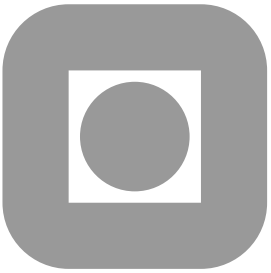
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Recursive computing for Markov random fields

by

Nial Friel and Håvard Rue

PREPRINT
STATISTICS NO. 3/2005



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2005/S3-2005.pdf
Håvard Rue has homepage: http://www.math.ntnu.no/∼hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

Recursive computing for Markov random fields

Nial Friel* and Håvard Rue†

August 29, 2005

Abstract

We present a recursive algorithm to compute a collection of normalising constants which can be used in a straightforward manner to sample a realisation from a Markov random field. Further we present important consequences of this result which renders possible tasks such as maximising Markov random fields, computing marginal distributions, exact inference for certain loss functions and computing marginal likelihoods.

Some key words: Autologistic distribution; exact sampling; hidden Markov random field; Ising model; normalising constant.

1 Introduction

Markov random fields (MRFs) play an important role in spatial statistics, for example, as prior distributions in image analysis. Early developments in Markov chain Monte Carlo arose from the need to sample from the Ising model, which is an important example of a Markov random field. Markov random fields are difficult to handle however since exact calculation of the normalisation constant is generally unavailable. Numerous approximate schemes have been presented in the literature, see for example, Geyer and Thompson (1992), Gelman and Meng (1998), Huang and Ogata (1999), Gu and Zhu (2001). More recently efficient schemes have appeared which allow normalising constants to be evaluated exactly for MRFs defined on relatively small lattices, most notably Bartolucci and Besag (2002) and Reeves and Pettitt (2004).

In this paper we present an algorithm, inspired by the recursive algorithm in Reeves and Pettitt (2004), which yields a collection of normalising constants, which can used in a very straightforward way to allow direct sampling of the entire lattice. This new method preserves the computational complexity of the algorithm presented by Reeves and Pettitt (2004). We argue that this new result opens doors to many possibilities - computing marginal distributions, exact inference for certain loss functions, computing modes of an MRF and computing marginal likelihoods for hidden MRF models. Exact inference is possible for binary lattices of relatively small size, with around 20 rows or columns, but we contend that these results are natural building blocks for carrying out inference for large lattices.

*Address for correspondence: Department of Statistics, University of Glasgow, Glasgow G12 8QW, UK. Email: nial@stats.gla.ac.uk

†Department of Mathematical Sciences, Norwegian University of Science and Technology.

2 Markov random fields

Consider a random process x defined on a rectangular $m \times m'$ lattice, with $m \leq m'$, where each lattice point takes values $\{-1, 1\}$. Define an index $i \in \{1, 2, \ldots, n\}$, $n = mm'$, where lattice points are ordered from top to bottom in each column and columns from left to right. The autologistic model (Besag 1974) may be defined as

$$\pi(x|\beta) \propto q(x|\beta) = \exp(\beta_0 V_0(x) + \beta_1 V_1(x)), \tag{1}$$

with normalising constant $z(\beta)$. Here,

$$V_0(x) = \sum_{i=1}^{n} x_i \text{ and } V_1(x) = \sum_{i \sim j} x_i x_j,$$

where '$i \sim j$' mean that x_j is one of the four directly adjacent neighbours of x_i, and further each neighbouring pair i, j enters the sum only once. The Ising model is a special case where β_0 takes the value 0. Many of the methods in the paper will also apply to the extended model

$$\pi(x|\beta) \propto \exp(\sum_i \beta_i x_i + \sum_{i \sim j} \beta_{ij} x_i x_j).$$

For ease of notation however we will concentrate on the case where the β_i's and β_{ij}'s both take constant values.

3 Generating samples from an MRF

Here we describe the means by which we may gather an exact sample from $\pi(x|\beta)$, however the same argument follows trivially if we condition on data y, where the likelihood of the data $\pi(y|x)$ preserves the Markov structure of x, for example if $\pi(y|x) = \prod_i \pi(y_i|x_i)$.

3.1 Background

The Markov property allows the un-normalised distribution in (1) to be written in a factorisable form as

$$q(x|\beta) = \prod_{i=1}^{n} q_i(x_i|x_{i+1:n}, \beta), \tag{2}$$

where

$$q_i(x_i|x_{i+1:n}, \beta) = \exp(\beta_0 x_i + \beta_1 x_i (x_{i+1} + x_{m+i})) \tag{3}$$

for all i, except when i corresponds to a lattice point on the last row, in which case,

$$q_i(x_i|x_{i+1:n}, \beta) = \exp(\beta_0 x_i + \beta_1 x_i x_{m+i}), \tag{4}$$

and finally when i corresponds to a point in the last column, where

$$q_i(x_i|x_{i+1:n}, \beta) = \exp(\beta_0 x_i + \beta_1 x_i x_{i+1}). \tag{5}$$

Throughout we use the notation $x_{i:j}$ to denote variables $\{x_i, \ldots, x_j\}$.

It is always possible to write the joint distribution of x as

$$\pi(x|\beta) = \pi(x_n|\beta) \prod_{i=1}^{n-1} \pi(x_i|x_{i+1:n}, \beta).$$

The Markov property implies that the conditional probability of each point x_i appearing on the right hand side above, just depends on the next m points in increasing index order

$$\pi(x_i|x_{i+1:n},\beta) = \pi(x_i|x_{i+1:m+i},\beta),$$

except for points x_i in the last column which just depend on the last $n-i$ points. It is possible to re-formulate this probability as

$$\pi(x_i|x_{i+1:n},\beta) = \frac{\pi(x_{i:n}|\beta)}{\pi(x_{i+1:n}|\beta)}. \tag{6}$$

Now,

$$\begin{aligned}\pi(x_{i:n}|\beta) &= \sum_{x_{1:i-1}} q(x|\beta)/z(\beta) \\ &= \sum_{x_{1:i-1}} q(x_{1:i-1}|x_{i:n},\beta)q(x_{i:n}|\beta)/z(\beta) \\ &= z_{i-1}(\beta,x_{i:n})\prod_{j=i}^{n} q_j(x_j|x_{j+1:n})/z(\beta). \end{aligned} \tag{7}$$

Here we introduce notation for the normalising constant $z_{i-1}(\beta,x_{i:n})$ indicating that variables $x_{1:i-1}$ have been summed out of $q(x_{1:i-1}|x_{i:n},\beta)$. We now re-write (6) using (7) as

$$\pi(x_i|x_{i+1:n},\beta) = \frac{q_i(x_i|x_{i+1:n},\beta)z_i(\beta,x_{i+1:n})}{z_{i-1}(\beta,x_{i:n})}. \tag{8}$$

In addition the marginal distribution for x_n can be written as,

$$\pi(x_n) = \frac{\exp(\beta_0 x_n)z_{n-1}(\beta,x_n)}{z(\beta)}. \tag{9}$$

Thus we see that to sample from the conditional distributions of $x_i|x_{i+1:n}$ requires knowledge of the corresponding normalising constants $z_i(\beta,x_{i+1:n})$. One of the main contributions of this paper is to show how these normalising constants can be calculated in an efficient manner.

3.2 The recursive algorithm

We propose a two-pass algorithm for sampling a realisation x. In the first pass we compute in turn each $z_i(\beta,x_{i+1:n})$, for $i=1,\ldots,n-1$, in a recursive manner, eventually computing $z(\beta)$. The second pass then amounts to sampling each lattice value from the conditional distributions in (8) and (9), using the normalising constants, z_i's calculated from the first pass.

Consider the following recursive scheme, which shares some of the aspects of the algorithm presented in Reeves and Pettitt (2004):

$$z_1(\beta,x_{2:n}) = \sum_{x_1} q_1(x_1|x_{2:n},\beta) \tag{10}$$

and

$$z_i(\beta,x_{i+1:n}) = \sum_{x_i} q_i(x_i|x_{i+1:n},\beta)z_{i-1}(\beta,x_{i:n}) \tag{11}$$

for $i = 2, n-m$. Here note that q_i is as defined in (3) for $i = 1, 2, \ldots, n-m$, except when lattice point x_i corresponds to the last row of the lattice, when $i = m, 2m, \ldots, n-m$, in which case q_i is as defined in (4). We complete the recursive scheme by computing

$$z_i(\beta, x_{i+1:n}) = \sum_{x_i} q_i(x_i|x_{i+1:n}, \beta) z_{i-1}(\beta, x_{i:n}), \tag{12}$$

for $i = n-m+1, \ldots, n$, where now q_i follows (5). Note that z_n is in fact the normalising constant for the complete lattice.

Note that this recursive scheme differs from that presented in Reeves and Pettitt (2004), principally in how the functions $q_i(\cdot)$ in (3), (4) and (5) above, are defined. In fact this algorithm could also be similarly extended to allow direct sampling however this possibility is not explored by Reeves and Pettitt (2004). Figure 1 illustrates graphically elements of the recursive scheme above.

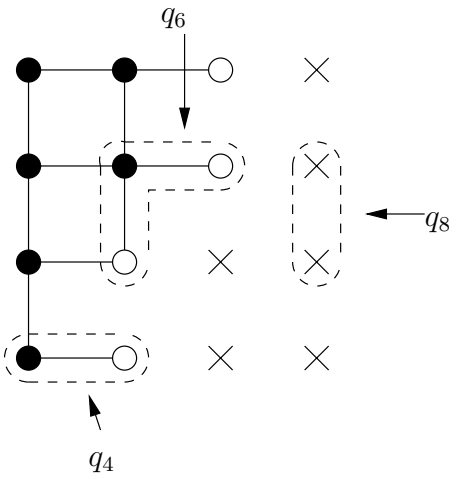


Figure 1: A 4×4 lattice where solid circles indicate nodes which appear in the summation $z_6(\beta, x_{7:10}) = \sum_{x_{1:6}} q(x_{1:6}|\beta, x_{7:10})$ and empty circles indicate conditioned points. Interaction terms from the V_1 statistics are indicated with solid lines, while abundance terms from the V_0 statistic are indicated with solid circles. In addition functions $q_i(\cdot)$ are highlighted for corresponding lattice points x_i.

The algorithm we propose can be thought of as a forward-backward algorithm. We first move in increasing index order through the lattice points calculating normalising constants, z_i, and then, once these have been calculated, we move in decreasing index order sampling each lattice point. Note that this algorithm can be naturally extended to MRFs with more than two state values, for example to a Potts model with unordered state variables, with obvious increased computational costs. In addition it is possible to extend the algorithm to larger neighbourhoods. For example if a second-order neighbour structure is used, where also diagonal interactions are included, then this would require computing normalising constants of lattice points conditional on the next $m+1$ values. In this case

$$q_i(x_i|x_{i+1:n}, \beta) = \exp(\beta_0 x_i + \beta_1 x_i (x_{i+1} + x_{m+i-1} + x_{m+i} + x_{m+i+1}))$$

with modifications for lattice points on the last row or column, see Figure 2.

The subsequent sections of this paper illustrate what is possible once exact sampling and normalising constants z_i, for $i = 1, 2, \ldots, n$ are available.

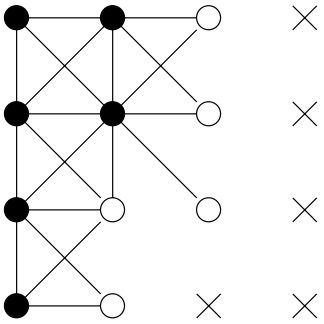


Figure 2: A 4×4 lattice illustrating the lattice points involved in the normalising constant $z_6(\beta, x_{7:11}) = \sum_{x_{1:6}} q(x_{1:6}|\beta, x_{7:11})$ for a second order neighbourhood. Again solid and empty circles represent terms in the summation and conditioned points respectively. Solid circles and straight lines represent abundance terms and interaction terms respectively.

4 Consequences and extensions of the methodology

4.1 Hidden Markov random fields

Suppose a MRF x with parameters β is corrupted by some observational noise process leaving data y. Typically it is assumed that the data y conditional on x preserves the Markov structure of x. The aim is to estimate all unknown quantities, see for example Friel, Pettitt, Reeves and Wit (2005).

Consider the posterior marginal distribution for β. It holds that

$$\pi(\beta|y) = \frac{\pi(x, \beta|y)}{\pi(x|\beta, y)},$$

for any realisation x. We can write this marginal distribution, up to proportionality constant, as

$$\pi(\beta|y) \propto \frac{\pi(y|x)\pi(x|\beta)\pi(\beta)}{\pi(x|\beta, y)}, \tag{13}$$

where the constant of proportionality is $\pi(y)$, the marginal likelihood of the data. Note however that every factor in the right-hand side of (13) is now available. In particular it is possible to sample x from $\pi(\cdot|\beta)$ and then to compute $\pi(x|\beta)$ and $\pi(x|\beta, y)$. Thus (13) can be used to estimate an unnormalised version of $\pi(\beta|y)$, which can then be normalised using numerical integration, provided the dimension of β is small. An estimate of the marginal likelihood, $\hat{\pi}(y)$ obtains by integrating the right hand side of (13) with respect to β.

It is possible to check the error in estimating $\hat{\pi}(\beta|y)$ and $\hat{\pi}(y)$ at any value β^* by re-using equation (13) by defining

$$\text{error}(\beta^*) = \frac{\pi(y|x)\pi(x|\beta^*)\pi(\beta^*)}{\hat{\pi}(\beta^*|y)\hat{\pi}(y)\pi(x|\beta^*, y)} - 1.$$

Note that, for this extension, exact sampling is not strictly needed. The reader is referred to Rue, Steinsland and Erland (2004) for an analysis of a similar approximation in the context of Gaussian Markov random fields.

Of course, computing the marginal likelihood allows the possibility to compute posterior model probabilities, which thus far for Markov random fields has received little attention in the literature. However the reader is referred to Robert, Ryden and Titterington (2000), where the number of latent states is a variable and the latent process is a Markov chain. We believe that the methodology presented here is a promising area of future research.

4.2 Computing marginal distributions

Recall from Section 3 that the recursive algorithm allows calculation of conditional probabilities $\pi(x_{1:i}|x_{i+1:n}, \beta)$. It follows then that it is possible to compute the joint distribution $\pi(x_{i+1:n}|\beta)$ since it holds that

$$\pi(x_{i+1:n}|\beta) = \frac{\pi(x|\beta)}{\pi(x_{1:i}|x_{i+1:n}, \beta)}, \tag{14}$$

for any configuration $x_{1:i}$. Consider now carrying out the forward pass in the recursive algorithm, but moving in decreasing index order through the lattice, from $i = n, n-1, \ldots, 2$. This computes conditional probabilities $\pi(x_{i+1:n}|x_{1:i}, \beta)$, and similarly to (14) allows joint probabilities $\pi(x_{1:i}|\beta)$ to be evaluated. This now makes it is possible to compute joint distributions

$$\pi(x_{i:m+i-1}|\beta) = \frac{\pi(x_{1:m+i-1}|\beta)}{\pi(x_{1:i-1}|x_{i:n}, \beta)},$$

since all probabilities on the right hand side above can be evaluated. Now marginal distributions of, for example, single lattice points x_i are available by summing $\pi(x_{i:m+i-1})$ over $x_{i+1:m+i-1}$. Indeed it is possible to calculate marginal distributions of $\pi(x_i, x_j|\beta)$ and thus $E(x_i x_j|\beta)$ for all pairs of neighbours $i \sim j$.

Moreover, returning to the context of hidden Markov random fields, note that it is in fact possible to estimate posterior marginals, for example,

$$\pi(x_i, x_j|y) = \int_\beta \pi(x_i, x_j|\beta, y)\pi(\beta|y)\, d\beta,$$

since estimates of $\pi(\beta|y)$ are available from (13). Numerical integration yields the estimate.

4.3 Computing the modal configuration of $\pi(x|\beta, y)$

Consider again the scenario presented in Section 4.1. Suppose that it was also of interest to compute the posterior mode of $\pi(x|\beta, y)$. An exact algorithm is addressed by Greig, Porteous and Seheult (1989). This algorithm can be implemented for larger lattices, but it is restricted to the case where the interaction parameters β_{ij} are all positive. The exact method which we now present poses no such restrictions.

For brevity of notation we will describe an algorithm for calculating the mode of $x|\beta$. However it will be seen to apply similarly if in addition we condition on y as would be the case for a hidden MRF. Consider the following recursive scheme, which shares much of the same flavour of that presented in Section 3. Define

$$f_2(x_{2:n}|\beta) = \max_{x_1} q_1(x_1|x_{2:n}, \beta) \tag{15}$$

and

$$f_i(x_{i:n}|\beta) = \max_{x_{i-1}} \left\{ f_{i-1}(x_{i-1:n}|\beta)\; q_{i-1}(x_{i-1}|x_{i:n}, \beta) \right\}, \tag{16}$$

for $i = 3, 4, \ldots, n$ where in each case above, q_i is as defined in equations (3), (4) and (5). Notice that

$$f_i(x_{i:n}|\beta) = \max_{x_{1:i}} q(x_{1:i}|x_{i+1:n}, \beta).$$

Note that for numerical stability it is preferable to work with $\log(q_i)$, replacing products above by sums. In effect these recursions give the probability of the modal configuration. A

backward recursive step is needed to actually find the modal configuration, which we denote by $\hat{x}$:

$$\hat{x}_n = \arg\max_{x_n} f_n(x_n|\beta)$$

and

$$\hat{x}_i = \arg\max_{x_i} f_i(x_i|\hat{x}_{i+1:n}, \beta),$$

for $i = n-1, n-2, \ldots, 1$. It is possible that expressions (15) and (16) do not yield a unique maximum over x_i, in which case a unique global maximum would not exist. Nevertheless it is possible to record where such ties occur and sample uniformly over these values in the backward step, with added computational costs ensuing.

4.4 Exact Bayesian inference for loss functions

It is well known that the posterior mean estimator corresponds to a mean squared loss function. In terms of visual restoration of a hidden binary image x, this means that inference is based on finding an image x which satisfies

$$\arg\min_z E_{x|y,\beta} L(x, z),$$

where $L(x, z) = \sum_i (x_i - z_i)^2$. In other words only error measured pointwise is accounted for, resulting potentially in a loss of fidelity in the restored image. The maximum a posteriori estimator corresponds to a zero-one loss function $L(x, z) = 0$, if $x = z$ and 1 otherwise. In both of these cases, spatial information modelled by the latent process $x|\beta$ is not accounted for in the loss function. See Rue (1995) for more details. However elements of the previous methodologies can all be brought together to allow exact inference to be carried out for improved loss functions.

Consider a loss function of the type

$$L(x, z) = \sum_i (x_i - z_i)^2 + \lambda \sum_{i \sim j} (x_i - z_i)(x_j - z_j), \tag{17}$$

which accounts for the error at each lattice point, together with a penalty for the simultaneous error at neighbouring lattice points. Following Section 4.2 it is clear that $E(x_i - z_i)$ and $E(x_i - z_i)(x_j - z_j)$ can both be calculated as a function of z, where the expectation is with respect to the distribution $x|\beta, y$. We outline this briefly. Define $M_i = Ex_i$ and $M_{ij} = Ex_i x_j$. Then for the loss function in (17) the expected posterior loss appears as,

$$EL(x, z) = \sum_i (M_{ii} - 2z_i M_i + 1) + \lambda \sum_{i \sim j} (M_{ij} - z_j M_i - z_i M_j + z_i z_j).$$

The point to note is that each of M_i and M_{ij} can be calculated exactly following the discuss in Section 4.2. Further the algorithm described in Section 4.3 can be used to minimise the expect loss, thus giving an exact optimal estimator for the given loss function.

Suppose now that we wish to find the optimal estimator for the joint distribution $x|y$ by maximising the corresponding expected loss. Here we can use ideas from Section 4.1. In particular,

$$E_{x|y}(x_i) = \int_\beta E_{x|\beta,y}(x_i)\ \pi(\beta|y)\ d\beta,$$

where following (13), we can estimate $\pi(\beta|y)$ on the right hand side above. We can similarly estimate $E_{x|y}(x_i x_j)$. Minimisation of the expected posterior loss follows as before giving an approximate optimal estimator. All of the above extends easily to the case where the latent process takes more than two colours.

5 An illustrative example

We consider a dataset consisting of measurements of soil phosphate content on a 16×16 grid at 10 metre intervals at a location in Greece. The dataset can be found in Buck, Cavanagh and Litton (1988) and was analysed in Besag, York and Mollié (1991).

We assume that the data y hides an underlying MRF, x. The aim here is to make inference on the order of the hidden MRF, either a first or a second order Ising model, using the methodology outlined in Section 4.1. Specifically we model the latent process x as

$$\pi(x|\beta) \propto \exp(\beta \sum_{i \sim j} x_i x_j),$$

where, as before, the notation $\sim$ denotes that locations i and j are neighbours of each other. We consider two possibilities:

Model $k = 1$: A first order neighbourhood where each point x_i has as neighbours the four nearest adjacent points.

Model $k = 2$: A 2nd order neighbourhood structure where in addition to the first order neighbours, the four nearest diagonal points also belong to the neighbourhood.

Following Besag *et al* (1991) we assume that the y's are conditionally independent given the x's and have normal distributions with mean $\mu(x_i)$ and common variance ν. It of course possible to also include μ as parameters in the analysis, but for ease of illustration we assume $\mu(-1) = 4$ and $\mu(1) = 4.5$. In fact these are values also chosen by Besag *et al* (1991). Our aim is now, following (13) to compute the right hand side of

$$\pi(\beta, \nu|y, k) \propto \frac{\pi(y|x, \nu)\pi(x|\beta, k)\pi(\beta)\pi(\nu)\pi(k)}{\pi(x|\beta, \nu, y, k)} \tag{18}$$

for $k = 1, 2$. A diffuse zero mean Gaussian prior was specified for β and a diffuse gamma prior for ν. The dataset contain 9 missing values, however it is entirely straightforward to modify the algorithm to compute $z(\beta, y, \nu, k)$, the normalising constant for the distribution $\pi(x|\beta, \nu, y, k)$ in (18).

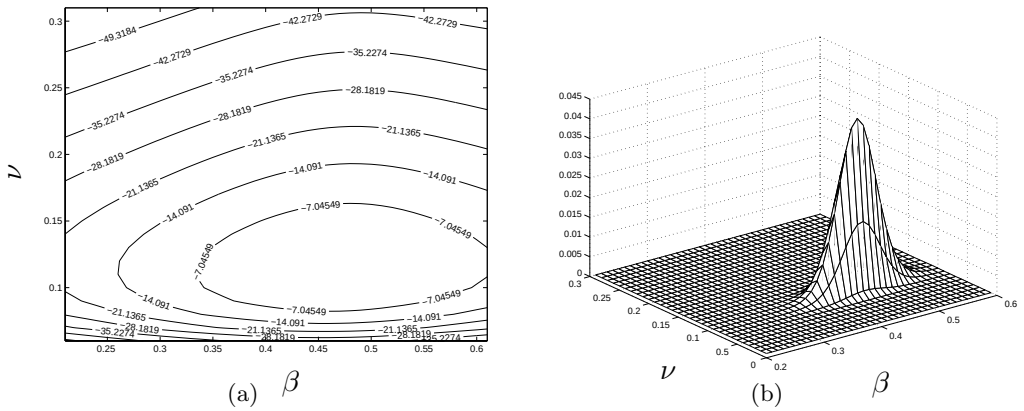


Figure 3: (a) Contour plot of un-normalised marginal distribution of $\log \pi(\beta, \nu|y, k = 1)$ with maximum value scaled to 0. (b) A normalised version of (a).

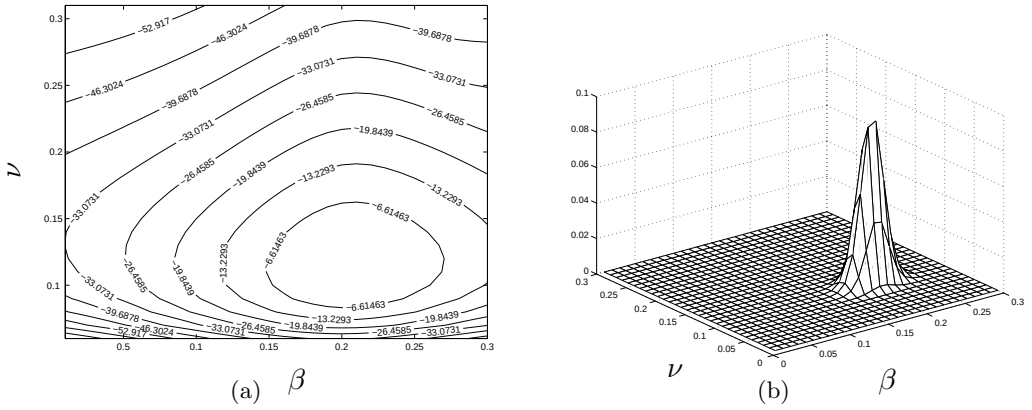


Figure 4: (a) Contour plot of un-normalised marginal distribution of $\log \pi(\beta, \nu|y, k = 2)$ scaled so that maximum value equals 0 (b) A normalised version of (a).

The un-normalised distribution $\pi(\beta, \nu|y, k)$ was evaluated at equal spaced points 0.01 apart in the (β, ν)-plane. Contour plots of the log un-normalised marginal distribution $\pi(\beta, \nu|y, k)$ and surface plots of $\pi(\beta, \nu|y, k)$ are presented in Figure 3 and Figure 4 for $k = 1, 2$ respectively. Both of these plots show that the posterior surface appears quite smooth and that most of the posterior mass is concentrated on quite a small region in the (β, ν)-plane. Maximum a posteriori parameters values were found at $(0.46, 0.10)$ for model $k = 1$ and at $(0.21, 0.10)$ for $k = 2$. Numerical integration of the un-normalised distribution $\pi(\beta, \nu|y, k)$ with respect to β and ν for $k = 1, 2$ yields the following marginal likelihoods estimates: $\log \pi(y|k = 1) = -110.168$ and $\log \pi(y|k = 2) = -114.075$. Assuming both models are equally weighted, a priori, yields posterior model probabilities, $\pi(k = 1|y) = 0.98$ and $\pi(k = 2|y) = 0.02$.

6 Discussion

This article has explained how it is possible to sample exactly from an Markov random field, using a recursive algorithm. Equally importantly we have shown that through knowledge of conditional normalising constants, and indeed knowledge of the normalising constant itself further exact inference and calculation is possible - computation of marginal distributions, including marginals of all pairs of variables entering the joint distribution of the Markov random field; exact calculation of the mode of the Markov random field; approximate inference for the posterior marginal of $\beta|y$; exact Bayesian inference for loss functions.

All of the methods presented in this article are restricted to relatively small lattices with the smaller dimension not greater than 20. This is important in its own right, but also as natural building blocks for inference for larger sized datasets. For example, our sampling algorithm could be embedded into a block-MCMC algorithm where each block is updated exactly conditional on all the remaining sites, the normalising constant is approximated similarly using a blocked version of pseudolikelihood, the marginal likelihood calculations is implemented using the normalising constant approximation and where marginal distributions for x_i and (x_i, x_j) for $i \sim j$ are approximated using a small lattice around each site. Finally the modal configuration could be approximated by successive maximisations of overlapping blocks similar to blocked MCMC. More sophisticated approximations are indeed possible.

Acknowledgements Nial Friel wishes to thank the Department of Mathematical Sciences, Norwegian University of Science and Technology for its hospitality during April 2005.

References

Bartolucci, F. and J. Besag (2002), a recursive algorithm for Markov random fields. *Biometrika* **89**, 724–730

Besag, J., J. York and A. Mollié (1991), Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59

Besag, J. E. (1974), Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* **36**, 192–236

Buck, C. E., W. G. Cavanagh and C. D. Litton (1988), The spatial analysis of site phosphate data. In S. Rahtz (ed.), *Computer Applications and Quantitative Methods in Archaeology*, pp. 151–160, British archaeological reports, International series 446

Friel, N., A. Pettitt, R. Reeves and E. Wit (2005), Comparison of MCMC methods for hidden Markov random field models for binary data. Technical report, University of Glasgow

Gelman, A. and X.-L. Meng (1998), Simulating normalizing contants: from importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185

Geyer, C. J. and E. A. Thompson (1992), Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B* **54**, 657–699

Greig, D., B. Porteous and A. Seheult (1989), Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B* **51**, 271–279

Gu, M. G. and H.-T. Zhu (2001), Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society, Series B* **63**, 339–355

Huang, F. and Y. Ogata (1999), Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics* **8**, 510–530

Reeves, R. and A. N. Pettitt (2004), Efficient recursions for general factorisable models. *Biometrika* **91**(3), 751–757

Robert, C. P., T. Ryden and D. M. Titterington (2000), Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B* **62**(1), 57–75

Rue, H. (1995), New loss functinos in Bayesian imaging. *Journal of the American Statistical Association* **90**, 900–908

Rue, H., I. Steinsland and S. Erland (2004), Approximating hidden Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B* **66**, 877–892