J. R. Statist. Soc. B (2003)
65, *Part* 1, *pp.* 235–246

Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice

A. N. Pettitt,

Queensland University of Technology, Brisbane, Australia

N. Friel

University of Glasgow, UK

and R. Reeves

Queensland University of Technology, Brisbane, Australia

[Received June 2000. Final revision July 2002]

Summary. Motivated by the autologistic model for the analysis of spatial binary data on the two-dimensional lattice, we develop efficient computational methods for calculating the normalizing constant for models for discrete data defined on the cylinder and lattice. Because the normalizing constant is generally unknown analytically, statisticians have developed various *ad hoc* methods to overcome this difficulty. Our aim is to provide computationally and statistically efficient methods for calculating the normalizing constant so that efficient likelihood-based statistical methods are then available for inference. We extend the so-called transition method to find a feasible computational method of obtaining the normalizing constant for the cylinder boundary condition. To extend the result to the free-boundary condition on the lattice we use an efficient path sampling Markov chain Monte Carlo scheme. The methods are generally applicable to association patterns other than spatial, such as clustered binary data, and to variables taking three or more values described by, for example, Potts models.

Keywords: Autologistic distribution; Gibbs distribution; Ising model; Markov chain Monte Carlo sampling; Path sampling; Potts model; Transition method

1. Introduction

The autologistic model of Besag (1972, 1974) is a popular choice for data analysis when a spatial component is involved, e.g. Preisler (1993), Augustin *et al.* (1996) and Wu and Huffer (1997). However, it is an awkward model because the normalizing constant is generally unknown analytically. To overcome this problem, approximate methods such as estimating equations, pseudo-likelihood or coding (Besag, 1986) are used. Our aim in this paper is to provide computationally and statistically efficient methods for calculating the normalizing constant, and hence efficient statistical methods for inference. The methods that we introduce can also be applied to related exponential family distributions, such as so-called quadratic exponential binary distributions (e.g. Zhao and Prentice (1990) and Molenberghs and Ryan (1999)), to association patterns other than spatial, such as clustered binary data, and to discrete or categorical variables taking three

Address for correspondence: A. N. Pettitt, School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane 4001, Australia.
E-mail: a.pettitt@fsc.qut.edu.au

 1369–7412/03/65235

or more values. For a non-spatial application of an exponential family model to deoxyribonucleic acid fingerprint data, Geyer and Thompson (1992) and discussants, Green (1992) and Smith (1992), give details of Markov chain Monte Carlo (MCMC) methods used for inference.

In statistical mechanics and statistical physics, where the normalizing constant is known as the partition or energy function, effort has been directed towards numerical schemes for its evaluation. Current statistical physics research concerns various non-regular shapes and boundary conditions for the related Ising model, e.g. Jensen *et al.* (1997). The torus boundary condition is extensively studied as it provides a stationary set-up for a finite lattice which is not available for the free-boundary condition (Besag, 1974). Here we consider the autologistic distribution on a regular lattice with two different boundary conditions, namely free and cylinder. For the cylinder boundary condition, the last column and first column are assumed neighbours. For the free-boundary condition, the case that is often considered in statistical applications, no such constraints are placed on boundary rows and columns.

We extend the transition method of Baxter (1982) for the Ising model to general discrete K-valued distributions, to find a feasible computational analytic matrix method of obtaining the normalizing constant for the cylinder condition. The matrix analytic method for an $m \times n$ (with m the smaller dimension) lattice has computational complexity $O(K^{3m})$ compared with direct computation which is $O(K^{mn})$. When n is large our method is $O(K^{2m})$.

We use the Monte Carlo computational scheme known as path sampling (e.g. Gelman and Meng (1998)) and analytic results for the cylinder to derive a highly efficient Monte Carlo scheme to obtain estimates of the normalizing constant for the lattice with the free-boundary condition. As a consequence, likelihood and Bayesian inference for the autologistic model and related exponential family distributions are feasible, avoiding the need for *ad hoc* methods. Our approach is distinct from the maximum likelihood method of Gu and Zhu (2001), where the computational challenge is to approximate derivatives of the log-normalizing constant by using MCMC methods. Our results should be useful in spatial data analysis, image processing and statistical modelling generally where autologistic or related models apply.

In Section 2 we review the autologistic distribution and path sampling. In Section 3 we give a proof for the normalizing constant matrix analytic result for the cylinder. In Section 4 we give efficiency results for the Gaussian conditional autoregressive model and show that the predicted orders of magnitude improvement in efficiency are realized. In Section 5 we demonstrate the efficiency of our method for the autologistic model and the efficiencies are found to exceed those for the Gaussian conditional autoregressive model. Section 6 gives an illustration of how the results could be used in statistical analysis and the paper concludes with a general discussion.

2. The autologistic model and path sampling

2.1. The autologistic model

Let $\mathbf{x} = \{x_{ij}\}, i = 1, \ldots, m, j = 1, \ldots, n$, be observations of a binary spatial process at mn sites on a regular lattice of size $m \times n$ taking values -1 or 1, rather than the usual 0 or 1. This allows for a more parsimonious parameterization and avoids problems of non-invariance when 0 and 1 are interchanged. We assume without loss of generality that m is less than or equal to n. We write the unnormalized autologistic distribution on the cylinder in exponential family form (Geyer and Thompson, 1992) as

$$\begin{aligned} q(\mathbf{x}|\Theta) &= \exp\{\Theta^{\mathrm{T}} V(\mathbf{x})\} \\ &= \exp\{\theta_0 V_0(\mathbf{x}) + \theta_{\mathrm{f}} V_{\mathrm{f}}(\mathbf{x}) + \theta_{\mathrm{c}} V_{\mathrm{c}}(\mathbf{x})\}. \end{aligned} \tag{1}$$

Here parameter $\Theta = (\theta_0, \theta_f, \theta_c)$ and sufficient statistic $V(\mathbf{x}) = (V_0(\mathbf{x}), V_f(\mathbf{x}), V_c(\mathbf{x}))$, with

$$V_0(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij},$$

$$V_f(\mathbf{x}) = \sum_{i=1}^{m-1} \sum_{j=1}^{n} x_{ij}x_{i+1,j} + \sum_{i=1}^{m} \sum_{j=1}^{n-1} x_{ij}x_{i,j+1},$$

$$V_c(\mathbf{x}) = \sum_{i=1}^{m} x_{i1}x_{in}.$$

$V_0(\mathbf{x})$ is the overall sum of variables. The first term of $V_f(\mathbf{x})$ is the sum of nearest neighbour products within columns, whereas the second term is the analogous sum within rows. The subscript 'f' denotes 'free-boundary' lattice points. Finally, $V_c(\mathbf{x})$ denotes the sum of neighbour products between the first and last columns. Here the subscript 'c' denotes 'cylindrical' lattice points. We could further generalize the model to allow for varying spatial association by multiplying each sum in $V_f(\mathbf{x})$ by a different parameter. Putting $\theta_c = 0$ gives the standard so-called isotropic autologistic model that is used by statisticians, the free-boundary model. Putting $\theta_c = \theta_f$ gives the cylinder condition, where the first and last columns of the lattice are neighbours. Extensions to a K-valued discrete variable $\mathbf{x}$ are also possible with $V_0(\mathbf{x})$ becoming a K-vector of counts of values of x_{ij} equal to the K possible values, and the $x_{ij}x_{i'j'}$ terms in $V_f(\mathbf{x})$ and $V_c(\mathbf{x})$ replaced by the indicator function $I(x_{ij}, x_{i'j'})$ giving 1 if $x_{ij} = x_{i'j'}$ and 0 otherwise. The normalizing constant is given by

$$z(\Theta) = \int_{\mathbf{x}} q(\mathbf{x}|\Theta)\, \mu(\mathrm{d}\mathbf{x}),$$

where μ is a measure, e.g. a counting measure. For the discrete autologistic model the integral becomes a sum over all possible outcomes of $\mathbf{x}$, which involves 2^{mn} terms. When $\theta_0 = 0$ and $\theta_c = 0$, equation (1) corresponds to the free-boundary Ising model.

2.2. Path sampling

Our goal is to compute the normalizing constant for the free-boundary case, i.e. the value $z(\theta_0, \theta_f = \theta_a, \theta_c = 0)$ for a fixed value of θ_0. For ease of notation we shall normally omit reference to θ_0 and refer to this by $z(\theta_a, 0)$.

The literature on finding normalizing constants for probability functions is substantial; see Evans and Swartz (2000), section 7.4.5, or especially Gelman and Meng (1998). Our approach is to note that

$$\log\left\{\frac{z(\theta_a, 0)}{z(0, 0)}\right\} = \int_0^{\theta_a} \mathbf{E}_{\mathbf{x}|(\theta_f, 0)} V_f(\mathbf{x})\, \mathrm{d}\theta_f; \tag{2}$$

see, for example, Ripley (1988), page 64, or Ogata (1989). The problem of evaluating this ratio, and hence $z(\theta_a, 0)$, amounts to evaluating an integral between the points (0, 0) and $(\theta_a, 0)$ in the (θ_f, θ_c) parameter space. $z(0, 0)$ is known, as this is the normalizing constant for the independent case. Any continuous path (a Feynman path) $\theta(t)$ in the (θ_f, θ_c) plane can also be used to calculate integral (2) in the θ_f-dimension. We first choose a path $\theta(t) = (\theta_f(t), \theta_c(t))$, for $t \in [0, 1]$ with $\theta(0) = (0, 0)$ and $\theta(1) = (\theta_a, 0)$. In this instance, equation (2) may be written as

$$\int_0^1 \mathbf{E}_{\mathbf{x}|\theta(t)} \left\{\frac{\mathrm{d}\theta_f(t)}{\mathrm{d}t} V_f(\mathbf{x}) + \frac{\mathrm{d}\theta_c(t)}{\mathrm{d}t} V_c(\mathbf{x})\right\} \mathrm{d}t.$$

Path sampling approximates this integral by drawing samples of $(\mathbf{x}, \Theta)$ along the path $\theta(t)$ and is explained in detail in Gelman and Meng (1998).

Crucially, however, having introduced θ_c, we may write

$$\log\{z(\theta_a, 0)\} = \log\left\{\frac{z(\theta_a, 0)}{z(\theta_a, \theta_a)}\right\} + \log\{z(\theta_a, \theta_a)\},$$

where $z(\theta_a, \theta_a)$ is the normalizing constant for the cylinder condition and can be calculated by using the computational analytic method to be described in Section 3. Selecting a path where $\mathrm{d}\theta_f(t)/\mathrm{d}t = 0$, we may write

$$\log\left\{\frac{z(\theta_a, 0)}{z(\theta_a, \theta_a)}\right\} = \int_{\theta_a}^{0} \mathbf{E}_{\mathbf{x}|(\theta_a, \theta_c)} V_c(\mathbf{x})\, \mathrm{d}\theta_c. \tag{3}$$

The major proposition of this paper is that use of the indirect path integral (3), combined with the matrix analytical result for the cylinder (Section 3), will yield computationally and statistically more efficient estimates of $\log z(\theta_a, 0)$ than the direct path integral (2). An indication of the efficiency of the indirect path can be obtained by considering the number of terms in the statistics which are averaged in the path sampling. The direct path involves the statistic $V_f(\mathbf{x})$, a sum of $(n-1)(m-1)$ terms. The indirect path involves $V_c(\mathbf{x})$, and thus a sum of m terms. Generally, therefore the variance of $V_c(\mathbf{x})$ will be somewhat less than the variance of $V_f(\mathbf{x})$ along their respective paths. If the terms x_{ij} and x_{kl} behave like independent terms then $\mathrm{var}\{V_f(\mathbf{x})\} = O(nm)$ and $\mathrm{var}\{V_c(\mathbf{x})\} = O(m)$ whereas, if the terms behave like positively correlated terms, $\mathrm{var}\{V_f(\mathbf{x})\} = O(n^2m^2)$ and $\mathrm{var}\{V_c(\mathbf{x})\} = O(m^2)$. It is only possible to use enumeration to investigate the situation for trivially small lattices. For larger lattices, we use two approaches. In Section 4, the equivalent Gaussian autoregressive model is used to investigate variances of equivalent statistics. In Section 5, multiple MCMC runs are used to estimate the variances.

3. Matrix methods for efficient normalizing constant calculation on the cylinder

We consider a general categorical K-valued distribution on an $m \times n$ lattice, with $m \leqslant n$, which can be factorized as a product of functions of adjacent columns, where the first and last columns are considered adjacent. We prove that the normalizing constant of such a distribution can be found by using computational matrix analytic methods. The autologistic model, defined by equation (1), as well as Potts models for categorical x_{ij}, where the sufficient statistics involve similarities between near neighbours on the array, are specific examples, as are K-valued association models in general. The method involves computing eigenvalues of a $K^m \times K^m$ matrix, so it is feasible if $K^m \lesssim 1024$. This is satisfied, for example, by $K = 2$ and $m = 10$, $K = 3$ and $m = 6$, and $K = 4$ and $m = 5$. The method of proof is motivated by Baxter (1982), chapter 7, where an exact result for the partition function for the two-parameter Ising model is obtained, involving analytic expressions for eigenvalues.

3.1. The theorem

Let $\mathbf{x} = \{x_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n\}$ be a K-valued array with x_{ij} taking K discrete values denoted by $B = \{b_1, b_2, \ldots, b_K\}$. Define an m-vector by $\mathbf{x}_j = (x_{1j}, \ldots, x_{mj})$, $j = 1, \ldots, n$. Let the set of all possible values of $\mathbf{x}_j$ be denoted by the set $A = \{a_1, \ldots, a_N\}$ with $N = K^m$. Define $\mathbf{x}_0$ to be identically equal to $\mathbf{x}_n$, imposing the cylinder boundary condition.

Theorem 1. Suppose that the unnormalized distribution $q(\mathbf{x}|\Theta)$ satisfies a factorization

$$q(\mathbf{x}|\Theta) = \prod_{j=1}^{n} h(\mathbf{x}_j, \mathbf{x}_{j-1}),$$

for a given positive real function $h(\cdot, \cdot)$ defined on the set $A \times A$. Then the normalizing constant for $q(\mathbf{x}|\Theta)$ is given by $\mathrm{tr}(Q^n)$ where Q is an $N \times N$ matrix with its kth row $(Q_{k1}, \ldots, Q_{kN})$ defined by

$$\{h(\mathbf{x}_1 = a_1, \mathbf{x}_0 = a_k), h(\mathbf{x}_1 = a_2, \mathbf{x}_0 = a_k), \ldots, h(\mathbf{x}_1 = a_N, \mathbf{x}_0 = a_k)\}$$

for $k = 1, \ldots, N$.

Proof. The normalizing constant for q is given by the sum over mn K-valued variables

$$\sum_{x_{11} \in B} \cdots \sum_{x_{mn} \in B} q(\mathbf{x}|\Theta).$$

We now partition the lattice into columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and, using the factorization of q, this sum then equals

$$\sum_{\mathbf{x}_1 \in A, \ldots, \mathbf{x}_n \in A} \prod_{j=1}^{n} h(\mathbf{x}_j, \mathbf{x}_{j-1}) \tag{4}$$

with $\mathbf{x}_0 = \mathbf{x}_n$. Write this in terms of the matrix Q and we obtain

$$\sum_{j_1=1, j_2=1, \ldots, j_n=1}^{N} Q_{j_0 j_1} Q_{j_1 j_2} \cdots Q_{j_{n-1} j_n}. \tag{5}$$

The cylinder condition $\mathbf{x}_0 = \mathbf{x}_n$ implies that $j_0 = j_n$, giving the required result that the normalizing constant is $\mathrm{tr}(Q^n)$. This completes the proof.

Remark 1. Computational simplification occurs because the matrix Q has strictly positive elements and therefore Q is irreducible. The Perron–Frobenius matrix theorem applies so Q can be diagonalized, $Q = H^{-1} D H$; see, for example, Cox and Miller (1965), section 3.10. Then $\mathrm{tr}(Q^n) = \mathrm{tr}(D^n)$. Thus all that is required is that the eigenvalues of Q be found.

Remark 2. The theorem applies to the distribution given by equation (1) but only if $\theta_f = \theta_c$ and equal to θ_1, say, in expression (4). We take $h(\cdot, \cdot)$ to be given by

$$h(\mathbf{x}_1, \mathbf{x}_0) = \exp\Big(\theta_0 \sum_{i=1}^{m} x_{i1} + \theta_1 \sum_{i=1}^{m-1} x_{i1} x_{i+1,1}\Big) \exp\Big(\theta_1 \sum_{i=1}^{m} x_{i0} x_{i1}\Big).$$

The first factor on the right-hand side above only involves $\mathbf{x}_1$ whereas the second involves both $\mathbf{x}_1$ and $\mathbf{x}_0$ or respectively within-column and between-column functions.

Remark 3. A general form for $h(\cdot, \cdot)$ is given by the factorization

$$h(\mathbf{x}_1, \mathbf{x}_0) = h_1(\mathbf{x}_1)\, h_2(\mathbf{x}_1, \mathbf{x}_0)$$

with h_1 giving within-column relationships and h_2 giving between-column relationships. The result can be applied more generally than just to spatial models where only near neighbours enter the relationships. For example, the array could represent clustered binary data $\mathbf{x}_j$ observed over time $j = 1, \ldots, n$ with

$$h_1(\mathbf{x}_1) = \exp\Big(\theta_0 \sum_{i=1}^{m} x_{i1} + \theta_1 \sum_{i=1}^{m-1} \sum_{i'=i+1}^{m} x_{i1} x_{i'1}\Big)$$

and

$$h_2(\mathbf{x}_1, \mathbf{x}_0) = \exp\Big(\theta_2 \sum_{i=1}^{m} x_{i0} x_{i1}\Big).$$

Here, h_1 gives relationships within a cluster and h_2 a relationship between a variable at two subsequent times. For example, Molenberghs and Ryan (1999) considered exponential family models for multivariate binary data.

Remark 4. To find the normalizing constant on the cylinder for the autologistic model, we need to find the elements of the diagonal matrix D or, equivalently, the eigenvalues of the $2^m \times 2^m$ transition matrix Q. For computational details see Press *et al.* (1992), sections 11.5 and 11.6, for reduction to Hessenberg form and then use of the QR algorithm. If a small set of the L largest eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_L$ (ordered in absolute value) suffices to determine $\mathrm{tr}(D^n)$ accurately by $\Sigma_j^L \lambda_j^n$ then these computations can be achieved in about $L\ O(2^{2m})$ number of operations. This would generally be the case for larger values of n where, for n tending to ∞, the value of L tends to 1.

4. Efficiencies of path sampling estimates for Gaussian autoregressions

The main proposition of this paper is that evaluating equation (3) by Monte Carlo sampling and using the cylinder result of Section 3 leads to far more efficient estimates of $\log\{z(\theta_a, 0)\}$ than evaluating equation (2) by Monte Carlo methods. However, to investigate how well we can estimate each expectation (other than by replicating MCMC chains) necessitates replacing the expectation by a variance, following Gelman and Meng (1998), section 4, equation (39). For the binary autologistic model, there is no analytic expression for the variance, necessitating multiple simulations. However, this is not so with the Gaussian conditional autoregressive model, e.g. Besag (1974) or Cressie (1993). In this case we can find explicit expressions for the variances of the sufficient statistics $V(\mathbf{x})$ by using standard multivariate Gaussian theory, and hence expressions for the variances of the path integrals (Gelman and Meng, 1998). Here we summarize the results, noting that the variance of the path sampling integrals are given by

$$\int_{\theta_a}^{0} \mathrm{var}_{\mathbf{x}|(\theta_a, \theta_c)} V_c(\mathbf{x})\ \mathrm{d}\theta_c$$

and

$$\int_{0}^{\theta_a} \mathrm{var}_{\mathbf{x}|(\theta_f, 0)} V_f(\mathbf{x})\ \mathrm{d}\theta_f,$$

and refer the reader to Pettitt and Friel (2002) for details.

The relative efficiency of the cylinder scheme compared with the free-boundary scheme can be defined as the ratio of the two average variances and is given in Table 1. The relative efficiency appears to depend only on the number of columns, with the relative efficiency being approximately equal to the number of columns for both choices of row size. The average variance for the cylinder path integral is independent of the number of columns, consistent with $V_c(\mathbf{x})$ being a sum over the first and last columns. The average variance for the free-boundary path integral increases linearly with the number of columns, n, for a given number of rows, consistent with $V_f(\mathbf{x})$ being a sum over the entire lattice. Since we expect a similar relationship between the path variances to hold for the autologistic model, we propose that substantially more efficient estimates of the autologistic normalizing constant can be made by using the cylinder path integral,

Table 1. Variances of path integrals for the Gaussian conditional autoregressive model with lattices of various sizes defined in terms of rows × columns

Lattice size	$\int var\{V_c(\mathbf{x})\}$	$\int var\{V_f(\mathbf{x})\}$	$\int var\{V_f(\mathbf{x})\}/\int var\{V_c(\mathbf{x})\}$
5×5	239.7	950.6	3.97
5×10	239.3	2149	8.98
5×20	239.3	4546	19.0
5×40	239.3	9340	39.0
10×10	481.9	4326	8.98
10×20	481.9	9154	19.0
10×40	481.9	18810	39.0

instead of the free-boundary path integral. In the next section we confirm this by using MCMC sampling.

5. Markov chain Monte Carlo approach

Following closely the discussion in Gelman and Meng (1998), section 5.1, we first consider the integral

$$\int_{\theta_a}^{0} \mathbf{E}_{\mathbf{x}|(\theta_a,\theta_c)} V_c(\mathbf{x})\, d\theta_c \tag{6}$$

and note that the approach to estimate equation (2) is trivially similar. Choosing a grid of equally spaced θ_c-values along the path of integration (parallel to the θ_c-axis), we construct for each such θ_c-value a Markov chain with stationary distribution $p(\mathbf{x}|\theta_a, \theta_c)$, using the Metropolis algorithm to update each site in the lattice successively. Then we estimate $\mathbf{E}_{\mathbf{x}|(\theta_a,\theta_c)} V_c(\mathbf{x})$ by an ergodic average of values $V_c(\mathbf{x})$ from this distribution. The integral (6) is then estimated via the trapezoidal rule

$$\int_{0}^{\theta_a} \mathbf{E}_{\mathbf{x}|(\theta_a,\theta_c)} V_c(\mathbf{x})\, d\theta_c \approx \frac{1}{2} \sum_{i=1}^{N-1} (\theta_{i+1} - \theta_i)\{\mathbf{E}_{\mathbf{x}|(\theta_a,\theta_{i+1})} V_c(\mathbf{x}) + \mathbf{E}_{\mathbf{x}|(\theta_a,\theta_i)} V_c(\mathbf{x})\}. \tag{7}$$

Initially, the Markov chain is burnt in for model parameters corresponding to the beginning of the path, with 1550 sweeps through the lattice. Then 500 samples are drawn, each after another full sweep through the lattice. The chain parameters are then adjusted for the next location along the path. Since this point is close by in parameter space, a burn-in of 550 sweeps through the lattice is sufficient, followed by 500 sample draws. This process is repeated until expectations have been computed for all points on the path.

A more efficient approach might be to draw samples $(\mathbf{x}, \theta_c)$ from the joint distribution

$$p(\mathbf{x}, \theta_c|\theta_f = \theta_a) = \frac{q\{\mathbf{x}|(\theta_a, \theta_c)\}}{z(\theta_a, \theta_c)} p(\theta_c|\theta_f = \theta_a).$$

If we assume that $p(\theta_c|\theta_f = \theta_a) \propto z(\theta_a, \theta_c)$, then the joint distribution is seen to be proportional to $q\{\mathbf{x}|(\theta_a, \theta_c)\}$. Thus we may write full conditionals for $\mathbf{x}$ and θ_c as

$$p\{\mathbf{x}|(\theta_a, \theta_c)\} \propto q\{\mathbf{x}|(\theta_a, \theta_c)\},$$
$$p(\theta_c|\mathbf{x}, \theta_f = \theta_a) \propto q\{\mathbf{x}|(\theta_a, \theta_c)\},$$

without knowing $z(\theta_a, \theta_c)$. Thus to obtain samples $(\mathbf{x}, \theta_c)$ we can alternately sample $\mathbf{x}$ and θ_c using, for example, a Gibbs-within-Metropolis type of algorithm. The obtained samples $(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \ldots, (\mathbf{x}_N, \theta_N)$, ordered increasingly by the θ_cs, could be used to estimate integral (6) numerically. The obvious advantage of this approach is that a single chain is used to calculate integral (6), as opposed to multiple chains, one for each value of the grid of Θ-values, and that the problem of burn-in, statistical precision and programming are consolidated within a single chain. However, for the free-boundary case, our experience is that such an MCMC sampler does not mix well. This agrees with the observations of Gelman and Meng (1998), who warned that $z(\theta_f)$ may vary over several orders of magnitude in the regions of θ_f of interest. Thus sampling from $(\mathbf{x}, \theta_f)$, where $p(\theta_f) \propto z(\theta_f)$, would lead to very few draws of Θ in regions of low marginal density. This in turn would lead to a poor estimation of $z(\theta_f)$ in such regions.

5.1. Results

Using a different seed for the random-number generator each time, the path integrals, with $N = 11$, were evaluated 100 times for lattices of size 6×10, 6×20, 6×40 and 6×80, with autologistic parameters $\theta_0 = \theta_f = 0.1$. The average value and standard deviation of the path integrals were then computed, the standard deviation taken as an estimate of the standard error of the path integral evaluated as outlined. The results are shown in Table 2 and replicate the earlier results given in Table 1, that the efficiency for fixed m is proportional to n, as n varies.

The computation time for each path integral is comparable, determined by the time taken to draw samples from the Markov chain. However, the standard errors and hence statistical efficiencies of the two integrals are substantially different. To compare the computation time required by each method fairly, we must do so at the same standard error. The standard error can be arbitrarily improved, by averaging independent repetitions of the estimation process, i.e., if the standard error is S after one estimation of the path integral, it will be $S/\sqrt{N}$ after N repetitions. Choosing an arbitrary small standard error of η, let S_f represent the standard error of the $E(V_f)$ integral and N_f represent the number of repetitions to achieve the standard error desired. Then $N_f = S_f^2/\eta^2$, and analogously for the $E(V_c)$ integral, $N_c = S_c^2/\eta^2$. Letting T_f and T_c represent the time taken to evaluate the path integrals, the times taken to achieve the standard error desired would be

$$T_f N_f = T_f S_f^2/\eta^2$$

Table 2. Estimates of path integrals by using MCMC sampling†

Lattice size	$\int \mathbf{E}\, V_c(\mathbf{x})\, d\theta_c$		$\int \mathbf{E}\, V_f(\mathbf{x})\, d\theta_f$		*Efficiency*
	Mean	*Standard error*	*Mean*	*Standard error*	
6×10	−0.042	0.0052	0.66	0.028	29
6×20	−0.041	0.0060	1.35	0.048	64
6×40	−0.040	0.0053	2.78	0.059	123
6×80	−0.041	0.0048	5.58	0.082	288

†Means and standard deviations of the estimates are calculated from 100 independent runs of the MCMC chain. Efficiency is the square of the ratio of the standard errors.

Table 3. Computation times for the cylinder analytic result, and for estimating the path sampling integrals, averaged over 100 MCMC runs with autologistic parameters $\theta_0 = \theta_f = 0.1$†

Lattice size	*Average computation time (ms)*			ζ
	T_{cyl}	T_c	T_f	
6×10	57.3	1549.1	1545.3	26
6×20	57.6	3110.0	3104.8	61
6×40	57.6	6245.5	6241.7	119
6×80	57.3	12530.5	12486.6	282

†Also shown is the estimated computation time for path sampling from the independent case as a multiple of the time for path sampling from the cylinder result (ζ) where a standard error of $\eta = 0.01$ is specified for both cases. Computations were performed on a Compaq–Digital Alphaserver 2100 computer with 512 Mbytes memory and four processors each running at 275 MHz.

and

$$T_c N_c = T_c S_c^2 / \eta^2$$

for the $E(V_f)$ and $E(V_c)$ path integrals respectively. The relative time taken for ideal computation is then

$$\frac{1}{\zeta} = \frac{T_c S_c^2/\eta^2 + T_{cyl}}{T_f S_f^2/\eta^2} \approx \frac{S_c^2}{S_f^2} + \frac{\eta^2 T_{cyl}}{S_f^2 T_c} \tag{8}$$

where T_{cyl} is the time taken to compute the cylinder analytic result and T_c and T_f are assumed to be equal for the approximation. Values for ζ based on average computation times are shown in Table 3.

At the same level of standard error, the cylinder analytic method followed by path sampling via equation (3) is 1–2 orders of magnitude faster than the standard method of path sampling via equation (2). The relative reduction in computation time increases with the number of columns. This effect is because the standard error of the path integral increases with the number of columns for the integral of $E(V_f)$, whereas the standard error for the integral of $E(V_c)$ is independent of the number of the columns.

Our empirical results suggest that because the normalizing constant ratio $z(\theta, \theta)/z(\theta, 0)$ is about 1 for the cylinder integral we could replace path sampling by acceptance ratio or bridge sampling; see Gelman and Meng (1998), section 3.2. Effectively, we can replace a continuous density for θ_c by a discrete density on 0 and θ.

6. Illustration of results

In Pettitt and Low Choy (1999) an experiment to investigate the effectiveness of six different chemical attractants for dingoes in the wild of western Queensland is described. Data were analysed which arose from observing signs of dingo presence or absence at 135 sites positioned at intervals of 500 m along a transect. Data were collected for seven consecutive days, giving rise to a 7×135 structure of data. For each day-by-site combination, binary response obser-

vations (signs of dingo presence) were available for two locations separated by 50 m. For each location, one of six chemical attractants was applied. The purpose of the experiment was to determine the best attractant. A feature of the experiment was that only about 15% of the sites were visited by dingoes so a model is introduced which considers dingo presence at a site as a possibly unobserved process. The observation model then conditions on the presence or absence of dingoes at a site. Given dingo presence, the observations, sign or no sign, at the two locations are considered independently distributed Bernoulli random variables. Given dingo absence, the observations are certain to be no sign of presence.

Low Choy (2001) considered various fully Bayesian analyses of the data. In one analysis the dingo presence–absence process is given by a three-parameter autologistic distribution with the free-boundary condition. The fully Bayesian analysis is effected by carrying out path sampling estimation of normalizing constants for the autologistic model. The autologistic parameters were discretized so that there were 27 points in the discrete parameter space. The computations involved running an MCMC chain at each point in the parameter space to estimate the sufficient statistics for the autologistic model on the 7×135 array: a substantial computational burden.

With the results of this paper, the free-boundary normalizing constant can be approximated by the cylinder result and its accuracy estimated or the approximation corrected by the path sampling method. The results of earlier sections suggest that the cylinder result by itself will be a very good approximation, as the difference between the free-boundary result and the cylinder result involves only one column statistic out of 135. However, it should be noted that this error is relative on the log-normalizing constant scale. What is important for the accuracy of likelihood calculations for inference is the change (over Θ-values) of the absolute error on the log-normalizing scale or, equivalently, the change (over Θ-values) of the relative error on the likelihood scale. Consequently the path sampling integral should be taken into account to estimate the error or to obtain a correction.

In Low Choy (2001) the posterior distribution is found to have its mode at $\theta_0 = -0.95$ and $\theta_f = 0.25$. We investigated values of the cylinder normalizing constant and the cylinder path sample estimate for parameter values with reasonable posterior support in the region of the mode. Results are given in Table 4, where it is seen that the values of the cylinder path sample estimate (third column) are considerably smaller than the values of the cylinder log-normalizing constant. However, when considering the likelihood, the path sampling correction multiplies the probability estimated by the cylinder normalizing constant, i.e.

$$p(\mathbf{x}|\Theta) = \frac{q(\mathbf{x}|\Theta)}{z(\theta_f, \theta_f)} \frac{z(\theta_f, \theta_f)}{z(\theta_f, 0)}.$$

Table 4. Normalizing constant for the autologistic model for values of (θ_0, θ_f) for a 7×135 array†

(θ_0, θ_f)	$\log\{z(\theta_f, \theta_f)\}$	$\log\{z(\theta_f, 0)/z(\theta_f, \theta_f)\}$
(−1.2, 0.25)	1587.04	0.20
(−0.95, 0.25)	1356.37	0.18
(−0.7, 0.25)	1141.94	0.16
(−0.95, 0.0)	1034.52	0.00
(−0.95, 0.5)	1778.85	0.82

†The second column is found by using the matrix result, whereas the third column is found by using path sampling.

In the case of $(\theta_0, \theta_f) = (-0.95, 0.5)$ this multiple is $1/\exp(0.82) = 0.44$, whereas for $(\theta_0, \theta_f) = (0, 0)$ the multiple is $1/\exp(0) = 1$, showing that the correction varies by a factor of 2 over the region of interest. Therefore, even though the path sampling correction appears relatively very small, it can have a significant effect on the posterior and, similarly, maximum likelihood parameter estimation, and so is necessary for accurate inference.

7. Discussion

In a typical statistical analysis we would only be interested in evaluating the normalizing constant where the likelihood or the posterior density was large. Thus, instead of drawing Θ from the marginal density which is uniform, we would want to draw it from the posterior density for the observed data. This could be achieved by using off-line estimation of $z(\Theta)$ or a process that computed both the posterior and the normalizing constant at the same time.

As mentioned previously the cylinder normalizing constant can be feasibly calculated for lattices where the smallest row or column is not greater than 10. However, it is straightforward to extend these results to larger lattices, by splitting the large lattice into smaller sublattices along the smallest row or column. Then path sampling, used here to go from the cylinder to the free-boundary lattice, may be used to go from the smaller lattices to the larger lattice. This would require the introduction of sufficient statistics connecting the rows or columns where the large lattice has been split. This is explained in detail in Friel and Pettitt (2002). Similar ideas apply if the boundary condition is given by fixing the values of boundary positions and therefore conditioning on them.

It would be interesting to compare maximum likelihood estimation based on the approaches outlined in the above paragraph with the MCMC maximum likelihood method given by Gu and Zhu (2001), where the computational challenge is to estimate the first and second derivatives of the log-normalizing constant efficiently to obtain an optimization scheme. The cylinder result and the path sampling extensions provide a method which is highly efficient both in statistical terms and in computational terms.

Finally, there is a problem of poor mixing when using MCMC sampling to draw from an autologistic distribution with parameter θ_f near a critical value which defines the so-called phase change, a transition to a region of parameter space where large contiguous sections of the lattice become single valued. In these regions, the cylinder analytic method behaves correctly, whereas MCMC methods will have problems with mixing and will be very inefficient.

Acknowledgements

We thank the Joint Editor and reviewers for their constructive comments on an earlier draft of this paper. Dr Nial Friel and Dr Robert Reeves were both supported by an Australian Research Council grant.

References

Augustin, N., Mugglestone, M. and Buckland, S. (1996) An autologistic model for spatial distribution of wildlife. *J. Appl. Ecol.*, **33**, 339–347.

Baxter, R. J. (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Besag, J. (1972) Nearest-neighbour systems and the auto-logistic model for binary data. *J. R. Statist. Soc.* B, **34**, 75–83.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *J. R. Statist. Soc.* B, **48**, 259–302.
Cox, D. R. and Miller, H. D. (1965) *The Theory of Stochastic Processes*. London: Chapman and Hall.
Cressie, N. (1993) *Spatial Statistics*, 2nd edn. New York: Wiley.
Evans, M. and Swartz, T. (2000) *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press.
Friel, N. and Pettitt, A. N. (2002) Likelihood estimation and inference for the Autologistic model. *J. Comput. Graph. Statist.*, to be published.
Gelman, A. and Meng, X. (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, **13**, 163–185.
Geyer, C. J. and Thompson, E. A. (1992) Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J. R. Statist. Soc.* B, **54**, 657–699.
Green, P. J. (1992) Discussion on 'Constrained Monte Carlo maximum likelihood for dependent data' (by C. J. Geyer and E. A. Thompson). *J. R. Statist. Soc.* B, **54**, 683–684.
Gu, M. G. and Zhu, H.-T. (2001) Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *J. R. Statist. Soc*. B, **63**, 339–355.
Jensen, I., Guttmann, A. J. and Enting, I. G. (1997) The Potts model on Kagome and honeycomb lattices. *J. Phys*. A, **30**, 8067–8083.
Low Choy, S. (2001) Hierarchical models for 2D presence/absence data having ambiguous zeroes. *PhD Thesis*. Queensland University of Technology, Brisbane.
Molenberghs, G. and Ryan, L. M. (1999) An exponential family model for clustered multivariate binary data. *Environmetrics*, **10**, 279–300.
Ogata, Y. (1989) A Monte Carlo method for high-dimensional integration. *Numer. Math.*, **55**, 137–157.
Pettitt, A. N. and Friel, N. (2002) Efficient calculation of the normalising constant of the autologistic and related models on the cylinder and lattice. *Technical Report*. School of Mathematical Sciences, Queensland University of Technology, Brisbane. (Available from `http://www.maths.qut.edu.au/~pettitt/research_papers/PF01pathsampling.ps`.)
Pettitt, A. N. and Low Choy, S. (1999) Bivariate binary data with missing values: analysis of a field experiment to investigate chemical attractants of wild dogs. *J. Agric. Biol. Environ. Statist.*, **4**, 57–76.
Preisler, H. K. (1993) Modelling spatial patterns of trees attacked by bark-beetles. *Appl. Statist.*, **42**, 501–514.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) *Numerical Recipes in Fortran 77*, vol. 1. Cambridge: Cambridge University Press.
Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
Smith A. F. M. (1992) Discussion on 'Constrained Monte Carlo maximum likelihood for dependent data' (by C. J. Geyer and E. A. Thompson). *J. R. Statist. Soc.* B, **54**, 684–686.
Wu, H. and Huffer, F. W. (1997) Modelling the distribution of plant species using the autologistic regression model. *Environ. Ecol. Statist.*, **4**, 49–64.
Zhao, L. P. and Prentice, R. L. (1990) Correlated binary regression using a quadratic exponential model. *Biometrika*, **77**, 642–648.