Identifying interacting pairs of sites in infinite range Ising models

Antonio Galves*, Enza Orlandi † and Daniel Yasumasa Takahashi ‡

June 23, 2010

Abstract

We consider Ising models (pairwise interaction Gibbs probability measures) in $\mathbb{Z}^d$ with an infinite range potential. We address the problem of identifying pairs of interacting sites from a finite sample of independent realizations of the Ising model. The sample contains only the values assigned by the Ising model to a finite set of sites in $\mathbb{Z}^d$. Our main result is an upperbound for the probability with our estimator to misidentify the pairs of interacting sites in this finite set.

1 Introduction

The class of Ising models is a popular family of exponential distributions on the set of binary configurations supported by a finite or a countable set of sites. The law of the Ising model is characterized by a pairwise potential, *i.e.*, a family of real numbers $J(i, j)$ indexed by pairs of sites (i, j). Originally introduced in statistical mechanics as a mathematical model for ferromagnetism, the Ising model has been extensively used, for instance, in computer vision (Woods, 1978, Besag, 1993), image processing (Cross and Jain, 1983), neuroscience (Schneidman et al., 2006), and as a general model in spatial statistics (Ripley, 1981). The references given above are just starting points of a huge literature. For a recent statistical physics oriented survey of rigorous mathematical results on Gibbs distributions, including Ising models, we refer the reader to Presutti (2009).

*AG was partially supported by CNPq grant 305447/2008-4. This work is part of CNPq projects 476501/2009-1 and 485999/2007-2.

†Partially supported by Prin07:20078XYHYS.

‡Supported by FAPESP grant 2008/08171-0.

In this article we consider Ising models having $\mathbb{Z}^d$ as set of sites and with an infinite range potential. This means that we may have pairs of interacting sites (i, j), with $|i - j|$ as big as we want, provided that they satisfy some minor requirements that will be given in Section 2. By interacting sites, we mean that $J(i, j) \neq 0$. We address the statistical problem of identifying these pairs of interacting sites, given a finite sample of independent realizations of the Ising model.

The Ising model is supported by the set of infinite configurations $\{-1, +1\}^{\mathbb{Z}^d}$. However, from an applied statistics point of view, we cannot observe more than the projection of the Ising model on a finite subset of sites. Therefore, our sample will be constituted by the values that a finite sequence of independent realizations of the Ising model assign to a fixed finite set of sites. As a consequence, we can only identify pairs of interacting sites belonging to the finite set of sites we observe. This finite set of sites is arbitrary. The idea is to make its size increase together with the size of the sample.

We introduce an estimator for the set of interacting pairs of sites belonging to the finite set we observe. This estimator can be informally described as follows. For each site i in the observed finite set we estimate the conditional probability of the model in i, given the remaining sites in the finite set. Then we compare this empirical conditional probability with the empirical conditional probability on the same site i given the remaining sites with the exception of another site j, with $j \neq i$. If the two conditional probabilities are statistically equal, we conclude that $J(i, j) = 0$. By *statistically equal* we mean that the weighted difference between the two empirical probabilities is smaller than a certain threshold value, which depends on the size of the sample and on the number of observed sites.

The main result of present article is an upperbound for the probability of misidentifying the pairs of interacting sites in this finite set using our estimator. The proof of this theorem has two ingredients, which are interesting by themselves.

The first ingredient is an upperbound for the probability of misidentification for the Ising model with finite range potential. This is the content of Theorem 3. This theorem applies to the usually considered situation (Ravikumar et al., 2010, Bento and Montanari, 2009, Bresler et al., 2008) where the totality of interacting sites are observed in the sample. The second ingredient in the proof of our main theorem is a coupling result given in Theorem 4. It says that we can couple together an Ising model with infinite range potential and an Ising model with truncated finite range potential in such a way that the probability of discrepancy at a fixed site i vanishes as the set of observed sites diverges to $\mathbb{Z}^d$. As a consequence of this result, we are able to bound above the probability of misidentification due to the fact that we are able to observe a finite set of sites, not the entire set of interacting sites. The proof of this result uses Dobrushin's contraction method but this result, as far as we know, has not been presented in the literature. For a presentation of the contraction method in its original framework we refer the reader to Dobrushin (1968).

As a byproduct of our results, we also obtain an identification theorem for the case of the Ising model on a finite set of points. This finite case was recently studied in several papers, including Ravikumar et al. (2010), Bento and Montanari (2009), Bresler et al. (2008). The case of the Ising model with finite range and homogenous potential was recently studied by Csiszar and Talata (2006) using a BIC like approach. A comparative discussion of our results versus the ones presented in the above mentioned papers will be done in Section 4.

This paper is organized as follows. Notation, definitions and results are presented in Section 2. In Section 3 we illustrate our theoretical results with a simulation experiment. Final remarks and a discussion of recent related results are given in Section 4. The proofs of the theorems are presented in the Appendix 1. Finally, Appendix 2 contains the pseudocode describing the algorithm to simulate the samples for the simulation study.

2 Notation, definitions and results

Definition 1. *A pairwise potential is a family* $J = \{J(i,j) : (i,j) \in \mathbb{Z}^d \times \mathbb{Z}^d\}$ *of real numbers which satisfy the conditions*

$$J(i,i) = 0, \qquad J(i,j) = J(j,i), \qquad \sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} |J(i,j)| < \infty. \tag{1}$$

Let $A = \{-1, 1\}$ be the basic binary set and $S = A^{\mathbb{Z}^d}$ be the set of spin configurations on the lattice $\mathbb{Z}^d$. Fixed configurations will be denoted by lower case letters $x, \ldots$ whereas the capital letter X will denote a random field. A point $i \in \mathbb{Z}^d$ will be called a site. For any $i \in \mathbb{Z}^d$, $x(i)$ will denote the value of the configuration x at site i. Given a subset F of $\mathbb{Z}^d$, we shall also denote $x(F) = \{x(i) : i \in F\}$ and similarly for X.

Definition 2. *The Ising model with pairwise potential J is a random element X with* *values on S, which distribution satisfies*

$$\mathbb{P}\left(X(i)=x(i)|X(j)=x(j)\,,\,j\neq i\right)=\frac{1}{1+\exp(-2\sum_{j\in\mathbb{Z}^d}J(i,j)x(i)x(j))},$$

for all $i\in\mathbb{Z}^d$ and for any fixed $x\in S$.

In the above definition the left hand side of the above equality denotes a version of the conditional probability of $X(i)$ given that $X(j)=x(j)\,,\,j\neq i$.

Let F be a finite subset of $\mathbb{Z}^d$. We shall use the shorthand notation $p(x(F))$ and $p(x(i)|x(F))$ to denote, respectively, the probability $\mathbb{P}(X(F)=x(F))$ and the conditional probability $\mathbb{P}(X(i)=x(i)|X(F\setminus\{i\})=x(F\setminus\{i\}))$.

Definition 3. *The interaction neighborhood $\mathcal{G}(i)$ of a site $i\in\mathbb{Z}^d$ is defined as*

$$\mathcal{G}(i)=\Big\{j\in\mathbb{Z}^d:J(i,j)\neq 0\Big\}\,.$$

Let $B^L(i)=\{j\in\mathbb{Z}^d:|i-j|\leq L\}$ be the ball of radius $L\geq 1$ centered at site $i\in\mathbb{Z}^d$, where $|\cdot|$ is the maximum norm. Our goal in this paper is to infer $\mathcal{G}(i)\cap B^L(i)$ from the values of $X_1(B^L(i)),\ldots,X_n(B^L(i))$, when n and L are suitably scaled, where $X_1,\ldots,X_n$ are independent realizations of the Ising model X.

To do this, let us introduce for any finite set F the empirical probability measure

$$\hat{p}_n(x(F))=\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}\{X_k(F)=x(F)\},$$

where $\mathbf{1}$ denotes the indicator function. Given any site $j\in\mathbb{Z}^d$, we will also define the empirical conditional probability

$$\hat{p}_n(x(j)|x(F\setminus\{j\}))=\frac{\hat{p}_n(x(F\cup\{j\}))}{\hat{p}_n(x(F\setminus\{j\}))}\,,$$

if $\hat{p}_n(x(F\setminus\{j\}))>0$ and $\hat{p}_n(x(j)|x(F\setminus\{j\}))=0$, otherwise.

For any fixed configuration $x(B^L(i))$ and any $j\in B^L(i)$ we define the empirical weighted distance between the conditional probabilities as follows

$$\hat{D}_n(x(B^L(i)),j)=\big|\hat{p}_n(x(i)|x(B^L(i)\setminus\{i\}))-\hat{p}_n(x(i)|x(B^L(i)\setminus\{i,j\}))\big|\,\hat{p}_n(x(B^L(i)\setminus\{i\})\,. \tag{2}$$

Note that $\hat{D}_n(x(B^L(i)),j)$ is a function of the sample $X_1,\ldots,X_n$ and is therefore a random variable.

We can now define our estimator.

Definition 4. *For any site i and range $L \geq 1$ the interaction neighborhood estimator is defined as follows*

$$\hat{V}_n^L(i) = \left\{ j \in B^L(i) : \max_{x(B^L(i))} \hat{D}_n(x(B^L(i)), j) > \epsilon \right\} .$$

where the threshold ϵ is a suitable positive real number.

Our first theorem shows that for a convenient threshold ϵ, defined as a function of L and n, the estimated interaction neighborhood $\hat{V}_n^L(i)$ recovers perfectly the set $\mathcal{G}(i)$, when L and n are suitably scaled.

Theorem 1. *Let $i \in \mathbb{Z}^d$ and $X_1(B^L(i)), \ldots, X_n(B^L(i))$ be the projections of independent realizations of an Ising model whose pairwise potential satisfies conditions*

$$\sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d} |J(i,j)| = r < 1 \tag{3}$$

and

$$\lim_{L \to \infty} \sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d, |j-i| > L} |J(i,j)| e^{(2L)^d} = 0. \tag{4}$$

Under these assumptions, if we choose the threshold $\epsilon = \epsilon(L, n)$ as

$$\epsilon(L,n) = C \sqrt{\frac{(2L)^d}{n} \left(\frac{1+e^2}{e^2} \right)^{-(2L)^d}}, \tag{5}$$

where $C > 0$ is an arbitrary constant and

$$L_n = \frac{1}{2}(0.79 \log n)^{1/d},$$

then for any $i \in \mathbb{Z}^d$ the following assertions hold

(i) for any $j \in \mathcal{G}(i)$, $\lim_{n \to \infty} \mathbb{P}\left(j \in \hat{V}_n^{L_n}(i) \right) = 1$;

(ii) for any $j \notin \mathcal{G}(i)$, $\lim_{n \to \infty} \mathbb{P}\left(j \notin \hat{V}_n^{L_n}(i) \right) = 1$.

Theorem 1 follows from an upperbound for the probability of misidentification of points in a suitably defined subset of $\mathcal{G}(i) \cap B^L(i)$, which holds for any choice of the sample size n, box size L, and threshold ϵ. This is the content of the next theorem. To state it we need some extra definitions.

Given an integer $L \geq 1$ and a pairwise potential J, we denote J^L the truncated potential defined as follows. For any pair of sites i and j

$$J^L(i,j) = \begin{cases} J(i,j) & \text{if } |i-j| \leq L \\ 0\,, & \text{otherwise.} \end{cases} \tag{6}$$

We shall also denote X^L the corresponding Ising model with pairwise potential J^L.

Let F be a finite subset of $\mathbb{Z}^d$. As before, we shall use the shorthand notation $p^L(x(F))$ and $p^L(x(i)|x(F))$ to denote, respectively, the probability $\mathbb{P}(X^L(F) = x(F))$ and the conditional probability $\mathbb{P}(X^L(i) = x(i)|X^L(F \setminus \{i\}) = x(F \setminus \{i\}))$.

Definition 5. *Given a site $i \in \mathbb{Z}^d$ and integers L and n, let*

$$D(x(B^L(i)), j) = \left|p^L(x(i)|x(B^L(i) \setminus \{i\})) - p^L\left(x(i)|x(B^L(i) \setminus \{i,j\})\right)\right| p^L(x(B^L(i) \setminus \{i\})).$$

The interaction neighborhood $V^L(i)$ of i is defined as

$$V^L(i) = \left\{ j \in B^L(i) : \max_{x(B^L(i))} D(x(B^L(i)), j) > 2\epsilon \right\}, \tag{7}$$

where ϵ denotes the same threshold appearing in Definition 4.

Theorem 2. *Let $i \in \mathbb{Z}^d, L \geq 1$, and $X_1(B^L(i)), \ldots, X_n(B^L(i))$ be the local projections of independent realizations of a Ising model whose pairwise potential satisfies*

$$\sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d} |J(i,j)| = r < 1. \tag{8}$$

Then for any site $i \in \mathbb{Z}^d$ and any threshold value $\epsilon > 0$, we have

$$\begin{aligned} &\mathbb{P}\left(\hat{V}_n^L(i) \neq V^L(i)\right) \\ &\leq 4 \exp\left(-\frac{n\epsilon^2}{8v + \frac{4}{3}\epsilon} + 2(2L)^d\right) + \frac{1}{1-r} n(2L)^d \left(\sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d, |j-i| > L} |J(i,j)|\right), \end{aligned} \tag{9}$$

where

$$v = \sup_{x(B^L(i))} \sup_{j \in B^L(i)} \left(1 - p^L\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) p^L(x(B^L(i) \setminus \{j\}))\right). \tag{10}$$

Theorem 2 applies also when the interaction graph is embedded in a finite set Λ and we analyze a sample with the values the Ising model assigns to a smaller subset of sites in Λ. This is the content of the following corollary.

Corollary 1. *Let $i \in \Lambda \subset \mathbb{Z}^d$, $L \geq 1$, and $X_1(B^L(i) \cap \Lambda), \ldots, X_n(B^L(i) \cap \Lambda)$ be the projections of independent realizations of a Ising model whose pairwise potential satisfies*

$$J(i,j) = 0 \ \text{ for } \ j \in \mathbb{Z}^d \setminus \Lambda \tag{11}$$

and

$$\sup_{i\in\Lambda} \sum_{j\in\Lambda} |J(i,j)| = r < 1. \tag{12}$$

Then for any site $i \in \Lambda$ and any threshold value $\epsilon > 0$, we have

$$\begin{aligned} &\mathbb{P}\left(\hat{V}_n^L(i) \neq V^L(i)\right) \\ &\leq 4\exp\left(-\frac{n\epsilon^2}{8v + \frac{4}{3}\epsilon} + 2(2L)^d\right) + \frac{1}{1-r} n(2L)^d \left(\sup_{i\in\Lambda} \sum_{j\in\Lambda, |j-i|>L} |J(i,j)|\right), \end{aligned} \tag{13}$$

where v is defined in (10).

The first ingredient in the proof of Theorem 2 is an upperbound for the probability of misidentification of interacting pairs in the case of a finite range interaction. This is given in the next theorem.

Theorem 3. *Let $i \in \mathbb{Z}^d, L \geq 1$, and $X_1(B^L(i)), \ldots, X_n(B^L(i))$ be the projections of independent realizations of a Ising model whose pairwise potential satisfies*

$$J(i,j) = 0, \ \ \text{if } \ |i-j| > L. \tag{14}$$

Then for any site $i \in \mathbb{Z}^d$ and any threshold value $\epsilon > 0$, we have

$$\mathbb{P}\left(\hat{V}_n^L(i) \neq V^L(i)\right) \leq 4\exp\left(-\frac{n\epsilon^2}{8v + \frac{4}{3}\epsilon} + 2(2L)^d\right),$$

where

$$v = \sup_{x(B^L(i))} \sup_{j\in B^L(i)} \left(1 - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right) p(x(B^L(i)\setminus\{j\}))\right).$$

The second ingredient in the proof of Theorem 2 is a coupling result. To state it we first need to introduce the notion of coupling. This is done as follows.

Let X be a Ising model with pairwise potential J. For a fixed integer $L \geq 1$, let J^L be the corresponding truncated potential defined as in (6).

Definition 6. *A coupling between X and X^L is a random element $(\tilde{X}, \tilde{X}^L)$ taking values on $S \times S$ such that*

1. $\tilde{X}$ *has the same law as* X*;*

2. $\tilde{X}^L$ *has the same law as* X^L*.*

The following theorem says that we can sample together X and X^L in such a way that the probability of discrepancy at the origin vanishes as L diverges.

Theorem 4. *If* J *is pairwise potential which satisfies condition (8), i.e.,*

$$\sup_{i\in\mathbb{Z}^d}\sum_{j\in\mathbb{Z}^d}|J(i,j)| = r < 1$$

and J^L *is defined as in (6), then there exist a coupling* $(\tilde{X},\tilde{X}^L)$ *such that for any* $i\in\mathbb{Z}^d$ *the following inequality holds*

$$\mathbb{P}\left(\tilde{X}(i)\neq\tilde{X}^L(i)\right) \leq \frac{1}{1-r}\sup_{i\in\mathbb{Z}^d}\sum_{j\in\mathbb{Z}^d\setminus B^L(i)}|J(i,j)|. \tag{15}$$

3 Simulation experiments

This section presents simulation experiments results that illustrates Theorem 2.

The experiment was done with the Ising model on $B^3(0)$, where $B^3(0)\in\mathbb{Z}^d$ is the box with width 7 centered at the origin. The set of pairs of interaction sites was chosen randomly as follows. Each pair of sites (i,j) inside $B^3(0)$ was included in the set of interacting pairs with probability 0.1, independently of the others, with the restriction that the degree of each site was bounded above by 4. For each pair of interaction site we took $J(i,j) = 0.2$. The set of interacting pairs considered in the experiment is represented in Figure 1.

Samples of the Ising model on this finite graph were generated using a Gibbs sampler. More precisely, we considered the Markov chain (Y_t, Y_t') taking values on $\{-1,+1\}^{2B^3(0)}$. The algorithm we used to generate the samples is described in the pseudocode Sample Generator presented in Appendix 2. We start with initial configurations $Y_0(i) = +1$ and $Y_0'(i) = -1$ for all $i\in B^3(0)$. We run the Markov chain until the first time $t\geq 1$ such that $Y_t = Y_t'$. At this time the Markov chain is in the stationary state and therefore the common configuration attained by Y_t and Y_t' is a realization of the Ising model.

The goal of simulation experiment was to identify the sites in $B^1(0)$ interacting with site $(0,0)$. Observe that there are two sites outside $B^1(0)$ which interact with $(0,0)$,

Figure 1: Graph representing the interacting pairs of sites. The vertices represent the sites and the presence of edges between the vertices represent the existence of interaction. The sites interacting with site $(0, 0)$ are indicated with bold grey edges.

namely $(0, 3)$ and $(2, 0)$. We considered 39 different sample sizes, $n = 500 + 250k$, where $k = 0, \ldots, 38$ for the estimator $\hat{V}_n^1(0)$. The threshold used in the estimator $\hat{V}_n^1(0)$ was

$$\epsilon(L, n) = C\sqrt{\frac{1}{n}}$$

with $C \in [0.06, 0.09]$. The probabilities of type 1 errors (false positives) and type 2 errors (false negatives) were estimated by repeating the experiments 2000 times for each sample size. The result is presented in Figure 2.

4 Discussion

Ising systems with infinite range potential are natural candidates to model population of neurons, protein networks, large computer networks, social networks, just to mention a few examples. This is due to the fact that the interaction potential describing the Ising model encodes the structure of conditional independence among the components of the system.

The examples mentioned above present all the following characteristics. They are supported by a very large number of components. The activity of all the components

Figure 2: Graph of estimated probabilities of false positive and false negative identification using the estimator $\hat{V}_n^1(0)$ for different sample sizes.

of the system cannot be observed simultaneously. Even worse, there is no clear *a priori* candidate for the range of interactions between pairs of components.

Summarizing, in those systems, only a small set of components is observed, even if these components are under the influence of the totality of non-observed components. This leads naturally to the following statistical question. Given a sample with the values assigned to a finite set of sites by a finite sequence of independent realizations of the Ising model on $\mathbb{Z}^d$, how can we identify pairs of interacting sites belonging to this finite subset? This is precisely the question we address here. For the best of our knowledge, the present article is the first one to address this problem.

We use $\mathbb{Z}^d$ to represent the set of components of the system. This is just to have a convenient definition of distance between sites. However, it should be clear from our proofs that many other metric structures between the sites could be used instead of the ℓ_∞ distance considered here without significant change in the results of this article. Also, instead of using the ℓ_∞-ball $B^L(i)$ as set of observed sites, we could use any increasing sequence of finite sets $\Lambda^L \subset \mathbb{Z}^d$ such that $\cup_{L=1}^{\infty} \Lambda^L = \mathbb{Z}^d$. It is an easy exercise to translate our results to this general framework.

Recently, Ravikumar et al. (2010), Bento and Montanari (2009), Bresler et al. (2008), and Csiszar and Talata (2006) addressed the problem of identification of interacting sites in the case of random fields, either on a finite set of sites or on an infinite set but with a finite range interaction. Let us briefly discuss these results.

Csiszar and Talata (2006) consider the case of homogeneous finite range Markov random fields in $\mathbb{Z}^d$. They use what they call the Pseudolikelihood Information Criterion (PIC) to identify the interaction neighborhood of a site. The homogeneity of the model makes it possible to use as a sample the values assigned to a finite set of sites by a unique realization of the Markov random field. The finiteness of the range of the Markov random field implies that the interaction neighborhood of a site will be entirely observed when the sample size increases. Therefore, they do not need to control the effect of non-observed sites as we do. In our Theorem 2 this corresponds to the second term of the upperbound. They prove the consistency of their identification procedure when the size of the set of observed sites diverges. They do not give any explicit upperbound for the probability of misidentification when we only observe a finite set of sites. The asymptotic nature of the result presented in Csiszar and Talata (2006) it is an important difference between this article and ours. An interesting feature of their result is that it holds even when the Markov random field has phase transition. It is enough to apply the algorithm to a pure phase of the model to correctly identify the interaction neighborhood. This feature is also shared by our approach in the case of finite range Ising model. We recall that our Theorem 3 does not assume Dobrushin's condition and therefore holds also for finite range potentials with arbitrary high interaction values. In statistical physics language, Theorem 3 holds also for low temperature Ising models exhibiting phase transition.

Bresler et al. (2008) propose an algorithm which is similar in spirit to ours but without the weighting parameter. This article studies the problem in the case of random field on a finite set of sites. This paper does not consider the situation in which the number of sites supporting the probability measure increases when the sites of samples diverges. The fact the set of sites is finite and fixed plays a crucial role in their approach and cannot be applied in an obvious way to the infinite range case considered in the present paper.

Ravikumar et al. (2010) consider the ℓ_1 pseudolikelihood estimator of the Ising model with increasing sample size and number of interacting sites. They assume that the totality of the interacting sites is observed. Their algorithm seems to be fast enough in practice. However, their incoherence condition assumption seems to be very restrictive and difficult to check even for simple models. The very interesting article Bento and Montanari (2009) shows that the probability that such a condition is satisfied for large random graphs is very small.

In Bento and Montanari (2009) it is conjectured that the existence of efficient low complexity algorithm seems to be related to the absence of phase transition at least in the case of Ising models. This discussion is outside of the scope of the present article. However, it is worth mentioning that in our Theorems 1 and 2 we assume Dobrushin's condition (8) which implies absence of phase transition (cf. Dobrushin (1968) or Presutti (2009)). In our case this condition is crucial to obtain an upper bound for the probability of discrepancy in a fixed finite region between the coupled versions of the infinite range and the truncated finite range potential Ising models. This makes it possible with high probability to identify pairs of interacting sites without inspecting distant pairs of sites. In particular, even in the case of a finite but very large set of sites our algorithm only needs to test pairs of sites belonging to a much smaller finite region. This is precisely the content of our Corollary 1. As a consequence our algorithm only needs to check a small set of sites, which is computationally less demanding than testing pairwise interaction in the whole set of sites.

To conclude, we observe that our coupling approach can be used to extend in a straightforward way the results of Ravikumar et al. (2010), Bento and Montanari (2009), Bresler et al. (2008), for the case of Ising model of infinite range interaction. To do this we just need to be sure that Dobrushin's condition (8) is satisfied.

5 Acknowledgements

The authors thank Luiz Lana and Matthieu Lerasle for many discussions during the period in which this paper was written. Part of this work was done during DYT stay at the Universitá di Roma Tre, Dipartmento di Matematica, Rome, Italy and during EO stay at the Universidade de São Paulo, Departamento de Estatstica, NUMEC, São Paulo, Brazil.

Appendix 1

Proof of Theorem 1

We will first prove that

$$\lim_{n\to\infty} V_n^{L_n}(i) = \mathcal{G}(i). \tag{16}$$

For this we observe that for any $i \in \mathbb{Z}^d$

$$\max_{x(B^L(i)\setminus\{i\})} p^L(x(B^L(i) \setminus \{i\})) \geq 2^{-(2L)^d+1}$$

and for any $j \in B^L(i)$ such that $J(i,k) \neq 0$ we have

$$\begin{aligned}&\left|p^L(x(i)|x(B^L(i) \setminus \{i\})) - p^L\left(x(i)|x(B^L(i) \setminus \{i,j\})\right)\right| \\ &\geq \frac{2e^{2r}}{(1+e^{2r})^2}|J(i,j)| \min_{x(j)\in\{-1,+1\}} p^L(x(j)|x(B^L(i) \setminus \{i,j\})).\end{aligned}$$

Also

$$\min_{x(j)\in\{-1,+1\}} p^L(x(j)|x(B^L(i) \setminus \{i,j\})) \geq \frac{1}{1+e^{2r}}.$$

From the above inequalities, we have

$$\max_{x(B^L(i))} D(x(B^L(i)),j) \geq \frac{4e^{2r}}{(1+e^{2r})^3}|J(i,j)|2^{-(2L)^d}.$$

If $j \in \mathcal{G}(i)$, the right hand side of the last inequality is strictly positive. Therefore, from the above inequality a sufficient condition for this site j to belong to $V^L(i)$ is that

$$\frac{4e^{2r}}{(1+e^{2r})^3}|J(i,j)|2^{-(2L)^d} \geq \epsilon(L,n).$$

Hence, to equality (16) holds it is enough that

$$\lim_{n\to\infty} \frac{\epsilon(L_n,n)}{2^{-(2L_n)^d}} = 0. \tag{17}$$

It is easy to check that the above condition holds for $L_n \leq \frac{1}{2}\left(0.79 \log n\right)^{1/d}$ and with the threshold

$$\epsilon(L_n,n) = C\sqrt{\frac{(2L_n)^d}{n}\left(\frac{1+e^2}{e^2}\right)^{-(2L_n)^d}}$$

as defined in (5).

Now, we will prove

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{V}_n^{L_n}(i) \neq V_n^L(i)\right) = 0. \tag{18}$$

For this, we have from inequality (9) in Theorem 2 that if we prove

$$\lim_{n\to\infty} -\frac{n\epsilon(L_n,n)^2}{8v+\frac{4}{3}\epsilon(L_n,n)} + 2(2L_n)^d = -\infty, \tag{19}$$

where v is defined in (10), and also that

$$\lim_{n\to\infty} n(2L_n)^d \left(\sup_{i\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d, |j-i|>L_n} |J(i,j)|\right) = 0, \tag{20}$$

we prove (18).

To prove (19), we first observe that

$$1 - p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) \geq \frac{1}{1+e^{2r}}$$

and

$$\sup_{x(B^L(i)\setminus\{j\})} p(x(B^L(i) \setminus \{j\})) \geq 2^{-(2L)^d+1}.$$

From these inequalities we have that

$$v \geq \left(\frac{2}{1+e^{2r}}\right) 2^{-(2L)^d}.$$

The above inequality implies that for $L \leq \frac{1}{2}(0.79 \log n)^{1/d}$ we have

$$v \;\geq\; \left(\frac{2}{1+e^{2r}}\right) 2^{-(2L)^d} \;\geq\; \frac{4}{3}\epsilon(L,n).$$

Therefore, we can bound below the first term in the exponent of the upperbound (9) as follows

$$\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n)} \geq \frac{n\epsilon(L,n)^2}{9v}. \tag{21}$$

From (21), to obtain (19) it is enough that

$$\lim_{L,n\to\infty} -\frac{n\epsilon(L,n)^2}{9v} + 2(2L)^d = -\infty.$$

This implies that

$$\epsilon(L,n) > C\sqrt{\frac{2(2L)^d v}{n}}. \tag{22}$$

We will now find an upper bound for v.

From condition (8) it follows that

$$1 - p\left(x(i)/x(B^L(i) \setminus \{i,j\})\right) \leq \frac{e^{2r}}{1+e^{2r}}$$

and

$$p(x(B^L(i) \setminus \{j\})) \leq \left(\frac{e^{2r}}{1+e^{2r}}\right)^{(2L)^d-1}.$$

These inequalities imply that

$$v \leq \left(\frac{1+e^{2r}}{e^{2r}}\right)^{-(2L)^d}.$$

From the above inequality and the fact that $r < 1$, the choice of $\epsilon(L,n)$ in (5) satisfies (19).

To prove (20), note that $L = \frac{1}{2}(0.79 \log n)^{1/d}$ implies that $n < e^{(2L)^d}$. This concludes the proof of Theorem 1. □

Proof of Theorem 2 and Corollary 1

Let the integer $L \geq 1$ be fixed and let $(\tilde{X}_1, \tilde{X}_1^L), \ldots, (\tilde{X}_n, \tilde{X}_n^L)$ be n independent copies of the pair $(\tilde{X}, \tilde{X}^L)$ which existence is guaranteed by Theorem 4. The random elements $\tilde{X}_1, \ldots, \tilde{X}_n$ are independent copies of the Ising model X with pairwise potential J. The random elements $\tilde{X}_1^L, \ldots, \tilde{X}_n^L$ are independent copies of the Ising model X^L with truncated pairwise potential J^L defined as in (6).

Let us indicate explicitly the sample in all the statistics and events appearing in Theorem 2 as functions either of the sample $\tilde{X}_1, \ldots, \tilde{X}_n$ or of the sample $\tilde{X}_1^L, \ldots, \tilde{X}_n^L$. We start with notation of the empirical probability measures $\hat{p}_n$, as follows

$$\hat{p}_n(x(F))[\tilde{X}_1, \ldots, \tilde{X}_n] = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{\tilde{X}_k(F) = x(F)\}$$
$$\hat{p}_n(x(F))[\tilde{X}_1^L, \ldots, \tilde{X}_n^L] = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{\tilde{X}_k^L(F) = x(F)\}.$$

To simplify the writing we shall use the short notation

$$\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \quad \text{and} \quad \tilde{\mathbf{X}}^L = (\tilde{X}_1^L, \ldots, \tilde{X}_n^L).$$

Now using either the empirical probability measures $\hat{p}_n(x(F))[\tilde{\mathbf{X}}]$ or $\hat{p}_n(x(F))[\tilde{\mathbf{X}}^L]$ we define the neighborhood estimators $\hat{V}_n^L(i)[\tilde{\mathbf{X}}]$ and $\hat{V}_n^L(i)[\tilde{\mathbf{X}}^L]$.

Now we are ready to conclude the proof. An upperbound for the probability of misidentification for the sample $\tilde{X}_1, \ldots, \tilde{X}_n$ is given by

$$\begin{aligned}
&\mathbb{P}\left(\hat{V}_n^L(i)[\tilde{\mathbf{X}}] \neq V^L(i)\right) \\
&\leq \mathbb{P}\left(\left\{\hat{V}_n^L(i)[\tilde{\mathbf{X}}] \neq V^L(i)\right\} \bigcap \bigcap_{k\in\{1,\ldots,n\}} \bigcap_{j\in B^L(i)} \left\{\tilde{X}_k^L(B^L(j)) = \tilde{X}_k(B^L(j))\right\}\right) \\
&+ \mathbb{P}\left(\bigcup_{k\in\{1,\ldots,n\}} \bigcup_{j\in B^L(i)} \left\{X_k^L(j) \neq X_k(j)\right\}\right).
\end{aligned}$$

By Theorem 4

$$\mathbb{P}\left(\bigcup_{k\in\{1,\ldots,n\}} \bigcup_{j\in B^L(i)} \left\{X_k^L(j) \neq X_k(j)\right\}\right) \leq n(2L)^d \sup\left\{\sum_{j\in\mathbb{Z}^d, |j-t|>L} |J(i,j)| : t \in \mathbb{Z}^d\right\}.$$

Now, we observe that in the set

$$\bigcap_{k\in\{1,\ldots,n\}} \bigcap_{j\in B^L(i)} \left\{\tilde{X}_k^L(B^L(j)) = \tilde{X}_k(B^L(j))\right\}$$

the following holds

$$\hat{V}_n^L(i)[\tilde{\mathbf{X}}] = \hat{V}_n^L(i)[\tilde{\mathbf{X}}^L].$$

Hence

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{V}_n^L(i)[\tilde{\mathbf{X}}] \neq V^L(i)\right\} \bigcap \bigcap_{k\in\{1,\ldots,n\}} \left\{\tilde{X}_k^L(B^L(i)) = \tilde{X}_k(B^L(i))\right\}\right) \\
&= \mathbb{P}\left(\left\{\hat{V}_n^L(i)[\tilde{\mathbf{X}}^L] \neq V^L(i)\right\} \bigcap \bigcap_{k\in\{1,\ldots,n\}} \left\{\tilde{X}_k^L(B^L(i)) = \tilde{X}_k(B^L(i))\right\}\right) \\
&\leq \mathbb{P}\left(\hat{V}_n^L(i)[\tilde{\mathbf{X}}^L] \neq V^L(i)\right). \qquad (23)
\end{aligned}$$

Since Theorem 3 provides an upperbound for the last term in (23), we have

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{V}_n^L(i)[\tilde{\mathbf{X}}] \neq V^L(i)\right\} \bigcap \bigcap_{k\in\{1,\ldots,n\}} \left\{\tilde{X}_k^L(B^L(i)) = \tilde{X}_k(B^L(i))\right\}\right) \\
&\leq 4\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n)} + 2(2L)^d\right).
\end{aligned}$$

This concludes the proof of Theorem 2.

□

To prove Corollary 1 we observe that the finite volume case is equivalent to consider the pairwise potential satisfying

$$J(i,j) = 0 \ \text{ for } \ j \in \mathbb{Z}^d \setminus V \qquad (24)$$

in the $\mathbb{Z}^d$ case.

Proof of Theorem 3

For convenience of the reader, before the proof let us recall the classical inequality of Bernstein which will be used in the sequence.

Bernstein inequality Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables with $|\xi_1| \leq b$ a.s. and $\mathbb{E}[\xi_1^2] \leq v < \infty$. Then the following inequality holds

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} \xi_k - \mathbb{E}[\xi_1]\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2(v + \frac{1}{3}b\epsilon)}\right).$$

For a proof of this inequality, we refer the reader to Massart (2003).

To begin the proof of Theorem 3, let us denote

$$\mathcal{O}_n^L(i) = \left\{ j \in \hat{V}_n^L(i) : j \in B^L(i) \setminus V^L(i) \right\} \tag{25}$$

the event of *false positive identification.*

The event of *false negative identification* is defined as

$$\mathcal{U}_n^L(i) = \left\{ j \in B^L(i) \setminus \hat{V}_n^L(i) : j \in V^L(i) \right\}. \tag{26}$$

We observe that

$$\{\hat{V}_n^L(i) \neq V^L(i)\} = \mathcal{O}_n^L(i) \cup \mathcal{U}_n^L(i).$$

We will first obtain an upperbound for the probability of event false positive identification. Observe that

$$\mathbb{P}\left(\mathcal{O}_n^L(i)\right) \leq \sum_{x(B^L(i))} \sum_{j \in B^L(i) \setminus V^L(i)} \mathbb{P}\left(\hat{D}_n(x(B^L(i)), j) > \epsilon(L, n)\right). \tag{27}$$

Let us fix $j \in B^L(i) \setminus V^L(i)$ and $x(B^L(i)) \in \{-1, +1\}^{B^L(i)}$. To obtain an upperbound for the right hand side of (27) we first observe that

$$\begin{aligned}
&\hat{D}_n(x(B^L(i)), j) \\
&\leq \left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) \hat{p}_n(x(B^L(i) \setminus \{i\}))\right| \\
&+ \left|\frac{\hat{p}_n(x(B^L(i) \setminus \{j\}))}{\hat{p}_n(x(B^L(i) \setminus \{i,j\}))} - p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right)\right| \hat{p}_n(x(B^L(i) \setminus \{i\})).
\end{aligned} \tag{28}$$

This inequality was obtained by adding and subtracting

$$p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) \hat{p}_n(x(B^L(i) \setminus \{i\}))$$

in expression (2).

Since

$$0 \leq \frac{\hat{p}_n(x(B^L(i) \setminus \{i\}))}{\hat{p}_n(x(B^L(i) \setminus \{i,j\}))} \leq 1,$$

we finally obtain the upperbound

$$\begin{aligned}
&\hat{D}_n(x(B^L(i)), j) \\
&\leq \left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) \hat{p}_n(x(B^L(i) \setminus \{i\}))\right| \\
&+ \left|\hat{p}_n(x(B^L(i) \setminus \{j\})) - p\left(x(i)|x(B^L(i) \setminus \{i,j\})\right) \hat{p}_n(x(B^L(i) \setminus \{i,j\}))\right|.
\end{aligned} \tag{29}$$

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\hat{D}_n(x(B^L(i)),j) > \epsilon(L,n)\right) \\
&\le \mathbb{P}\left(\left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)\hat{p}_n(x(B^L(i)\setminus\{i\}))\right| > \frac{1}{2}\epsilon(L,n)\right) \\
&+ \mathbb{P}\left(\left|\hat{p}_n(x(B^L(i)\setminus\{j\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)\hat{p}_n(x(B^L(i)\setminus\{i,j\}))\right| > \frac{1}{2}\epsilon(L,n)\right).
\end{aligned}
$$

The classical Bernstein inequality provides the following upperbounds for the terms in the right hand side of the above equation

$$
\begin{aligned}
&\mathbb{P}\left(\left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)\hat{p}_n(x(B^L(i)\setminus\{i\}))\right| > \frac{1}{2}\epsilon(L,n)\right) \\
&\le 2\exp\left(-\frac{n\epsilon(L,n)^2}{8v_1 + \frac{4}{3}\epsilon(L,n))}\right),
\end{aligned}
\tag{30}
$$

where

$$
v_1 = \sup_{x(B^L(i))}\sup_{j\in B^L(i)}(1 - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)p(x(B^L(i)))). \tag{31}
$$

Also

$$
\begin{aligned}
&\mathbb{P}\left(\left|\hat{p}_n(x(B^L(i)\setminus\{j\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)\hat{p}_n(x(B^L(i)\setminus\{i,j\}))\right| > \frac{1}{2}\epsilon(L,n)\right) \\
&\le 2\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n))}\right),
\end{aligned}
\tag{32}
$$

where

$$
v = \sup_{x(B^L(i))}\sup_{i,j\in B^L(i)}(1 - p\left(x(i)|x(B^L(i)\setminus\{i,j\}\right)p(x(B^L(i)\setminus\{j\}))). \tag{33}
$$

Summing up inequalities (30) and (32) for all configurations $x(B^L(i))$ and all sites $j \in B^L(i)\setminus V^L(i)$ we obtain the following upperbound for the probability of false positive identification

$$
\begin{aligned}
&\mathbb{P}\left(\mathcal{O}_n^L(i)\right) \\
&\le 4(|B^L(i)| - |V^L(i)|)\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n))}\right) \\
&\le 4(|B^L(i)| - |V^L(i)|)\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n))}\right).
\end{aligned}
\tag{34}
$$

We will now obtain an upperbound for the probability of false negative identification. For any $j \in V^L(i)$ we have

$$
\mathbb{P}\left(j \notin \hat{V}_n^L(i)\right) = \mathbb{P}\left(\bigcap_{x(B^L(i))}\left\{\hat{D}_n(x(B^L(i),j)) \le \epsilon(L,n)\right\}\right). \tag{35}
$$

To obtain an upperbound for (35), it is enough to obtain an upperbound for

$$\mathbb{P}\left(\hat{D}_n(x(B^L(i),j)) \le \epsilon(L,n)\right) \tag{36}$$

where $x(B^L(i))$ is any fixed configuration. In particular, we can take a configuration which maximizes

$$\left|p(x(i)|x(B^L(i)\setminus\{i\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\right| p(x(B^L(i)\setminus\{i\}). \tag{37}$$

To do this, we first obtain a lower bound for $\hat{D}_n(x(B^L(i)),j)$ in the same way we obtained the upperbound in (29).

$$\begin{aligned}
&\hat{D}_n(x(B^L(i)),j)\\
&\ge \left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i\}))\right|\\
&- \left|\hat{p}_n(x(B^L(i)\setminus\{j\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i,j\}))\right| \frac{\hat{p}_n(x(B^L(i)\setminus\{i\}))}{\hat{p}_n(x(B^L(i)\setminus\{i,j\}))}.
\end{aligned}$$

Observing again that

$$0 \le \frac{\hat{p}_n(x(B^L(i)\setminus\{i\}))}{\hat{p}_n(x(B^L(i)\setminus\{i,j\}))} \le 1$$

we finally obtain the lower bound

$$\begin{aligned}
&\hat{D}_n(x(B^L(i)),j)\\
&\ge \left|\hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i\}))\right|\\
&- \left|\hat{p}_n(x(B^L(i)\setminus\{j\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i,j\}))\right|.
\end{aligned} \tag{38}$$

To make formulas shorter let us call for the moment

$$W = \hat{p}_n(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i\}))$$

and

$$R = \hat{p}_n(x(B^L(i)\setminus\{j\})) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)\hat{p}_n(x(B^L(i)\setminus\{i,j\})).$$

With this new notation, using inequalities (29) and (38) we obtain

$$\left|\hat{D}_n(x(B^L(i))) - |W|\right| \le |R|. \tag{39}$$

A straightforward computation shows that

$$\mathbb{E}[W] = p(x(B^L(i))) - p\left(x(i)|x(B^L(i)\setminus\{i,j\})\right)p(x(B^L(i)\setminus\{i\})).$$

Assuming that $j \in V^L(i)$ and that configuration $x(B^L(i))$ maximizes (37), we have that

$$|\mathbb{E}[W]| \geq 2\epsilon(L,n).$$

Therefore to bound (36) for $j \in \Gamma_n^L(i)$, it is enough to have an upperbound for

$$\mathbb{P}\left(\left|\hat{D}_n(x(B^L(i)),j) - |\mathbb{E}[W]|\right| \geq \epsilon(L,n)\right).$$

To do this, we observe that

$$\left|\hat{D}_n(x(B^L(i)),j) - |\mathbb{E}[W]|\right| \leq \left|\hat{D}_n(x(B^L(i)),j) - |W|\right| + ||W| - |\mathbb{E}[W]||.$$

Then, using inequality (39) we have

$$\left|\hat{D}_n(x(B^L(i)),j) - |\mathbb{E}[W]|\right| \leq |R| + |W - \mathbb{E}[W]|. \tag{40}$$

Now, by (40)

$$\mathbb{P}\left(\left|\hat{D}_n(x(B^L(i)),j) - |\mathbb{E}[W]|\right| \geq \epsilon(L,n)\right) \tag{41}$$

$$\leq \mathbb{P}\left(|R| \geq \frac{1}{2}\epsilon(L,n)\right) + \mathbb{P}\left(|W - \mathbb{E}[W]| \geq \frac{1}{2}\epsilon(L,n)\right). \tag{42}$$

Note that $\mathbb{E}[R] = 0$, thus by Bernstein inequality

$$\mathbb{P}\left(|R| \geq \frac{1}{2}\epsilon(L,n)\right) \leq 2\exp\left(-\frac{3n\epsilon(L,n)^2}{4(6v + \epsilon(L,n))}\right), \tag{43}$$

where v is the same in (33). By Bernstein inequality also we have

$$\mathbb{P}\left(|W - \mathbb{E}[W]| \geq \frac{1}{2}\epsilon(L,n)\right) \leq 2\exp\left(-\frac{3n\epsilon(L,n)^2}{4(6v_1 + \epsilon(L,n))}\right), \tag{44}$$

where v_1 is the same in (31).

Combining (43) and (44) we have for $j \in V^L(i)$

$$\mathbb{P}\left(j \notin \hat{V}_n^L(i)\right) \leq 4\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n))}\right).$$

From this, it follows that

$$\mathbb{P}\left(\mathcal{U}_n^L(i)\right) \leq 4|V^L(i)|\exp\left(-\frac{n\epsilon(L,n)^2}{8v + \frac{4}{3}\epsilon(L,n))}\right). \tag{45}$$

Adding (35) and (45) we conclude the proof of Theorem (3).

Proof of Theorem 4

Let $z, z' \in \{-1,+1\}^{\mathbb{Z}^d}$ be two fixed configurations. For a fixed $i \in \mathbb{Z}^d$ and an integer $L \geq 1$, let $(Y_t^z, Y_t^{z',L})$ be a discrete time Markov chain taking values on $\{-1,+1\}^{2B^L(i)}$ with the following features.

1. The Ising model on $\{-1,+1\}^{B^L(i)}$ with pairwise potential J and boundary condition $z(B^L(i)^c)$ is reversible with respect to the first marginal Y_t^z.
2. The Ising model on $\{-1,+1\}^{B^L(i)}$ with pairwise potential J^L and boundary condition $z'(B^L(i)^c)$ is reversible with respect to the first marginal $Y_t^{z',L}$.
3. The coupling chain $(Y_t^z, Y_t^{z',L})$ is irreducible and aperiodic, and has an unique invariant probability measure. Taking into the account items (1) and (2), this unique invariant probability measure is a coupling between the Ising models on $\{-1,+1\}^{B^L(i)}$ with interaction potentials J and J^L and boundary conditions $z(B^L(i)^c)$ and $z'(B^L(i)^c)$ respectively.

We now construct $(Y_t^z, Y_t^{z',L})$ with $t \in \mathbb{N}$. This can be done as follows. Let $(I_t)_{t\geq 1}$ be an independent sequence of random variables uniformly distributed on $B^L(i)$. For any $j \in B^L(i)$ and $y \in \{-1,+1\}^{B^L(i)}$, let also the probabilities $p_j(\cdot \mid y)$ and $p_j^L(\cdot \mid y)$ on $\{-1,+1\}$ be defined as follows.

$$p_j(+1 \mid y) = \left\{1 + e^{-2\sum_{k\in B^L(i)} J(j,k)y(k) - 2\sum_{k\notin B^L(i)} J(j,k)z(k)}\right\}^{-1},$$
$$p_j^L(+1 \mid y) = \left\{1 + e^{-2\sum_{k\in B^L(i)} J^L(j,k)y(k) - 2\sum_{k\notin B^L(i)} J^L(j,k)z'(k)}\right\}^{-1}.$$

For any pair $(y, y') \in \{-1,+1\}^{2B^L(i)}$, let $(\xi_t^{j,y,y'})_{t\geq 1}$, be an i.i.d. sequence of random variables taking values on $\{-1,+1\}^2$ with distribution

$$\mathbb{P}\left(\xi_t^{j,y,y'} = (s,s)\right) = \min\left\{p_j(s|y), p_j^L(s|y')\right\},$$
$$\mathbb{P}\left(\xi_t^{j,y,y'} = (s,-s)\right) = \max\left\{p_j(s|y) - p_j^L(s|y'), 0\right\}, \tag{46}$$

for any $s \in \{-1,+1\}$.

Finally, let us assume that the sequences $(I_t)_{t\geq 1}$ and $(\xi_t^{j,y,y'})_{t\geq 1}$, with $(y,y') \in \{-1,+1\}^{2B^L(i)}$ and $j \in B^L(i)$ are all independent. The Markov chain $(Y_t^z, Y_t^{z',L})$ is constructed as follows. For any $t \geq 1$ and any $j \in B^L(i)$

$$(Y_t^z(j), Y_t^{z',L}(j)) = (Y_{t-1}^z(j), Y_{t-1}^{z',L}(j)), \text{ if } j \neq I_t$$

and

$$(Y_t^z(j), Y_t^{z',L}(j)) = \xi_{t-1}^{j, Y_{t-1}^z, Y_{t-1}^{z',L}} \ , \ \text{if } j = I_t \ . \tag{47}$$

We stress the fact that the probabilities $p_j(\cdot|y)$ and $p_j^L(\cdot|y)$ depend on the fixed configurations $z(B^L(i)^c) \in \{-1,+1\}^{B^L(i)^c}$ and $z' \in \{-1,+1\}^{B^L(i)^c}$ respectively. As a consequence the law of the Markov chain $(Y_t^z, Y_t^{z',L})$ depends on the pair of fixed configurations $(z(B^L(i)^c), z'(B^L(i)^c))$. Moreover these probabilities do not depend on the value of $y(j)$; they depend only on the values of $y(k), k \in B^L(i) \setminus j$ and $z(B^L(i)^c)$ and $z(B^L(i)^c)$. Therefore, a more explicit notation should mention all these details. This would produce cumbersome things like $p_j(s|y(B^L(i) \setminus j), z(B^L(i)^c))$, $p_j^L(s|y(B^L(i) \setminus j), z'(B^L(i)^c))$. Hence we decided to use a simplified notation.

Let us assume that the initial value $(Y_0^z, Y_0^{z',L})$ of the chain is chosen according to its unique invariant probability measure. For every integer $t \geq 1$ we have

$$\begin{aligned} \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i)\right) &= \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t \neq i\right) \\ &\quad + \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t = i\right) . \end{aligned} \tag{48}$$

For the first term in the left hand side of the above equation we have

$$\mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t \neq i\right) = \frac{|B^L(i)| - 1}{|B^L(i)|} \mathbb{P}\left(Y_{t-1}^z(i) \neq Y_{t-1}^{z',L}(i)\right) . \tag{49}$$

Substituting (49) in (48), we obtain

$$\frac{1}{|B^L(i)|} \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i)\right) = \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t = i\right) . \tag{50}$$

Now, we have

$$\begin{aligned} &\mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \, , \, I_t = i\right) \\ &= \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t = i \ , \ Y_{t-1}^z(j) = Y_{t-1}^{z',L}(j) \ \text{ for all } \ j \in B^L(i)\right) \\ &+ \mathbb{P}\left(Y_t^z(i) \neq Y_t^{z',L}(i) \ , \ I_t = i \ , \ Y_{t-1}^z(j) \neq Y_{t-1}^{z',L}(j) \text{ for some } j \in B^L(i)\right) . \end{aligned} \tag{51}$$

Using (46) and (47), the first term in the right hand side of (51) is bounded above by

$$2 \sup_{y \in \{-1,1\}^{B^L(i)}} \left(p_j(s|y) - p_j^L(s|y)\right) \mathbb{P}\left(I_t = i \ , \ Y_{t-1}^z(j) = Y_{t-1}^{z',L}(j) \ \text{ for all } \ j \in B^L(i)\right) . \tag{52}$$

Using the Mean Value Theorem, we have

$$\sup_{y \in \{-1,1\}^{B^L(i)}} \left(p_j(s|y) - p_j^L(s|y)\right) \leq \frac{1}{2} \sum_{l \notin B^L(i)} |J(i,l)|. \tag{53}$$

Therefore an upperbound for expression (52) is given by

$$\frac{1}{|B^L(i)|}\sum_{l\notin B^L(i)}|J(i,l)|\mathbb{P}\left(Y^z_{t-1}(j)=Y^{z',L}_{t-1}(j)\ \text{ for all }\ j\in B^L(i)\right)$$
$$=\frac{1}{|B^L(i)|}\sum_{l\notin B^L(i)}|J(i,l)|\left[1-\mathbb{P}\left(Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\text{ for some } j\in B^L(i)\right)\right] \tag{54}$$

Let now study the second term of the right hand side of (51). We first rewrite it as

$$\sum_{\substack{U\subset B^L(i)\\ U\neq\emptyset}}\mathbb{P}\left(Y^z_t(i)\neq Y^{z',L}_t(i)\ ,\ I_t=i\ ,\ \bigcap_{j\in U}\{Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\}\ ,\ \bigcap_{j\in B^L(i)\setminus U}\{Y^z_{t-1}(j)=Y^{z',L}_{t-1}(j)\}\right).$$

Therefore, proceeding as in (52) and (53), we obtain the following upperbound for the second term in the right hand side of (51)

$$\frac{1}{|B^L(i)|}\sum_{\substack{U\subset B^L(i)\\ U\neq\emptyset}}\sum_{l\notin B^L(i)}|J(i,l)|\mathbb{P}\left(\bigcap_{j\in U}\{Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\}\ ,\ \bigcap_{k\in B^L(i)\setminus U}\{Y^z_{t-1}(k)=Y^{z',L}_{t-1}(k)\}\right)$$
$$+\frac{1}{|B^L(i)|}\sum_{\substack{U\subset B^L(i)\\ U\neq\emptyset}}\sum_{l\in U}|J(i,l)|\mathbb{P}\left(\bigcap_{j\in U}\{Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\}\ ,\ \bigcap_{k\in B^L(i)\setminus U}\{Y^z_{t-1}(k)=Y^{z',L}_{t-1}(k)\}\right). \tag{55}$$

The first part of the right hand side of (55) can be rewritten as

$$\frac{1}{|B^L(i)|}\sum_{l\notin B^L(i)}|J(i,l)|\mathbb{P}\left(Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\text{ for some } j\in B^L(i)\right). \tag{56}$$

The second part of (55) can be rewritten as

$$\frac{1}{|B^L(i)|}\sum_{l\in B^L(i)}\sum_{U\subset B^L(i):l\in U}|J(i,l)|\mathbb{P}\left(\bigcap_{j\in U}\{Y^z_{t-1}(j)\neq Y^{z',L}_{t-1}(j)\}\bigcap_{k\in B^L(i)\setminus U}\{Y^z_{t-1}(k)=Y^{z',L}_{t-1}(k)\}\right)$$

and this is equal to

$$\frac{1}{|B^L(i)|}\sum_{l\in B^L(i)}|J(i,l)|\mathbb{P}\left(Y^z_{t-1}(l)\neq Y^{z',L}_{t-1}(l)\right). \tag{57}$$

Collecting together (51), (54), (56), (57), we finally get the upperbound

$$\mathbb{P}\left(Y^z_t(i)\neq Y^{z',L}_t(i)\right)\le\sum_{l\notin B^L(i)}|J(i,l)|+\sum_{l\in B^L(i)}|J(i,l)|\mathbb{P}\left(Y^z_{t-1}(l)\neq Y^{z',L}_{t-1}(l)\right). \tag{58}$$

To conclude the proof of the theorem, let Z and Z' be two independent copies of the Ising models on $\{-1,+1\}^{\mathbb{Z}^d}$ with potentials J and J^L, respectively. For a fixed realization of the pair Z and Z', construct as before the coupled chains $(Y_t^Z, Y_t^{Z',L})$ taking values on $\{-1,+1\}^{2B^L(i)}$, and having $Z(B^L(i)^c)$ and $Z'(B^L(i)^c)$ as boundary conditions.

Using inequality (58) and taking the expectation with respect to (Z, Z'), we have

$$\mathbb{E}\left[\mathbb{P}\left(Y_t^Z(i) \neq Y_t^{L,Z'}(i)\right)\right] \leq \sum_{l \notin B^L(i)} |J(i,l)| + \sum_{l \in B^L(i)} |J(i,l)| \mathbb{E}\left[\mathbb{P}\left(Y_{t-1}^Z(l) \neq Y_{t-1}^{L,Z'}(l)\right)\right]. \tag{59}$$

Now observe that

$$\mathbb{E}\left[\mathbb{P}\left(Y_t^Z(j) \neq Y_t^{L,Z'}(j)\right)\right] = \mathbb{P}\left(Y(j) \neq Y^L(j)\right),$$

for any $j \in B^L(i)$, where $Y(j)$ and $Y^L(j)$ are the projections on site j of realizations of the Ising model with pairwise potential J and J^L, respectively. From this identity and inequality (59), it follows that

$$\mathbb{P}\left(Y(i) \neq Y^L(i)\right) \leq \sum_{l \notin B^L(i)} |J(i,l)| + \sum_{l \in B^L(i)} |J(i,l)| \mathbb{P}\left(Y(l) \neq Y^L(l)\right).$$

Finally, taking the supremum for all $i \in \mathbb{Z}^d$ we have

$$\sup_{i \in \mathbb{Z}^d} \mathbb{P}\left(Y(i) \neq Y^L(i)\right) \leq \sup_{i \in \mathbb{Z}^d} \sum_{l \notin B^L(i)} |J(i,l)| + r \sup_{i \in \mathbb{Z}^d} \mathbb{P}\left(Y(i) \neq Y^L(i)\right),$$

which concludes the proof.

Appendix 2

Pseudocode of the algorithm Sample Generator

Algorithm Sample generator

for $k : 1 \to n$ **do**

  $t = 0$ and $Y_t(i) = +1$ and $Y'_t(i) = -1$ for all $i \in B^L(0)$.

  while $Y_t \neq Y'_t$ **do**

    $t \leftarrow t + 1$

    Generate a random variable I_t uniformly distributed on $B^L(0)$.

    for $j \in B^L(0)$ **do**

      if $j = I_t$ **then**

        Choose $(s, s') \in \{-1, 1\}^2$ with probability

$$P_j\left(s, s'|y, y'\right) = \min\left\{p_j(s|y), p_j(s|y')\right\}, \text{ if } s = s'$$
$$P_j\left(s, s'|y, y'\right) = \max\left\{p_j(s|y) - p_j(s|y'), 0\right\}, \text{ if } s \neq s'$$

        where for any $y \in \{-1, +1\}^{B^L(0)}$ is given by

$$p_j(s \mid y) = \left\{1 + e^{-2s\sum_{k\in B^L(0)} J(j,k)y(k)}\right\}^{-1}.$$

        $Y_t(j) \leftarrow s$ and $Y'_t(j) \leftarrow s'$,

      else

        $Y_t(j) \leftarrow Y_{t-1}(j)$ and $Y'_t(j) \leftarrow Y'_{t-1}(j)$

      end if

    end for

  end while

  $X_k = Y_t$

end for

return Sample $X_1, \ldots, X_n$.

References

J. Bento and A. Montanari. Which graphical models are difficult to learn? *avalilable Markov random field texture modelsat http://arxiv.org/pdf/0910.5761*, 2009.

L. Besag. Statistical analysis of dirty pictures. *Journal of Applied Statistics*, 20(5):63–87, 1993.

G. Bresler, E. Mossel, and A. Sly. *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, chapter Reconstruction of Markov Random Fields from Samples: Some Easy Observations and Algorithms, pages 343–356. Springer, 2008.

G. Cross and A. Jain. Markov random field texture models. *IEEE Trans. PAMI*, 5:25–39, 1983.

I. Csiszar and Z. Talata. Consistent estimation of the basic neighborhood of markov random fields. *Annals of Statistics*, 34:123–145, 2006.

R. Dobrushin. The description of the random field by its conditional distributions and its regularity conditions. *Theory of Probability and its Applications*, 13:197–224, 1968.

P. Massart. *Concentration Inequalities and Model Selection*. Springer, Berlim, 2003.

E. Presutti. *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*. Springer, Berlin, 2009.

P. Ravikumar, M.J. Wainwright, and J.D. Lafferty. High-dimensional ising model selection using ℓ_1-regularized logistic regression. *Annals of Statistics*, 38:1287–1319, 2010.

B.D. Ripley. *Spatial Statistics*. Wiley, New York, 1981.

E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012, 2006.

J. Woods. Markov image modeling. *IEEE Trans. Automat. Control*, 23:846–850, 1978.