An estimator of the coupling coefficient suitable for partially known Ising lattices*

Tommy Norberg†

January 18, 1999

Abstract

The paper introduces an estimator of the coupling coefficient in the Ising lattice, which is based on counting transitions, and suitable for lattices with missing data. A central limit result is proved for the case of no missing data. The performance of the estimator is studied by means of Bernoulli-thinned simulated Ising lattices.

1 Introduction

The Ising model goes back all the way to 1925 [9]. It is a model for ferromagnetism in statistical mechanics, which have found numerous applications in image processing. Each vertex of a lattice (or tree) is supposed to contain an atom, the state of which can be in either of the two spin orientations $-$ or $+$. The material is said to be magnetised if a large majority of atoms have the same spin. At high temperatures there are approximately as many atoms with positive as with negative spin. The atoms change spin in a random manner, but in such a way that the approximate equality of the number of atoms with negative and positive spin is kept. Nearby atoms, however, tend to be oriented in the same direction. This is referred to as clustering. The tendency to cluster becomes stronger when the temperature is lowered and when it falls below a critical value suddenly a gross majority of atoms have identical spins and spontaneous magnetisation occurs. The interested reader is referred to Georgii [6] and/or Liggett [11] for more details on this interesting phenomenon, referred to as phase transition.

The Ising model is a Markov random field, and as such it is one of the most well-known. See, e g, Kinderman & Snell [10] and Guyon [8].

*Research supported by the Swedish Natural Science Research Council and by the Swedish Nuclear Fuel and Waste Management Company

†Mathematical Statistics, Chalmers and Göteborg University, S-412 96 Göteborg, Sweden; email: tommy@math.chalmers.se; www: http://www.math.chalmers.se/~tommy

The inverse of the temperature will in this paper be referred to as the coupling coefficient. Sometimes people refer to it as the coefficient of clustering. A high value of the coupling coefficient corresponds to a low temperature, and clustering of atoms with identical spins.

In this paper, we introduce an estimator of the coupling (or clustering) coefficient, which is based on counting transitions. A main reason for its introduction is that it can be effectively calculated also for Ising lattices that are only partially known. A drawback is that it becomes more and more biased as the coupling coefficient increases towards the critical value.

The paper is organised as follows: First, we lay down our notations and terminology in Section 1.1. Then, in Section 2, the Ising model is described first as a Gibbs distribution (Section 2.1.1) and then as a Markov random field (Section 2.1.2). This section moreover contains a description of the maximum likelihood (Section 2.2.1), the pseudo-likelihood (Section 2.2.2) and a method of moments estimator (Section 2.2.3) of the coupling coefficient. The latter two estimators cannot without a lot of 'pain' be adapted to the case when the Ising lattice is only partially known.

In Section 3 we introduce our new estimator. We first discuss the case of a general Markov random field (Section 3.1). We then specialise to the Ising case in Section 3.2. This section splits into two parts—one (Section 3.2.1) that treats the case when the Ising lattice is completely known and one (Section 3.2.3) that treats the case when the lattice is only partially known. In Section 3.2.2 we prove a central limit result valid in the case when the lattice is completely known.

We use a standard Markov chain Monte Carlo technique to simulate Ising lattices and to obtain realisations of estimated Ising distributions. See Section 4. Section 5 describes how we did restorations of thinned Ising lattices. In the two following sections we describe our obtained simulations results—in Section 6.1 we treat the completely known case and in Section 6.2, we discuss the results obtained for partially known lattices. Finally, in Section 7 we discuss the pros and cons of the restoration technique of this paper.

1.1 Notations and terminology

We deal only with the Ising model on a finite rectangular part of Z^2 (Z denotes the integers), to be denoted $R = \{1, \ldots, r\} \times \{1, \ldots, c\}$, although much of what is said below do extend to any finite number of dimensions (and also to Ising models on a large class of bipartite lattices). The members of R are called *sites* or *pixels*. Any element $\mathbf{x} = (x_{(i,j)},\ (i,j) \in R) \in \{-1,1\}^R$ is referred to as a *configuration*, *image* or *lattice*. An *Ising lattice* is a realisation of the Ising model. Notice that $x_{(i,j)} \in \{-1,1\}$ for all $(i,j) \in R$. Occasionally we will have to deal with $x_{(i,j)}$ for which $(i,j) \in Z^2 \setminus R$. Our convention in all such cases is $x_{(i,j)} = 0$.

A configuration $\mathbf{x} \in \{-1,1\}^R$ may be partially known only, in which

case there is a (sometimes implicit) partition of R into two subsets A, B, say, and we write $\mathbf{x} = \mathbf{x}_A\mathbf{x}_B$. Here $\mathbf{x}_A = (x_{(i,j)},\ (i,j) \in A)$ and similarly for $\mathbf{x}_B$.

The probability of $\mathbf{x} \in \{-1,1\}^R$, to be denoted $p(\mathbf{x})$ will be written as follows in terms of its *energy* $\Psi(\mathbf{x})$:

$$p(\mathbf{x}) = \frac{1}{Z}\exp\left(-\Psi(\mathbf{x})\right)$$

(Typically, the energy of a configuration may range from $-\infty$ to $+\infty$, and configurations with low energy are the more likely ones.) The normalising constant Z is referred to as the *partition function*. Clearly the energy function Ψ defines a unique probability measure p on $\{-1,1\}^R$, while Ψ is given from p only modulo some additive constant. Notice that it is generally impossible to calculate the partition function Z by pure enumeration because of the very high cardinality of $\{-1,1\}^R$ also for moderate values of r, c.

2 Background

2.1 Ising model

Below we will describe two illuminating ways of defining the Ising model.

2.1.1 Viewed as a Gibbs distribution

Let $e(\mathbf{x}) = e_h(\mathbf{x}) + e_v(\mathbf{x})$, where $e_h(\mathbf{x})$ denotes the number of horizontal pairs of sites (i,j) and $(i,j+1)$ such that $x_{(i,j)} = x_{(i,j+1)}$ and $e_v(\mathbf{x})$ similarly denotes the number of vertical site pairs (i,j) and $(i+1,j)$ such that $x_{(i,j)} = x_{(i+1,j)}$. Moreover, let $d(\mathbf{x}) = d_h(\mathbf{x}) + d_v(\mathbf{x})$, where $d_h(\mathbf{x})$ denotes the number of horizontal site pairs (i,j) and $(i,j+1)$ such that $x_{(i,j)} \neq x_{(i,j+1)}$ and $d_v(\mathbf{x})$ similarly denotes the number of vertical site pairs (i,j) and $(i+1,j)$ such that $x_{(i,j)} \neq x_{(i+1,j)}$.

The *Ising model* on R is the probability measure which is defined by the energy function

$$\Psi(\mathbf{x}) = 2\beta d(\mathbf{x}), \quad \mathbf{x} \in \{-1,1\}^R$$

Another energy function that defines the Ising model is

$$\Psi(\mathbf{x}) = \beta(d(\mathbf{x}) - e(\mathbf{x})), \quad \mathbf{x} \in \{-1,1\}^R$$

since $d(\mathbf{x}) + e(\mathbf{x}) = 2rc - (r+c)$ does not depend on the particular configuration $\mathbf{x}$.

The parameter β is be referred to as *the coupling coefficient*. It is a measure of the strength of the spatial dependence. Nearby sites tend to have the same value if β is positive, while, if β is negative, adjacent sites tend to have opposite values. In the former case observations tend to form

clusters, the size of which depends on the magnitude of β. In the latter case one may observe local checkerboard patterns in the configurations. Notice also that the case $\beta = 0$ corresponds to independence between sites.

This heuristic description of the behaviour of the Ising model breaks down if β is too big (above a critical value β_c) or too low (below $-\beta_c$). The phenomenon is often referred to as a *phase transition*, although this terminology is strict only in the case of Markov random field models on infinite lattices or trees. Phase transitions occur in many Markov random fields and the Ising model on Z^2 is one of the few for which the critical value is fully known. The calculation is due to Onsager [13], who showed that $\beta_c = \frac{1}{2}\log(1+\sqrt{2}) \approx 0.441$. For more facts, refer, e g, to Georgii [6].

2.1.2 Viewed as a Markov random field

For $A \subseteq R$, write

$$\partial A = \{(i,j) \in Z^2 \setminus A : \exists (k,l) \in A \text{ such that } |i-k| + |j-l| = 1\}$$

This set contains the *neighbours* of A. The neighbours of a singleton $A = \{(i,j)\}$ will be denoted $\partial(i,j)$. Notice that $\partial A \subseteq Z^2 \setminus A$. Thus there may very well be neighbours of a set A that does not belong to R.

The Markov property says that

$$p(\mathbf{x}_A | \mathbf{x}_{R \setminus A}) = p(\mathbf{x}_A | \mathbf{x}_{\partial A})$$

It is easy to see that this is satisfied by the Ising model as defined in the above section. In particular, we get for any singleton $A = \{(i,j)\}$ that

$$\partial(i,j) = \{(i-1,j),(i,j+1),(i+1,j),(i,j-1)\}$$

and

$$p(x_{(i,j)} | \mathbf{x}_{\partial(i,j)}) = \frac{e^{\beta x_{(i,j)}(x_{(i-1,j)}+x_{(i,j+1)}+x_{(i+1,j)}+x_{(i,j-1)})}}{e^{\beta(x_{(i-1,j)}+x_{(i,j+1)}+x_{(i+1,j)}+x_{(i,j-1)})} + e^{-\beta(x_{(i-1,j)}+x_{(i,j+1)}+x_{(i+1,j)}+x_{(i,j-1)})}}$$

Recall our convention, which applies here and throughout, that $x_{(i,j)} = 0$ if $(i,j) \in Z^2 \setminus R$.

Thus the conditional probability distribution for $x_{(i,j)}$ given $\mathbf{x}_{\partial(i,j)}$ depends only on the sum $s_{(i,j)} = x_{(i-1,j)} + x_{(i,j+1)} + x_{(i+1,j)} + x_{(i,j-1)}$:

$$p(x_{(i,j)} | \mathbf{x}_{\partial(i,j)}) = \frac{e^{\beta x_{(i,j)} s_{(i,j)}}}{e^{\beta s_{(i,j)}} + e^{-\beta s_{(i,j)}}}$$

Suppressing the dependence on the site, we may write this in the following way

$$p(x|s) = \frac{e^{\beta x s}}{e^{\beta s} + e^{-\beta s}}$$

for $x \in \{-1, 1\}$ and $s \in \{-4, -2, 0, 2, 4\}$.

Below, we will refer to the Ising model with coupling coefficient β as the Ising(β)-distribution.

2.2 Statistics

Here we mention a few alternatives to our estimator (which will be discussed in Section 3 below) of the coupling coefficient. In all cases except, perhaps, the first, we see no straightforward way to implementation in a missing data situation. Our review is not meant to be exhaustive.

2.2.1 The maximum likelihood estimator of β

Below, we apply to the Ising model a well-known method for Monte Carlo estimation of the maximum likelihood estimate. The log likelihood for β, given the observation $\mathbf{x}$, is

$$\mathcal{L}(\beta|\mathbf{x}) = \beta(e(\mathbf{x}) - d(\mathbf{x})) - \log Z(\beta)$$

Clearly,

$$Z(\beta) = \sum_{\mathbf{y}} \exp\left(\beta(e(\mathbf{y}) - d(\mathbf{y}))\right)$$

This can be rewritten

$$Z(\beta) = Z(\eta) E_\eta \exp\left((\beta - \eta)(e(\mathbf{y}) - d(\mathbf{y}))\right)$$

where E_η denotes expectation w r t the Ising(η)-distribution. Hence, maximising the log likelihood is equivalent to maximising

$$\beta(e(\mathbf{x}) - d(\mathbf{x})) - \log E_\eta \exp\left((\beta - \eta)(e(\mathbf{y}) - d(\mathbf{y}))\right)$$

for a fixed value of η.

The idea now is to guess a value η which hopefully is near the true maximum likelihood estimate $\widehat{\beta}_{\mathrm{ML}}$ and then approximate the expectation to the right using a suitable Markov chain Monte Carlo method. The drawback is that this can become extremely time consuming, since very large Monte Carlo sample sizes may be needed. For details on a reasonable implementation method, see Geyer & Thompson [7].

2.2.2 The pseudo-likelihood estimator of β

Besag [3] introduced the following alternative to the likelihood, which has become known as the pseudo-likelihood,

$$\mathrm{PL}(\beta|\mathbf{x}) = \prod_{(i,j)} \frac{e^{\beta x_{(i,j)} s_{(i,j)}}}{e^{\beta s_{(i,j)}} + e^{-\beta s_{(i,j)}}}$$

The product is over all interior sites (i e, sites with all neighbours in R). The log pseudo-likelihood,

$$\mathcal{PL}(\beta|\mathbf{x}) = \sum_{(i,j)} \left(\beta x_{(i,j)} s_{(i,j)} - \log \left(e^{\beta s_{(i,j)}} + e^{-\beta s_{(i,j)}} \right) \right)$$

is concave, thus has a unique maximum, to be referred to as the pseudo-likelihood estimate of β and written $\widehat{\beta}_{\mathrm{PL}}$. It is a consistent estimator of β, see Geman & Graffigne [5].

Notice that the terms with $s_{(i,j)} = 0$ do not contribute to the argument of the maximum. We split the log pseudo-likelihood into four terms, depending on the value of $s_{(i,j)}$. Thus, for $s = -4, -2, 2, 4$, write

$$\bar{x}_s = \frac{1}{n_s} \sum_{i,j:s_{(i,j)}=s} x_{(i,j)}$$

where n_s is the total number of terms in the sum. (The sum is over all interior sites (i,j) such that $s_{(i,j)} = s$.) Finally, maximising the pseudo-likelihood is equivalent to maximising

$$\sum_{s \neq 0} \left(s\beta n_s \bar{x}_s - n_s \log \left(e^{\beta s} + e^{-\beta s} \right) \right)$$

2.2.3 A method of moments estimator of β

Sherman & Seymore [15] introduced a method of moments estimator of β, which is closely related to Besag's pseudo-likelihood estimator. By solving

$$p(1|s) = \frac{e^{\beta s}}{e^{\beta s} + e^{-\beta s}}, \quad s \neq 0$$

for β, we get

$$\beta = \frac{1}{2s} \log \frac{p(1|s)}{p(-1|s)}, \quad s = -4, -2, 2, 4$$

Sherman & Seymore [15] used this relation to define the following four method of moments estimators of β:

$$\widehat{\beta}_{\mathrm{MM}_s} = \frac{1}{2s} \log \frac{\hat{p}(1|s)}{\hat{p}(-1|s)}, \quad s = -4, -2, 2, 4$$

where

$$\hat{p}(1|s) = \frac{|\{(i,j) \in R : x_{(i,j)} = 1, s_{(i,j)} = s\}|}{|\{(i,j) \in R : s_{(i,j)} = s\}|}$$

and

$$\hat{p}(-1|s) = 1 - \hat{p}(1|s)$$

($|A|$ denotes the cardinality of A). They further studied the mean of these four estimators:

$$\widehat{\beta}_{\mathrm{MM}} = \frac{1}{4}\sum_s \widehat{\beta}_{\mathrm{MM}_s}$$

Now, notice that

$$\bar{x}_s = \hat{p}(1|s) - \hat{p}(-1|s)$$

from which

$$\widehat{\beta}_{\mathrm{MM}_s} = \frac{1}{2s}\log\frac{1+\bar{x}_s}{1-\bar{x}_s}$$

follows. The reader now easily shows that the term

$$s\beta n_s \bar{x}_s - n_s \log\left(e^{\beta s} + e^{-\beta s}\right)$$

in the log pseudo-likelihood is maximised by $\widehat{\beta}_{\mathrm{MM}_s}$. Thus the pseudo-likelihood estimator $\widehat{\beta}_{\mathrm{PL}}$ is a weighted mean of Sherman & Seymore's four method of moments estimators $\widehat{\beta}_{\mathrm{MM}_s}$.

3 A canonical minimum energy estimator

3.1 The general Markov random field case

In the Ising model that we study in this paper, there are nearest neighbour interactions in the horizontal and vertical directions only. (This holds true also for many more general Markov random fields.) We therefore write the energy function as follows:

$$\Psi(\mathbf{x}) = \Psi_i(\mathbf{x}) + \Psi_h(\mathbf{x}) + \Psi_v(\mathbf{x})$$

where

$$\Psi_i(\mathbf{x}) = \sum_{(i,j)} \psi_i(x_{(i,j)})$$

while

$$\Psi_h(\mathbf{x}) = \sum_{(i,j)} \psi_h(x_{(i,j)}, x_{(i,j+1)})$$

and, similarly,

$$\Psi_v(\mathbf{x}) = \sum_{(i,j)} \psi_v(x_{(i,j)}, x_{(i+1,j)})$$

Convention: The value of a function is zero if an argument vanishes. Notice that the densities ψ_h and ψ_v coincide in the isotropic case.

Next, notice that

$$\Psi_i(\mathbf{x}) + \Psi_h(\mathbf{x}) = \sum_i \sum_j (\psi_i(x_{(i,j)}) + \psi_h(x_{(i,j)}, x_{(i,j+1)}))$$

The idea now, is to think of each sum

$$\sum_j (\psi_i(x_{(i,j)}) + \psi_h(x_{(i,j)}, x_{(i,j+1)}))$$

as the negative of the log-likelihood (i e, canonical energy) of a homogeneous Markov chain with transition matrix $P_h = (p_h(x,y))$ and invariant distribution π satisfying $\pi P_h = \pi$. A simple argument then shows that this sum can be written

$$\sum_j (-\log \pi(x_{(i,j)}) + \log \pi(x_{(i,j+1)}) - \log p_h(x_{(i,j)}, x_{(i,j+1)}))$$

By identification of terms, we see that

$$\psi_i(x) = -\log \pi(x)$$

and

$$\psi_h(x,y) = \log \pi(y) - \log p_h(x,y)$$

An analogous argument for the vertical direction yields

$$\psi_v(x,y) = \log \pi(y) - \log p_v(x,y)$$

where also the p_v's form a transition matrix $P_v = (p_v(x,y))$. Notice that P_h and P_v must have the same invariant distribution π. Notice also that ψ_i, ψ_h and ψ_v are defined modulo additive constants. In the isotropic case, $P_h = P_v$.

We use formal maximum likelihood technique to estimate the horizontal and vertical transition matrices. In the case of a complete observation, the maximum likelihood estimate of $p_h(x,y)$ is

$$\hat{p}_h(x,y) = \frac{f_h(x,y)}{f_h(x,\cdot)}$$

where $f_h(x,y)$ denotes the number of transitions from state x into state y in the horizontal direction, and $f_h(x,\cdot) = \sum_y f_h(x,y)$ is the total number of jumps out of state x. Notice here that each row $(x_{(i,j)}, 1 \leq j \leq c)$ contains the same amount of information. That is why we sum over all rows above. We finally get the following estimate of ψ_h:

$$\widehat{\psi}_h(x,y) = \log \hat{\pi}_h(y) - \log \hat{p}_h(x,y)$$

where $\hat{\pi}_h$ is the invariant distribution corresponding to the estimated transition matrix $\widehat{P}_h = (\hat{p}_h(x,y))$.

Similarly,

$$\hat{p}_v(x,y) = \frac{f_v(x,y)}{f_v(x,\cdot)}$$

where $f_v(x,y)$ denotes the number of transitions from state x into state y in the vertical direction, and $f_v(x,\cdot) = \sum_y f_v(x,y)$ is the total number of jumps out of state x. Moreover, ψ_v is estimated by

$$\widehat{\psi}_v(x,y) = \log \hat{\pi}_v(y) - \log \hat{p}_v(x,y)$$

where $\hat{\pi}_v$ is the invariant distribution corresponding to the transition matrix $\widehat{P}_v = (\hat{p}_v(x,y))$.

Notice that, this approach gives us two estimates $\hat{\pi}_h$ and $\hat{\pi}_v$ of the invariant distribution π. Practise has shown that they are not very different. Each is used to calculate an estimate of ψ_i. As our final estimate of ψ_i, we use the mean of these two estimators.

In the isotropic case, we write $f(x,y) = f_v(x,y) + f_h(x,y)$ and estimate the common transition probability $p(x,y) = p_h(x,y) = p_v(x,y)$ by means of

$$\hat{p}(x,y) = \frac{f(x,y)}{f(x,\cdot)}$$

where again $f(x,\cdot) = \sum_y f(x,y)$. We then proceed as above. Notice that in this case we have just one estimate of the invariant distribution.

In the case of a non-complete observation, we express the 'likelihood' as a function of data. However, typically there is no closed form solution. This calls for a numerical maximisation. There is moreover no guarantee that the likelihood is concave. That is to say, there may be many local maxima. In order to find the best (or one which is nearly as good) we have used simulated annealing [2].

We refer to these kind of estimators as *canonical minimum energy estimators*, since they are formed by writing the energy in a canonical way and then finding the function that minimises the energy. This may be done parametrically in the Ising model (as in this paper) as well as non-parametrically in general Markov random field models [12].

3.2 The Ising model

Let us now see what the above yields for the isotropic Ising model. The only reasonable parametrisation of the transition matrices is the following

$$P_h = P_v = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

with invariant distribution

$$\pi = [\, \tfrac{1}{2} \quad \tfrac{1}{2} \,]$$

since the two states -1 and 1 are completely symmetric in the Ising model. This yields an energy function specified by means of

$$\psi_i(-1) = \psi_i(1) = \log 2$$

and

$$\begin{array}{rcrcl} \psi_h(-1,-1) & = & \psi_h(1,1) & = & -\log 2 - \log(1-p) \\ \psi_h(-1,1) & = & \psi_h(1,-1) & = & -\log 2 - \log p \end{array}$$

Under isotropocy, $\psi_v = \psi_h$. It is convenient to write ψ_i as a vector and ψ_h and ψ_v as matrices. We then get (using the fact that these densities are defined modulo additive constants)

$$\psi_i = [\,0 \quad 0\,]$$

and

$$\psi_h = \psi_v = \begin{bmatrix} 0 & \log\frac{1-p}{p} \\ \log\frac{1-p}{p} & 0 \end{bmatrix}$$

showing that

$$\Psi(\mathbf{x}) = \left(\log\frac{1-p}{p}\right) d(\mathbf{x})$$

Hence,

$$\beta = \frac{1}{2}\log\frac{1-p}{p}$$

which clearly holds true if, and only if,

$$p = \frac{1}{1+e^{2\beta}}$$

One may look upon these two latter formulae as a reparametrization of the Ising model. While the coupling coefficient β measures the amount of dependence between adjacent sites, the transition probability p is a measure of the willingness to change state. Configurations with $d(\mathbf{x})$ small are favoured if $\beta > 0$. Now $\beta > 0$ holds true, if, and only if, $p < \frac{1}{2}$ in which case the tendency to change state is low in the process, hence clustering occurs. Similarly, if $\beta < 0$ configurations with $d(\mathbf{x})$ large are favoured. Also, $\beta < 0$ if, and only if, $p > \frac{1}{2}$, in which case the process has a high tendency to change state, hence local checkerboard patterns will be seen. Finally, $p = \frac{1}{2}$ if, and only if, the sites act independently.

Notice also that the critical transition probabilities (that corresponds to β_c) are

$$p_c = 1 - \frac{\sqrt{2}}{2} \approx 0.2929$$

and

$$1 - p_c = \frac{\sqrt{2}}{2} \approx 0.7071$$

Our motivation for this reparametrization of the Ising model is of course that p has a natural (canonical minimum energy) estimator to be discussed next.

3.2.1 Case of complete observation

In the case of an observation without missing data, the maximum likelihood technique described above yields

$$\hat{p} = \frac{d(\mathbf{x})}{e(\mathbf{x}) + d(\mathbf{x})} = \frac{d(\mathbf{x})}{2rc - (r + c)}$$

showing that the canonical minimum energy estimator of β is

$$\widehat{\beta} = \frac{1}{2} \log \frac{e(\mathbf{x})}{d(\mathbf{x})}$$

3.2.2 Asymptotics

Pickard [14] proved under the torus convention, a law of large numbers and a central limit theorem for the nearest neighbour sample correlation $S = Q/2rc$, where

$$Q = e(\mathbf{x}) - d(\mathbf{x})$$

The central limit result is of the form

$$\sqrt{2rc}\,(S - B') \xrightarrow{\mathcal{D}} \mathcal{N}(0, B'')$$

The constants B' and B'' are difficult to explain. We therefore refer the reader to [14]. Notice now, that

$$S = \frac{e(\mathbf{x}) - d(\mathbf{x})}{e(\mathbf{x}) + d(\mathbf{x})} = 1 - 2\hat{p}$$

showing that we may transform Pickard's two limit assertions into a law of large numbers and a central limit theorem for $\hat{p}$. Using well-known facts for the logarithm, we may further prove a law of large numbers

$$\widehat{\beta} \to \beta_0 = \frac{1}{2} \log \frac{1 + B'}{1 - B'}$$

and a central limit result of the form

$$\sqrt{2rc}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1/2B'')$$

(for details on how to do the latter derivation, refer, e g, to [1, p 422]). Notice also that there are restrictions on the relative rates at which $r, c \to \infty$ [14].

Thus, our canonical minimum energy estimator do converge to some value β_0. It is quite clear that $\beta_0 \neq \beta$ (unless $\beta = 0$). We will see below (Table 1) that the difference between β_0 (or rather the mean of our estimate $\widehat{\beta}$) and the true value β can be quite big.

3.2.3 Case of missing data

The idea behind the canonical minimum energy estimator is to study transitions in the horizontal and the vertical directions, using the simplifying (and not correct) Markov chain assumption. Thus, for each row i and each column j, we write down a likelihood L_i^r and L_j^c, resp, thinking as if the row (or column) is a finite realization of a Markov chain with transition matrix

$$P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

having invariant distribution

$$\pi = [\, \tfrac{1}{2} \quad \tfrac{1}{2} \,]$$

Notice that the likelihoods are functions of one parameter p, only. Nearby rows (columns) are clearly not independent. Still, they contain precisely the same amount of information, so one reasonable way of weighting them together is to multiply the likelihoods, thus obtaining

$$L(p) = \prod_i L_i^r(p) \times \prod_j L_j^c(p)$$

In the case when all realizations are complete (treated above in Section 3.2.1) this yields (as noted) the estimator

$$\hat{p} = \frac{d(\mathbf{x})}{2rc - (r+c)}$$

of p, from which our estimator $\widehat{\beta} = (1/2)\log(e(\mathbf{x})/d(\mathbf{x}))$ of the coupling coefficient β may be derived.

In the case of non-complete observations, there is no closed form solution to the problem of maximising $L(p)$, so the natural thing to do is to resort to numerical methods. However, in theory the likelihood $L(p)$ need not be concave. (For an example, see [2].) Thus there may be more than one local maximum. Therefore we employed simulating annealing [2] to find a value $\hat{p}_{\mathrm{SA}}$ of p near the global maximum. Starting with that value, we then ran a standard numerical maximisation algorithm to find the local maximum that is closest to $\hat{p}_{\mathrm{SA}}$. We also started from some randomly chosen values of p. We did not find a case with more than one local maximum, so employing simulated annealing in this one-parameter case seems to be to over-do it. The found value is, at least a good approximation of the canonical minimum energy estimate $\hat{p}$ of p. Finally, the canonical minimum energy estimator of the coupling coefficient β is given by means of

$$\widehat{\beta} = \frac{1}{2}\log\frac{1-\hat{p}}{\hat{p}}$$

Figure 1: An Ising simulation with coupling coefficient $\beta = 0.20$. The canonical minimum energy estimator of the coupling coefficient is $\widehat{\beta} = 0.2138$

4 Simulations

All simulations of the Ising model employed the Gibbs sampler [4], which updates one site at a time (see also [8, p 211]). An iteration consists of one sweep through all sites. The site visiting scheme was randomly permuted before each new sweep began. All simulated configurations has $r = c = 50$, which means that they contain 2500 pixels. For an example, see Fig 1.

5 Restorations

In our restoration experiments R is partitioned into a known region A and an unknown region B of sites. The problem is to find the configuration $\widehat{\mathbf{x}}_B$ making the conditional probability $p(\mathbf{x}_B|\mathbf{x}_A)$ maximal. For the Ising model, this is equivalent to maximising

$$e^{-2\beta d(\mathbf{x}_A \mathbf{x}_B)}$$

holding $\mathbf{x}_A$ fixed. If $\beta > 0$, the maximum clearly occurs when $d(\mathbf{x}_A \mathbf{x}_B)$ is minimal. It is interesting to note that we do not need to know the true value of the coupling coefficient. It is enough to know whether it is positive or negative.

The unknown region B was randomly selected using an independent thinning mechanism conditioned on the total number of remaining sites.

We used ordinary simulated annealing to find a sub-configuration $\mathbf{x}_B$ with $d(\mathbf{x}_A \mathbf{x}_B)$ minimal or near

$$\widehat{d}_A = \min_{\mathbf{x}_B} d(\mathbf{x}_A \mathbf{x}_B)$$

For a description of simulated annealing for the Gibbs sampler dynamic, see [8, p 213]. We used an exponential cooling schedule of the form

$$T_k = T_0 \times 2^{-k/k_{1/2}}, \quad k = 0, 1, \ldots, k_s$$

true β	$\widehat{\beta}$		$\widehat{\beta}_{\mathrm{MM}}$	
	$\overline{x}$	s	$\overline{x}$	s
0.05	0.0483	0.01507	0.0479	0.01522
0.10	0.1021	0.01413	0.1003	0.01401
0.15	0.1591	0.01495	0.1516	0.01497
0.20	0.2172	0.01692	0.2006	0.01637
0.25	0.2874	0.01919	0.2520	0.01504
0.30	0.3621	0.01939	0.2982	0.01431

Table 1: A comparison between the canonical minimum energy estimator $\widehat{\beta}$ and Sherman & Seymore's [15] method of moments estimator $\widehat{\beta}_{\mathrm{MM}}$ for different values of the true coupling coefficient β. The number of consecutive sweeps used to calculate $\overline{x}$ and s is in each case $n = 100$ ($\overline{x}$ and s denote the sample mean and standard deviation, resp)

where T_k denotes the temperature of sweep (or iteration) number k, T_0 the initial temperature, $k_{1/2}$ the number of sweeps needed to half the temperature and k_s the total number of sweeps. Most restorations used the schedule defined by $T_0 = 4$, $k_{1/2} = 25$ and $k_s = 200$.

6 Results

6.1 Completely known Ising lattices

Although the main idea of this paper is to present an estimator of the coupling coefficient specially design to deal with partially observed configurations, it is of course of great interest to see how this estimator works for complete observations.

Table 1 shows the results of comparisons between the canonical minimum energy estimator $\widehat{\beta}$ and Sherman & Seymore's [15] method of moments estimator $\widehat{\beta}_{\mathrm{MM}}$ for a couple of simulations using the Gibbs sampler with random sweep as described above. As can be seen in the table, the canonical minimum energy estimator $\widehat{\beta}$ works well when the dependence is not too strong. For stronger dependence $\widehat{\beta}$ is severely biased. The reason for this bias when the coupling coefficient is large, is of course that two-dimensional effects become more prominent. When β is too close to the critical value β_c, the amount of clustering is larger than the one-dimensional theory predicts. Thus, $d(\mathbf{x})$ is smaller and $e(\mathbf{x})$ is larger than predicted by the one-dimensional theory (which lie behind the canonical minimum energy estimator), making $\widehat{\beta} = (1/2) \log e(\mathbf{x})/d(\mathbf{x})$ positively biased.

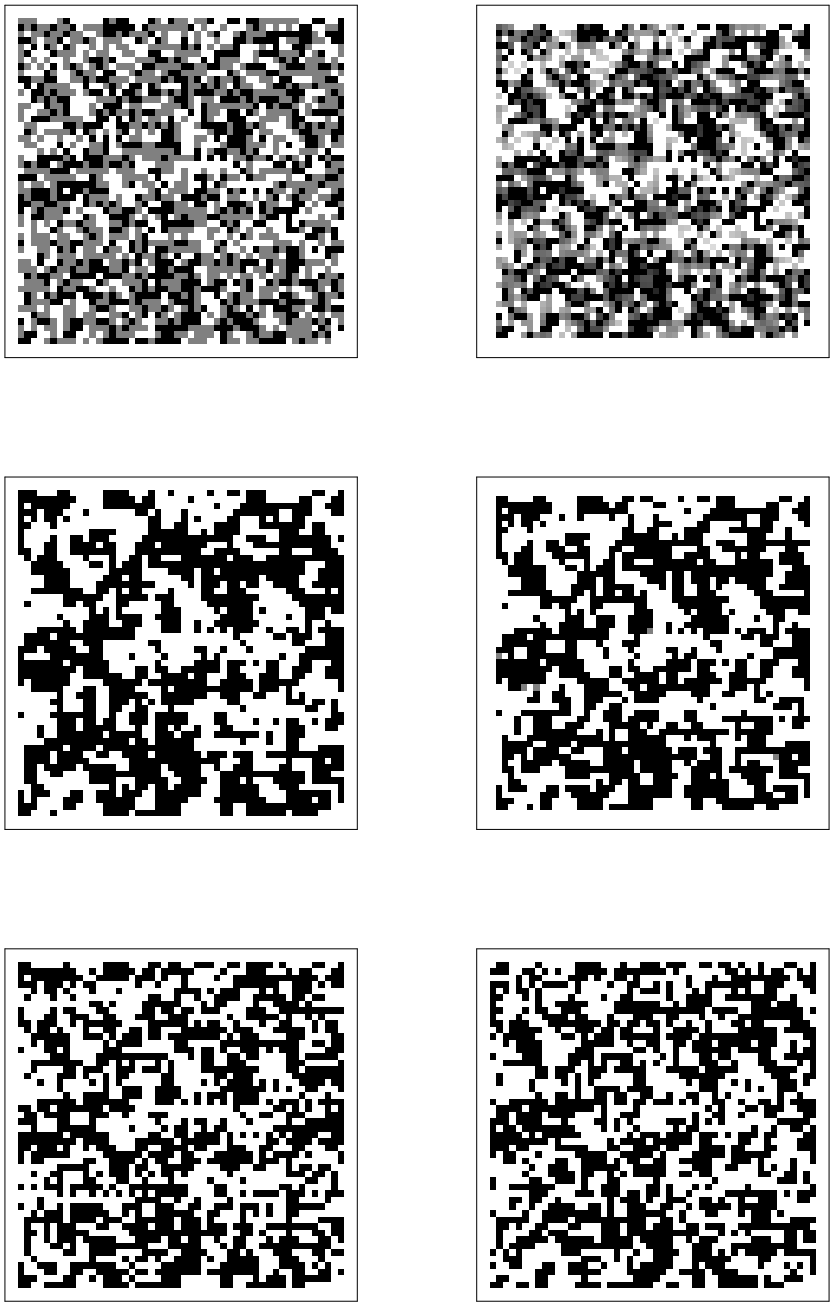

Figure 2: Upper left: 40% thinned Ising lattice; upper right: mean lattice; second row left and right: restoration and marginal mode plot, resp; third row: two conditional simulations in the estimated model ($\widehat{\beta} = 0.182$)

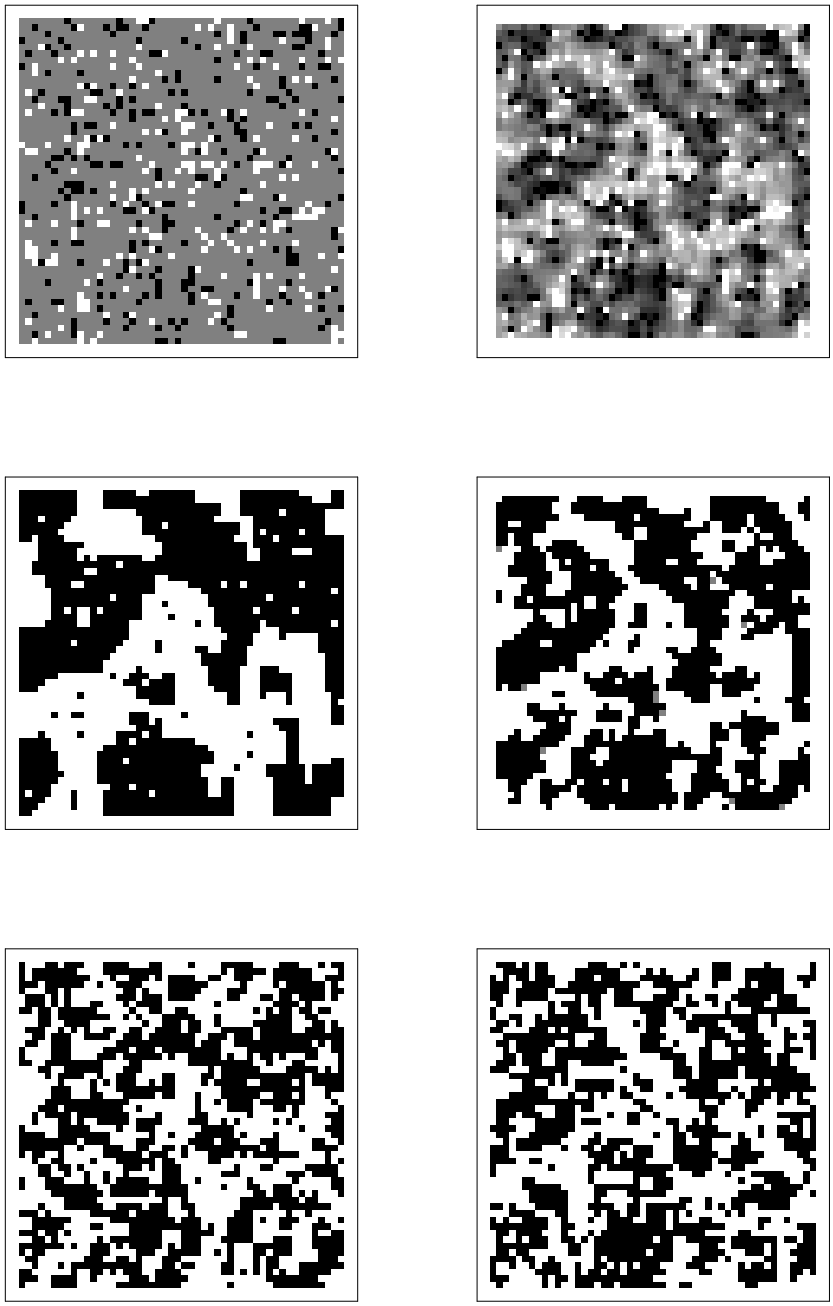

Figure 3: Upper left: 80% thinned Ising lattice; upper right: mean lattice; second row left and right: restoration and marginal mode plot, resp; third row: two conditional simulations in the estimated model ($\widehat{\beta} = 0.305$)

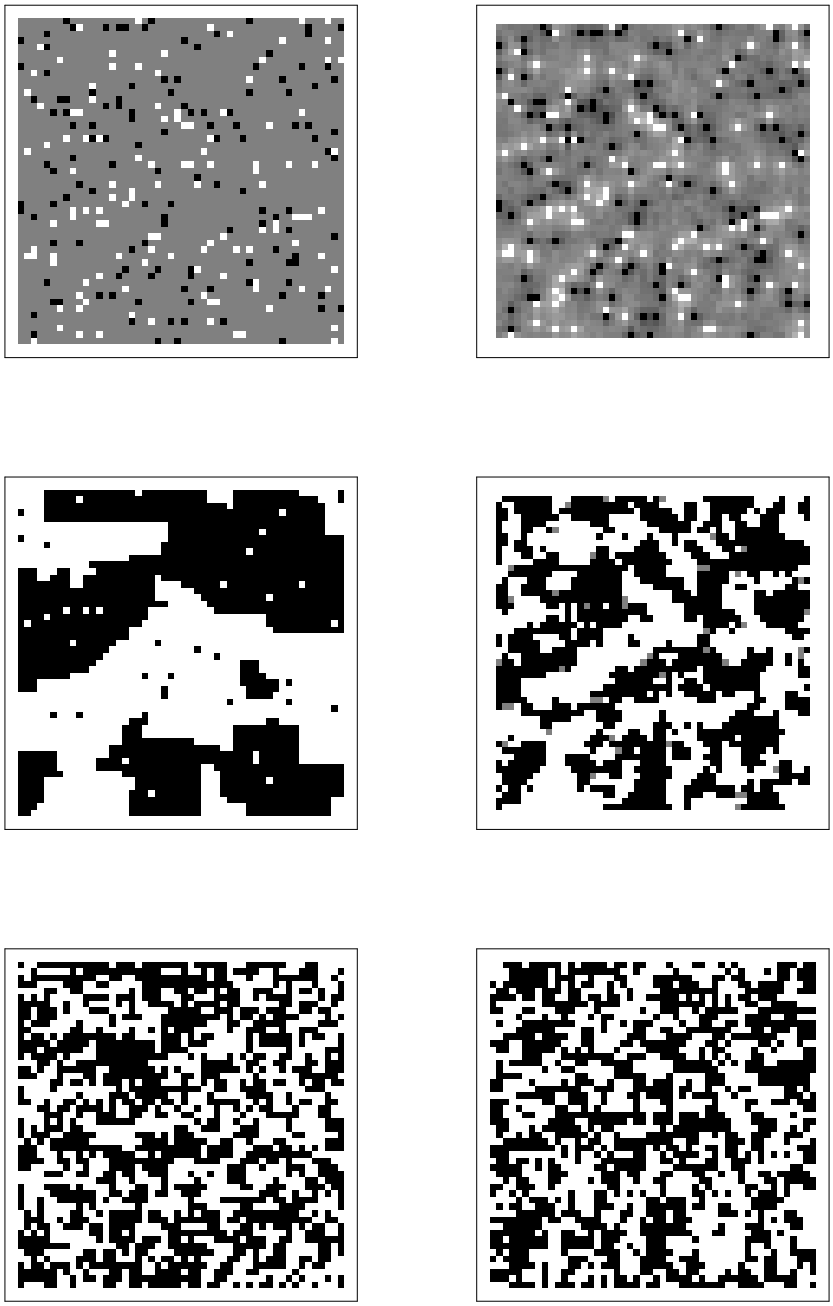

Figure 4: Upper left: 90% thinned Ising lattice; upper right: mean lattice; second row left and right: restoration and marginal mode plot, resp; third row: two conditional simulations in the estimated model ($\widehat{\beta} = 0.187$)

amount of thinning	observations of $\widehat{\beta}$ calculated in thinned configuration					$\overline{x}$	s	$2s/\sqrt{n}$
40%	0.214	0.152	0.215	0.200	0.211	0.1984	0.0266	0.0238
80%	0.324	0.216	0.145	0.085	0.295			
	0.091	0.184	0.079	0.226	0.249	0.1894	0.0880	0.0556
90%	0.177	0.185	0.459	0.233	0.259	0.2626	0.1149	0.1028

Table 2: Calculation of $\widehat{\beta}$ for three different amounts of thinning of the Ising lattice in Fig 1 with true $\beta = 0.20$ (the estimate based on the complete configuration is $\widehat{\beta} = 0.2138$; $\overline{x}$ and s denote the sample mean and standard deviation, resp, and $s/\sqrt{n}$ the standard error)

6.2 Partially known Ising lattices

Figures 2, 3 and 4 show the results of three thinning experiments. We randomly discarded 40%, then yet another 40% and then further 10% of the 2500 pixels of the Ising lattice in Fig 1 with true coupling coefficient $\beta = 0.20$. The thinned configurations are shown in the upper left positions of the three figures. The removed data are marked grey. The canonical minimum energy estimator of the coupling coefficient is for the three thinned configurations $\widehat{\beta} = 0.182$, $\widehat{\beta} = 0.305$ and $\widehat{\beta} = 0.187$, resp. In the upper right positions are marginal conditional mean lattices, each obtained by running a long conditional simulation given the known sub-configuration in the estimated model. The results of one restoration (using simulated annealing typically with $T_0 = 4$, $k_{1/2} = 25$ and $k_s = 200$) are shown underneath to the left and to its right is a simulated conditional marginal mode lattice shown. These mode plots come from the same run as the mean plots on the first row. In the third row are two conditional simulations from the estimated model shown.

The mean lattice gives quite a good insight into the true configuration in the two first cases, in which the amount of thinning is not too big (i e, 90%). In all three cases the mode lattice is a valuable supplement, but notice that its quality depends very much on the outcome of the estimation of β. This remark is of course also valid for the mean plots. Notice also that the mode lattice is a much more realistic restoration than the true maximal aposteriori prediction to its left. The conditional simulations in the last row does not help much in understanding the true configuration. Instead they could be looked upon as 'alternative realities'.

Table 2 shows the result of calculating $\widehat{\beta}$ for independent 40%, 80% and 90% thinnings of the Ising lattice in Fig 1. Notice that the sample standard deviation s increases with the amount of thinning. This experiment strongly supports the statements $\sigma_{.4} < \sigma_{.8}$ and $\sigma_{.4} < \sigma_{.9}$, while it does not give

true β	'true' $\widehat{\beta}$	observations of $\widehat{\beta}$ calculated in thinned configuration					$\overline{x}$	s	$2s/\sqrt{n}$
0.15	0.1673	0.194	0.083	0.213	0.180	0.307			
		0.189	0.259	0.122	0.105	0.275	0.1927	0.0742	0.0469
0.20	0.2202	0.232	0.211	0.228	0.377	0.306			
		0.360	0.306	0.129	0.289	0.160	0.2598	0.0817	0.0517
0.25	0.3064	0.266	0.329	0.266	0.325	0.305			
		0.395	0.299	0.269	0.197	0.315			
		0.332	0.324	0.387	0.289	0.402	0.3133	0.0545	0.0282

Table 3: Calculation of $\widehat{\beta}$ for three different values of the true β and 80% thinning (the 'true' $\widehat{\beta}$ is the value calculated in the complete configuration; $\overline{x}$ and s denote the sample mean and standard deviation, resp, and $s/\sqrt{n}$ the standard error)

any conclusive evidence as to weather $\sigma_{.8} < \sigma_{.9}$ or not. Here, for $p = .4, .8, .9$, σ_p^2 denotes the variance of any of the independent and identically distributed measurements of $\widehat{\beta}$ for a p-thinning of the configuration in Fig 1. The standard variance ratio test, based on the normal assumption, rejects strongly the two null hypotheses $H_1 : \sigma_{.4} = \sigma_{.8}$ and $H_2 : \sigma_{.4} = \sigma_{.9}$ in favour of the ordered alternatives $K_1 : \sigma_{.4} < \sigma_{.8}$ and $K_2 : \sigma_{.4} < \sigma_{.9}$. The one-sided p-values are 0.017 and 0.0076, resp. The one-sided p-value of the null hypothesis $H_3 : \sigma_{.8} = \sigma_{.9}$ is 0.232. Notice also that the estimate based on the complete configuration, $\widehat{\beta} = 0.2138$, in all three cases is within two standard errors from its 'estimate' $\overline{x}$.

Table 3 shows the result of calculating $\widehat{\beta}$ for independent 80% thinnings of three Ising lattices with true $\beta = 0.15$, $\beta = 0.20$ and $\beta = 0.25$. One would expect that the sample standard deviation s should decrease as the true β increases, since larger values of β typically give larger clusters. The data displayed does not support this or any alternative to this belief. The standard variance ratio test, based on the normal assumption, do not reject any of the three null hypotheses $H_1 : \sigma_{.15} = \sigma_{.20}$, $H_2 : \sigma_{.20} = \sigma_{.25}$ and $H_3 : \sigma_{.15} = \sigma_{.25}$ (here $\sigma_{.15}^2$ denotes the variance of any of the 10 independent and identically distributed observations of $\widehat{\beta}$ in the first case with true $\beta = 0.15$, etc). Notice also that the estimate based on the complete configuration in all three cases is within two standard errors from its 'estimate' $\overline{x}$.

7 Discussion

We have seen above that the canonical minimum energy estimator $\widehat{\beta}$ of β can be severely biased. One way to overcome this is to estimate the difference

between $E\widehat{\beta}$ and the true value β for various values of β. Table 1 is an embryo to this function for the case of 50×50 configurations. It is of course also possible (and probably better) to invoke such a procedure directly into the estimation process, using, e g, a Newton-Raphson like algorithm.

Our main purpose, however, with this research is not to introduce a new estimator of the coupling coefficient β in the Ising model. Rather it is to find new methods for doing restorations of partially observed discrete geological configurations. The method of this paper consists of the following three steps:

1. predict the value of the statistic $d(\mathbf{x})$;
2. use the prediction to estimate the model; and
3. find the most likely configuration in the estimated model.

We have seen that the method of this paper for doing step 2 is (sometimes severely) biased. Also step 3 is biased, since the most likely configurations are the ones with $d(\mathbf{x})$ minimal (as long as $\beta > 0$ which is the case treated in this paper) (see Section 5). This difficulty can however be overcome as we did above, by also calculating the marginal mode of repeated simulations of the estimated model. Notice also that the bias of step 3 adds to the bias of step 2.

In a follow up to this paper we will study step 1 more thoroughly. One reason for doing so is that it is possible to make restorations without invoking steps 2 and 3, by employing the fact that

$$p(\mathbf{x}|d(\mathbf{x}) = k) = \frac{1}{|\{\mathbf{x} : d(\mathbf{x}) = k\}|}$$

One difficulty with this approach is to sample from the uniform probability distribution $p(\mathbf{x}|d(\mathbf{x}))$.

Acknowledgements

I like to thank Olle Häggström, Olle Nerman and Geoff Nicholls for stimulating conversations regarding the topic of this paper.

References

[1] Agresti, A. (1990): *Categorical Data Analysis.* New York. Wiley.

[2] Á. Baran & S. Baran (1997): An application of simulated annealing to ML-estimation of a partially observed Markov chain. Report. Centre for Applied Mathematics and Statistics. Chalmers & Göteborg University, Sweden.

[3] J. Besag (1974): Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B*, **36**, 192-236.

[4] D. Geman & S. Geman (1984): Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. PAMI* **6**, 721-741.

[5] S. Geman & C. Graffigne (1986): Markov random field image models and their applications to compute vision. In *Proceedings of the Int. Congress of Mathematicians*. Berkeley.

[6] Georgii, H.-O. (1988): *Gibbs measures and phase transitions*. Walter de Gruyter.

[7] C. J. Geyer & E. A. Thompson (1992): Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J. Roy. Statist. Soc. Ser. B*, **54**, 657-699.

[8] X. Guyon (1995): *Random Fields on a Network*. Springer-Verlag.

[9] E. Ising (1925): Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift Physik*, **31**, 253-258.

[10] R. Kinderman & J. L. Snell (1980): *Markov Random Fields and their Applications*. American mathematical Society, Providence, Rhode Island.

[11] T. M. Liggett (1985): *Interacting particle systems*. Springer-Verlag.

[12] T. Norberg, L. Rosén, Á. Baran & S. Baran (1998): On modelling discrete geological structures as Markov random fields. Preliminary report of work in progress.

[13] L. Onsager (1944): Crystal statistics: A two-dimensional model with order-disorder transition. *Phys. Rev.*, **65**, 117-149.

[14] D. K. Pickard (1976): Asymptotic inference for the Ising lattice. *J. Appl. Prob.*, **13**, 486-497.

[15] M. Sherman & L. Seymore (1994): Method of moments and method of moments-type estimators in image processing and statistical mechanics models. Technical report no 93-26. Department of Statistics. The University of Georgia.