

Recursive computing and simulation-free inference for general factorizable models

BY NIAL FRIEL

Department of Statistics, University of Glasgow, Glasgow G12 8QW, U.K.
nial@stats.gla.ac.uk

AND HÅVARD RUE

*Department of Mathematical Sciences, Norwegian University of Science and
Technology, N-7491 Trondheim, Norway*
havard.rue@math.ntnu.no

SUMMARY

We illustrate how the recursive algorithm of Reeves & Pettitt (2004) for general factorizable models can be extended to allow exact sampling, maximization of distributions and computation of marginal distributions. All of the methods we describe apply to discrete-valued Markov random fields with nearest neighbour integrations defined on regular lattices; in particular we illustrate that exact inference can be performed for hidden autologistic models defined on moderately sized lattices. In this context we offer an extension of this methodology which allows approximate inference to be carried out for larger lattices without resorting to simulation techniques such as Markov chain Monte Carlo. In particular our work offers the basis for an automatic inference machine for such models.

Some key words: Autologistic distribution; Exact sampling; Hidden Markov random field; Ising model; Normalizing constant.

1. INTRODUCTION

Numerous examples exist of distributions for which calculation of the normalization constant is intractable, rendering the joint likelihood unavailable. This is often the situation in spatial statistics, where the autologistic distribution (Besag, 1974) is one such example. Schemes for approximating the normalizing constant include Geyer & Thompson (1992), Gelman & Meng (1998), Huang & Ogata (1999) and Gu & Zhu (2001). More recently, efficient schemes have appeared which allow the joint likelihood to be computed for small lattices, most notably Bartolucci & Besag (2002) and Reeves & Pettitt (2004). The work in Bartolucci & Besag (2002) allows the joint distribution to be calculated when its full conditional distributions are specified, whereas the work of Reeves & Pettitt (2004) works directly with the unnormalized distribution and presents a recursive algorithm for calculating the corresponding normalizing constant. The purpose of our paper is to consider extensions of the approach of Reeves & Pettitt (2004).

2. GENERAL FACTORIZABLE MODELS

2.1. *Review of general factorizable models*

Here we present a brief review of the work by Reeves & Pettitt (2004) and suggest an approach which allows exact sampling and further important extensions. Throughout the paper we use the notation $x_{i:j}$ to denote variables (x_i, \dots, x_j) . Consider an unnormalized distribution of a discrete-valued vector $x = (x_1, \dots, x_n)$ given by $q(x)$, and suppose that $q(x)$ can be factorized as

$$q(x) = \prod_{i=1}^k q_i(x_{i:i+r}),$$

where $k = n - r$ and $r < n$. This is called a lag- r model. The normalizing constant, z , of q can be calculated recursively from

$$z_1(x_{2:r+1}) = \sum_{x_1} q_1(x_{1:r+1}), \quad (1)$$

$$z_i(x_{i+1:i+r}) = \sum_{x_i} q_i(x_{i:i+r}) z_{i-1}(x_{i:i+r-1}), \quad i = 2, \dots, k, \quad (2)$$

$$z = \sum_{x_{k+1:n}} z_k(x_{k+1:n}). \quad (3)$$

Reeves & Pettitt (2004) also illustrate how the stochastic backwards algorithm (Scott, 2002) can be applied to a general factorizable model to calculate the joint likelihood in terms of a product of conditional probabilities,

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

in the following recursive manner:

$$\pi(x_{k+1:n}) = \frac{z_k(x_{k+1:n})}{z}, \quad (4)$$

$$\pi(x_i | x_{i+1:i+r}) = \frac{q_i(x_{i:i+r}) z_{i-1}(x_{i:i+r-1})}{z_i(x_{i+1:i+r})}, \quad i = k, k-1, \dots, 2, \quad (5)$$

$$\pi(x_1 | x_{2:r+1}) = \frac{q_1(x_{1:r+1})}{z_1(x_{2:r+1})}. \quad (6)$$

2.2. *Exact sampling*

Reeves & Pettitt (2004) point out that only a forwards recursion is needed to calculate z , and hence the joint likelihood. However we observe, that if in the forwards recursion (1)–(3) each of the conditional normalizing constants, $z_i(x_{i+1:i+r})$, is stored, then these may be used to draw values from the conditional distributions given in equations (4)–(6) in reverse index order, and this in turn yields an exact sample x , a point evidently missed by Reeves & Pettitt. However we illustrate, in the subsequent sections, that many more statistical tasks are also possible.

2.3. Computing the modal configuration of $\pi(x)$

Consider the following recursive scheme, which is of the same flavour as that presented above. Define

$$f_2(x_{2:r+1}) = \max_{x_1} q_1(x_{1:r+1}), \tag{7}$$

$$f_i(x_{i:r+i-1}) = \max_{x_{i-1}} \{f_{i-1}(x_{i-1:r+i-2}) q_{i-1}(x_{i-1:r+i-1})\}, \quad i = 3, \dots, k + 1. \tag{8}$$

These recursions can be used to give the probability of the modal configuration, which we denote by \hat{x} . This probability can be written as

$$p(\hat{x}) = \frac{\max_{x_{k+1:n}} f_{k+1}(x_{k+1:n})}{z}.$$

A backwards recursive step is needed to find the actual modal configuration:

$$\begin{aligned} \hat{x}_{k+1:n} &= \arg \max_{x_{k+1:n}} f_{k+1}(x_{k+1:n}), \\ \hat{x}_i &= \arg \max_{x_i} f_i(x_i, \hat{x}_{i+1:r+i-1}), \quad i = k, k - 1, \dots, 2, \\ \hat{x}_1 &= \arg \max_{x_1} q_1(x_1, \hat{x}_{2:r+1}). \end{aligned}$$

It is possible that expressions (7) and (8) 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 backwards step.

2.4. Computing marginal distributions

It is possible to compute the joint distribution $\pi(x_{i:n})$ from the forwards-backwards recursions using the stored z_i 's since it holds that

$$\pi(x_{i:n}) = \sum_{x_{1:i-1}} q(x)/z = \frac{z_{i-1}(x_{i:i+r-1}) \prod_{j=i}^k q_j(x_{j:j+r})}{z}. \tag{9}$$

However, the conditional probability $\pi(x_{1:i-1}|x_{i:i+r})$ is also available since, for any $x_{i:n}$,

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

and both terms on the right-hand side are available. Consider now carrying out the forward pass in the recursive algorithm, but moving in decreasing index order through the lattice, for $i = n, n - 1, \dots, 2$. Similarly to (9) this allows marginal probabilities $\pi(x_{1:i})$ to be computed. This in turn allows computation of

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

since all probabilities on the right-hand side above can be evaluated. Now marginal distributions of, for example, variables x_i are available by summing $\pi(x_{i:i+r})$ over $x_{i+1:i+r}$. We remark that our algorithm is useful in practice since it amounts to reusing the same algorithm twice, first a forwards recursion in increasing index order, and then a forwards recursion in decreasing index order.

3. APPLICATION TO AUTOLOGISTIC MODELS

3.1. *Autologistic models written as general factorizable models*

Consider a random process x defined on a rectangular $m \times m'$ lattice, with $m \leq m'$, where each lattice point takes a value from $\{-1, 1\}$. Define an index $i \in \{1, \dots, n\}$, $n = mm'$, 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

$$\pi(x|\beta) \propto \exp\left(\sum_i \beta_i x_i + \sum_{i \sim j} \beta_{ij} x_i x_j\right), \quad (10)$$

with normalizing constant $z(\beta)$. Here ' $i \sim j$ ' means 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. A more parsimonious and widely used version of the autologistic model is

$$\pi(x|\beta) \propto q(x|\beta) = \exp\left(\beta_0 \sum_i x_i + \beta_1 \sum_{i \sim j} x_i x_j\right). \quad (11)$$

The Ising model is a special case in which $\beta_0 = 0$. In practice, a noisy version, y , of the autologistic distribution is often observed. Here it is typically assumed that the data y are conditionally independent given x ; that is,

$$\pi(y|x) = \prod_i p(y_i|x_i). \quad (12)$$

The full conditional for the hidden field x is then given by

$$\pi(x|\beta, y) \propto \exp\left(\sum_i \{\log p(y_i|x_i) + \beta_0 x_i\} + \beta_1 \sum_{i \sim j} x_i x_j\right). \quad (13)$$

It is straightforward to see that each of the models (11), (10) and (13) can be written as a general factorizable model, with optimal lag m . For example, the autologistic model in (11) can be written as

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

where each factor may be written as

$$q_i(x_{i:i+m}, \beta) = \exp\{\beta_0 x_i + \beta_1 x_i(x_{i+1} + x_{i+m})\} \quad (14)$$

for all i , except when i corresponds to a lattice point on the last row or last column, in which case x_{i+1} or x_{m+1} , respectively, drops out of (14). Factors for the autologistic model (10) and its hidden version (13) can be similarly described. In effect, expressing the unnormalized autologistic distribution as a lag- m model allows the 2-dimensional field to be viewed as a linearly ordered process. In terms of implementing the algorithm described earlier, in the forwards step we compute normalizing constants, $z_i(\beta, x_{i+1:i+m})$, of the unnormalized distribution $q(x_{1:i}|x_{i+1:i+m}, \beta)$, for all possible realizations of the conditioned values $x_{i+1:i+m}$, for $i = 1, \dots, n - m$, and $x_{i+1:n}$, for $i = n - m + 1, \dots, n - 1$. For an autologistic model, this number of realizations totals $(n - m)2^m + \sum_{i=1}^{m-1} 2^i$. Clearly, for the forwards step, the collection of z_i 's must be stored, and this determines the computational complexity of the algorithm. For our C computer code implementation of the sampling algorithm in §2.2, we are able to sample lattices where m is at most 19. For example, to sample a lattice of size 19×19 takes 100 seconds of computation time on a 2 Ghz laptop.

3.2. Marginal distributions for the hidden autologistic model

Consider a hidden autologistic distribution, x , described in equations (12) and (13). The posterior marginal distribution for β is

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

for any realization x . This is nothing more than the ‘candidate’s formula’ (Besag, 1989). We can write this marginal distribution, up to a proportionality constant, as

$$\pi(\beta|y) \propto \frac{\pi(y|x)\pi(x|\beta)\pi(\beta)}{\pi(x|\beta, y)}, \tag{15}$$

where $\pi(y)$, the marginal likelihood of the data, is the normalizing constant for the right-hand side of (15). Note however that every factor on the right-hand side of (15) is now available. In particular it is possible to compute the maximal realization x from $\pi(x|\beta, y)$, or indeed to fix x to any configuration, and then to compute $\pi(x|\beta)$ and $\pi(x|\beta, y)$. Thus (15) can be used to compute exactly the unnormalized version of $\pi(\beta|y)$ at selected values β_l , for $l = 1, \dots, L$, and they can then be used to estimate the normalized posterior marginal of β , $\hat{\pi}(\beta|y)$ say, using numerical integration, provided the dimension of β is small. An estimate of the marginal likelihood, $\hat{\pi}(y)$, is given by

$$\hat{\pi}(y) = \sum_{l=1}^L \frac{\pi(y|x)\pi(x|\beta_l)\pi(\beta_l)}{\pi(x|\beta_l, y)}.$$

It is possible to check the error in estimating $\hat{\pi}(\beta|y)$ and $\hat{\pi}(y)$ at any value β^* by reusing equation (15) by defining

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

We make the important observation that we estimate the marginal distribution, $\hat{\pi}(\beta|y)$, and the estimated marginal likelihood, $\hat{\pi}(y)$, without recourse to any simulation techniques, such as Markov chain Monte Carlo. Indeed it is clear that we can obtain increasingly better estimates of $\pi(\beta|y)$ and $\pi(y)$ which will always outperform their Markov chain Monte Carlo counterparts by choosing a finer grid of β values. A sensible strategy therefore might be first to choose a coarse grid of β values, and then to choose a fine grid of β values in the regions of $\pi(\beta|y)$ which contain most probability mass. The reader is referred to Rue et al. (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 of computing posterior model probabilities, which until now for autologistic models has received little attention in the literature. However, the reader is referred to Robert et al. (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.

We remark that the exact sampling algorithm in §2.2 allows sampling from $\pi(x|\beta, y)$. In addition the algorithm in §2.3 can be easily extended to find the modal configuration of $\pi(x|\beta, y)$. Finally the ideas in §2.4 can be extended to estimate posterior marginals, such as

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

since estimates of $\pi(\beta|y)$ are available from (15). Numerical integration yields the estimate.

3.3. 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 noisy image y , this means that inference is based on finding an image z 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

$$L(x, z) = \sum_i (x_i - z_i)^2 + \lambda \sum_{i \sim j} (x_i - z_i)(x_j - z_j), \quad (16)$$

which accounts for the error at each lattice point, together with a penalty for the simultaneous error at neighbouring lattice points. From §2.4 it is clear that $E(x_i - z_i)$ and $E(x_i - z_i)(x_j - z_j)$ can each be calculated as a function of z , where the expectation is with respect to the distribution $x|\beta, y$. We outline this briefly. Define $M_i = E(x_i)$ and $M_{ij} = E(x_i x_j)$. Then for the loss function in (16) 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 discussion in §2.4. Furthermore, the algorithm described in §2.3 can be used to minimize the expected 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 maximizing the corresponding expected loss. Here we can use ideas from §3.2. In particular,

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

in which, following (15), we can estimate $\pi(\beta|y)$ on the right-hand side above. We can similarly estimate $E_{x|y}(x_i x_j)$. Minimization of the expected posterior loss follows as before giving an approximately optimal estimator. All of the above extends easily to the case where the latent process takes more than two colours, up to a maximum of four colours for practical purposes.

4. ILLUSTRATIVE EXAMPLES

4.1. Archaeology dataset: a lattice of moderate size

Here we consider a dataset for which our methods can be carried out exactly. The data consist 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 et al. (1988) and was analyzed in Besag et al. (1991).

We assume that the data y hide an underlying Markov random field, x . The aim here is to make inference about the order of the hidden Markov random field, either a first- or a second-order Ising model, using the methodology outlined in §3.2. 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. The $k = 1$ model corresponds to a first-order neighbourhood where each point x_i has as neighbours the four nearest adjacent points. The $k = 2$ model includes a second-order neighbourhood structure in which, in addition to the first-order neighbours, the four nearest diagonal points also belong to the neighbourhood of a given point.

Following Besag et al. (1991) we assume that the y 's are conditionally independent given the x 's and have normal distributions with, for y_i , mean $\mu(x_i)$ and common variance κ . It is possible also to include the $\mu(x_i)$'s as parameters in the analysis, but for ease of illustration we assume that $\mu(-1) = 4$ and $\mu(1) = 4.5$. These values were chosen by Besag et al. (1991). Our aim is now, following (15), to compute the right-hand side of

$$\pi(\beta, \kappa|y, k) \propto \frac{\pi(y|x, \kappa)\pi(x|\beta, k)\pi(\beta)\pi(\kappa)\pi(k)}{\pi(x|\beta, \kappa, y, k)}, \tag{17}$$

for $k = 1, 2$. A diffuse zero-mean Gaussian prior was specified for β and a diffuse gamma prior for κ .

The unnormalized distribution $\pi(\beta, \kappa|y, k)$ was evaluated at equally spaced points 0.01 apart in the (β, κ) -plane. Surface plots of $\pi(\beta, \kappa|y, k)$ are presented in Fig. 1 for $k = 1, 2$ respectively.

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, κ) -plane. Maximum a posteriori values for (β, κ) were found at (0.46, 0.10) for model $k = 1$ and at (0.21, 0.10) for $k = 2$. Numerical integration of the unnormalized distribution $\pi(\beta, \kappa|y, k)$ with respect to β and κ for $k = 1, 2$ yields the following marginal likelihood 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 $\pi(k = 2|y) = 0.02$. Note that this dataset was analyzed by Besag et al. (1991) using model $k = 2$, but our analysis shows that model $k = 1$ would have sufficed.

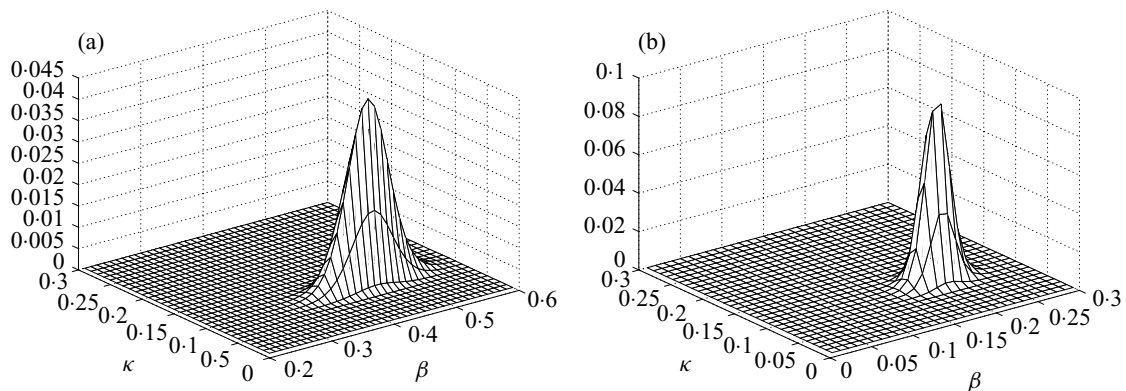


Fig. 1. Archaeology dataset. Normalized marginalized distribution of $\pi(\beta, \kappa|y, k)$ for (a) $k = 1$ and (b) $k = 2$.

4.2. Time course microarray dataset: a lattice of large size

In this example our methods offer approximate inference. Gene expressions were measured across the whole genome of *Plasmodium falciparum*, the organism that causes human malaria, for 46 1-hour consecutive intervals, in a recent experiment carried out by Bozdech et al. (2003) The experiment was conducted over the complete asexual intraerythrocytic development cycle in order to establish which genes might be potential drug targets for deregulating the organism in order to prevent malaria. The *Plasmodium falciparum* genome consists of fourteen linear chromosomes, a circular genome and a linear mitochondrial genome. This example focuses on the relatively short mitochondrial chromosome, which consists of 72 genes and about which relatively little is known.

The data, y , consist of observations on a 46×72 spatial-temporal rectangular lattice, and y_{tg} is the log-expression of the g th gene at time point t . Figure 2 displays the data y .

Here we assume that the data hide a lattice of latent states x modelled as a nonhomogeneous Ising distribution with 2 states $\{-1, 1\}$ corresponding to ‘up-regulation’ and ‘down-regulation’. Thus the likelihood of x given model parameters β is

$$\pi(x|\beta) \propto \exp\{\beta_t V_t(x) + \beta_g V_g(x)\},$$

where $V_t(x)$ measures the interactions between neighbouring lattice points corresponding to the same gene in the ‘time’ direction, while $V_g(x)$ similarly measures interactions at the same time point between neighbouring genes. The parameters β_t and β_g allow the strength of the interaction to be not the same in both directions. Of course other models could be proposed to capture more information, for instance extending this model to a 3-state model including a state of ‘no differential expression’. However, Bozdech et al. (2003) suggest that a vast majority of the genes are active, and so we ignore this possibility here. Further extensions might include time- or gene-specific interaction parameters.

Here the size of the lattice on which the data are defined is too large for the exact results to apply, and we offer the following approximation to $\pi(x|\beta)$ and $\pi(x|\beta, y)$. We describe the approximation for $\pi(x|\beta)$; the approximation for $\pi(x|\beta, y)$ follows trivially. Given the

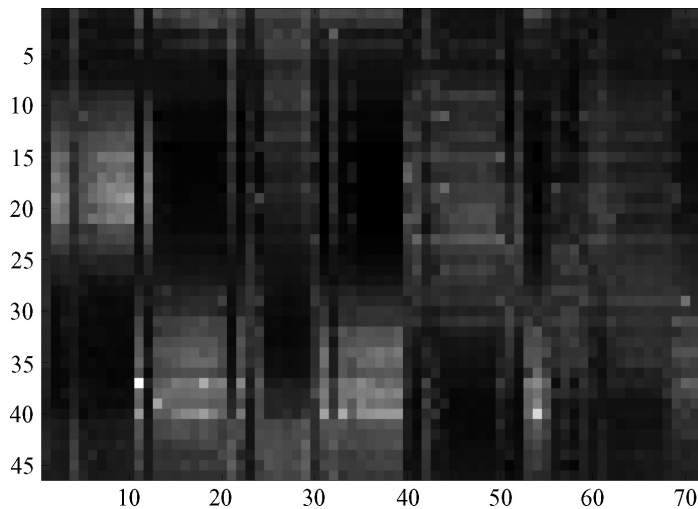


Fig. 2. Log-differential expressions levels for the mitochondrial genome across 46 1-hour time intervals. Columns are genes and rows are time-points.

marginal distribution of a column, x_c , of a lattice, x , the sublattices immediately to the left and right of x_c , which we denote by x_A and x_B respectively, are conditionally independent, so that

$$\pi(x|\beta) = \pi(x_A|\beta, x_c)\pi(x_c|\beta)\pi(x_B|\beta, x_c).$$

Assume for the moment that we can calculate normalizing constants for the unnormalized distributions, $q(x_A|\beta, x_c)$ and $q(x_B|\beta, x_c)$. The problem remains to compute the marginal distribution of the column x_c . We suggest the following strategy. Consider a sublattice, x^* , covering x_c , where x_S and x_T are the sublattices of x^* immediately to the left and right of x_c , respectively. Here the lattice points of x^* are partitioned into x_S , x_T and x_c . We use the approximation

$$\pi(x_c|\beta) = \frac{\pi(x|\beta)}{\pi(x_A|x_c, \beta)\pi(x_B|x_c, \beta)} \tag{18}$$

$$\simeq \frac{\pi(x^*|x_{c1}, x_{c2}, \beta)}{\pi(x_S|x_c, x_{c1}, \beta)\pi(x_T|x_c, x_{c2}, \beta)}, \tag{19}$$

where x_{c1} and x_{c2} are the columns immediately to the left and right of the sublattice x^* , respectively. Figure 3 illustrates this idea graphically.

Note that in computing $\pi(x|\beta)$ or $\pi(x|\beta, y)$ the only approximation we make is for the marginal distribution of the column of points. This approach can be extended straightforwardly to larger lattices by partitioning the lattice into sublattices separated by single columns and/or rows of lattice points, where the marginal distribution of each intermediate column and/or row is estimated similarly to the above idea. Of course more sophisticated approximations are possible.

Here we investigate the performance of the approximation for a small dataset for which we can carry out exact computations. A realization from an isotropic Ising model with parameter $\beta = 0.4$ defined on a lattice of size 19×19 was exactly sampled using the recursion method described earlier. Gaussian noise with zero mean and unit variance was added to each state value. Here we can compute the unnormalized marginal distribution $\pi(\beta|y)$ exactly for a fine grid of β values leading to a very precise estimate of the normalized posterior marginal distribution for β . Our interest is to see how our approximations to $\pi(x|\beta)$ and $\pi(x|\beta, y)$ perform. Using the above approximation, we partition the 19×19 lattice into 2 sublattices of size 19×9 , separated by a single column of lattice points. The lattice x on which we condition (19) was a thresholded version, at threshold level 0, of y . The approximation results from approximating the marginal distribution of this column. Following (19), we cover this column with a sublattice of size 19×5 . In this case the

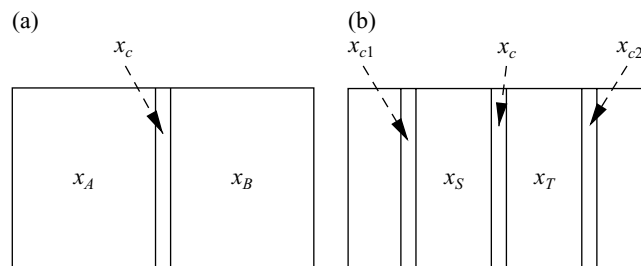


Fig. 3. Microarray example. (a) Displays the partition of x ; (b) displays the sublattice and columns used in the approximation of $\pi(x_c|\beta)$.

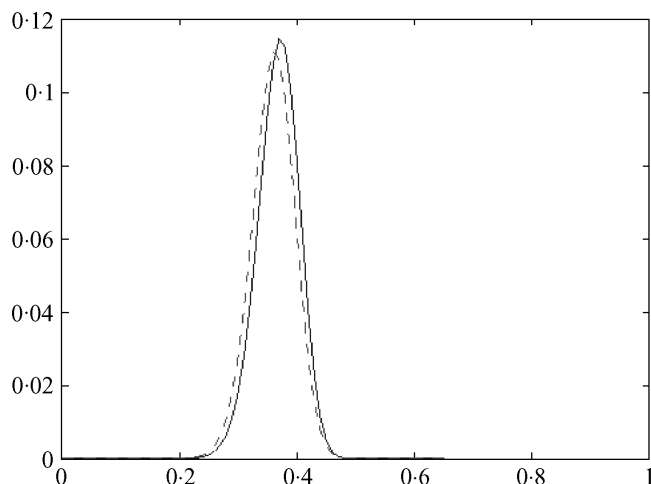


Fig. 4. Hidden Ising model. Posterior marginal $\pi(\beta|y)$. Solid line indicates the exact method, dashed line indicates the approximate method where a sublattice of size 19×5 covers the column of lattice points.

marginal column is covered by only 2 columns of lattice points on both sides, in addition to conditioning on bordering lattice points. This is quite a challenging situation, since in addition the value of $\beta = 0.4$ induces strong dependence through the Markov random field.

Figure 4 shows that the difference between the true posterior marginal for β and the approximation is negligible. Clearly, increasing the size of the covering sublattice improves the approximation. In fact, when the covering sublattice is of size 19×11 or bigger, the two distributions are indistinguishable.

For the gene-expression example, the likelihood of y given x is modelled as independent Gaussian noise, with a known mean μ_- or μ_+ conditional on the corresponding state variable being equal to -1 or $+1$ respectively, with common variance σ^2 . Here we fixed the values of μ_- and μ_+ at 0.8 and 2.0 , respectively, and we took $\sigma = 0.5$. Diffuse Gaussian zero-mean priors were chosen for each of the β parameters.

In order to compute the normalizing constants for the unnormalized distributions $q(x|\beta)$ and $q(x|\beta, y)$, the 46×72 lattice was partitioned into 3 disjoint sublattices of dimension 46×17 and a final sublattice of dimension 46×18 , each separated by a column of length 46. To compute the marginal distribution of the columns of lattice points, following (19), we used a lattice of size 17×46 to cover each column. Each of the covering sublattices used in the approximation (19) was taken from a thresholded version of y , where the threshold level is chosen as 1.4 , the midpoint of the underlying means of the noise distributions.

A grid of 80×40 equally spaced points in the β_t and β_g ranges from $[1.0, 1.8]$ and $[0.0, 0.2]$, respectively, was used to estimate $\pi(\beta|y)$. The surface plot of $\pi(\beta|y)$ is shown in Fig. 5.

Figure 6 displays posterior marginal distributions for each of β_t and β_g . Clearly this indicated a strong time interaction, which is unsurprising. However, of more biological interest perhaps is that the gene interaction parameter, β_g , shows a positive interaction, indicating that expression levels of genes are influenced by their geographical neighbours on the genome.

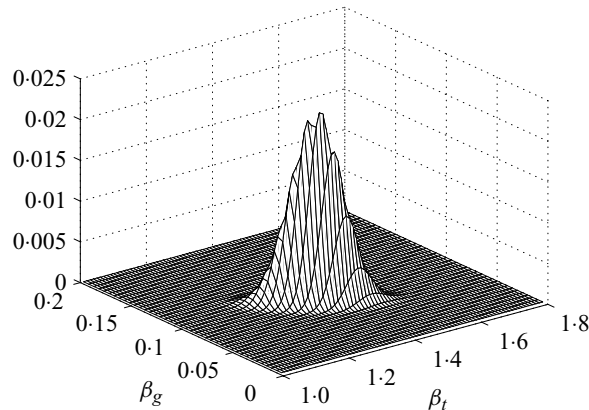


Fig. 5. Microarray example. Approximation to the distribution $\pi(\beta|y)$.

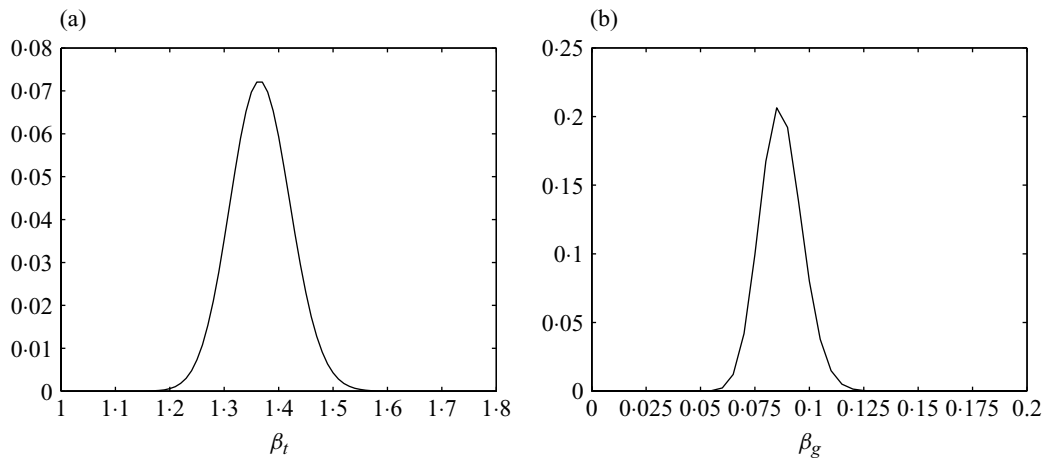


Fig. 6. Microarray example. (a) Plot of marginal posterior distribution $\pi(\beta_t|y)$ illustrating strong positive association between consecutive time points. (b) Plot of marginal posterior distribution $\pi(\beta_g|y)$ suggesting positive interaction of neighbouring genes on the genome..

5. DISCUSSION

The methods in this paper apply directly to distributions expressible as general factorizable models. Our methods do not apply to the situation in Bartolucci & Besag (2002) where the full conditional distributions are specified, but not the joint distribution. From the point of view of implementation, our methods are well suited to Markov random field models when the number of field parameters in β is small. In addition for such models, the factors $q_i(x_{i:i+m})$ of the unnormalized distribution are typically similar for each index i , and thus these may be pre-computed once, giving a large computational saving. Our algorithms have complexity $O(k^{m+1})$, where k is the number of latent states and m is the smaller of the number of rows or columns. For binary Markov random field models our algorithms are feasible for lattices of size up to $m = 19$, while this reduces to $m = 12$ and $m = 9$ for $k = 3$ and $k = 4$, respectively. Finally we note that our inference methods for Markov random field models provide an appealing simulation-free alternative to the usual Markov chain Monte Carlo approaches.

ACKNOWLEDGEMENT

Nial Friel wishes to thank the Department of Mathematical Sciences, Norwegian University of Science and Technology for their hospitality during April 2005. We wish to thank Tony Pettitt for comments on an earlier draft of this paper. We are grateful to anonymous referees for comments which have considerably improved the presentation of the paper.

REFERENCES

- BARTOLUCCI, F. & BESAG, J. E. (2002). A recursive algorithm for Markov random fields. *Biometrika* **89**, 724–30.
- 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. (1989). A candidate's formula: A curious result in Bayesian prediction. *Biometrika* **76**, 183.
- BESAG, J. E., YORK, J. C. & MOLLIE, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with Discussion). *Ann. Inst. Statist. Math.* **43**, 1–59.
- BOZDECH, Z., LLINÁS, M., PULLIAM, B. L., WONG, E. D., ZHU, J. & DERISI, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, 85–100.
- BUCK, C. E., CAVANAGH, W. G. & LITTON, C. D. (1988). The spatial analysis of site phosphate data. In *Computer Applications and Quantitative Methods in Archaeology*, International Series 446, Ed. S. Rahtz, pp. 151–60. Oxford: British Archaeological Reports.
- GELMAN, A. & MENG, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13**, 163–85.
- GEYER, C. J. & THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with Discussion). *J. R. Statist. Soc. B* **54**, 657–99.
- GU, M. G. & ZHU, H.-T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *J. R. Statist. Soc. B* **63**, 339–55.
- HUANG, F. & OGATA, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Comp. Graph. Statist.* **8**, 510–30.
- REEVES, R. & PETTITT, A. N. (2004). Efficient recursions for general factorisable models. *Biometrika* **91**, 751–7.
- ROBERT, C. P., RYDÉN, T. & TITTERINGTON, 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.
- RUE, H. (1995). New loss functions in Bayesian imaging. *J. Am. Statist. Assoc.* **90**, 900–8.
- RUE, H., STEINSLAND, I. & ERLAND, S. (2004). Approximating hidden Gaussian Markov random fields. *J. R. Statist. Soc. B* **66**, 877–92.
- SCOTT, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Am. Statist. Assoc.* **97**, 337–51.

[Received August 2005. Revised February 2007]