

A Recursive Algorithm for Markov Random Fields

Francesco Bartolucci

Julian Besag



NRCSE

Technical Report Series

NRCSE-TRS No. 073

February 28, 2002

The NRCSE was established in 1997 through a cooperative agreement with the United States
Environmental Protection Agency which provides the Center's primary funding.



A recursive algorithm for Markov random fields

By FRANCESCO BARTOLUCCI

*Department of Statistics, Box 1315, University of Perugia,
Perugia 06100, Italy*

bart@stat.unipg.it

and JULIAN BESAG

*Department of Statistics, Box 354322, University of Washington,
Seattle, Washington 98195, U.S.A.*

julian@stat.washington.edu

Summary

We propose a recursive algorithm as a more useful alternative to the Brook expansion for the joint distribution of a vector of random variables when the original formulation is in terms of the corresponding full conditional distributions, as occurs for Markov random fields. Usually, in practical applications, the computational load will still be excessive but then the algorithm can be used to obtain the componentwise full conditionals of a system after marginalizing over some variables or the joint distribution of subsets of the variables, conditioned on values of the remainder, which is required for block Gibbs sampling. As an illustrative example, we apply the algorithm in the simplest nontrivial setting of hidden Markov chains. More important, we demonstrate how it applies to Markov random fields on regular lattices and to perfect block Gibbs sampling for binary systems.

Some key words: Autologistic distribution; Binary system; Full conditional; Graphical model; Hidden Markov chain; Markov chain Monte Carlo; Markov random field; Monotone coupling from the past.

1 Introduction

Let $X = (X_1, \dots, X_n)$ denote a vector of n random variables, which here we take to be discrete. Let S_i denote the minimal sample space for X_i and S that for X . Assume the positivity condition (e.g. Besag, 1974) that $S = S_1 \times \dots \times S_n$, which implies that any conditional probabilities we may wish to consider are well defined. A Markov random field for X is then a corresponding probability distribution $\{\pi(x) : x \in S\}$ that is specified via its n local characteristics, more recently termed full conditionals, $\{\pi(x_i|x_{-i}) : x \in S\}$, where x_{-i} denotes all components of x other than x_i . Markov random field formulations originated in spatial statistics and there it is typical to choose the full conditional for each X_i to depend only on a few of the other X_j 's. A well-known example is the Ising model that occurs in statistical physics (Newman & Barkema, 1999, Ch. 1, 3, 4) but Markov random fields also play a central role in graphical models (e.g. Lauritzen, 1996), in random graphs (e.g. Frank and Strauss, 1986), in Markov chain Monte Carlo methods (e.g. Besag and Green, 1993; Smith and Roberts, 1993) and elsewhere.

The requirements for a self-consistent specification of $\pi(\cdot)$ via its full conditionals are not at all obvious but are identified by the Hammersley–Clifford theorem (Besag, 1974). Then, if a is any fixed vector in S , the Brook (1964) expansion determines $\pi(\cdot)$ up to scale:

$$\pi(x) / \pi(a) = \prod_{i=1}^n \{\pi(x_i|x_{<i}, a_{>i}) / \pi(a_i|x_{<i}, a_{>i})\}, \quad x \in S, \quad (1)$$

where, for example, $x_{<i} = \{x_j : j < i\}$. Although the normalising constant for $\pi(\cdot)$ is available in theory as an n -fold summation, its direct evaluation is usually prohibitive unless n is very small. Even for a simple binary Markov random field, such as the Ising model on the physically trivial 10×10 rectangular array, naive calculation of $\pi(\cdot)$ cannot be implemented because it involves summation over 2^{100} terms. Indeed, it was for such reasons that the Metropolis et al. (1953) and other Markov chain Monte Carlo algorithms were first developed.

It is of course always true that

$$\pi(x) = \pi(x_1) \prod_{i=2}^n \pi(x_i|x_{<i}), \quad (2)$$

and this is the natural factorisation in time series analysis, with the index i referring to temporal order. However, for Markov random fields, the terms in the product are not easily available. There are also lesser tasks for which solutions in terms of the full conditionals of X are important but awkward to obtain. Examples include the joint distribution of a particular subset of the X_i 's, given the values of the remainder, and the componentwise full conditionals of such a subset, irrespective of the other components. It is not obvious how to make such calculations in a reasonably efficient manner. This is especially important in Markov chain Monte Carlo algorithms where the computations need to be repeated a very large number of times. The Brook expansion (1) is not well suited to such tasks.

Such difficulties motivate the new recursion in §2 for the individual terms in (2). This allows explicit calculation of $\pi(\cdot)$ when the conditional probability structure is sufficiently straightforward and is also useful more generally in tackling the simpler subtasks that are described above. A mainly illustrative example is given in §3, where we apply the recursion to a hidden Markov chain and obtain a new algorithm of comparable complexity to that of Baum et al. (1970). We outline some extensions to more complex systems. In §4, we describe implementation of the recursion specifically for a first-order Markov random field on a finite rectangular array and evaluate the corresponding complexity. We also discuss the corresponding algorithm for second-order Markov random fields. Larger systems, whether or not they are on a regular array, can be broken down into subsystems which are conditioned by their current boundary values, so that block Gibbs samplers (e.g. Besag et al., 1995, §2.4.5) can be devised. Finally, in §5, we show that, for multivariate binary distributions satisfying total positivity (Karlin and Rinott, 1980), our implementation of block Gibbs satisfies the monotonicity condition of Propp and Wilson (1996), so that perfect block samplers can be constructed. For example, this applies to autologistic distributions with nonnegative interactions.

2 An algorithm for Markov random fields

In this section, we derive an algorithm for evaluating the individual terms on the right-hand side of (2) as functions of the full conditionals $\{\pi(x_k|x_{<k})\}$ for $k = 1, \dots, n$. We achieve this by interpreting $\pi(x_i|x_{<i})$ as the full conditional of X_i in the Markov random field induced by marginalizing over $X_{>i}$ and using the following simple result.

Lemma. Let U , V and W denote discrete random variables. Then, in an obvious notation,

$$\text{pr}(u|v) = \left\{ \sum_w \text{pr}(w|u, v) / \text{pr}(u|v, w) \right\}^{-1}, \quad (3)$$

provided the conditional probabilities are well defined.

If we now interpret the random variable V as a random vector, then the relevance of the lemma is that it identifies the full conditional of U , and similarly of any component, in the reduced system (U, V) from the full conditionals in the system (U, V, W) . For example, the term for $i = n - 1$ in (2) is obtained immediately from equation (3) by the substitution $U = X_{n-1}$, $V = X_{<n-1}$ and $W = X_n$. Of course, additional full conditionals in the system (X_1, \dots, X_{n-1}) are required to evaluate the term for $i = n - 2$ in (2), and so on, but these can also be obtained from (3). This process leads to a recursion for $\pi(\cdot)$, based on the following general result, which is an immediate consequence of the Lemma.

Theorem. Let $X = (X_1, \dots, X_n)$ denote a vector of discrete random variables with joint distribution $\{\pi(x) : x \in S\}$, defined in terms of its full conditionals and satisfying positivity. Write $X^k = (X_1, \dots, X_k)$ for $k = 1, \dots, n$, so that, for example, $x_{-i}^k = \{x_j : j \leq k, j \neq i\}$ for $i = 1, \dots, k$. Then

$$\pi(x_i|x_{-i}^{k-1}) = \left\{ \sum_{x_k} \pi(x_k|x_{<k}) / \pi(x_i|x_{-i}^k) \right\}^{-1}, \quad i = 1, \dots, k - 1, \quad (4)$$

for all $x \in S$.

As a trivial example, suppose that $n = 3$. First, we can evaluate

$$\begin{aligned} \pi(x_1|x_2) &= \left\{ \sum_{x_3} \pi(x_3|x_1, x_2) / \pi(x_1|x_2, x_3) \right\}^{-1} \\ \pi(x_2|x_1) &= \left\{ \sum_{x_3} \pi(x_3|x_1, x_2) / \pi(x_2|x_1, x_3) \right\}^{-1} \end{aligned}$$

for any x_1 and x_2 and then

$$\pi(x_1) = \left\{ \sum_{x_2} \pi(x_2|x_1) / \pi(x_1|x_2) \right\}^{-1}.$$

Note that quite generally the indexing of the n random variables is arbitrary and that, as with (1), the $n!$ available choices may lead to factorizations that look distinct but which, of course, must lead to the same joint distribution.

In the worst case, no conditional independence relations hold between X_1, \dots, X_n and there is no preferred indexing. Then, assuming for simplicity that each S_i has the same cardinality s , the number of operations required to evaluate $\pi(x)$ is of order s^n and, as one would expect, there is no benefit. However, when the Markov random field has a simplified structure, judicious indexing leads to massive reductions in the computational load, as we demonstrate in §3 and §4.

Nevertheless, it is unlikely that $\pi(x)$ can be evaluated explicitly and so we turn now to the intermediate task of constructing block Gibbs samplers. Such samplers are intended to increase mobility around S by updating sets of conditionally dependent X_i 's rather than single components. Without loss of generality, consider any particular block B of b components, labelled $i = 1, \dots, b$. Then the analogue of (2) is

$$\pi(x_B|x_{-B}) = \prod_{i=1}^b \pi(x_i|x_{<i}, x_{-B}), \quad (5)$$

which can be evaluated from the full conditionals as above. For a simple example, suppose that $n = 5$ and that we require

$$\pi(x_1, x_2, x_3|x_4, x_5) = \pi(x_1|x_4, x_5) \pi(x_2|x_1, x_4, x_5) \pi(x_3|x_1, x_2, x_4, x_5).$$

Then, as one of six possible decompositions, we have

$$\pi(x_1|x_4, x_5) = \left\{ \sum_{x_2} \pi(x_2|x_1, x_4, x_5) / \pi(x_1|x_2, x_4, x_5) \right\}^{-1},$$

with

$$\pi(x_1|x_2, x_4, x_5) = \left\{ \sum_{x_3} \pi(x_3|x_1, x_2, x_4, x_5) / \pi(x_1|x_2, x_3, x_4, x_5) \right\}^{-1}$$

and a corresponding formula for $\pi(x_2|x_1, x_4, x_5)$. We return to block Gibbs in §5.

3 Hidden Markov chains

Hidden Markov chains are relevant in quite diverse areas, including speech recognition, neurophysiology, protein modelling and climatology; see, for example, MacDonald and Zucchini (2000). Here we use hidden Markov chains to illustrate the ease with which the theorem in §2 can be applied. Thus, consider random variables X_1, \dots, X_n , where X_1 has distribution $\{p_1(x_1)\}$ and the subsequent X_2, \dots, X_n follow a Markov chain with transition probabilities $\{p(x_i|x_{i-1})\}$. The complication in a hidden Markov chain is that the outcomes x_1, \dots, x_n cannot be observed and only a degraded version y_i of each x_i is available, generated via conditional probabilities $\{f(y_i|x_i)\}$. It is implicit in our formulation that $p(\cdot|\cdot)$ and $f(\cdot|\cdot)$ are time homogeneous but this is merely a notational convenience. Also, although we appear to assume that $p_1(\cdot)$, $p(\cdot|\cdot)$ and $f(\cdot|\cdot)$ are known, they can change dynamically as part of a wider simulation algorithm.

The standard tool for analyzing hidden Markov chains is the Baum et al. (1970) algorithm. The original goal was to evaluate $\text{pr}(y)$ but the underlying recursions are also used for restoration and for simulation, whether in a frequentist or Bayesian paradigm; see, for example, Robert et al. (2000). Here we focus on the calculation of $\{\pi(x|y) : x \in S\}$, where π now refers to the distribution of X given $Y = y$. Equation (2) is replaced by

$$\pi(x|y) = \pi(x_1|y) \prod_{i=2}^n \pi(x_i|x_{i-1}, y), \quad x \in S, \quad (6)$$

using the Markov property. Elementary probability manipulations and the Markov structure imply that

$$\pi(x_1|y) \propto f(y_1|x_1) \text{pr}(y_{>1}|x_1) p_1(x_1), \quad (7)$$

$$\pi(x_i|x_{i-1}, y) \propto f(y_i|x_i) \text{pr}(y_{>i}|x_i) p(x_i|x_{i-1}), \quad i = 2, \dots, n, \quad (8)$$

where proportionality is with respect to the first argument, and with the convention that $\text{pr}(y_{>n}|x_n) \equiv 1$. The Baum et al. (1970) algorithm uses the recursion

$$\text{pr}(y_{>i}|x_i) = \sum_{x_{i+1}} \text{pr}(y_{>i+1}|x_{i+1}) f(y_{i+1}|x_{i+1}) p(x_{i+1}|x_i), \quad (9)$$

evaluated in reverse order $i = n - 1, \dots, 1$, then substitutes into (7) and (8), and finally normalizes the results with respect to each x_i . Note that dummy renormalisations are

generally required during the backwards recursion because $\text{pr}(y_{>i}|x_i)$ becomes vanishingly small as i decreases. Equations (7), (8) and (9) enable X given y to be simulated via (6) and other quantities of interest to be calculated. For further details, see MacDonald and Zucchini (2000, Ch. 2 and Appendix A).

We now consider the analogue of the above results, based on the lemma in §2. If, in equation (3), we condition on y throughout, let $U = X_i$, $V = X_{i-1}$ and $W = X_{i+1}$, and apply the Markov property to $\pi(x_{i+1}|x_{i-1}, x_i, y)$, we immediately obtain the backwards recursion,

$$\pi(x_i|x_{i-1}, y) = \left\{ \sum_{x_{i+1}} \pi(x_{i+1}|x_i, y) / \pi(x_i|x_{i-1}, x_{i+1}, y) \right\}^{-1}, \quad (10)$$

for $i = n - 1, \dots, 2$, with

$$\pi(x_1|y) = \left\{ \sum_{x_2} \pi(x_2|x_1, y) / \pi(x_1|x_2, y) \right\}^{-1}. \quad (11)$$

In both (10) and (11), the denominator in the summation can be evaluated explicitly in terms of $f(\cdot)$, $p(\cdot)$ and p_1 because it depends on y only through y_i and so

$$\pi(x_i|x_{i-1}, x_{i+1}, y_i) \propto f(y_i|x_i) p(x_{i+1}|x_i) p(x_i|x_{i-1}),$$

for $i = 2, \dots, n - 1$, and

$$\pi(x_1|x_2, y_1) \propto f(y_1|x_1) p(x_2|x_1) p_1(x_1),$$

with normalizing constants found by summing the right-hand sides over x_i . The number of floating point operations required in calculating the data-dependent conditional probabilities (6) is of order n , as it is in the Baum et al. (1970) algorithm. The multiplicative factor is generally greater but the new recursion is more direct, does not require dummy renormalisations and generalizes very easily, as we illustrate below.

Suppose, for example, that the underlying Markov chain is second-order rather than first-order. Then the theorem in §2 immediately yields the backwards recursion,

$$\begin{aligned} \pi(x_i|x_{i-2}, x_{i-1}, y) &= \left\{ \sum_{x_{i+1}} \pi(x_{i+1}|x_{i-1}, x_i, y) / \pi(x_i|x_{i-2}, x_{i-1}, x_{i+1}, y) \right\}^{-1} \\ \pi(x_i|x_{i-2}, x_{i-1}, x_{i+1}, y) &= \left\{ \sum_{x_{i+2}} \pi(x_{i+2}|x_i, x_{i+1}, y) / \pi(x_i|x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}, y) \right\}^{-1} \end{aligned}$$

in which

$$\pi(x_i|x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}, y) \propto f(y_i|x_i) p(x_{i+2}|x_i, x_{i+1}) p(x_{i+1}|x_{i-1}, x_i) p(x_i|x_{i-2}, x_{i-1}).$$

Now suppose, in addition, that y_i depends on x_{i-2} and x_{i-1} , as well as x_i . Then the general recursion is unchanged and $f(y_i|x_i)$ in the final equation is merely replaced by

$$f(y_{i+2}|x_i, x_{i+1}, x_{i+2}) f(y_{i+1}|x_{i-1}, x_i, x_{i+1}) f(y_i|x_{i-2}, x_{i-1}, x_i).$$

Other forms of generalization are also available but below we discuss applications to Markov random fields.

4 Markov random fields on regular lattices

Consider a rectangular lattice, with r rows and c columns, the rc sites of which are labelled lexicographically. Suppose that, associated with each site i , there is a random variable X_i . We assume that $X = (X_1, \dots, X_{rc})$ is a first-order Markov random field (Besag, 1974), so that the full conditional of each X_i depends only on the x_j 's at sites directly adjacent to i . In practice, X might be conditioned on a fixed or, in block Gibbs sampling, a temporarily fixed border.

It follows from the global Markov property (Besag, 1974) that $X_{\geq i}$, given $X_{< i}$, depends only on $X_{\rho i}$, where ρi comprises the c sites previous to i , except for simple further reductions when i is in the first and last rows. Therefore, the same is true of X_i itself and equation (2) becomes

$$\pi(x) = \prod_{i=1}^{rc} \pi(x_i|x_{\rho i}). \quad (12)$$

We evaluate the terms on the right-hand side of (12) in reverse lexicographic order, in which case $\pi(x_i|x_{\rho i})$ is determined by applying (4) successively for $k = i + c, \dots, i + 1$, except for reductions in row r . To be more explicit, consider any i not in the first or last rows or columns. First note that each use of (4) requires $2s$ floating point operations.

When $k = i + c$,

$$\pi(x_i|x_i^{i+c-1}) \equiv \pi(x_i|x_{i-c}, x_{i-1}, x_{i+1}, \dots, x_{i+c-1}),$$

so that a total of $2s^{c+3}$ operations is involved in calculating all possible values of the left-hand side. However, for each of the other $c - 1$ values of k ,

$$\pi(x_i|x_i^{k-1}) \equiv \pi(x_i|x_{i-c}, x_{k-c}, \dots, x_{i-1}, x_{i+1}, \dots, x_{k-1}),$$

involving the lesser number $2s^{c+2}$ of operations. Hence, the total number of operations in evaluating all terms on the right-hand side of (12) is bounded above by $2rc(c+s-1)s^{c+2}$. This also applies when there is a border. Note that, if $r < c$, it is generally better to interchange rows and columns. We have used the algorithm for binary systems with $c \leq 12$. For large binary systems, one can work with say 10×10 blocks.

The algorithm extends immediately to second-order Markov random fields, in which the full conditional of each X_i depends additionally on the x_j 's at sites diagonally adjacent to i . Then ρ_i in (12) comprises the $c + 1$ sites previous to i , except for reductions in the first and last rows and in the first column. Now $\pi(x_i|x_{\rho_i})$ is determined by applying (4) for $k = i + c + 1, \dots, i + 1$, apart from reductions in special cases. Finally, note that versions for other lattices can be constructed, especially for the first-order field on a hexagonal lattice.

5 Perfect block Gibbs for binary distributions

In standard Markov chain Monte Carlo algorithms, it is often difficult to determine an adequate burn-in period during which output should be discarded. In principle, the problem can be eliminated using Propp and Wilson's (1996) "coupling from the past". In practice, the method is limited by its reliance on a stringent monotonicity condition and by the possible need for a prohibitive run length. Nevertheless, Propp and Wilson (1996) show that their methodology can be used in several important cases, including both the Gibbs sampler and Sweeny's (1983) algorithm applied to the two-dimensional ferromagnetic Ising model, even at its critical temperature. More generally, their method extends to any autologistic distribution (Besag, 1974) having non-negative interactions; and, indeed, to the Gibbs sampler for any binary distribution that satisfies what is known as the FKG condition (Fortuin, Kasteleyn and Ginibre, 1971) in statistical physics and as multivariate total positivity (Karlin and Rinott, 1980) in statistics.

Definition. Let $\pi(\cdot)$ denote a (discrete) distribution over a sample space S . Then $\pi(\cdot)$ satisfies multivariate total positivity if, for any pair of vectors $y, z \in S$,

$$\pi(\min\{y, z\}) \pi(\max\{y, z\}) \geq \pi(y) \pi(z), \quad (13)$$

where the minimum and maximum functions act elementwise.

It is known (Karlin and Rinott, 1980) that, for binary distributions, (13) is equivalent to the requirement that the full conditional distribution of each pair of components, given any particular values of the others, has a non-negative log-odds ratio. This is clearly satisfied by but is not restricted to any binary distribution having non-negative interaction parameters in its Bahadur expansion, including autologistic distributions with non-negative interactions. It is also standard, and can be shown directly with some care, that, if $\pi(\cdot)$ satisfies (13), then so does any marginalization of $\pi(\cdot)$. This is crucial below. Note here that marginalizations of autologistic distributions are not generally autologistic themselves.

Although the monotonicity of the Gibbs sampler for binary distributions satisfying (13) is implicit in Propp and Wilson (1996), we briefly recall the main idea. For n components in $\pi(\cdot)$, imagine 2^n runs of the Gibbs sampler, in which each run is initialized by a different configuration but starts with the same random seed. Suppose that the usual inverse distribution function method is employed to update each variable. Then (13) implies that, at any stage, a component that takes the value 1 in the run initialized by all 0's must also take the value 1 in all other runs at that stage; and correspondingly a component that takes the value 0 in the run initialized by all 1's must also take the value 0 in every run. This monotonicity implies that all realizations must have coalesced by the time those initialized by all 0's and by all 1's are observed to do so. Thus, it is necessary merely to run the two extreme cases to determine the time at which all 2^n realizations are coalescent. Because this coupling time is random, it does not itself produce a random draw from the target distribution and it is this problem that can be fixed by coupling from the past. The required monotonicity condition remains the same.

Now suppose $\pi(\cdot)$ is a multivariate binary distribution satisfying (13) and that, despite monotonicity or indeed because of it, a single-component Gibbs sampler produces

coalescence times that are prohibitive. For applications in statistics, Sweeny's algorithm may either be irrelevant or counter-productive, in which case a block Gibbs sampler may provide a viable alternative, provided that block updates are reasonably efficient. In fact, since $\pi(\cdot)$ satisfies (13), so do the marginal distributions $\pi(x_{\leq i}, x_{-B})$, corresponding to the terms on the right-hand side of (5), and it follows that the block updates and the sampler itself inherit the required monotonicity.

Acknowledgements

We are grateful to the editor and a referee for their helpful suggestions. JB acknowledges support from the National Research Center for Statistics and the Environment at the University of Washington.

References

- Bahadur, R. R. (1960). Some approximations to the binomial distribution function. *Ann. of Math. Statist.* **31**, 43–54.
- Baum, L. E., Petrie, T., Soules, G. & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. of Math. Statist.* **41**, 164–71.
- Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc. B* **36**, 192–236.
- Besag, J. E. and Green, P. J. (1993). Spatial statistics and Bayesian computation (with Discussion). *J. R. Statist. Soc. B* **55**, 25–37, 53–102.
- Besag, J. E., Green, P. J., Higdon, D. M. & Mengersen, K. L. (1995). Bayesian computation and stochastic systems (with Discussion). *Statist. Sci.* **10**, 3–66.
- Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*

51, 481–3.

Fortuin, C. M., Kasteleyn, P. W. & Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22**, 89–103.

Frank, O. & Strauss, D. (1986). Markov graphs. *J. Am. Statist. Assoc.* **81**, 832–42.

Karlin, S. & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Mult. An.* **10**, 467–498.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press.

MacDonald, I. L. & Zucchini, W. (2000). *Hidden Markov and Other Models for Discrete-valued Time Series*. London: Chapman and Hall.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. of Chem. Phys.* **21**, 1087–92.

Newman, M. E. J. & Barkema, G. T. (1999). *Monte Carlo Methods in Statistical Physics*. Oxford: Clarendon Press.

Propp, J. G. and Wilson, B. M. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.

Robert, C. P., Rydén, T. & Titterton, D. M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. R. Statist. Soc. B* **62**, 57–75.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with Discussion). *Journal of the Royal Statistical Society B*, **55**, 39–52.

Sweeny, M. (1983). Monte Carlo study of weighted percolation clusters relevant to the Potts model. *Physical Review Letters*, **58**, 86–88.