

## Identifying interacting pairs of sites in Ising models on a countable set

Antonio Galves<sup>a</sup>, Enza Orlandi<sup>b</sup> and Daniel Y. Takahashi<sup>c</sup>

<sup>a</sup>*Instituto de Matemática e Estatística, Universidade de São Paulo*

<sup>b</sup>*Dipartimento di Matematica, Università di Roma Tre*

<sup>c</sup>*Institute of Neuroscience, Princeton University*

### *Para Errico, com um grande abraço*

**Abstract.** This paper address the problem of identifying pairs of interacting sites from a finite sample of independent realizations of the Ising model. We consider Ising models in a infinite countable set of sites under Dobrushin uniqueness condition. The observed sample contains only the values assigned by the Ising model to a finite set of sites. Our main result is an upperbound for the probability of misidentification of the pairs of interacting sites in this finite set.

## 1 Introduction

In this article, we address the statistical problem of identifying the pairs of interacting sites of an Ising model on a countable set of sites (possibly infinite) when we only observe the values assigned on a finite subset of sites (partially observed). Our sample consists of a finite number of independent realizations of the Ising model observed at this finite set. We introduce a statistical procedure to identify the interacting pair of sites given the observations. Our main result is an upper bound for the probability of misidentifying the pairs of interacting sites.

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, 1986), 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).

When the set of sites is not finite, the Ising model is supported by a continuous set of infinite configurations. 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. 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 difference between the conditional probabilities are small, we conclude that interacting weight between the sites is null.

The proof of the main theorem has two ingredients, which are interesting by themselves. The first ingredient is an upperbound for the probability of misidentification for the Ising model on a finite set of sites. This is the content of Theorem 2. The second ingredient is a coupling result given in Theorem 3. It says that we can couple together an Ising model in which each site directly interacts with possibly an infinite number of sites and an Ising model in which each site interacts only with a finite subset of sites, in such a way that the probability of a discrepancy at a fixed site vanishes as the set of interacting sites increases. 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 only a finite set of sites, not the entire set of interacting sites. The proof of this result uses a constructive version of classical Dobrushin's contraction method. For a nice presentation of the contraction method in its original framework, we refer the reader to [Presutti \(2009\)](#).

It is important to note that we do not need to assign a metric to the set of sites to state and prove our results. It turns out that in several applied contexts, a pre-defined metric is unwarranted. For instance, for the problem of inferring the presence/absence of interactions between pairs of neurons, it is not clear a priori that there is any consistent relationship between the strength of the interaction and the physical distance (or any other metric) between the neurons. Moreover, in several situations, the experimenter does not know if the recorded neurons are physically close or far. This justifies the approach taken here.

Let conclude this section with short comments on the recent related literature. The case of random field on a finite set of sites, which is entirely observed was considered in [Ravikumar et al. \(2010\)](#), [Bento and Montanari \(2009\)](#), [Bresler et al. \(2008\)](#). The infinite case was also considered in [Löcherbach and Orlandi \(2011\)](#), [Csiszar and Talata \(2006\)](#), [Lerasle and Takahashi \(2015, 2011\)](#). The first two use a BIC like approach for an homogeneous random field using a single observation and the last two use an oracle approach, which solves a problem that is different from the identification problem considered in this article.

This paper is organized as follows. Notation, definitions and main results are presented in Section 2. The proofs of the theorems are presented in Section 3.

## 2 Notation, definitions, and main results

Let  $S$  be a countable set of sites.

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

$$J(i, i) = 0, \quad J(i, j) = J(j, i), \quad \sup_{i \in S} \sum_{j \in S} |J(i, j)| < \infty. \quad (1)$$

Let  $\mathcal{X} = \{-1, 1\}^S$  be the set of configurations on the set of sites  $S$ . A fixed configurations will be denoted by lower case letter  $x$  whereas the capital letter  $X$  will denote a random configuration taking values on  $\mathcal{X}$  and probability measure  $\mathbb{P}$ . For any  $i \in S$ ,  $x(i)$  will denote the value of the configuration  $x$  at site  $i$ . Given a subset  $F$  of  $S$ , 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 configuration  $X$  with values on  $\mathcal{X}$ , which probability satisfies

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

for all  $i \in S$  and for  $\mathbb{P}$ -a.e.  $x \in \mathcal{X}$ .

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

Let  $F$  be a finite subset of  $S$ . We 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  $\mathbb{P}(X(i) = x(i) | X(F) = x(F))$ .

**Definition 3.** For any site  $i \in S$ , the interaction neighborhood  $\mathcal{G}(i)$  is defined as

$$\mathcal{G}(i) = \{j \in S : J(i, j) \neq 0\}.$$

In general, we cannot observe the entire random configuration on  $S$ , but only the values on some finite subset  $F \subset S$ . Moreover, we observe only a finite number of samples, *that is*, the observations are i.i.d. samples  $X_1(F), \dots, X_n(F)$ . In this situation, we do not expect to recover  $\mathcal{G}(i)$ , but we might be able to identify  $\mathcal{G}(i) \cap F$ . In this article, we show how we can do it. The following family of sets will be useful for the rest of the article.

**Definition 4.** A family  $\mathcal{F}$  of finite subsets  $F_i \subset S$  indexed by  $i \in S$  is called a truncation class if for any  $i, j \in S$  we have that (a)  $i \in F_i$  and (b)  $j \in F_i \iff i \in F_j$ .

Using a truncation class, we can introduce the following truncated version of the Ising model.

**Definition 5.** Given a truncation class  $\mathcal{F}$ , we denote by  $J^{\mathcal{F}}$  the truncated potential defined as follows

$$J^{\mathcal{F}}(i, j) = \begin{cases} J(i, j), & \text{if } j \in F_i, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We also denote by  $X^{\mathcal{F}}$  the corresponding Ising model with pairwise potential  $J^{\mathcal{F}}$ .

From now on,  $\mathcal{F}$  will always denote a truncation class and  $F_i$  will be an element of  $\mathcal{F}$  with index  $i$ . Let  $F$  be a finite subset of  $S$ . As before, we use the shorthand notation  $p^{\mathcal{F}}(x(F))$  and  $p^{\mathcal{F}}(x(i)|x(F))$  to denote, respectively, the probability  $\mathbb{P}(X^{\mathcal{F}}(F) = x(F))$  and the conditional probability  $\mathbb{P}(X^{\mathcal{F}}(i) = x(i)|X^{\mathcal{F}}(F) = x(F))$ .

Given a site  $i \in S$ , a set  $F_i \in \mathcal{F}$ , and  $x \in \mathcal{X}$ , let

$$D(x, F_i, i, j) = |p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))|p^{\mathcal{F}}(x(F_i \setminus \{i\})).$$

**Definition 6.** Given a truncation class  $\mathcal{F}$ , and a positive  $\varepsilon$ , for each  $i \in S$ , the interaction neighborhood  $V^{\mathcal{F}}(i)$  is defined as

$$V^{\mathcal{F}}(i) = \left\{ j \in F_i : \max_{x(F_i)} D(x, F_i, i, j) > 2\varepsilon \right\}, \tag{3}$$

where  $F_i \in \mathcal{F}$ .

We prove the following lowerbound.

**Proposition 1.** Let  $\sup_{i \in S} \sum_{j \in S} |J(i, j)| = \gamma$ . Given a truncation class  $\mathcal{F}$ , for each  $i \in S$  and  $j \in F_i$ , where  $F_i \in \mathcal{F}$ , we have that

$$\max_{x(F_i)} D(x, F_i, i, j) \geq \frac{2^{-|F_i|+2}e^{2\gamma}}{(1 + e^{2\gamma})} |J(i, j)|. \tag{4}$$

Observe that if  $j \in \mathcal{G}(i) \cap F_i$ , we have  $|J(i, j)| > 0$ , therefore for small enough  $\varepsilon$ , we will have  $V^{\mathcal{F}}(i) = \mathcal{G}(i) \cap F_i$ .

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  $j \in S$ , 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  $i \in S$ ,  $F_i \in \mathcal{F}$ , any configuration  $x(F_i)$ , and  $j \in F_i$  we define the empirical weighted distance between the conditional probabilities as follows

$$\begin{aligned} \hat{D}_n(x, F_i, i, j) &= |\hat{p}_n(x(i)|x(F_i \setminus \{i\})) - \hat{p}_n(x(i)|x(F_i \setminus \{i, j\}))| \hat{p}_n(x(F_i \setminus \{i\})). \end{aligned} \tag{5}$$

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

We can now define our estimator.

**Definition 7.** Given a truncation class  $\mathcal{F}$ , for each  $i \in S$ , the interaction neighborhood estimator is defined as

$$\hat{V}_n(i) = \left\{ j \in F_i : \max_{x(F_i)} \hat{D}_n(x, F_i, i, j) > \varepsilon \right\},$$

where  $F_i \in \mathcal{F}$  and the threshold  $\varepsilon$  is the same as in Definition 6.

We can now state our main result.

**Theorem 1.** Given a truncation class  $\mathcal{F}$  and  $i \in S$ , let  $X_1(F_i), \dots, X_n(F_i)$ , where  $F_i \in \mathcal{F}$ , be the local projections of independent realizations of an Ising model whose pairwise potential satisfies

$$\sup_{k \in S} \sum_{j \in S} |J(k, j)| = r < 1. \tag{6}$$

Then, for any threshold value  $\varepsilon > 0$ , we have

$$\begin{aligned} &\mathbb{P}(\hat{V}_n(i) \neq V^{\mathcal{F}}(i)) \\ &\leq 4 \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon} + 2|F_i|\right) + \frac{1}{1-r} n|F_i| \left(\sup_{k \in S} \sum_{j \in S \setminus F_k} |J(k, j)|\right), \end{aligned} \tag{7}$$

where

$$v = \sup_{x(F_i)} \sup_{j \in F_i} (1 - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))) p^{\mathcal{F}}(x(F_i)). \tag{8}$$

The condition (6) is known as Dobrushin uniqueness condition in the statistical physics literature (Presutti, 2009).

The first ingredient in the proof of Theorem 1 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 2.** Given a truncation class  $\mathcal{F}$ , and  $i \in S$ , let  $X_1^{\mathcal{F}}(F_i), \dots, X_n^{\mathcal{F}}(F_i)$ , where  $F_i \in \mathcal{F}$ , be the projections of independent realizations of an Ising model with pairwise potential  $J^{\mathcal{F}}$ .

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

$$\mathbb{P}(\hat{V}_n(i) \neq V^{\mathcal{F}}(i)) \leq 4 \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon} + 2|F_i|\right),$$

where

$$v = \sup_{x(F_i)} \sup_{j \in F_i} (1 - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))) p^{\mathcal{F}}(x(F_i)).$$

The second ingredient in the proof of Theorem 1 is a coupling result. To state it, we first need to introduce the definition of coupling.

**Definition 8.** Given a truncation class  $\mathcal{F}$ , let  $X$  and  $X^{\mathcal{F}}$  be Ising models with pairwise potentials  $J$  and  $J^{\mathcal{F}}$ , respectively. A coupling between  $X$  and  $X^{\mathcal{F}}$  is a random element  $(\tilde{X}, \tilde{X}^{\mathcal{F}})$  taking values on  $S \times S$  such that:

1.  $\tilde{X}$  has the same law as  $X$ ;
2.  $\tilde{X}^{\mathcal{F}}$  has the same law as  $X^{\mathcal{F}}$ .

The following theorem says that we can sample together  $X$  and  $X^{\mathcal{F}}$  and gives an upper bound for the probability of discrepancy between  $X(i)$  and  $X^{\mathcal{F}}(i)$ .

**Theorem 3.** If  $J$  is pairwise potential which satisfies condition (6), that is,

$$\sup_{k \in S} \sum_{j \in S} |J(k, j)| = r < 1 \tag{9}$$

and  $J^{\mathcal{F}}$  is defined as in (2), then there exists a coupling  $(\tilde{X}, \tilde{X}^{\mathcal{F}})$  such that for any  $i \in S$  the following inequality holds

$$\mathbb{P}(\tilde{X}(i) \neq \tilde{X}^{\mathcal{F}}(i)) \leq \frac{1}{1-r} \sup_{k \in S} \sum_{j \in S \setminus F_k} |J(k, j)|. \tag{10}$$

**Remark.** To exemplify the application of the above theorem, let  $d \geq 1$ , and  $X$  be an Ising model with  $S = \mathbb{Z}^d$  and potential  $J$  such that  $|J(k, j)| < C\|k - j\|^{-2d}$ , where  $\|\cdot\|$  is the  $\ell^\infty$  norm and  $C$  is sufficiently small so that the potential satisfies (9). For each  $k \in \mathbb{Z}^d$  and positive integer  $n$ , we define  $F_k(n) = \{j \in \mathbb{Z}^d : \|k - j\| \leq n\}$  and  $\mathcal{F}(n) = \{F_k(n) \subset \mathbb{Z}^d : k \in \mathbb{Z}^d\}$ . Using Theorem 3, it is clear that for sufficiently large  $n$ ,  $X^{\mathcal{F}(n)}$  will be a good approximation to  $X$ . In practice, the exact metric structure is rarely known to the experimenter and the choice of  $\mathcal{F}$  is usually given by experimental constraints. For example, in neuroscience, the size of the sets in  $\mathcal{F}$  can be decided by the number of electrodes that the experimenter can measure simultaneously in the brain. Thus Theorem 3 gives us a simple guideline, showing that if the size of the observed sites is limited, as it is always the case in practice, the experimenter, if possible, has to choose the sites with strongest interaction, that is, largest  $|J(i, j)|$  to obtain a good approximation.

### 3 Proofs of the results

#### Proof of Proposition 1

Let  $i, j \in F_i$ ,  $x(F_i \setminus \{i\}) \in \{-1, +1\}^{|F_i|-1}$ , and  $y(F_i \setminus \{i\}) \in \{-1, +1\}^{|F_i|-1}$  with  $y(j) = -x(j)$ . Using the mean value theorem, we have

$$|p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i\})) - p^{\mathcal{F}}(x(i)|y(F_i \setminus \{i\}))| \geq \frac{e^{2\gamma}}{(1 + e^{2\gamma})^2} |J(i, j)|.$$

Hence, for any  $j \in F_i$  such that  $J(i, j) \neq 0$  we have

$$\begin{aligned} & |p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))| \\ & \geq \frac{2e^{2\gamma}}{(1 + e^{2\gamma})^2} |J(i, j)| \min_{x(j) \in \{-1, +1\}} p^{\mathcal{F}}(x(j)|x(F_i \setminus \{i, j\})). \end{aligned}$$

Also

$$\min_{x(j) \in \{-1, +1\}} p^{\mathcal{F}}(x(j)|x(F_i \setminus \{i, j\})) \geq \frac{1}{1 + e^{2\gamma}}.$$

We observe that for any  $i \in F$

$$\max_{x(F_i \setminus \{i\})} p^{\mathcal{F}}(x(F_i \setminus \{i\})) \geq 2^{-|F_i|+1}.$$

Combining the above inequalities, we have

$$\max_{x(F_i)} D(x, F_i, i, j) \geq \frac{2^{-|F_i|+2} e^{2\gamma}}{(1 + e^{2\gamma})^3} |J(i, j)|,$$

as we wanted to show.

#### Proof of Theorem 1

Let  $\mathcal{F}$  be a truncation class. Let  $(\tilde{X}_1, \tilde{X}_1^{\mathcal{F}}), \dots, (\tilde{X}_n, \tilde{X}_n^{\mathcal{F}})$  be  $n$  independent copies of the pair  $(\tilde{X}, \tilde{X}^{\mathcal{F}})$  which existence is guaranteed by Theorem 3. The random elements  $\tilde{X}_1, \dots, \tilde{X}_n$  are independent copies of the Ising model  $X$  with pairwise potential  $J$ . The random elements  $\tilde{X}_1^{\mathcal{F}}, \dots, \tilde{X}_n^{\mathcal{F}}$  are independent copies of the Ising model  $X^{\mathcal{F}}$  with truncated pairwise potential  $J^{\mathcal{F}}$  defined as in (2).

Let us indicate explicitly the sample in all the statistics and events appearing in Theorem 1 as functions either of the sample  $\tilde{X}_1, \dots, \tilde{X}_n$  or of the sample  $\tilde{X}_1^{\mathcal{F}}, \dots, \tilde{X}_n^{\mathcal{F}}$ . We start with notation of the empirical probability measures  $\hat{p}_n$ , as follows

$$\begin{aligned} \hat{p}_n(x(F_i))[\tilde{X}_1, \dots, \tilde{X}_n] &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}\{\tilde{X}_k(F_i) = x(F_i)\}, \\ \hat{p}_n(x(F_i))[\tilde{X}_1^{\mathcal{F}}, \dots, \tilde{X}_n^{\mathcal{F}}] &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}\{\tilde{X}_k^{\mathcal{F}}(F_i) = x(F_i)\}. \end{aligned}$$

To simplify the writing, we shall use the short notation

$$\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \quad \text{and} \quad \tilde{\mathbf{X}}^{\mathcal{F}} = (\tilde{X}_1^{\mathcal{F}}, \dots, \tilde{X}_n^{\mathcal{F}}).$$

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

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

$$\begin{aligned} & \mathbb{P}(\hat{V}_n(i)[\tilde{\mathbf{X}}] \neq V^{\mathcal{F}}(i)) \\ & \leq \mathbb{P}\left(\{\hat{V}_n(i)[\tilde{\mathbf{X}}] \neq V^{\mathcal{F}}(i)\} \cap \bigcap_{k \in \{1, \dots, n\}} \bigcap_{j \in F_i} \{\tilde{X}_k^{\mathcal{F}}(j) = \tilde{X}_k(j)\}\right) \\ & \quad + \mathbb{P}\left(\bigcup_{k \in \{1, \dots, n\}} \bigcup_{j \in F_i} \{\tilde{X}_k^{\mathcal{F}}(j) \neq \tilde{X}_k(j)\}\right). \end{aligned}$$

By Theorem 3

$$\mathbb{P}\left(\bigcup_{k \in \{1, \dots, n\}} \bigcup_{j \in F_i} \{\tilde{X}_k^{\mathcal{F}}(j) \neq \tilde{X}_k(j)\}\right) \leq n|F_i| \sup_{k \in \mathcal{S}} \sum_{j \in \mathcal{S} \setminus F_k} |J(k, j)|.$$

Now, we observe that in the set

$$\bigcap_{k \in \{1, \dots, n\}} \bigcap_{j \in F} \{\tilde{X}_k^{\mathcal{F}}(j) = \tilde{X}_k(j)\}$$

the following holds

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

Hence,

$$\begin{aligned} & \mathbb{P}\left(\{\hat{V}_n(i)[\tilde{\mathbf{X}}] \neq V^{\mathcal{F}}(i)\} \cap \bigcap_{k \in \{1, \dots, n\}} \{\tilde{X}_k^{\mathcal{F}}(F_i) = \tilde{X}_k(F_i)\}\right) \\ & = \mathbb{P}\left(\{\hat{V}_n(i)[\tilde{\mathbf{X}}^{\mathcal{F}}] \neq V^{\mathcal{F}}(i)\} \cap \bigcap_{k \in \{1, \dots, n\}} \{\tilde{X}_k^{\mathcal{F}}(F_i) = \tilde{X}_k(F_i)\}\right) \quad (11) \\ & \leq \mathbb{P}(\hat{V}_n(i)[\tilde{\mathbf{X}}^{\mathcal{F}}] \neq V^{\mathcal{F}}(i)). \end{aligned}$$

Since Theorem 2 provides an upperbound for the last term in (11), we have

$$\begin{aligned} & \mathbb{P}\left(\{\hat{V}_n(i)[\tilde{\mathbf{X}}] \neq V^{\mathcal{F}}(i)\} \cap \bigcap_{k \in \{1, \dots, n\}} \{\tilde{X}_k^{\mathcal{F}}(F_i) = \tilde{X}_k(F_i)\}\right) \\ & \leq 4 \exp\left(-\frac{n\varepsilon(F_i, n)^2}{8v + (4/3)\varepsilon(F_i, n)} + 2|F_i|\right). \end{aligned}$$

This concludes the proof of Theorem 1.



**Proof of Theorem 2**

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  $\xi_1, \dots, \xi_n$  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

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

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

To begin the proof of Theorem 2, let us denote

$$\mathcal{O}_n^{\mathcal{F}}(i) = \{j \in \hat{V}_n^{\mathcal{F}}(i) : j \in F_i \setminus V^{\mathcal{F}}(i)\} \tag{12}$$

the event of *false positive identification*.

The event of *false negative identification* is defined as

$$\mathcal{U}_n^{\mathcal{F}}(i) = \{j \in F_i \setminus \hat{V}_n(i) : j \in V^{\mathcal{F}}(i)\}. \tag{13}$$

We observe that

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

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

$$\mathbb{P}(\mathcal{O}_n^{\mathcal{F}}(i)) \leq \sum_{x(F_i)} \sum_{j \in F_i \setminus V^{\mathcal{F}}(i)} \mathbb{P}(\hat{D}_n(x, F_i, i, j) > \varepsilon). \tag{14}$$

Let us fix  $j \in F_i \setminus V^{\mathcal{F}}(i)$  and  $x(F_i) \in \{-1, +1\}^{F_i}$ . To obtain an upperbound for the right-hand side of (14), we first observe that

$$\begin{aligned} & \hat{D}_n(x, F_i, i, j) \\ & \leq \left| \hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\})) \hat{p}_n(x(F_i \setminus \{i\})) \right| \\ & \quad + \left| \frac{\hat{p}_n(x(F_i \setminus \{j\}))}{\hat{p}_n(x(F_i \setminus \{i, j\}))} - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\})) \right| \hat{p}_n(x(F_i \setminus \{i\})). \end{aligned} \tag{15}$$

This inequality was obtained by adding and subtracting

$$p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\})) \hat{p}_n(x(F_i \setminus \{i\}))$$

in expression (5).

Since

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

we finally obtain the upperbound

$$\begin{aligned} \hat{D}_n(x, F_i, i, j) &\leq |\hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))| \\ &\quad + |\hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\}))|. \end{aligned} \tag{16}$$

Therefore,

$$\begin{aligned} \mathbb{P}(\hat{D}_n(x, F_i, i, j) > \varepsilon) &\leq \mathbb{P}\left(|\hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))| > \frac{1}{2}\varepsilon\right) \\ &\quad + \mathbb{P}\left(|\hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\}))| > \frac{1}{2}\varepsilon\right). \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}\left(|\hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))| > \frac{1}{2}\varepsilon\right) \\ \leq 2 \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon}\right), \end{aligned} \tag{17}$$

where

$$v = \sup_{x(F_i)} \sup_{j \in F_i} (1 - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))p^{\mathcal{F}}(x(F_i))). \tag{18}$$

Also

$$\begin{aligned} \mathbb{P}\left(|\hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\}))| > \frac{1}{2}\varepsilon\right) \\ \leq 2 \exp\left(-\frac{n\varepsilon^2}{8v' + (4/3)\varepsilon}\right), \end{aligned} \tag{19}$$

where

$$v' = \sup_{x(F_i)} \sup_{j \in F_i} (1 - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))p^{\mathcal{F}}(x(F_i \setminus \{j\}))). \tag{20}$$

Summing up inequalities (17) and (19) for all configurations  $x(F_i)$  and all sites  $j \in F_i \setminus V^{\mathcal{F}}(i)$ , we obtain the following upperbound for the probability of false positive identification

$$\begin{aligned} \mathbb{P}(\mathcal{O}_n^{\mathcal{F}}(i)) &\leq 4(|F_i| - |V^{\mathcal{F}}(i)|) \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon}\right) \\ &\leq 4(|F_i| - |V^{\mathcal{F}}(i)|) \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon}\right). \end{aligned} \tag{21}$$

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

$$\mathbb{P}(j \notin \hat{V}_n(i)) = \mathbb{P}\left(\bigcap_{x(F_i)} \{\hat{D}_n(x, F_i, i, j) \leq \varepsilon\}\right). \tag{22}$$

To obtain an upperbound for (22), it is enough to obtain an upperbound for

$$\mathbb{P}(\hat{D}_n(x, F_i, i, j) \leq \varepsilon), \tag{23}$$

where  $x(F_i)$  is any fixed configuration. In particular, we can take a configuration which maximizes

$$|p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))|p(x(F_i \setminus \{i\})). \tag{24}$$

To do this, we first obtain a lower bound for  $\hat{D}_n(x, F_i, i, j)$  in the same way we obtained the upperbound in (16).

$$\begin{aligned} &\hat{D}_n(x, F_i, i, j) \\ &\geq |\hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))| \\ &\quad - |\hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\}))| \\ &\quad \times \frac{\hat{p}_n(x(F_i \setminus \{i\}))}{\hat{p}_n(x(F_i \setminus \{i, j\}))}. \end{aligned}$$

Observing again that

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

we finally obtain the lower bound

$$\begin{aligned} &\hat{D}_n(x, F_i, i, j) \\ &\geq |\hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))| \\ &\quad - |\hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\}))|. \end{aligned} \tag{25}$$

To make formulas shorter, let us call for the moment

$$W = \hat{p}_n(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i\}))$$

and

$$R = \hat{p}_n(x(F_i \setminus \{j\})) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))\hat{p}_n(x(F_i \setminus \{i, j\})).$$

With this new notation, using inequalities (16) and (25) we obtain

$$|\hat{D}_n(x, F_i, i, j) - |W|| \leq |R|. \tag{26}$$

A straightforward computation shows that

$$\mathbb{E}[W] = p(x(F_i)) - p^{\mathcal{F}}(x(i)|x(F_i \setminus \{i, j\}))p^{\mathcal{F}}(x(F_i \setminus \{i\})).$$

Assuming that  $j \in V^{\mathcal{F}}(i)$  and that configuration  $x(F)$  maximizes (24), we have that

$$|\mathbb{E}[W]| \geq 2\varepsilon.$$

Therefore to bound (23) for  $j \in V^{\mathcal{F}}(i)$ , it is enough to have an upperbound for

$$\mathbb{P}(|\hat{D}_n(x, F_i, i, j) - |\mathbb{E}[W]|| \geq \varepsilon).$$

To do this, we observe that

$$|\hat{D}_n(x, F_i, i, j) - |\mathbb{E}[W]|| \leq |\hat{D}_n(x, F_i, i, j) - |W|| + ||W| - |\mathbb{E}[W]||.$$

Then, using inequality (26) we have

$$|\hat{D}_n(x, F_i, i, j) - |\mathbb{E}[W]|| \leq |R| + |W - \mathbb{E}[W]|. \tag{27}$$

Now, by (27)

$$\mathbb{P}(|\hat{D}_n(x, F_i, i, j) - |\mathbb{E}[W]|| \geq \varepsilon) \tag{28}$$

$$\leq \mathbb{P}\left(|R| \geq \frac{1}{2}\varepsilon\right) + \mathbb{P}\left(|W - \mathbb{E}[W]| \geq \frac{1}{2}\varepsilon\right). \tag{29}$$

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

$$\mathbb{P}\left(|R| \geq \frac{1}{2}\varepsilon\right) \leq 2 \exp\left(-\frac{3n\varepsilon^2}{4(6v + \varepsilon)}\right), \tag{30}$$

where  $v$  is the same in (20). By Bernstein inequality also we have

$$\mathbb{P}\left(|W - \mathbb{E}[W]| \geq \frac{1}{2}\varepsilon\right) \leq 2 \exp\left(-\frac{3n\varepsilon^2}{4(6v_1 + \varepsilon)}\right), \tag{31}$$

where  $v_1$  is the same in (18).

Combining (30) and (31), we have for  $j \in V^{\mathcal{F}}(i)$

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

From this, it follows that

$$\mathbb{P}(\mathcal{U}_n^{\mathcal{F}}(i)) \leq 4|V^{\mathcal{F}}(i)| \exp\left(-\frac{n\varepsilon^2}{8v + (4/3)\varepsilon}\right). \tag{32}$$

Adding (22) and (32) we conclude the proof of Theorem 2.

**Proof of Theorem 3**

Let  $z, z' \in \{-1, +1\}^S$  be two fixed configurations. For  $i \in S$ ,  $F_i \in \mathcal{F}$ , let  $(Y_t^z, Y_t^{z', \mathcal{F}})$  be a discrete time Markov chain taking values on  $\{-1, +1\}^{F_i}$  with the following features:

- (1) The Ising model on  $\{-1, +1\}^{F_i}$  with pairwise potential  $J$  and boundary condition  $z(F_i^c)$  is reversible with respect to the first marginal  $Y_t^z$ .
- (2) The Ising model on  $\{-1, +1\}^{F_i}$  with pairwise potential  $J^{\mathcal{F}}$  and boundary condition  $z'(F_i^c)$  is reversible with respect to the second marginal  $Y_t^{z', \mathcal{F}}$ .
- (3) The coupling chain  $(Y_t^z, Y_t^{z', \mathcal{F}})$  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  $\{-1, +1\}^{F_i}$  with interaction potentials  $J$  and  $J^{\mathcal{F}}$  and boundary conditions  $z(F_i^c)$  and  $z'(F_i^c)$ , respectively.

We now construct  $(Y_t^z, Y_t^{z', \mathcal{F}})$  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  $F_i$ . For any  $j \in F_i$  and  $y \in \{-1, +1\}^{F_i}$ , let also the probabilities  $p_j(\cdot|y)$  and  $p_j^{\mathcal{F}}(\cdot|y)$  on  $\{-1, +1\}$  be defined as follows

$$p_j(+1|y) = \{1 + e^{-2 \sum_{k \in F_j} J(j,k)y(k) - 2 \sum_{k \notin F_j} J(j,k)z(k)}\}^{-1},$$

$$p_j^{\mathcal{F}}(+1|y) = \{1 + e^{-2 \sum_{k \in F_j} J^{\mathcal{F}}(j,k)y(k) - 2 \sum_{k \notin F_j} J^{\mathcal{F}}(j,k)z'(k)}\}^{-1}.$$

For any pair  $(y, y') \in \{-1, +1\}^{F_i} \times \{-1, +1\}^{F_i}$ , 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

$$\mathbb{P}(\xi_t^{j,y,y'} = (s, s)) = \min\{p_j(s|y), p_j^{\mathcal{F}}(s|y')\},$$

$$\mathbb{P}(\xi_t^{j,y,y'} = (s, -s)) = \max\{p_j(s|y) - p_j^{\mathcal{F}}(s|y'), 0\},$$
(33)

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  $(y, y') \in \{-1, +1\}^{2F_i}$  and  $j \in F_i$  are all independent. The Markov chain  $(Y_t^z, Y_t^{z', \mathcal{F}})$  is constructed as follows. For any  $t \geq 1$  and any  $j \in F_i$

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

and

$$(Y_t^z(j), Y_t^{z', \mathcal{F}}(j)) = \xi_{t-1}^{j, Y_{t-1}^z, Y_{t-1}^{z', \mathcal{F}}} \quad \text{if } j = I_t. \tag{34}$$

We stress the fact that the probabilities  $p_j(\cdot|y)$  and  $p_j^{\mathcal{F}}(\cdot|y)$  depend on the fixed configurations  $z(F_i^c) \in \{-1, +1\}^{F_i^c}$  and  $z'(F_i^c) \in \{-1, +1\}^{F_i^c}$ , respectively.

As a consequence, the law of the Markov chain  $(Y_t^z, Y_t^{z', \mathcal{F}})$  depends on the pair of fixed configurations  $(z(F_i^c), z'(F_i^c))$ . Therefore, a more explicit notation should mention all these details. This would produce cumbersome things like  $p_j(s|y(F_i \setminus j), z(F_i^c)), p_j^{\mathcal{F}}(s|y(F_i \setminus j), z'(F_i^c))$ . Hence, we decided to use a simplified notation  $p_j(\cdot|y)$  and  $p_j^{\mathcal{F}}(\cdot|y)$ , respectively.

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

$$\begin{aligned} \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i)) &= \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t \neq i) \\ &\quad + \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i). \end{aligned} \tag{35}$$

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

$$\mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t \neq i) = \frac{|F_i| - 1}{|F_i|} \mathbb{P}(Y_{t-1}^z(i) \neq Y_{t-1}^{z', \mathcal{F}}(i)). \tag{36}$$

Substituting (36) in (35), and using the fact that the Markov chain is stationary, we obtain

$$\frac{1}{|F_i|} \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i)) = \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i). \tag{37}$$

Now, we have

$$\begin{aligned} &\mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i) \\ &= \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i, Y_{t-1}^z(j) = Y_{t-1}^{z', \mathcal{F}}(j) \text{ for all } j \in F_i) \\ &\quad + \mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i, Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j) \text{ for some } j \in F_i). \end{aligned} \tag{38}$$

Using (33) and (34), the first term in the right-hand side of (38) is bounded above by

$$2 \sup_{y \in \{-1, 1\}^{F_i}} (p_i(s|y) - p_i^{\mathcal{F}}(s|y)) \mathbb{P}(I_t = i, Y_{t-1}^z(j) = Y_{t-1}^{z', \mathcal{F}}(j) \text{ for all } j \in F_i). \tag{39}$$

Using the mean value theorem, we have

$$\sup_{y \in \{-1, 1\}^{F_i}} (p_i(s|y) - p_i^{\mathcal{F}}(s|y)) \leq \frac{1}{2} \sum_{l \notin F_i} |J(i, l)|. \tag{40}$$

Therefore an upperbound for expression (39) is given by

$$\begin{aligned} &\frac{1}{|F_i|} \sum_{l \notin F_i} |J(i, l)| \mathbb{P}(Y_{t-1}^z(j) = Y_{t-1}^{z', \mathcal{F}}(j) \text{ for all } j \in F_i) \\ &= \frac{1}{|F_i|} \sum_{l \notin F_i} |J(i, l)| [1 - \mathbb{P}(Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j) \text{ for some } j \in F_i)]. \end{aligned} \tag{41}$$

Let now study the second term of the right-hand side of (38). We first rewrite it as

$$\sum_{\substack{U \subset F_i \\ U \neq \emptyset}} \mathbb{P} \left( Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i), I_t = i, \bigcap_{j \in U} \{Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j)\}, \right. \\ \left. \bigcap_{j \in F_i \setminus U} \{Y_{t-1}^z(j) = Y_{t-1}^{z', \mathcal{F}}(j)\} \right).$$

Therefore, proceeding as in (39) and (40), we obtain the following upperbound for the second term in the right-hand side of (38)

$$\frac{1}{|F_i|} \sum_{\substack{U \subset F_i \\ U \neq \emptyset}} \sum_{l \notin F_i} |J(i, l)| \mathbb{P} \left( \bigcap_{j \in U} \{Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j)\}, \bigcap_{k \in F_i \setminus U} \{Y_{t-1}^z(k) = Y_{t-1}^{z', \mathcal{F}}(k)\} \right) \\ + \frac{1}{|F_i|} \sum_{\substack{U \subset F_i \\ U \neq \emptyset}} \sum_{l \in U} |J(i, l)| \mathbb{P} \left( \bigcap_{j \in U} \{Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j)\}, \right. \\ \left. \bigcap_{k \in F_i \setminus U} \{Y_{t-1}^z(k) = Y_{t-1}^{z', \mathcal{F}}(k)\} \right). \tag{42}$$

The first part of (42) can be rewritten as

$$\frac{1}{|F_i|} \sum_{l \notin F_i} |J(i, l)| \mathbb{P}(Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j) \text{ for some } j \in F_i). \tag{43}$$

The second part of (42) can be rewritten as

$$\frac{1}{|F_i|} \sum_{l \in F_i} \sum_{U \subset F_i: l \in U} |J(i, l)| \mathbb{P} \left( \bigcap_{j \in U} \{Y_{t-1}^z(j) \neq Y_{t-1}^{z', \mathcal{F}}(j)\}, \right. \\ \left. \bigcap_{k \in F_i \setminus U} \{Y_{t-1}^z(k) = Y_{t-1}^{z', \mathcal{F}}(k)\} \right)$$

and this is equal to

$$\frac{1}{|F_i|} \sum_{l \in F_i} |J(i, l)| \mathbb{P}(Y_{t-1}^z(l) \neq Y_{t-1}^{z', \mathcal{F}}(l)). \tag{44}$$

Collecting together (38), (41), (43), (44), we finally get the upperbound

$$\mathbb{P}(Y_t^z(i) \neq Y_t^{z', \mathcal{F}}(i)) \leq \sum_{l \notin F_i} |J(i, l)| + \sum_{l \in F_i} |J(i, l)| \mathbb{P}(Y_{t-1}^z(l) \neq Y_{t-1}^{z', \mathcal{F}}(l)). \tag{45}$$

To conclude the proof of the theorem, let  $Z$  and  $Z'$  be two independent copies of the Ising models on  $\{-1, +1\}^S$  with potentials  $J$  and  $J^{\mathcal{F}}$ , respectively. For a fixed

realization of the pair  $Z$  and  $Z'$ , construct as before the coupled chains  $(Y_t^Z, Y_t^{Z', \mathcal{F}})$  taking values on  $\{-1, +1\}^{2F_i}$ , and having  $Z(F_i^c)$  and  $Z'(F_i^c)$  as boundary conditions.

Using inequality (45) and taking the expectation with respect to  $(Z, Z')$ , we have

$$\begin{aligned} & \mathbb{E}[\mathbb{P}(Y_t^Z(i) \neq Y_t^{Z', \mathcal{F}}(i))] \\ & \leq \sum_{l \notin F_i} |J(i, l)| + \sum_{l \in F_i} |J(i, l)| \mathbb{E}[\mathbb{P}(Y_{t-1}^Z(l) \neq Y_{t-1}^{Z', \mathcal{F}}(l))]. \end{aligned} \tag{46}$$

Now observe that

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

for any  $j \in F$ , where  $Y(j)$  and  $Y^{\mathcal{F}}(j)$  are the projections on site  $j$  of realizations of the Ising model with pairwise potential  $J$  and  $J^{\mathcal{F}}$ , respectively. From this identity and inequality (46), it follows that

$$\mathbb{P}(Y(i) \neq Y^{\mathcal{F}}(i)) \leq \sum_{l \notin F_i} |J(i, l)| + \sum_{l \in F_i} |J(i, l)| \mathbb{P}(Y(l) \neq Y^{\mathcal{F}}(l)).$$

Finally, taking the supremum for all  $i \in S$  we have

$$\sup_{i \in S} \mathbb{P}(Y(i) \neq Y^{\mathcal{F}}(i)) \leq \sup_{i \in S} \sum_{l \notin F_i} |J(i, l)| + r \sup_{i \in S} \mathbb{P}(Y(i) \neq Y^{\mathcal{F}}(i)),$$

which concludes the proof.

### Acknowledgments

The authors would like to thank an anonymous referee who pointed an error in the statement of Theorem 3 in a earlier version of the article. The authors also thank Luiz Lana and Matthieu Lerasle for many illuminating discussions.

This article was produced as part of the activities of FAPESP Center for Neromathematics (FAPESP grant 2013/ 07699-0). This work is part of USP project “Mathematics, computation, language, and the brain,” and CNPq project “Stochastic modeling of the brain activity” (grant 480108/2012-9). A. Galves was partially supported by a CNPq fellowship (grant 309501/2011-3). D. Y. Takahashi was supported by FAPESP grant 2008/08171-0 and Pew Latin American Fellowship. E. Orlandi was partially supported by Prin07:20078XYHYS. D. Y. Takahashi thanks the hospitality of Università di Roma Tre, Dipartimento di Matematica. E. Orlandi thanks the hospitality of NUMEC, USP.



## References

- Bento, J. and Montanari, A. (2009). Which graphical models are difficult to learn? Available at <http://arxiv.org/pdf/0910.5761>.
- Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B* **48**, 259–302. [MR0876840](#)
- Bresler, G., Mossel, E. and Sly, A. (2008). Approximation, randomization and combinatorial optimization. Algorithms and techniques. In *Reconstruction of Markov Random Fields from Samples: Some Easy Observations and Algorithms* 343–356. Berlin: Springer. [MR2538799](#)
- Cross, G. and Jain, A. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**, 25–39.
- Csiszar, I. and Talata, Z. (2006). Consistent estimation of the basic neighborhood of Markov random fields. *Annals of Statistics* **34**, 123–145. [MR2275237](#)
- Lerasle, M. and Takahashi, D. Y. (2011). An oracle approach for interaction neighborhood estimation in random fields. *Electronic Journal of Statistics* **5**, 534–571. [MR2813554](#)
- Lerasle, M. and Takahashi, D. Y. (2015). Sharp oracle inequalities and slope heuristic for specification probabilities estimation in general random fields. *Bernoulli*. To appear.
- Löcherbach, E. and Orlandi, E. (2011). Neighborhood radius estimation for variable-neighborhood random fields. *Stochastic Processes and Their Applications* **121**, 2151–2185. [MR2819245](#)
- Massart, P. (2003). *Concentration Inequalities and Model Selection*. Berlin: Springer. [MR2319879](#)
- Presutti, E. (2009). *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*. Berlin: Springer. [MR2460018](#)
- Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). High-dimensional Ising model selection using  $\ell_1$ -regularized logistic regression. *Annals of Statistics* **38**, 1287–1319. [MR2662343](#)
- Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley. [MR0624436](#)
- Schneidman, E., Berry, M. J., Segev, R. and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012.
- Woods, J. (1978). Markov image modeling. *IEEE Transactions on Automatic Control* **23**, 846–850.

A. Galves  
 Instituto de Matemática e Estatística  
 Universidade de São Paulo  
 Caixa Postal 66281  
 05315-970 São Paulo  
 Brasil  
 E-mail: [galves@usp.br](mailto:galves@usp.br)

E. Orlandi  
 Dipartimento di Matematica  
 Università di Roma Tre  
 L.go S. Murialdo 1  
 00146 Roma  
 Italy  
 E-mail: [orlandi@mat.uniroma3.it](mailto:orlandi@mat.uniroma3.it)

D. Y. Takahashi  
 Institute of Neuroscience  
 Princeton University  
 Princeton, New Jersey 08548  
 USA  
 E-mail: [takahashiyd@gmail.com](mailto:takahashiyd@gmail.com)