potts box interaction model, and let $\phi_{p,q}$ be the corresponding random cluster measure. Let $S_n(Y)$ be defined as in Definition 12.2. Then $S_n(Y)$ is unbiased as an estimator of $f_{avg}^{a,b}(\mathbf{0})$, the average connection probability. If $\beta < \beta_c^I$ then $S_n(Y)$ is consistent and we have

$$\frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)}\ (\ S_n(Y) - f_{avg}^{a,b}(p)\ ) \xrightarrow{\mathcal{D}} Z$$

for some $p \in [0,1]$, where $Z \stackrel{\mathcal{D}}{=} N(0,1)$.

Proof:
Let $A = q(q-1)^{-1}$ and $S_n(Y)$ be as defined above. Then $S_n(Y)$ is indeed unbiased.

$$\begin{aligned}
\mathbb{E}[S_n(Y)] &= \mathbb{E}\left[A\ \left(\frac{1}{2\ p_s\ |\mathbb{B}_n|}\ \sum_{v\in\mathbb{B}_n} Y(v) + \frac{1}{2} - \frac{1}{q}\right)\right] \\
&= A\ \left(\frac{1}{2\ p_s}\ \mathbb{E}[Y(\mathbf{0})] + \frac{1}{2} - \frac{1}{q}\right) \\
&= A\ \left(\frac{1}{2\ p_s}\ p_s\left(\frac{2}{q} - 1 + \frac{2}{A}\ \frac{1}{|E(v)|}\ \sum_{e_l\in E(\mathbf{0})} \phi_{p,q}(l \leftrightarrow l')\right) + \frac{1}{2} - \frac{1}{q}\right) \\
&= A\ \left(\frac{1}{2\ p_s}\ p_s\left(\frac{2}{q} - 1 + \frac{2}{A}\ f_{\text{avg}}^{a,b}(p)\right) + \frac{1}{2} - \frac{1}{q}\right) \\
&= f_{\text{avg}}^{a,b}(p)
\end{aligned}$$

From Bolthausen's central limit theorem we have

$$\sum_{v\in\mathbb{B}_n} \frac{(Y(v) - \mathbb{E}[Y(v)])}{\sigma_Y|\mathbb{B}_n|^{1/2}} \xrightarrow{\mathcal{D}} Z_0$$

where Z_0 follows the standard normal distribution. Then

$$\begin{aligned}
&\sum_{v\in\mathbb{B}_n} \frac{(Y(v)-\mathbb{E}[Y(v)])}{\sigma_Y|\mathbb{B}_n|^{1/2}} \\
&= \frac{1}{\sigma_Y|\mathbb{B}_n|^{1/2}}\left(\sum_{v\in\mathbb{B}_n} Y(v) - |\mathbb{B}_n|p_s\left(\frac{2}{q}-1+2\ \frac{1}{A}\ f^{a,b}_{\text{avg}}(p)\right)\right) \\
&= \frac{p_s|\mathbb{B}_n|^{1/2}}{\sigma_Y}\left(\frac{1}{p_s|\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} Y(v) - \frac{2}{q}+1-2\ \frac{1}{a}\ f^{a,b}_{\text{avg}}(p)\right) \\
&= \frac{1}{A}\ \frac{2p_s|\mathbb{B}_n|^{1/2}}{\sigma_Y}\left(\left(A\ \frac{1}{2p_s|\mathbb{B}_n|}\sum_{v\in\mathbb{B}_n} Y(v) - \frac{1}{q}+\frac{1}{2}\right) - f^{a,b}_{\text{avg}}(p)\right) \\
&= A\ \frac{2p_s|\mathbb{B}_n|^{1/2}}{\sigma_Y}\left(S_n(Y) - f^{a,b}_{\text{avg}}(p)\right)
\end{aligned}$$

and

$$\frac{q-1}{q}\ \frac{2p_s|\mathbb{B}_n|^{1/2}}{\sigma_Y}\ (\ S_n(Y) - f^{a,b}_{\text{avg}}(\mathbf{0})\ ) \xrightarrow{\mathcal{D}} Z$$

follows where $Z \overset{\mathcal{D}}{=} N(0,1)$. Now, as a consequence of the above and unbiasedness

$$|S_n(Y) - f^{a,b}_{\text{avg}}(\mathbf{0})| \xrightarrow{P} 0$$

follows and also consistency, and we are done.
□

We use the same estimated average connection probabilities $\hat{f}^{a,b}_{\text{avg}}$ as for the Ising model.

The point and interval estimators

Given the statistic $S_n(Y)$ our estimate of β is

$$\hat{\beta} = -\ \frac{1}{2}\ \log(\ 1-(\hat{f}^{a,b}_{\text{avg}})^{-1}(\ S_n(Y)\ )\ )$$

The confidence interval is given by the following result.

Theorem 12.4 ***Confidence interval for the inverse temperature*** *Let Y be distributed according to the marked Potts box interaction model on a finite box $\mathbb{B}_n$ at inverse temperature β and some $q \geq 2$. Let β_c^I be the critical temperature for the Ising model. Let $S_n(Y)$ and $f_{avg}^{a,b}$ be defined as above. Then $[\beta_{low}, \beta_{up}]$ is a level α confidence interval for the inverse temperature β. The bounds are given by*

$$\beta_{lower} = -\frac{1}{2} \log\left(1 - (f_{avg}^{a,b})^{-1}\left(S_n(Y) - z_{(\sqrt{\alpha}+1)/2} \frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)} + \frac{D_{\sqrt{\alpha}}}{\sqrt{m}} \right)\right)$$

and

$$\beta_{upper} = -\frac{1}{2} \log\left(1 - (f_{avg}^{a,b})^{-1}\left(S_n(Y) + z_{(\sqrt{\alpha}+1)/2} \frac{q\ \sigma_Y}{2p_s|\mathbb{B}_n|^{1/2}(q-1)} - \frac{D_{\sqrt{\alpha}}}{\sqrt{m}} \right)\right)$$

where z_α and D_α are quantiles from the standard normal distribution and the Kolmogorov-Smirnov distribution respectively. The confidence interval is valid if $\beta_c^I \notin [\beta_{low}, \beta_{up}]$.

As for a proof see the argumentation before Theorem 11.2, page 149.

12.5 Simulation study: box interactions

We simulate the statistic $S_n(Y)$ for boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$ and $\mathbb{B}_{40}$ using $a = b = 2$ and selection probabilities, $p_s = 0.1, 0.2$, and also for $p_s = 0.4$, 0.6 0.8 and $p_s = 1.0$ on $\mathbb{B}_{20}$. We generate 100 omniparametric samples Y_k for each box size, calculate statistics

$$S_n(Y_1), ..., S_n(Y_{100})$$

and then proceed as before with validation, precision and reliability, see Section 11.4 for a detailed description of the procedure.

Validation

We see in Figure 12.6 and B.6 the results of the validation procedure for $p_s = 0.1$ and $p_s = 0.2$ on boxes $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$ and $\mathbb{B}_{40}$. In all cases the confidence intervals are good, there are no deviations from the expected behaviour. In Figure 12.7 we see the corresponding results for $p_s = 0.2$, 0.4, 0.6, 0.8 and $p_s = 1.0$ on $\mathbb{B}_{20}$, and as in the previous simulation studies we see a drop in estimated confidence level when $\beta \geq 0.25$, with its lowest level at $\beta \approx 0.37$. For β around the critical value we have acceptable

estimated confidence levels again, but what else is there to expect when $\mathbb{E}[W_{n,\alpha,p_s}]$ increase dramatically for $\beta \geq 0.4$ (Figure 12.9).

In Figure 12.8 (right) we see the reason for the deviation. The curve for $\mathbb{E}[S_n(Y)]$ barely falls within the confidence region around the sample mean. For detailed discussion of the subject see under "Validation" in Section 11.4 on page 151.

Precision

We see the results of our simulations in Figure 12.9. These are rather similar to the corresponding results in previous section (Figure 12.4). We extended to box interaction to get more out of our data, and we do get smaller confidence intervals than we did using only pairwise interactions. Our best estimate still comes from simulation on box $\mathbb{B}_{50}$ using $p_s = 0.2$ resulting in estimated expected width less than 0.05 for $\beta \leq 0.4$. For fixed β we also have a reduced risk of getting one-sided intervals.

Reliability

We estimate the probability of having a confidence interval covering the critical value β_c^I, see Figure 12.10 for results. By using enough data, at least $p_s = 0.2$ on $\mathbb{B}_{20}$, we have a very high probability of not getting faulty confidence intervals for $\beta < 0.35$, this approach requires on average 336 data points (see Table 11.1, page 156).

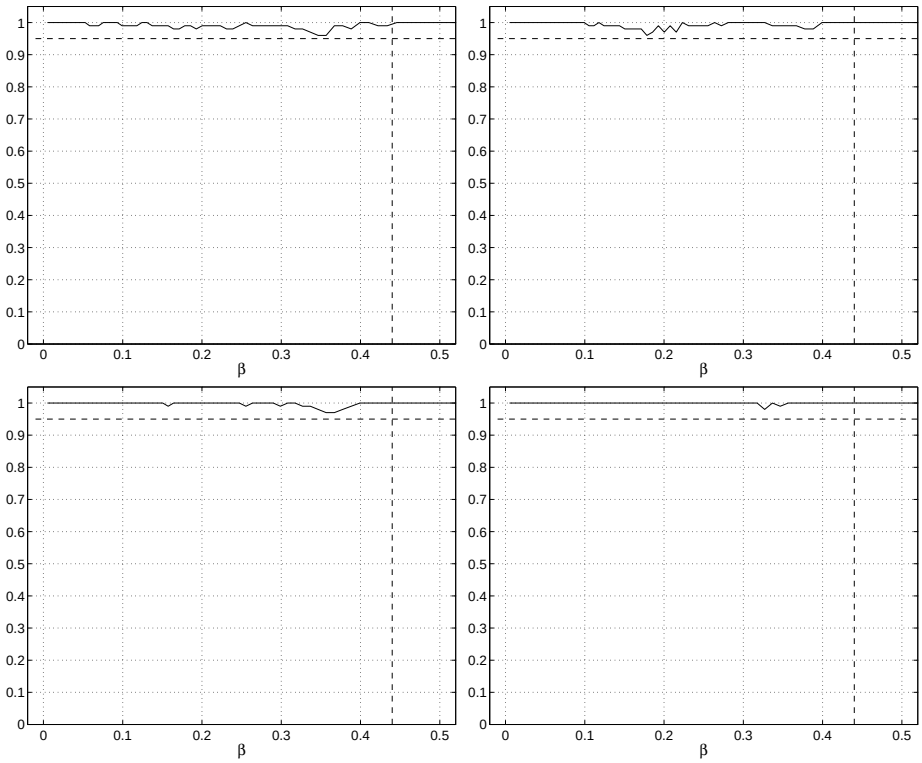

Figure 12.6: *Estimated confidence level on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (lower left) and $\mathbb{B}_{40}$ (lower right). The desired confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.10$. For corresponding results with $p_s = 0.20$ see Figure B.6 (page 215).*

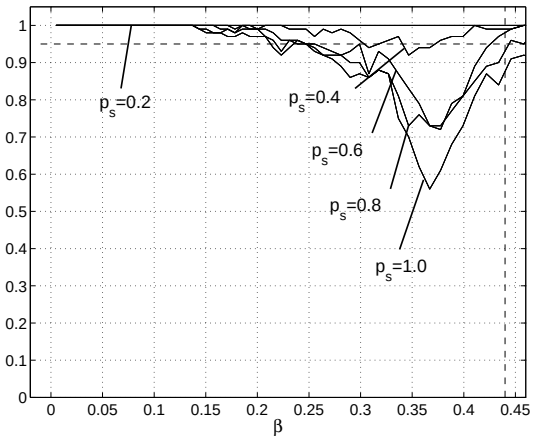

Figure 12.7: *Estimated confidence levels $\hat{\alpha}$ for confidence intervals based on data within box $\mathbb{B}_{20}$, for different choices of selections probabilities, $p_s = 0.2, 0.4, 0.6, 0.8$ and $p_s = 1.0$.*

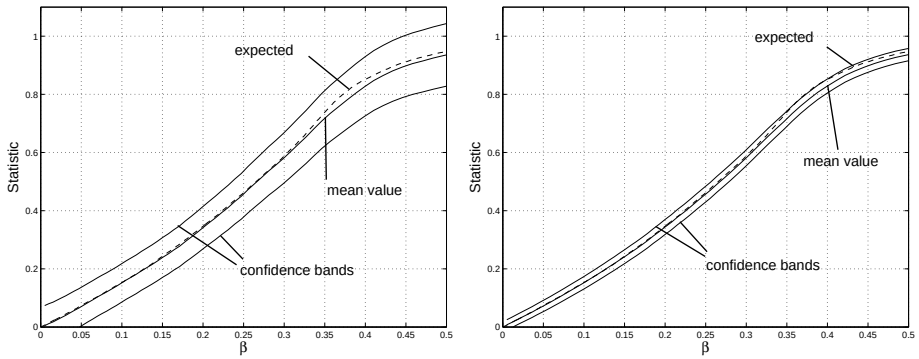

Figure 12.8: *In the left diagram we see the mean value of the statistic S (solid line) and confidence bands for the true value using the central limit theorem (Theorem 9.6) and confidence level $\alpha = 0.95$ for $p_s = 0.1$ on $\mathbb{B}_{20}$. The left diagram shows the corresponding information for $p_s = 0.8$, also on $\mathbb{B}_{20}$.*

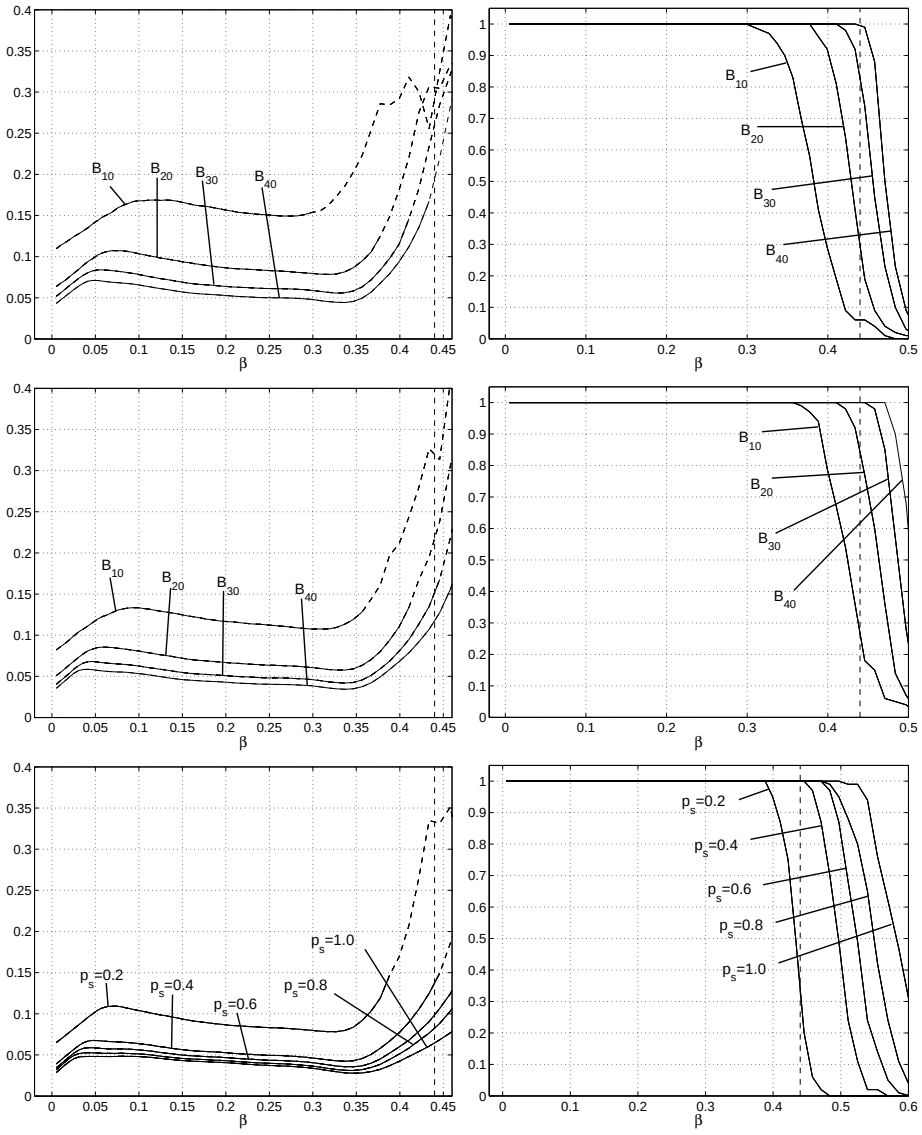

Figure 12.9: *The width of level* 0.95 *confidence intervals for different boxes, using* $p_s = 0.1$ *(upper left) and* $p_s = 0.2$ *(middle left) for boxes* $\mathbb{B}_{10}$ *and* $\mathbb{B}_{20}$*. In the lower left diagram we see the corresponding information for* $p_s = 0.2, 0.4, 0.6, 0.8$ *and* 1.0 *on box* $\mathbb{B}_{50}$*. Solid lines represent* $\mathbb{E}[W_{n,\alpha,p_s}]$ *and dashed lines* $\mathbb{E}[W_{n,\alpha,p_s}|W_{n,\alpha,p_s} < \infty]$*. The diagrams in the right column show* $\mathbb{P}(W_{n,\alpha,p_s} < \infty)$ *for the corresponding data sets.*

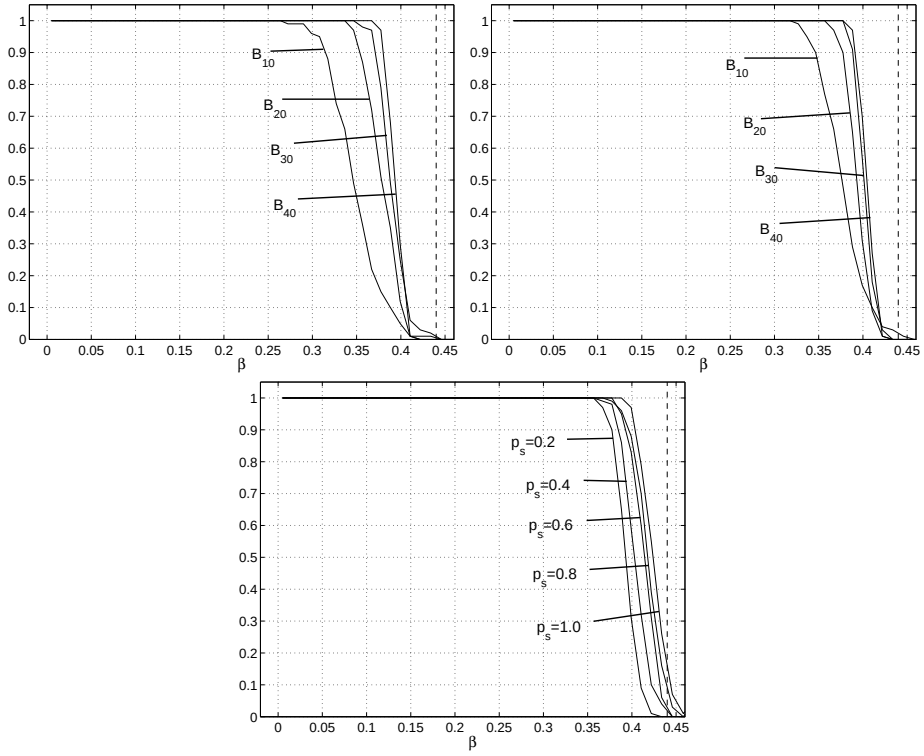

Figure 12.10: *Estimations of* $\mathbb{P}(V)$ *as functions of* β *for boxes* $\mathbb{B}_{10}$, $\mathbb{B}_{20}$ *using* $p_s = 0.1$ *(left) and* $p_s = 0.2$ *(right), both for* $\alpha = 0.95$.

12.6 A non-asymptotic method

We now adapt the material presented in Section 11.7 (page 165) to the Potts model. We make the necessary adjustments for the Potts model. For motivations and proofs see Section 11.7.

The point and interval estimators

Given data X and the calculated statistic $T(X)$ we let

$$\hat{\beta} = -\frac{1}{2}\,\log(\,1 - f_{\text{avg}}^{-1}(\,T(X)\,)\,)$$

be out point estimate of β. A level α conservative confidence interval for β is $[\beta_{\text{low}}, \beta_{\text{up}}]$ as described below.

Theorem 12.5 ***Confidence interval for the inverse temperature*** *Let Y be distributed according to the Potts model on a finite box $\mathbb{B}_n$ at inverse temperature β and let β_c^I be the critical temperature for the Ising model. Let T be any statistic with finite first and second moments. Generate omniparametric Potts samples $X_1, ..., X_n$ and compute quantile functions $f_{1-\alpha,low}$ and $f_{1-\alpha,up}$ as described in Section 11.7.2. Then $[\beta_{low}, \beta_{up}]$ is a level α confidence interval for the inverse temperature β, with bounds as follows.*

$$\beta_{low} = -\frac{1}{2}\,\log(1 - p_{low}), \quad p_{low} = \min\{p : f_{\alpha,up}(p) = T(X)\}$$

$$\beta_{up} = -\frac{1}{2}\,\log(1 - p_{up}), \quad p_{up} = \max\{p : f_{\alpha,low}(p) = T(X)\}$$

12.7 Simulation study: the non-asymptotic method

We simulate 1000 data sets for each combination of selection probability and box size and use these for generating percentile functions. For the non-asymptotic method the percentile function replace the confidence bands we computed using central limit theorems. This give us a cruder (read non-smooth) estimate of the confidence region for the statistic which in the end leads to more variations in the estimated confidence levels (see "Validation" below).

Validation

We generate for each combination of selection probability and box size 100 data sets. The validation procedure is carried out in the same manner as for the Ising model in Section 11.8, for results see figures 12.11,12.11 and B.8 (pages 193, 216, 217). In all cases the desired confidence levels are met. There are more variations than for the asymptotic methods but nothing that casts any doubt over the method. There are no phenomenons resulting in low (estimated) confidence levels encountered for pairwise and box interactions in previous sections.

Precision

We are now able to get confidence intervals with width between 0.05 and 0.1 without using too much data. For example combination $(p_s, \mathbb{B}_n) = (0.2, \mathbb{B}_{10})$ gives us a width of ≈ 0.1 for a subcritical process. The most narrow interval comes from the combination $(p_s, \mathbb{B}_n) = (0.2, \mathbb{B}_{50})$ which gives us $W \leq 0.05$ for a subcritical process. In Figure 12.13 we see $\mathbb{E}[W_{n,\alpha,p_s}]$ in detail for simulations on box $\mathbb{B}_{20}$ for a subcritical process. The minimum width is well below 0.05.

In all cases the situation is worse for the supercritical process, we do not only get wider intervals but also suffer a risk of getting one-sided intervals. On the $\mathbb{B}_{10}$ using only $p_s = 0.1$ we may get one-sided intervals throughout the supercritical phase, while on $\mathbb{B}_{50}$ using $p_s = 0.2$ we get two-sided intervals with very high probability as long as $p \leq 0.8$. We see in Figure 12.12 (right column) that the risk of getting one-sided intervals increase rather quick with p once the risk is present. For large enough p we will get one-sided intervals no matter how much data we use.

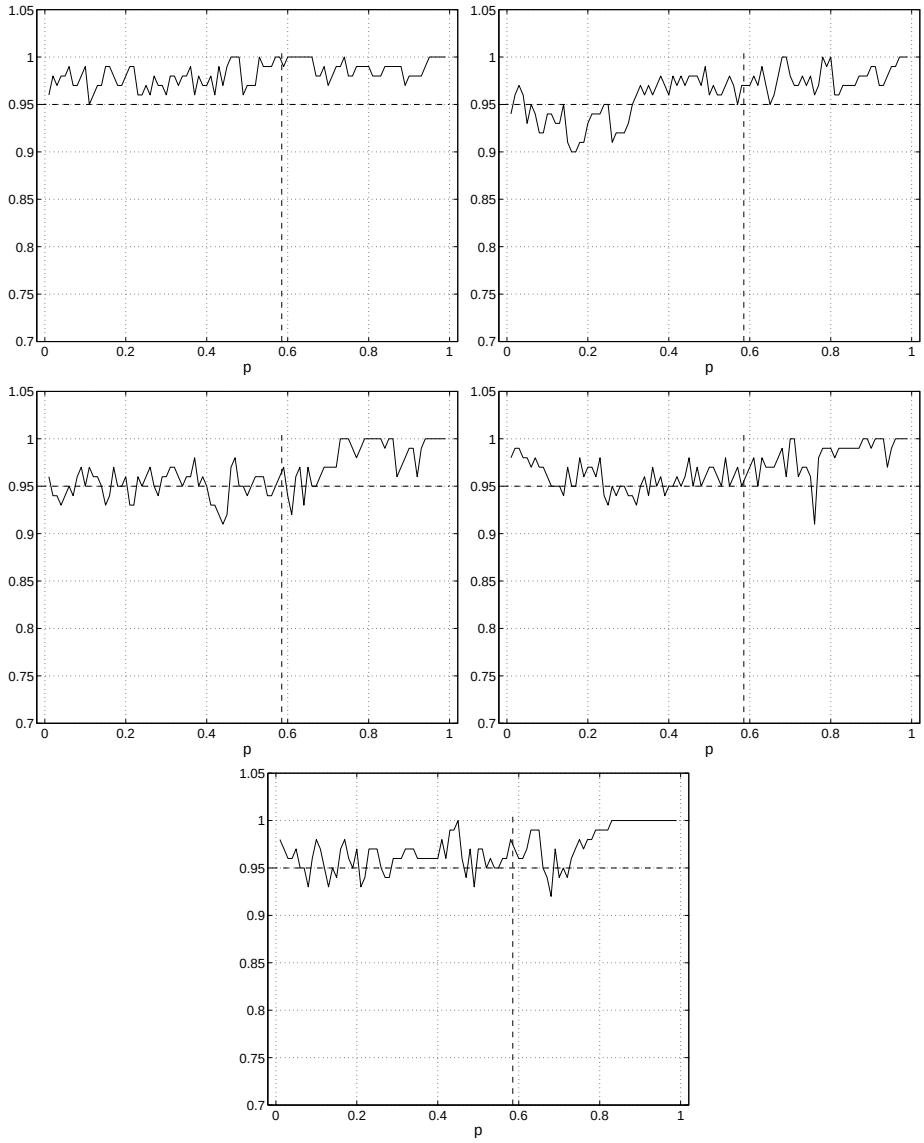

Figure 12.11: *Estimated confidence level on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower) with $q = 4$. The desired level is $\alpha = 0.95$. The results are given for $p_s = 0.10$. For the corresponding results using $p_s = 0.20$ see Figure B.7 , page 216.*

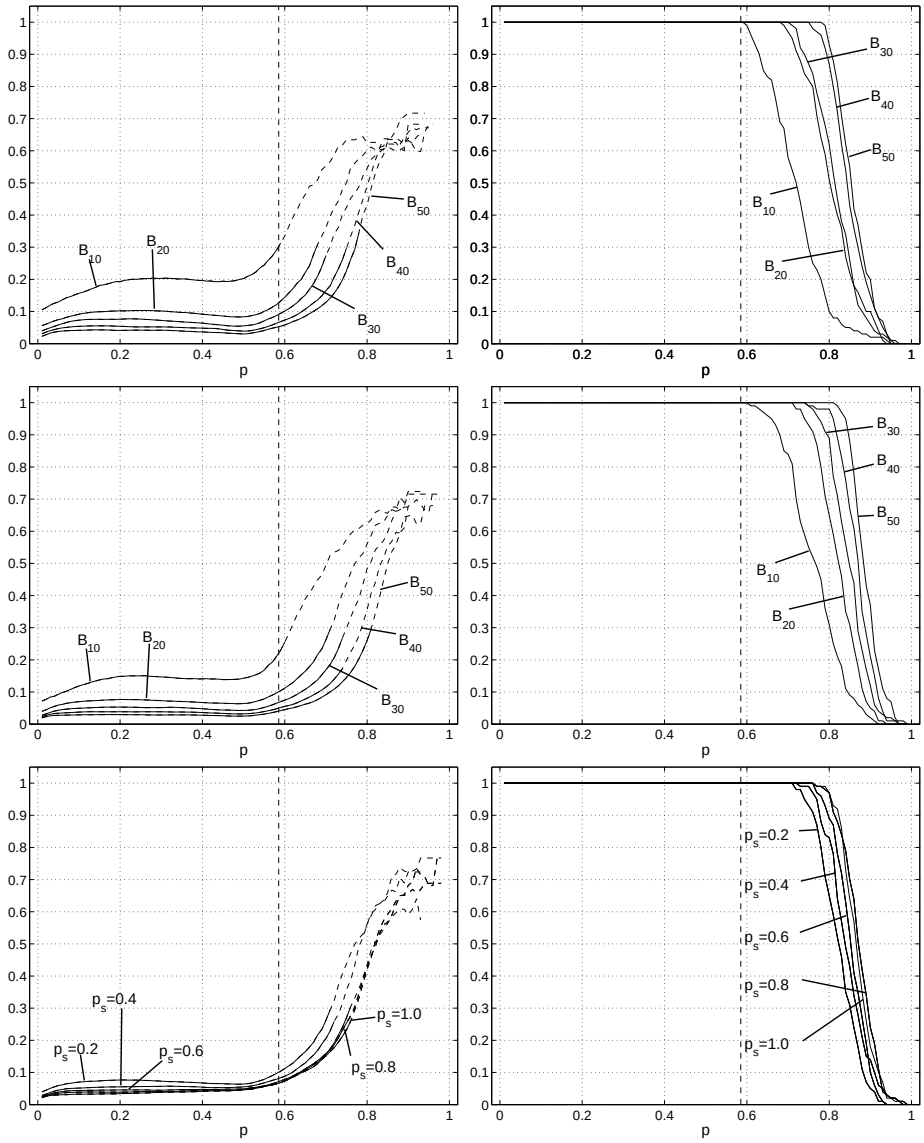

Figure 12.12: *The width of level* 0.95 *confidence intervals for different boxes, using* $q = 4$, $p_s = 0.1$ *(upper left) and* $p_s = 0.2$ *(middle left) for boxes* $\mathbb{B}_{10}$, $\mathbb{B}_{20}$, $\mathbb{B}_{30}$, $\mathbb{B}_{40}$ *and* $\mathbb{B}_{50}$, *and for* $p_s = 0.2$, 0.4, 0.6, 0.8 *and* $p_s = 1.0$ *on* $\mathbb{B}_{20}$ *(lower left). Solid lines represent* $\mathbb{E}[W_{n,\alpha,p_s}]$, *and dashed lines* $\mathbb{E}[W_{n,\alpha,p_s}|W_{n,\alpha,p_s} < \infty]$. *The right column shows* $\mathbb{P}(V)$ *in the corresponding situations. The desired confidence level is* $\alpha = 0.95$.

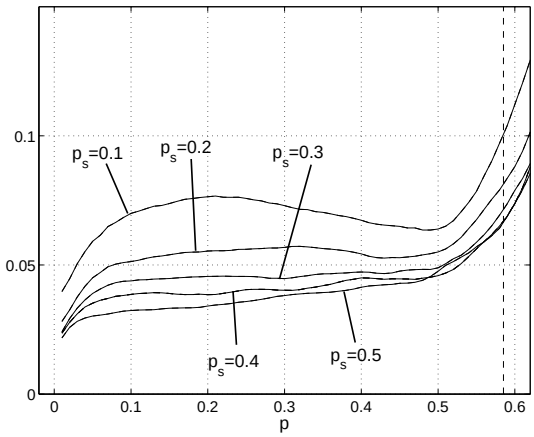

Figure 12.13: *A detailed diagram of the estimated width on box $\mathbb{B}_{20}$, the full diagram is shown in the lower left of Figure 12.12.*

CHAPTER 13

Parameter estimation: concluding remarks

We have presented three methods for parameter estimation for the Ising and Potts models, two asymptotic and one non-asymptotic. The theoretical background differs, especially if we compare the asymptotic methods and the non-asymptotic method. If choosing an asymptotic method there are a few choices we have to make before applying the procedure, and we need to estimate the method characteristics such as susceptibility and connection probability. For the non-asymptotic method we estimate percentile functions instead of susceptibility. There are also issues concerning which model the data set is taken from, it might seem simple but there may be problems.

13.1 Comparison of methods

We have made a few design choices during the development of the procedures. In this section we compare and reflect on them, and also linger on the ideal versus real situation, discussing how constraints in computer resources, and above all time may affect the final result. Precision is vital in parameter estimation and we discuss the most important task in applying the methods in a real world situation, capture the variance of the statistic. Finally we discuss which method is best.

13.1.1 Theoretical background and design choices

The asymptotic models both depend on how closely the distribution of a certain statistic approximates a normal distribution. The basic assumption is that we estimate the inverse temperature β for the Ising or Potts model on the whole two-dimensional lattice. Since we can not do simulation on the entire $\mathbb{Z}^2$ lattice we perform them on boxes $\mathbb{B}_n$ and make n as large as we can manage, especially when estimating the two required theoretical quantities, or model characteristics.

- The connection probabilities, $\phi_{p,q}(\mathbf{0} \leftrightarrow \mathbf{1})$ or $f^{a,b}_{\text{avg}}$.
- The susceptibility, that is, the sum of covariances, $\text{Cov}[Y(\mathbf{0}), Y(v)]$, for all v in $\mathbb{Z}^2$.

Depending on computer resources and time these can both be estimated to arbitrary precision. In the thesis we have used the Kolmogorov-Smirnov distribution to capture the remaining variance of the connection probability. For the susceptibility however, there is no theorem for the remaining variance of the estimate, so we consider it known at its value.

In the non-asymptotic approach we view the models on a finite box $\mathbb{B}_n$, as if the rest of the infinite lattice does not exist. This fits well into the ordinary boundary conditioned orientated approach with the free boundary condition. We simulate data and get a sample of the statistic and generate mean value and percentile functions. Here we capture the variance of a statistic by pure simulation instead of using a central limit theorem accompanied by simulations.

13.1.2 Ideal versus real situation

In a perfect world we estimate the true connection probability on $\mathbb{Z}^2$ with the technique described in Chapter 2. Since the susceptibility in all four asymptotic models can be rewritten as a sum of random cluster connection probabilities we use the same technique again. In the non-asymptotic method we estimate the connection probability on $\mathbb{B}_n$ to arbitrary precision using as many simulations as we need, the percentile functions are handled in the same manner. Once the connection probabilities and the susceptibility (on $\mathbb{Z}^2$) are estimated to arbitrary precision and data collected from within some finite region the open question is how well that data describes the behaviour on the whole lattice.

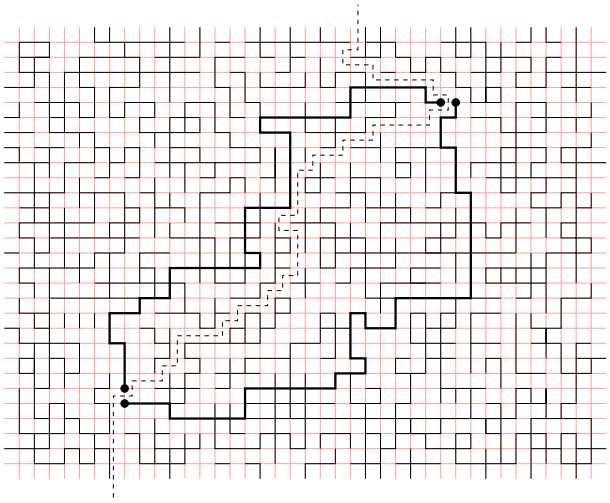

Figure 13.1: *An illustration of a rare event where two pairs of neighbouring vertices are unconnected but two other pairs are connected over longer paths of open vertices. The two clusters are separated by the dashed line.*

In the real world however we have limited time and computer resources, so we have to decide which compromises to apply. In the thesis we have (for the asymptotic models) based connection probabilities on simulations on as large boxes as possible (in our case B_{50}) in order to generate 10000 simulations. The susceptibility estimates are based on 100000 simulations on various box sizes (see chapters 11 and 12). We have also in chapters 11 and 12, due to limited computer resources, been forced to estimate the connection probabilities on smaller boxes than desired. As a consequence we experience deviations between the statistic's sample mean and its expected value, as see in figures 11.6, 11.12, 12.3 and 12.8. A simple remedy for this problem is the use of larger boxes when generating model characteristics, as seen in Figure 11.7 (page 153).

In applying the non-asymptotic methods we have used the same estimate of the connection probability as in the pairwise interaction models to achieve a better estimate. A more theoretically motivated approach would be to estimate a new connection probability for each simulated box size, as we have done with the percentile functions (generated using

1000 simulations for each box size), and use that instead.

With all the results at hand other choices would perhaps be preferable. A better approach might be to apply also the asymptotic methods on the graph $\mathbb{B}_n$ and generate all characteristics using the same n. This would get rid of the deviations between the statistics sample mean and its expected value for some β. The question regarding how close the distribution of the statistic approximates a normal distribution remains open.

13.1.3 The role of susceptibility

We have used two different approaches to estimate the susceptibility in the marked interaction models, one based on random cluster connection probabilities and one based on omniparametric Potts samples.

For the marked Ising interaction model we have shown that the covariance can be written as a sum of connection probabilities in the random cluster model, and estimated these. Some of these connection probabilities are however difficult to estimate since they describe rare events, and non-monotone probabilities, see Figure 13.1. We have used as many as 100000 simulations and still the estimated susceptibility have a ragged curve (Figure 10.5, page 136). If we instead base our estimate on omniparametric Ising samples, as in the marked Ising box interaction, marked Potts interaction model and marked Potts box interaction model we get a much smoother curve, see figures 10.6 and 10.7 (pages 137 and 138) for a comparison. There is also reasons to believe that the extra randomisation used when generating omniparametric Ising or Potts samples has a smoothing effect. We have found no reason to believe that the true susceptibility is described by something else than a rather smooth curve, at least throughout the subcritical region.

13.1.4 Which method is best?

We would recommend the non-asymptotic method, for several reasons. Considering precision the non-asymptotic method is not dramatically better than the two others. It is more precise than method one (pairwise interactions) but approximately equal to method two (box interactions). Method two and three are not directly comparable since they require different kind of data sets, while method one and three can be directly applied to the same type of data.

Another advantage with method three is that it gives us valid or reli-

able confidence intervals for all β while any of the asymptotic methods only work in parts of the subcritical regime (i.e. $\beta < \beta_C^I$, see Section 12.1). Without any prior information on the value of β this is a major drawback for the asymptotic methods.

However, we feel that the strongest argument for method three is speed. All three methods require a lot of computer resources so speed is relative. But instead of spending cpu time on estimating the susceptibility for the asymptotic methods we can generate functions for the statistic's expected value and percentiles in the non-asymptotic approach using less resources and still get good results. Also method three allows us to use virtually any statistic, as long as it give us reasonable function estimates (expected value, percentiles). From a theoretical point of view we need finite first and second moments, nothing else.

13.2 Theoretical parameter estimation issues

Parameter estimation for Gibbs fields such as the Ising and Potts models can be a difficult task. Given a data set from the Ising or Potts model we may produce an estimate of the inverse temperature. By assuming q known and zero external field the only parameter left is β. So if we put strong confidence in our estimate $\hat{\beta}$ one would like to think that the model assumption and $\hat{\beta}$ together give us a close description of the real situation. However both the Ising and the Potts model experience a phase transition, that is, for β above some threshold β_c we do not have unique probabilistic behaviour. Even though the real world is finite and no sharp phase transition ever occur we may encounter smeared out effects of a phase transition if the subset of $\mathbb{Z}^2$ we use is large enough. So our data could be generated from several different measures with qualitatively different behaviour.

Let $\mathcal{P}(\beta)$ be the set of plausible measures for the model at hand at some inverse temperature β. Then $|\mathcal{P}(\beta)| = 1$ if $\beta \leq \beta_c$ and $|\mathcal{P}(\beta)| > 1$ if $\beta > \beta_c$ holds for both Ising and Potts models. In the Ising case there are two extremal measures (originating from different boundary conditions) in $\mathcal{P}(\beta)$ and every other element therein is a mix of these two. For the q-state Potts model there are q extremal measures, and by mixing them we may produce all other measures in $\mathcal{P}(\beta)$.

Let us return to $\hat{\beta}$ and the information it carries. If $\hat{\beta}$ indicates $\beta \leq \beta_c$ we are done since the model assumption and estimate give us a full description of the situation. If $\hat{\beta}$ indicates $\beta > \beta_c$ we need extra informa-

tion to determine which measure in $\mathcal{P}(\beta)$ is responsible for producing our data set. This measure estimation problem is rather complicated and separate from estimating β, we do treat it further. We mention it here since one otherwise may believe that the underlying measure is known, based on $\hat{\beta}$, in situations when clearly $\beta > \beta_c$. It also affects the generation of model characteristics (connection probabilities, susceptibility, percentile functions) since we always assume that they are generated using the correct model.

Part IV

Appendix

CHAPTER A

Notation

The lattice

$\mathbb{Z}^d$	The integer lattice in d dimensions.
$\mathbb{E}^d$	Edge set of the integer lattice in d dimensions.
$\mathbb{L}^d$	The graph having $\mathbb{Z}^d$ as vertex set, and edges between vertices with distance 1.
$\mathbb{B}_n^d$	The subgraph of $\mathbb{L}^d$, restricted to the box $[-n, n]^d$ ...
$\partial\mathbb{B}_n$	... and its boundary $\partial\mathbb{B}_n = \mathbb{B}_{n+1} \setminus \mathbb{B}_n$.
$\mathbf{0}$	Short for the origin $(0, ..., 0) \in \mathbb{Z}^d$.
$\mathbf{1}$	Short for $(1, 0) \in \mathbb{Z}^2$.

Graphs

$G = (V, E)$	A graph with vertex set V and edge set E.
$V(G)$	The vertex set of a graph G.
$E(G)$	The edge set of a graph G.

Vertices, edges and configurations

$u, v, w, x, y, ...$	Vertices
$\langle x, y\rangle, e, f, g, h, ...$	Edges
$\Gamma_{p,q}$	The random cluster model process with parameters p and q.
Σ_β	The Ising model at inverse temperature β
$\Omega_{\beta,q}$	The Potts model with parameters β and q
$\gamma, \sigma, \omega, ...$	Fixed parameter configurations
$\gamma_o, \sigma_o, \omega_o, ...$	Omniparametric configurations
$\gamma(e),\ \gamma^e,\ \gamma^{e,k}$	Given γ the configuration at e is denoted $\gamma(e)$, γ^e is the configuration on the graph $(V, E \setminus \{e\})$ and $\gamma^{e,k}$ is the configuration with $\gamma^{e,k}(e) = k$ and $\gamma^{e,k}(f) = \gamma(f)$ whenever $f \neq e$. The same notation applies for edges as well as vertices.
$G(\gamma)$	Given a random cluster sample γ we let $G(\gamma) = (V, \{e \in E : \lambda(e) = 1\})$.
$\kappa(\gamma)$	The number of clusters in the graph $G(\gamma)$.

The two-type Richardson model

$\Xi_v^\lambda(t)$	The state of vertex $v \in \mathbb{Z}^d$ at time t when the intensity of the type two infection is λ
$\Xi^\lambda(t)$	The configuration of $\mathbb{Z}^d$ at time t.
$\eta_i^\lambda(t)$	The subset of $\mathbb{Z}^d$ having type i at time t, $i = 1, 2$.
ξ_i^λ	The subset of $\mathbb{Z}^d$ having type i at $t = 0$, $i = 1, 2$.

The omniparametric two-type Richardson model

$\Theta(t)$	The configuration of $\mathbb{Z}^d$ at time t.
$\Theta_v(t)$	The state of vertex $v \in \mathbb{Z}^d$ at time t.

Special events

$G_{i,n}^\lambda$	Then event that infection type $i = 1, 2$ survives until reaching the boundary of a box B_n.
G_n^λ	Then event that both infection types survives until reaching the boundary of a box B_n.
G^λ	Then event that both infection types grows to infinity.

Measures

$\mathbb{P}^{\lambda_1,\lambda_2}_{\xi^1,\xi^2}$	The probability measure for the two-type Richardson model with starting configuration (ξ^1, ξ^2) having intensities λ_1 and λ_2 respectively for the two infection types.
$\mathbb{P}^{\lambda}_{\xi^1,\xi^2}$	Short for $\mathbb{P}^{\lambda_1,\lambda_2}_{\xi^1,\xi^2}$ with $\lambda_1 = 1$ and $\lambda_2 = \lambda$.
$\phi_{p,q}$	The fixed parameter random cluster measure with parameters p and q.
ϕ_q	The omniparametric random cluster measure with parameter q fixed.
$\mu_{\beta,q}$	The Potts measure with parameters β and q.
μ_β	The Ising measure with inverse temperature β.
$\rho_{p,q}$	The Edwards-Sokal measure.

Miscellanea

$C(u)$	The connected component containing the vertex u. We write $C(u) = C(v)$ for the event that u and v are in the same connected component.
$I_A(\lambda)$	Indicator variable for the event A in configuration λ.

Operators

$\mathbf{P}^{model}_{param}$	A projection mapping from omniparametric a space to the corresponding fixed parameter space

CHAPTER B

Additional simulation results

This appendix show some of the simulation results from chapter 11 and 12. We do not make any comments on them here, instead see the respective section given for each figure.

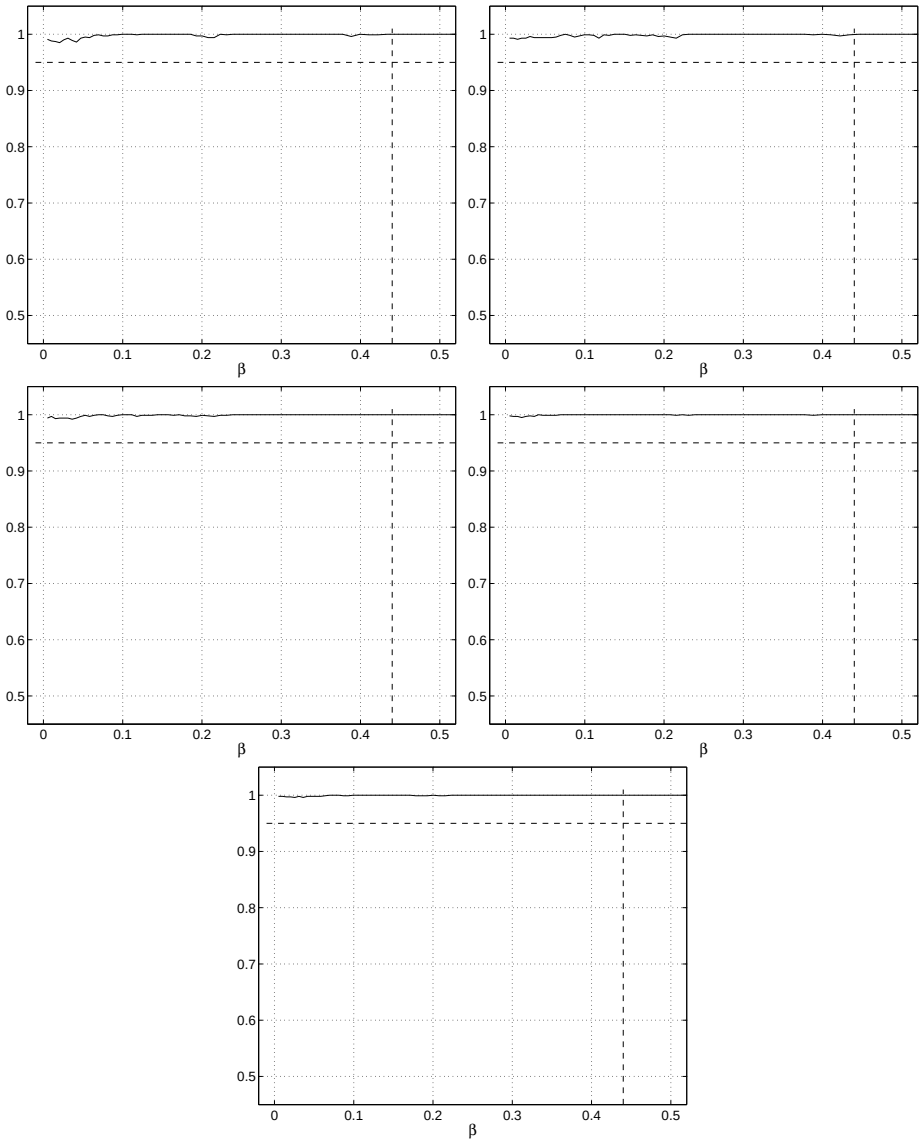

Figure B.1: *Pairwise interactions for the Ising model, Section 11.4, page 149. Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{150}$ (lower). The correct confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.20$.*

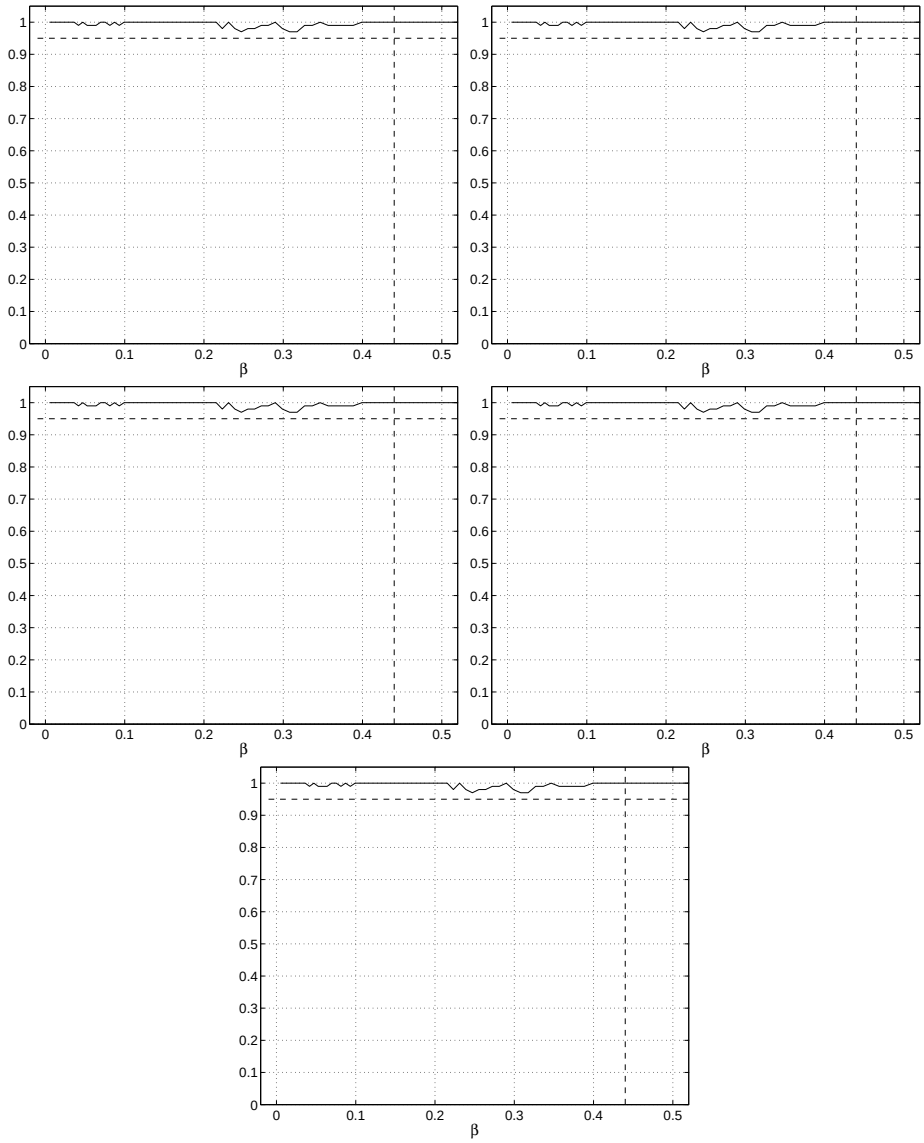

Figure B.2: *Box interactions for the ising model, Section 11.6, page 159. Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). The desired confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.20$. For corresponding results with $p_s = 0.10$ see Figure 11.11 (page 161).*

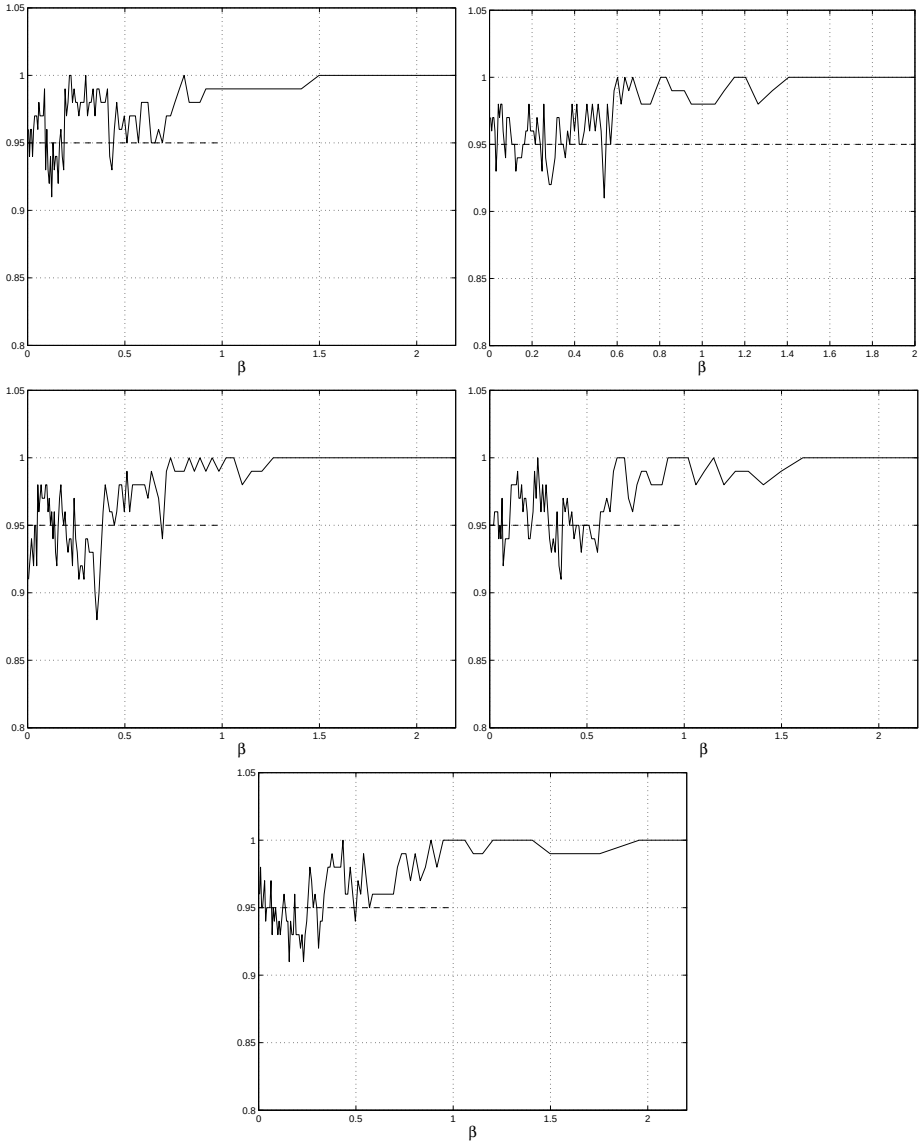

Figure B.3: *The nonasymptotic method for the Ising model, Section 11.8, page 168. Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right), and $\mathbb{B}_{50}$ (lower). The correct confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.20$.*

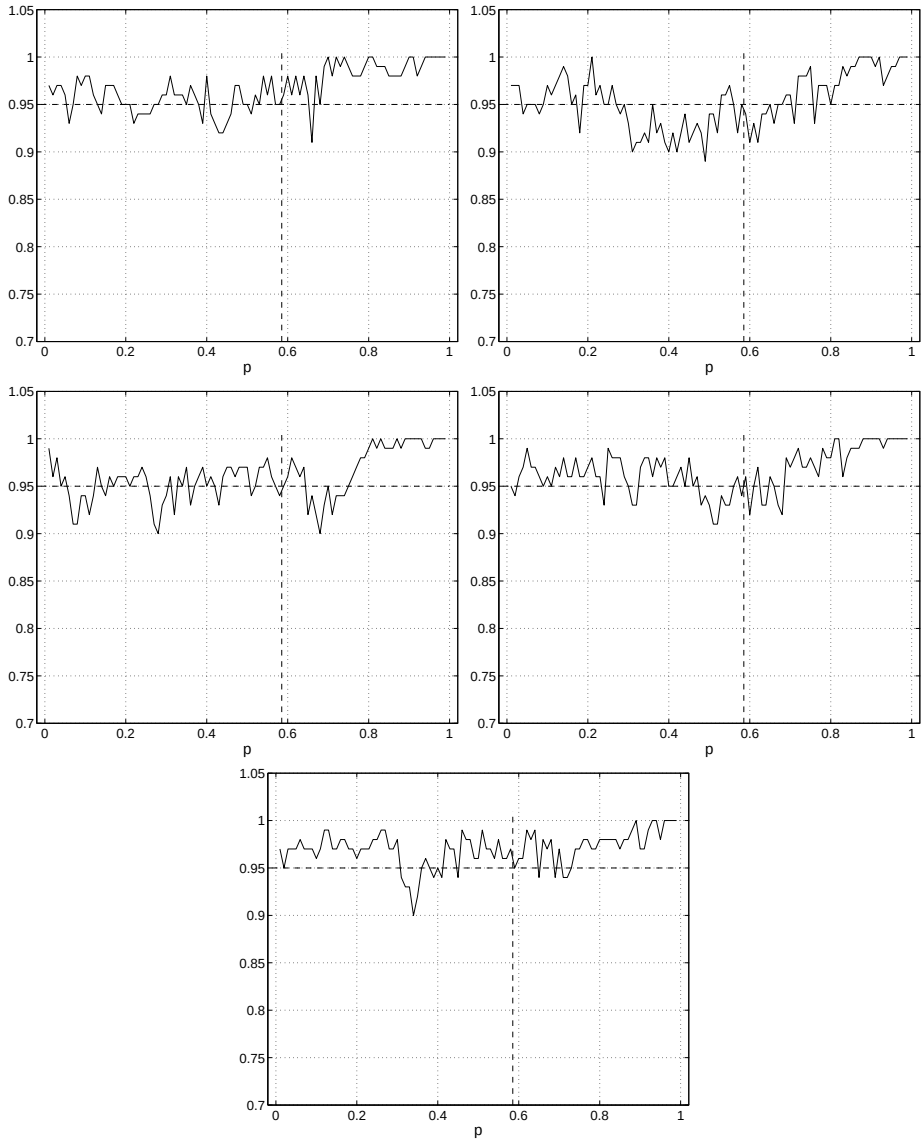

Figure B.4: *The nonasymptotic method for the Ising model, Section 11.8, page 168. Proportion of the confidence intervals that actually cover the correct value of β on box $\mathbb{B}_{20}$ for different values of the selection probability. The upper left diagram is for $p_s = 0.20$, upper right for $p_s = 0.40$, middle left for $p_s = 0.60$, middle right for $p_s = 0.80$ and lower for $p_s = 1.00$.*

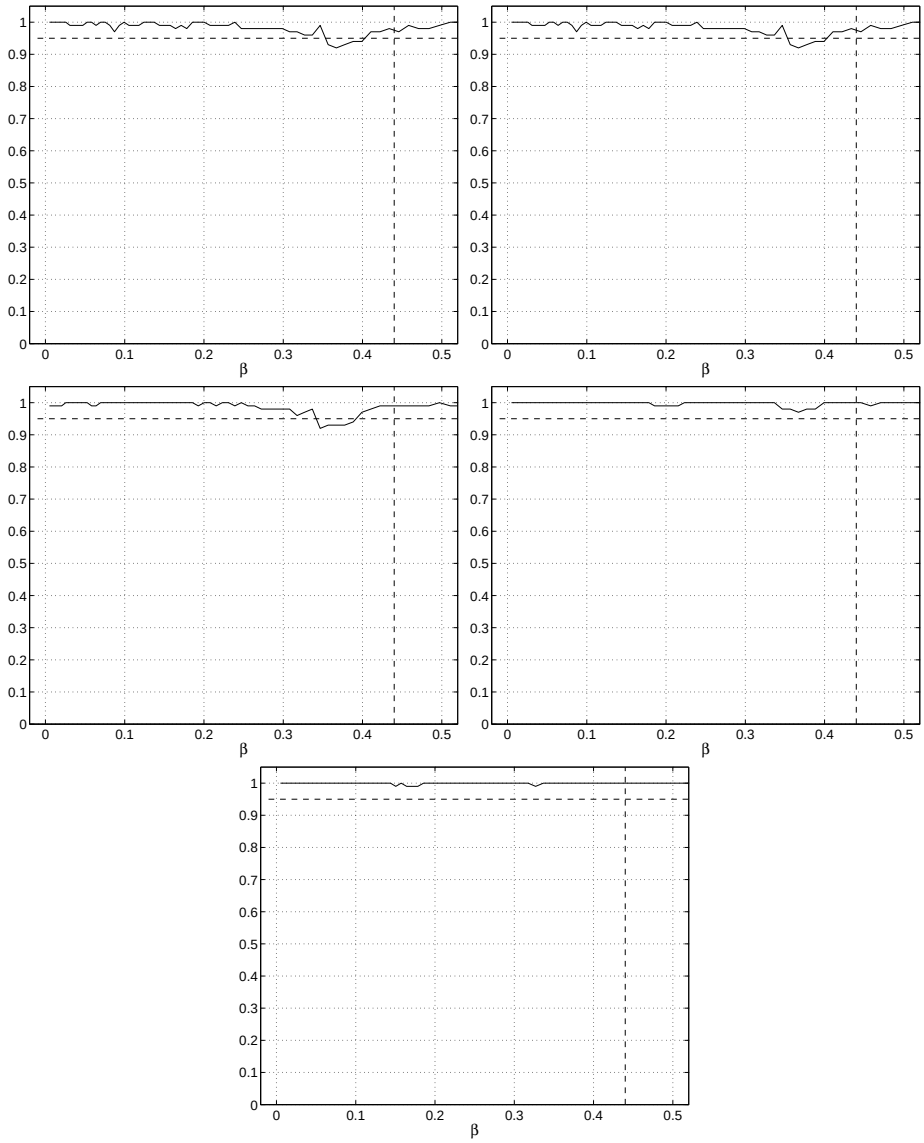

Figure B.5: *Pairwise interaction for the Potts model, Section 12.3, page 176. Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (middle left), $\mathbb{B}_{40}$ (middle right) and $\mathbb{B}_{50}$ (below). The correct confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.20$ and $q = 4$.*

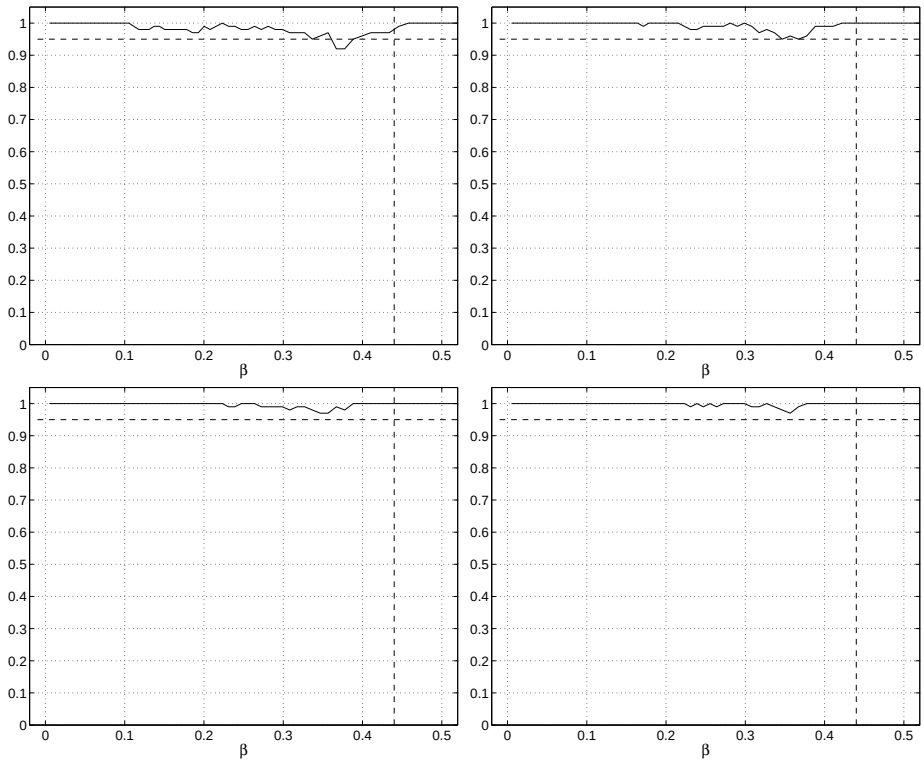

Figure B.6: *Box interactions for the Potts model, Section 12.5, page 185.Proportion of the confidence intervals that actually cover the correct value of β on boxes $\mathbb{B}_{10}$ (upper left), $\mathbb{B}_{20}$ (upper right), $\mathbb{B}_{30}$ (lower left) and $\mathbb{B}_{40}$ (lower right). The desired confidence level is $\alpha = 0.95$. The results are given for selection probability $p_s = 0.20$. The corresponding diagrams for $p_s = 0.10$ are shown in Figure 12.6 on page 187.*

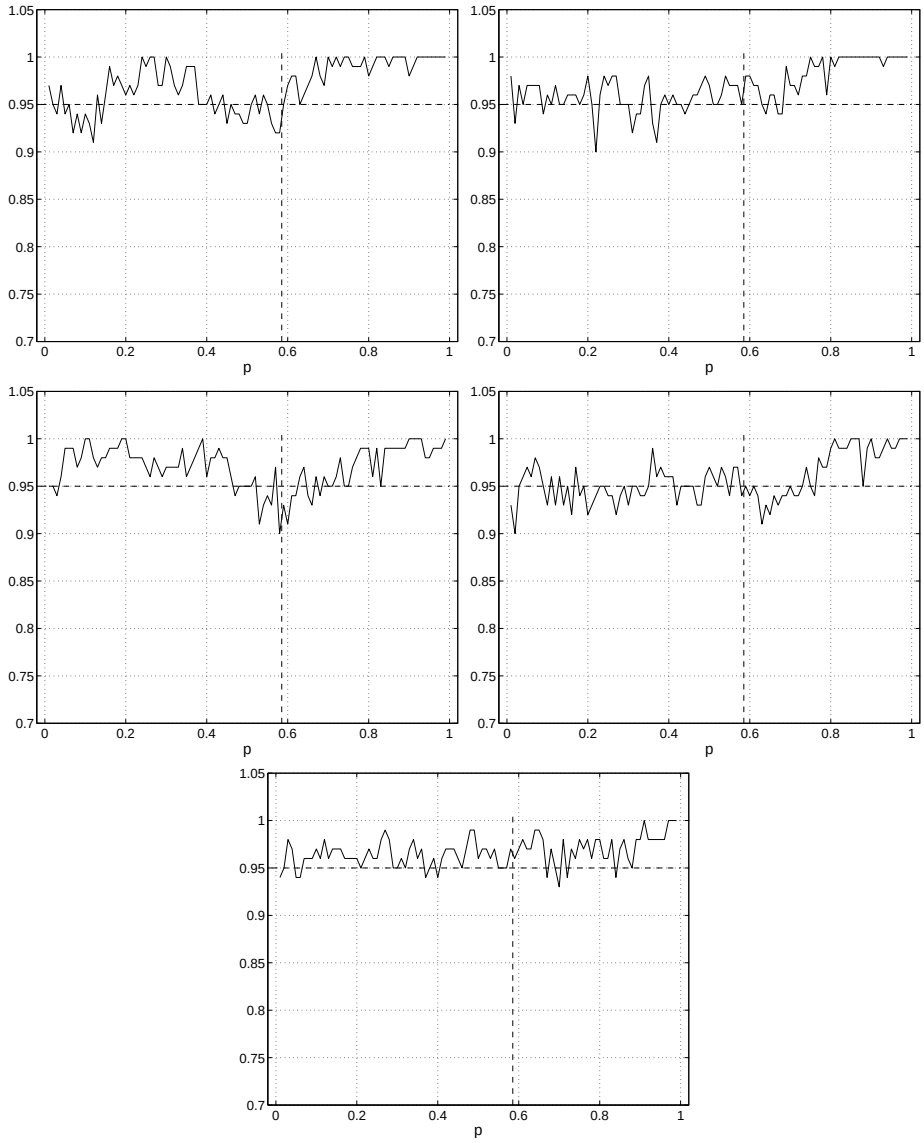

Figure B.7: *The nonasymptotic method for the Potts model, Section 12.7, page 191. Proportion of the confidence intervals that actually cover the correct value of* β *on boxes* $\mathbb{B}_{10}$ *(upper left),* $\mathbb{B}_{20}$ *(upper right),* $\mathbb{B}_{30}$ *(middle left),* $\mathbb{B}_{40}$ *(middle right), and* $\mathbb{B}_{50}$ *(lower) with* $q = 4$*. The desired confidence level is* $\alpha = 0.95$*. The results are given for selection probability* $p_s = 0.20$*.*

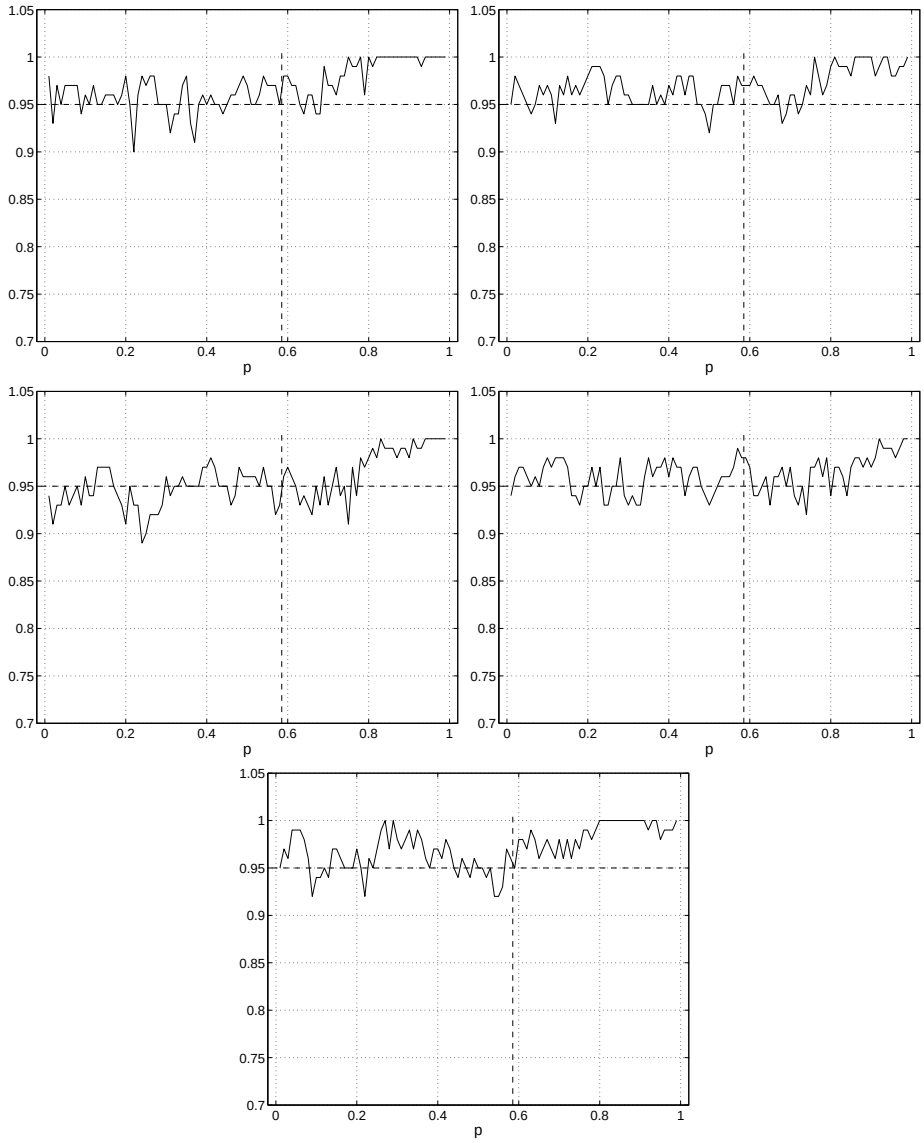

Figure B.8: *The nonasymptotic method for the Potts model, Section 12.7, page 191. Proportion of the confidence intervals that actually cover the correct value of β on box $\mathbb{B}_{20}$ for different values of the selection probability with $q = 4$. The upper left diagram is for $p_s = 0.20$, upper right for $p_s = 0.40$, middle left for $p_s = 0.60$, middle right for $p_s = 0.80$ and lower for $p_s = 1.00$.*

Bibliography

[ACCN87] M. Aizenman, J.T. Chayes, L. Chayes, and C.M. Newman. The phase boundary in dilute and random ising and potts ferromagnets. *Journal of* P*hysics A*, 30:L313–L318, 1987.

[ACCN88] M. Aizenman, J.T. Chayes, L. Chayes, and C.M. Newman. Discontinuity of the magnetization in one-dimensional $1/|x-y|^2$ ising and potts models. *Journal of* S*tatistical* P*hysics*, 50(1-2):1–40, 1988.

[Ale95] K.S. Alexander. Simultaneous uniqueness of infinite clusters in stationary random labeled graphs. *Communications in Mathematical Physics*, 168:39–55, 1995.

[Ale01] K.S. Alexander. The asymmetric random cluster model and comparison of Ising and Potts models. *Probability Theory and Related Fields*, 120(3):395–444, 2001.

[AML73] D.B. Abraham and A. Martin-Löf. The transfer matrix for a pure phase in the two-domensional Ising model. *Communications in Mathematical Physics*, 32:245–268, 1973.

[AT43] J. Ashkin and E. Teller. Mixing properties and exponential decay for lattice systems in finite volumes. *Physical Review*, 64(5-6):178–184, 1943.

[BC00] M.B. Bouabci and C. E. I. Carneiro. Random-cluster representation for the Blume-Capel model. *Journal of statistical physics*, 100(5-6):805–827, 2000.

[Boc83] N. Boccara. Dilute Ising models: a simple theory. *Physics Letters A*, 94(3-4):185–187, 1983.

[Bol82] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *Annals of Probability*, 10(4):1047–1050, 1982.

[CG92] F. Comets and B. Gidas. Parameter estimation for Gibbs distributions from partially observed data. *Annals of Applied Probability*, 2(1):142–170, 1992.

[DH05] M. Deijfen and O. Häggström. Nonmonotonic coexistence regions for the two-type Richardson model on graphs. *Electronic Journal of Probability, to appear*, 2005.

[DL81] R. Durrett and T.M. Liggett. The shape of the limit set in the Richardson growth model. *The Annals of Probability*, 9:2:186–193, 1981.

[Dob68] R.L. Dobrushin. Gibbsian random fields for lattice systems and pairwise interactions. *Functional analysis and its applications (in translation)* , 2:292–301, 1968.

[Dob72] R.L. Dobrushin. Gibbs state describing co-existence of phases for a three-dimensional Ising model. *Theory of Applied probability*, 17:582–600, 1972.

[Doe38] W. Doeblin. Exposé de la Théorie des Chaînes simples constantes de Markoff à un nombre fini d'´Etats. *Revue Mathematique de l'Union Interbalkanique*, 2:77–105, 1938.

[Dur96] R. Durrett. *Probability: Theory and examples (2:nd ed.)*. Duxbury Press, 1996.

[ES88] R.G. Edwards and A.D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physics Review D*, 38:2009–2012, 1988.

[Fel48] W. Feller. On the Kolmogorov-Smirnov distribution for empirical distributions. *Annals of Mathematical Statistics*, 19:177–189, 1948.

[FK72] C.M. Fortuin and P.W. Kasteleyn. On the random cluster model I. Introduction and relation to other models. *Physica*, 57:536–564, 1972.

[FKG71] C.M. Fortuin, P. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22:89–103, 1971.

[For72a] C.M. Fortuin. On the random cluster model II. The percolation model. *Physica*, 58:393–418, 1972.

[For72b] C.M. Fortuin. On the random cluster model III. The simple random-cluster model. *Physica*, 59:545–570, 1972.

[GHM01] H.O. Georgii, O. Häggström, and C. Maes. The random geometry of equilibrium phases. *Phase Transitions and Critical Phenomena*, 18:1–142, 2001.

[Gid93] B. Gidas. *Markov random fields: Theory and applications*, chapter 17, Parameter estimation for Gibbs distributions from fully observed data. Academic Press, 1993.

[GM05] O. Garet and R. Marchand. Coexistence in two-type first-passage percolation models. *Annals of Applied Probability*, 15(1A):298–330, 2005.

[GMS72] G. Gallavotti and S. Miracle-Sole. Equilibrium states of the Ising model in the two-phase region. *Physics Review B*, 5:2555–2559, 1972.

[GP97] G. Grimmett and M. Piza. Decay of correlations in subcritical Potts and random-cluster models. *Communications in Mathematical Physics*, 189(2):465–480, 1997.

[Gri95] G. Grimmett. The stochastic random-cluster process and the uniqueness of random cluster measures. *Annals of Probability*, 23(4):1461–1510, 1995.

[Gri99] G. Grimmett. *Percolation*. Springer-Verlag, 1999.

[Gri03] G. Grimmett. The random-cluster model. *Encyclopedia of Mathematical Sciences*, 110:73–123, 2003.

[Hig91] Y. Higuchi. Level set representation for the Gibbs states of the ferromagnetic Ising model. *Probability theory and related fields*, 90:203–221, 1991.

[Hof05] C. Hoffman. Coexistence for Richardson type competing spation growth models. *Annals of Applied Probability*, 15(1B):739–747, 2005.

[Hol74] R. Holley. Remarks on the FKG inequality. *Communications in Mathematical Physics*, 36:227–231, 1974.

[HP98] O. Häggström and R. Pemantle. First passage percolation and a model for competing spatial growth. *Journal of Applied Probability*, 35:683–692, 1998.

[HP00] O. Häggström and R. Pemantle. Absence of mutual unbounded growth for almost all parameter values in the two-type Richardson model. *Stochastic Processes and their Applications*, 90(2):207–222, 2000.

[Häg98] O. Häggström. Random-cluster representations in the study of phase transitions. *Markov Processes and Related Fields*, 4:275–321, 1998.

[Isi25] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

[Kau49] B. Kaufmann. Crystal statistics II: Partition function evaluated by spinar analysis. *Physics Review*, 76:1244–1252, 1949.

[Kes80] H. Kesten. The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Communications in Mathematical Physics*, 74:41–59, 1980.

[KS82] R. Kotečky and S. Schlosman. First-order phase transitions in large entropy lattice systems. *Communications in Mathematical Physics*, 83:493–550, 1982.

[Len20] W. Lenz. Beitrag zum verständnis der magnetischen eigenschaften in festen körpern. *Phys. Zeitschr*, 21:613–615, 1920.

[Lin92] T. Lindvall. *Lectures on the coupling method*. John Wiley and Sons, 1992.

[LMMS+91] L. Laanait, A. Messager, S. Miracle-Solé, J. Ruiz, and S. Shlosman. Interfaces in the Potts model. I. Pirogov-Sinai theory of the Fortuin-Kasteleyn representation. *Communications in Mathematical Physics*, 140(1):81–91, 1991.

[LMR86] L. Laanait, A. Messager, and J. Ruiz. Phase coexistence and surface tensions for the potts model. *Communications in Mathematical Physics*, 105:527–545, 1986.

[LR69] O.E. Lanford and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics*, 13:194–215, 1969.

[MM04] E. Mossel and M.Steel. A phase transition for the random cluster model on a phylogenetic tree. *Mathematical Biosciences*, 187(2):189–203, 2004.

[MMS75] A. Messager and S. Miracle-Sole. Equilibrium states of the two-domensional Ising model in the two-phase region. *Communications in Mathematical Physics*, 40:187–196, 1975.

[MV95] C. Maes and K. Vande Velde. The fuzzy Potts model. *Journal of physics* A, 28:4261–4271, 1995.

[Nor97] J.R. Norris. *Markov Chains*. Cambride University Press, 1997.

[Ons44] L. Onsager. Crystal statistics I. a two-dimensional model with order-disorder transformation. *Physics review*, 65:117–149, 1944.

[Pei36] R. Peierls. On Ising's model on ferromagnetism. *Proceedings of the* Cambridge Philosophical Society, 32:477–481, 1936.

[Pic76] D.K. Pickard. Asymptotic inference for an Ising lattice. *Journal of Applied Probability*, 13(3):486–497, 1976.

[Pic77] D.K. Pickard. Asymptotic inference for an Ising lattice II. *Advances in Applied Probability*, 9(3):476–501, 1977.

[Pic79] D.K. Pickard. Asymptotic inference for an Ising lattice III. nonzero field and ferromagnetic states. *Journal of Applied Probability*, 16(1):12–24, 1979.

[Pot52] R.B. Potts. Some generalized order-disorder transformations. *Proceedings of the* Cambridge Philosophical Society, 48:106–109, 1952.

[PV97] C. E. Pfister and Y. Velenik. Random-cluster representation of the Ashkin-Teller model. *Journal of statistical physics*, 88(5-6):1295–1331, 1997.

[PW96] J. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[Ric73] D. Richardson. Random growth in a tessellation. *Proceedings of the Cambridge Philosophical Society*, 74:515–528, 1973.

[Smi48] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, 19:279–281, 1948.

[Ste70] V. E. Stepanov. On the probability of connectedness of a random graph $\mathcal{G}_m(t)$. *Theory of probability and its applications*, 15(1):55–67, 1970.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36:423–439, 1965.

[SW87] R.H. Swendsen and J.S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.

[Wu80] F. Y. Wu. Exact results for a dilute Potts model. *Journal of Statistical Physics*, 23(6):773–782, 1980.

[Wu81] F. Y. Wu. Dilute Potts model, duality and site-bound percolation. *Journal of Physics A*, 14(2):L39–L44, 1981.

Index