Forward recursions and normalizing constant for Gibbs fields

By Xavier GUYON and Cécile HARDOUIN
SAMOS / Université Paris 1,
90 rue de Tolbiac, 75634 Paris Cedex 13, France.
guyon@univ-paris1.fr, hardouin@univ-paris1.fr

Summary

Maximum likelihood parameter estimation is frequently replaced by various techniques because of its intractable normalizing constant. In the same way, the literature displays various alternatives for distributions involving such unreachable constants.

In this paper, we consider a Gibbs distribution π and present a recurrence formula allowing a recursive calculus of the marginals of π and in the same time its normalizing constant.

The numerical performance of this algorithm is evaluated for several examples, particularly for an Ising model on a lattice.

Some key words: Gibbs distribution; interaction potential; Markov Chain; Markov field; marginal law, normalizing constant; Ising model.

Corresponding author.
Cécile HARDOUIN
SAMM/Université Paris 1
90 rue de Tolbiac
75634 Paris Cedex
France
Email: Cecile.Hardouin@univ-paris1.fr

1 Introduction

Usually, the normalizing constant of a discrete probability π distribution involves high dimensional summation such that the direct evaluation of these sums becomes quickly infeasible in practice. For example, the direct calculation of the normalizing constant of the Ising model on a 10×10 grid involves summation over 2^{100} terms. The problem is bypassed for instance by evicting the distribution of interest by an alternative as, for example in spatial statistics, replacing the likelihood for the conditional pseudo likelihood (see Besag (1974)). Another solution consists of estimating the normalizing constant; a number of techniques have been proposed for this approximate evaluation, see for example Pettitt *et al.* (2003) and Moeller *et al.* (2006) for efficient Monte Carlo methods. Bartolucci and Besag (2002) present a recursive algorithm computing the likelihood of a Markov random field in the form of a product of conditional probabilities. Reeves (2004) propose efficient computation of the normalizing constant for a factorisable model, that is when π can be written as a product of factors.

In this paper we give more specific results when we specify π as a Gibbs distribution. We derive some results of Khaled (see Khaled 1 (2008), Khaled 2 (2008)) giving an original linear recursive calculation of the marginal distributions in the case of a particular distribution π of $Z = (Z_1, Z_2, \cdots, Z_T)$ used in econometrics for the modelling of a hidden regime (see Lovinson (2006)). An interesting consequence is to ease the calculation of the normalizing constant of π. We generalize Khaled's results, noticing that if π is a Gibbs distribution on $\mathcal{T} = \{1, 2, \cdots, T\}$, therefore π is a Markov field on $\mathcal{T}$ and it is easy to manipulate its conditional distributions. This approach allows to extend the recurrences given by Khaled to general Gibbs distribution π (either spatial or temporal) on a general finite state space E; those recursions yield exact calculation of the marginal distributions π_t of $(Z_1, Z_2, \cdots, Z_t)$, $1 \leq t \leq T$ as well as the normalization constant C of π.

First, we recall in section 1 some basic properties about Gibbs and Markov fields for $Z = (Z_1, Z_2, \cdots, Z_T)$, a sequence of joint distribution π on E^T where E is a finite state space. The main result of this paper is presented in section 2, where we give forward recursion for the marginal distributions and the application to the calculation of the normalization constant. We present some simple examples and give the computing times for carrying out the normalizing constant. In section 3, we extend the results to general Gibbs fields, in the sense of temporal Gibbs fields as well as spatial Gibbs fields. Finally we evaluate the numerical performance of our method to compute the normalizing constant for a spatial Ising model on a lattice $m \times T$.

2 Markov chain and Markov field properties for a Gibbs field

Let $T > 0$ be a fix positive integer, $E = \{e_1, e_2, \cdots, e_N\}$ a finite state space with N elements, $Z = (Z_1, Z_2, \cdots, Z_T)$ a temporal sequence with joint distribution π on E^T. We assume that π is a Gibbs distribution with an energy U_T described by singletons potentials $(\theta_s)_{s=1,T}$ and pairs potentials $(\Psi_s)_{s=2,T}$, that is, denoting $z(t) = (z_1, z_2, \cdots, z_t)$ for $1 \leq t \leq T$,

$$\begin{aligned} \pi(z(T)) &= C \exp U_T(z(T)) \text{ with } C^{-1} = \sum_{z(T) \in E^T} \exp U_T(z(T)) \text{ where} \qquad (2{\cdot}1) \\ U_t(z(t)) &= \sum_{s=1,t} \theta_s(z_s) + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s) \text{ for } 2 \leq t \leq T, \text{ and } U_1(z_1) = \theta_1(z_1). \end{aligned}$$

2·1 *Markov field property*

The neighbourhood graph on $\mathcal{T} = \{1, 2, \cdots, T\}$ associated to π is the so-called 2 nearest neighbours system. Moreover, we consider π as a bilateral Markov random field equipped with the 2 nearest neighbours system (see Kindermann and Snell (1980), Guyon (1995)): if $1 < t < T$,

$$\pi(z_t \mid z_s, 1 \leq s \leq T \text{ and } s \neq t) = \pi(z_t \mid z_{t-1}, z_{t+1}) \qquad (2{\cdot}2)$$

Indeed:

$$\begin{aligned} \pi(z_t \mid z_s, 1 \leq s \leq T \text{ and } s \neq t) &= \frac{\pi(z_1, z_2, \cdots, z_T)}{\sum_{u \in E} \pi(z_1, z_2, \cdots, z_{t-1}, u, z_{t+1}, ..z_T)} \\ &= \frac{\exp\{\theta_t(z_t) + \Psi_t(z_{t-1}, z_t) + \Psi_{t+1}(z_t, z_{t+1})\}}{\sum_{u \in E} \exp\{\theta_t(u) + \Psi_t(z_{t-1}, u) + \Psi_{t+1}(u, z_{t+1})\}} = \pi(z_t \mid z_{t-1}, z_{t+1}). \end{aligned}$$

Therefore, the non causal conditional distribution $\pi(z_t \mid z_{t-1}, z_{t+1})$ can be easily computed as soon as N, the cardinal of E, remains rather small.

2·2 *Markov chain property*

Proposition 1 *Z is a Markov chain: $\pi(z_t \mid z_s, s \leq t-1) = \pi(z_t \mid z_{t-1})$ if $1 < t \leq T$.*

Proof : $\pi(z_t \mid z_s, s \leq t-1) = \frac{\pi_t(z_1, z_2, \cdots, z_t)}{\pi_{t-1}(z_1, z_2, \cdots, z_{t-1})}$. Let's identify $\pi_t(z_1, z_2, \cdots, z_t)$. For $1 \leq s < t \leq T$, and using the notation $u_s^t = (u_s, u_{s+1}, \cdots, u_t)$, we write:

$$\begin{aligned} \pi_t(z_1, z_2, \cdots, z_t) &= \sum_{u_{t+1}^T \in E^{T-t}} \pi(z_1, z_2, \cdots, z_t, u_{t+1}^T) \qquad (2{\cdot}3) \\ &= C \exp\{\sum_{s=1,t} \theta_s(z_s) + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s)\} \times \exp\{\theta_t^*(z_t)\} \text{ where} \end{aligned}$$

$$\exp\{\theta_t^*(z_t)\} = \sum_{u_{t+1}^T} \exp\{ \sum_{s=t+1}^{T} \theta_s(u_s) + \Psi_{t+1}(z_t, u_{t+1}) + \sum_{t+2}^{T} \Psi_s(u_{s-1}, u_s)\}. \quad (2{\cdot}4)$$

We deduce that the conditional (to the past) distribution of z_t is given by:

$$\pi(z_t \mid z_s, s \leq t-1) = \exp\{\theta_t(z_t) + \theta_t^*(z_t) - \theta_{t-1}^*(z_{t-1}) + \Psi_t(z_{t-1}, z_t)\} = \pi(z_t \mid z_{t-1}). \quad (2{\cdot}5)$$

■

Remarks:

1. Following the proof above, we write again the marginal distribution $\pi_t(z_1, z_2, \cdots, z_t)$ as

   $$\pi_t(z(t)) = C \exp\{ \sum_{s=1,t-1} \theta_s(z_s) + [\theta_t(z_t) + \theta_t^*(z_t)] + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s)\}.$$

   Therefore, the marginal field $(Z_1, \cdots, Z_t)$ is also a Gibbs field with the 2 nearest neighbours system, associated to the same potentials, except for the last singleton potential which equals $\widetilde{\theta}_t(z_t) = \theta_t(z_t) + \theta_t^*(z_t)$.

2. An important difference appears between formula (2·5) and (2·2): indeed, (2·2) is computationally feasible, when (2·5) is not, because of the summation over u_{t+1}^T which is of complexity N^{T-t}.

3 Recursion over marginal distributions

3·1 *Future-conditional contribution* $\Gamma_t(z(t))$

For $t \leq T-1$, the distribution conditionally to the future $\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}, z_{t+2}, \cdots, z_T)$, depends only on z_{t+1}:

$$\begin{aligned}
\pi(z_1, z_2, \cdots, z_t \quad | \quad z_{t+1}, z_{t+2}, \cdots, z_T) &= \frac{\pi(z_1, z_2, \cdots, z_T)}{\sum_{u_1^t \in E^t} \pi(u_1^t, z_{t+1}, ..z_T)} \\
&= \frac{\exp\{U_t(z_1, ..., z_t) + \Psi_{t+1}(z_t, z_{t+1})\}}{\sum_{u_1^t \in E^t} \exp\left\{\sum_{s=1,t} \theta_s(u_s) + \Psi_{t+1}(u_t, z_{t+1}) + \sum_{s=2,t} \Psi_s(u_{s-1}, u_s)\right\}} \\
&= \pi(z_1, z_2, \cdots, z_t \mid z_{t+1}).
\end{aligned}$$

We can write this conditional distribution on the following feature:

$$\pi(z_1, z_2, \cdots, z_t \quad | \quad z_{t+1}) = \frac{\exp\{U_t(z_1, ..., z_t) + \Psi_{t+1}(z_t, z_{t+1})\}}{\sum_{u_1^t \in E^t} \exp\{U_t(u_1, ..., u_t) + \Psi_{t+1}(u_t, z_{t+1})\}}, \text{ i.e.}$$

$$\pi(z_1, z_2, \cdots, z_t \quad | \quad z_{t+1}) = C_t(z_{t+1}) \exp U_t^*(z_1, z_2, \cdots, z_t; z_{t+1})$$

where U_t^* is the *future-conditional energy* defined by:

$$U_t^*(z_1, z_2, \cdots, z_t; z_{t+1}) = U_t(z_1, z_2, \cdots, z_t) + \Psi_{t+1}(z_t, z_{t+1}), \tag{3·1}$$

and $C_{t+1}(z_{t+1})^{-1} = \sum_{u_1^t \in E^t} \exp\{U_t^*(u_1, ..., u_t; z_{t+1})\}$. Then, for $i = 1, N$:

$\pi(z_1, z_2, \cdots, z_t \mid z_{t+1} = e_i) = C_t(e_i)\gamma_t(z_1, z_2, \cdots, z_t; e_i)$ where $\gamma_t(z(t); e_i) = \exp U_t^*(z(t); e_i)$.

Thus, we define $\gamma_t(z(t); e_i)$ as the contribution to the distribution π_t of $Z(t)$ conditionally to the future $z_{t+1} = e_i$.

Definition 1 *For $t \leq T-1$, the vector $\Gamma_t(z(t))$ of the future-conditional contributions is defined by the vector of $\mathbb{R}^N$ with i-th component, $1 \leq i \leq N$:*

$$(\Gamma_t(z(t)))_i = \gamma_t(z(t); e_i).$$

For $t = T$, there is no conditional future and $\Gamma_T(z(T))$ is the constant vector of components $\gamma_T(z(T)) = \exp U_T(z(T))$. The definition of $\Gamma_T(z(T))$ is analogous to the one of Γ_t for $t \leq T-1$ with the convention $\Psi_{T+1} \equiv 0$. Still with this convention, we define for $1 \leq t \leq T$ the matrix A_t of size $N \times N$ with general term:

$$A_t(i, j) = \exp\{\theta_t(e_j) + \Psi_{t+1}(e_j, e_i)\}, \text{ et } i, j = 1, N. \tag{3·2}$$

Let us note that A_T has constant columns $A_T(i, j) = \exp \theta_T(e_j)$ for $i, j = 1$ to N. Then we get the fundamental recurrence:

Proposition 2 *For all $2 \leq t \leq T$, $z(t) = (z_1, z_2, \cdots, z_t) \in E^t$ and $e_i \in E$, we have:*

$$\gamma_t(z(t-1), e_j; e_i) = A_t(i, j) \times \gamma_{t-1}(z(t-1); e_j) \ , \tag{3·3}$$

and

$$\sum_{z_t \in E} \Gamma_t(z(t-1), z_t) = A_t \Gamma_{t-1}(z(t-1)). \tag{3·4}$$

Proof : (i) Let us consider the case $2 \leq t \leq T-1$. The energy U_t verifies $U_t(z(t-1), z_t) = U_{t-1}(z(t-1)) + \theta_t(z_t) + \Psi_t(z_{t-1}, z_t)$; therefore (3·1) ensures for all $(z_t, z_{t+1}) = (a, b) \in E^2$:

$$\begin{aligned} U_t^*(z(t-1), a; b) &= U_{t-1}(z(t-1)) + \theta_t(a) + \Psi_t(z_{t-1}, a) + \Psi_{t+1}(a, b) \\ &= U_{t-1}^*(z(t-1); a) + \{\theta_t(a) + \Psi_{t+1}(a, b)\}. \end{aligned}$$

This implies the recurrence (3·3), with $(e_j, e_i) = (a, b) = (z_t, z_{t+1})$. The summation over $z_t = e_j$ gives the component i on the left hand side of (3·4). That ensures the result.

(ii) For $t = T$, since $\Psi_{T+1} \equiv 0$, then $U_T^*(z(T-1), a) = U_{T-1}^*(z(T-1); a) + \theta_T(a)$. Therefore we obtain (3·4) for $A_T(i, j) = \exp \theta_T(e_j)$. ■

3·2 *Forward recursion for marginal distributions and normalization constant*

Let us define the following row vectors $1 \times N$: $E_1 = B_T = (1, 0, \cdots, 0)$, and $(B_t)_{t=T,2}$ the sequence specified by the forward recursion:

$$B_{t-1} = B_t A_t \text{ if } t \leq T.$$

We also denote $K_1 = \sum_{z_1 \in E} \Gamma_1(z_1) \in \mathbb{R}^N$. The evaluation of K_1 is easy if N is not too large, since its i-th component is $K_{1i} = \sum_{z_1 \in E} \exp\{\theta_1(z_1) + \Psi_2(z_1, e_i)\}$. We give below the main result of this paper.

Proposition 3 *Marginal distributions π_t and calculation of the normalization constant C.*
(1) For $1 \leq t \leq T$:

$$\pi_t(z(t)) = C \times B_t \Gamma_t(z(t)). \tag{3·5}$$

(2) The normalization constant C of the joint distribution π verifies:

$$C^{-1} = E_1 A_T A_{T-1} \cdots A_2 K_1. \tag{3·6}$$

Proof :

(1) Let us prove (3·5) by descending recurrence. For $t = T$, the equality is verified since,

$$\pi(z_1, z_2, \cdots, z_T) = \pi_T(z(T)) = C \exp U_T(z(T)) = C \times E_1 \Gamma_T(z(T)).$$

Let us assume that (3·5) is verified for t, $2 \leq t \leq T$. We use (3·4) which gives:

$$\begin{aligned} \pi_{t-1}(z(t-1)) &= \sum_{z_t \in E} \pi_t(z(t-1), z_t) = C \times B_t \{ \sum_{z_t \in E} \Gamma_t(z(t-1), z_t) \} \\ &= C \times B_t A_t \Gamma_{t-1}(z(t-1)) = C \times B_{t-1} \Gamma_{t-1}(z(t-1)). \end{aligned}$$

(2) Refering to (3·5), we have $\pi_1(z_1) = C \times B_1\ \Gamma_1(z_1)$. The summation over z_1 gives $C \times B_1 K_1 = 1$. The result ensures from the equality $B_1 = E_1 A_T A_{T-1} \cdots A_2$. ■

Remarks :

1 - The formula (3·6) allowing the calculus of C reduces in the case of time invariant potentials. In that case, we have $A_t \equiv A$ for $1 \leq t \leq T-1$, with $A(i,j) = \exp\{\theta(e_j) + \Psi(e_j, e_i)\}$, $A_T(i,j) = \exp \theta(e_j)$ and

$$C^{-1} = E_1 A_T A^{T-2} K_1. \tag{3·7}$$

Therefore, if the size N of E allows the diagonalization of the matrix A, we can achieve the calculus of the constant C independently of the temporal dimension T.

We give below two examples to illustrate our method. In each example, we consider Gibbs models with time independent potentials. For increasing values of T and various methods, we compare the times necessary for the computation of the normalizing constant C^{-1}. When possible, we also give the computing times for the ordinary summation method. In all examples, we have used the Matlab software and times are given in seconds.

Example 1 : *binary temporal model*
The state space is $E = \{0, 1\} = \{e_1, e_2\}$ with $N = 2$ states, $\theta_t(z_t) = \alpha z_t$ and $\Psi_{t+1}(z_t, z_{t+1}) = \beta z_t z_{t+1}$ for $t \leq T-1$. We have :

$A_t = A = \begin{pmatrix} 1 & e^{\alpha} \\ 1 & e^{\alpha+\beta} \end{pmatrix}$ for $t = 1, T-1$, $A_T = \begin{pmatrix} 1 & e^{\alpha} \\ 1 & e^{\alpha} \end{pmatrix}$, $E_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}$, $K_1 = \begin{pmatrix} 1+e^{\alpha} \\ 1+e^{\alpha+\beta} \end{pmatrix}$.

We present in Table 1 the times for the computation of C^{-1} for various values of T and the following methods : (1) calculation of $C^{-1} = E_1 A_T A^{T-2} K_1$; *(2)* C^{-1} is obtained by direct summation on E^T using a simple loop, or using a bitmap dodge which computes simultaneously the 2^T elements of E^T *(2 bis)*. We stopped computing C^{-1} by summation for $T > 25$.

	Meth. 1	*Meth. 2*	*Meth. 2 bis*	Value C^{-1}
$T = 10$	0	0.4690	0.0150	3.3441e+004
$T = 20$	0	744.6570	33.8120	8.6756e+008
$T = 25$	0	$\sim$ 6 hours	1315.0	1.3974e+011
$T = 690$	0			4.7610e+304
$T \geq 700$	0			∞

Table 1 : Computing times of C^{-1} for the binary temporal Gibbs distribution with parameters $\alpha = 1$, $\beta = -0.8$.

We see that the computing times of $C^{-1} = E_1 A_T A^{T-2} K_1$ are negligible for $T < 700$ while the summation method becomes quickly infeasible.

Example 2 : *bivariate binary temporal model*
$E = \{0, 1\}^2$ ($N = 4$ states), and $Z(T)$ is the anisotropic Ising model:

$$\pi_2((x_1, y_1), (x_2, y_2)) = C \exp\{\alpha x_1 + \beta y_1 + \gamma x_1 y_1 + \alpha x_2 + \beta y_2 + \gamma x_2 y_2 + \delta(x_1 x_2 + y_1 y_2)\}.$$

$E_1 A_T = A_T(1,\cdot)$ is A_T's first row with $A_T(1,j) = \exp\{\alpha x_j + \beta y_j + \gamma x_j y_j\}$; A is of size 4×4 defined by $A(i,j) = \exp\{\alpha x_j + \beta y_j + \gamma x_j y_j + \delta(x_i x_j + y_i y_j)\}$, and K_1's $i-$th component equals $\sum_{z=(x,y)\in E} \exp\{\alpha x + \beta y + \gamma xy + \delta(x x_i + y y_i)\}$. We computed C^{-1} in two ways, first using the the powers of A i.e. $C^{-1} = E_1 A_T A^{T-2} K_1$, then using its diagonalization $C^{-1} = E_1 A_T P D^{T-2} P^{-1} K_1$. We took parameters $\alpha = 1,\ \beta = -0.8,\ \gamma = -0.5,\ \delta = 0.04$. We were able to calculate $C^{-1} = 9.9491e+006$ for $T = 430$ and then stopped for larger T since the software treats C^{-1} as equals infinity. The computing times are null for both methods, which means that we get instantaneously C^{-1}'s value, and the size of A is still too small to distinguish computations using power or diagonalization of A.

4 General Gibbs fields

4·1 *Temporal Gibbs fields*

The previous results can be extended to general temporal Gibbs models. As an illustration, let us consider the following model, characterized by the energy:

$$U_T(z(T)) = \sum_{s=1,T} \theta_s(z_s) + \sum_{s=2,T} \Psi_{1,s}(z_{s-1}, z_s) + \sum_{s=3,T} \Psi_{2,s}(z_{s-2}, z_s).$$

The joint distribution π defines a bilateral Markov field with the 4 nearest neighbours system, and conditional distributions:

$$\begin{aligned}
\pi(z_t \quad | \quad & z_s, 1 \le s \le T \text{ et } s \ne t) = \\
= \quad & \frac{\exp\{\theta_t(z_t) + \Psi_{1,t}(z_{t-1}, z_t) + \Psi_{1,t+1}(z_t, z_{t+1}) + \Psi_{2,t}(z_{t-2}, z_t) + \Psi_{2,t+2}(z_t, z_{t+2})\}}{\sum_{u\in E} \exp\{\theta_t(u) + \Psi_{1,t}(z_{t-1}, u) + \Psi_{1,t+1}(u, z_{t+1}) + \Psi_{2,t}(z_{t-2}, u) + \Psi_{2,t+2}(u, z_{t+2})\}} \\
= \quad & \pi(z_t \mid z_{t-1}, z_{t+1}, z_{t-2}, z_{t+2}).
\end{aligned}$$

Z is also a Markov chain of order 2 with:

$$\begin{aligned}
\pi(z_t \quad | \quad & z_s, s \le t-1) \\
= \quad & \exp\{\theta_t(z_t) + \theta_t^*(z_t) - \theta_{t-1}^*(z_{t-1}) + \Psi_{1,t}(z_{t-1}, z_t) + \Psi_{2,t}(z_{t-2}, z_t) \\
& + \theta_t^{**}(z_{t-1}, z_t) - \theta_{t-1}^{**}(z_{t-2}, z_{t-1})\} \\
= \quad & \pi(z_t \mid z_{t-1}, z_{t-2})
\end{aligned}$$

where $\theta_t^*(z_t)$ is given by (2·4) and

$$\theta_t^{**}(z_{t-1}) = \sum_{u_{t+1}^T} \exp\{\Psi_{2,t+1}(z_{t-1}, u_{t+1}) + \Psi_{2,t+2}(z_t, u_{t+2}) + \sum_{t+3}^{T} \Psi_{2,s}(u_{s-2}, u_s)\}.$$

For $t \leq T-2$, the conditional distribution $\pi(z(t) \mid z_{t+1}, z_{t+2}, \cdots, z_T)$ depends only on (z_{t+1}, z_{t+2}):

$$\pi(z(t) \mid z_{t+1}, z_{t+2}) = C_t(z_{t+1}, z_{t+2}) \exp U_t^*(z(t); z_{t+1}, z_{t+2}),$$

where we name $U_t^*(z(t); z_{t+1}, z_{t+2})$ the *future conditional energy*:

$$\begin{aligned} U_t^*(z(t); z_{t+1}, z_{t+2}) &= U_t(z(t)) + \Psi_{1,t+1}(z_t, z_{t+1}) + \Psi_{2,t+1}(z_{t-1}, z_{t+1}) + \Psi_{2,t+2}(z_t, z_{t+2}), \\ \text{and } C_t(z_{t+1}, z_{t+2})^{-1} &= \sum_{u_1^t \in E^t} \exp U_t^*(u_1, .., u_t; z_{t+1}, z_{t+2}). \end{aligned}$$

For a, b and $c \in E$, it is easy to verify that:

$$U_t^*(z(t-1), a; (b, c)) = U_{t-1}^*(z(t-1); (a, b)) + \theta_t(a) + \Psi_{1,t+1}(a, b) + \Psi_{2,t+2}(a, c).$$

With the convention $\Psi_{1,s} \equiv \Psi_{2,s} \equiv 0$ for $s > T$, we define:

- for $t \leq T-2$, the vector $\Gamma_t(z(t))$ of the conditional contributions, conditionally to the future $(z_{t+1}, z_{t+2}) = (e_i, e_j), i, j = 1, N$, by the $N^2 \times 1$ vector of components (i, j):

$$\gamma_t(z(t); (e_i, e_j)) = \exp U_t^*(z(t); e_i, e_j);$$

- $\Gamma_{T-1}(z(T-1))$, the vector of the contributions conditionally to the future $z_T = e_i$:

$$(\Gamma_{T-1}(z(T-1))_i = \exp\{U_{T-1}(z(T-1)) + \Psi_{1,T}(z_{T-1}, e_i) + \Psi_{2,T}(z_{T-1}, e_i)\};$$

- $\Gamma_T(z(T))$ the constant vector of components $\exp\{U_T(z(T)\}$.

In the same way, we define the matrix A of size $N^2 \times N^2$ whose non zero components are:

$$A_t((i, j), (k, i)) = \exp\{\theta_t(e_k) + \Psi_{1,t+1}(e_k, e_i) + \Psi_{2,t+2}(e_k, e_j)\}$$

Like in section 3·2, we obtain a recurrence formula on the γ_t :

$$\gamma_t(z(t-1), e_k; (e_i, e_j)) = A_t((i, j), (k, i)) \times \gamma_{t-1}(z(t-1); (e_k, e_i))$$

together with the statement (3·4) on the contributions $\Gamma_t(z(t))$. In this context, $(Z_t, t = 1, T)$ is a Markovian process with memory 2 and $Y_t = (Z_t, Z_{t+1})$ a bivariate Markov chain for which we get the results (3·5) and (3·6).

4·2 *Spatial Gibbs fields*

Let us consider $Z_t = (Z_{(t,i)},\ i \in \mathcal{I})$, where $\mathcal{I} = \{1, 2, \cdots, m\}$, and $Z_{(t,i)} \in F$ ($Z_t \in E = F^m$). Then $Z = (Z_s,\ s = (t,i) \in \mathcal{S})$ is a spatial field on $\mathcal{S} = \mathcal{T} \times \mathcal{I}$. We note again $z_t = (z_{(t,i)},\ i \in \mathcal{I})$, $z(t) = (z_1, .., z_t)$ and $z = z(T)$.

Without loss of generality, we suppose that the distribution π of Z is a Gibbs distribution with translation invariant potentials $\Phi_{A_k}(\bullet),\ k = 1, K$ associated to a family of subsets $\{A_k, k = 1, K\}$ of $\mathcal{S}$, $\Phi_{A_k}(z)$ depending only on z_{A_k}, the layout of z over A_k. Then π is characterized by the energy:

$$U(z) = \sum_{k=1,K} \sum_{s \in S(k)} \Phi_{A_k+s}(z), \text{ with } S(k) = \{s \in \mathcal{S} \text{ s.t. } A_k + s \subseteq \mathcal{S}\}.$$

For $A \subseteq \mathcal{S}$, we define the height of A by $H(A) = \sup\{|u - v|, \exists (u,i) \text{ and } (v,j) \in A\}$, and $H = \sup\{H(A_k),\ k = 1, K\}$ the biggest height of the potentials. With this notation, we write the energy U as the following**:**

$$U(z) = \sum_{h=0}^{H} \sum_{t=h+1}^{T} \Psi(z_{t-h}, \cdots, z_t) \text{ with } \Psi(z_{t-h}, \cdots, z_t) = \sum_{k:H(A_k)=h} \sum_{s \in S_t(k)} \Phi_{A_k+s}(z)$$

where $S_t(k) = \{s = (u,i) : A_k + s \subseteq \mathcal{S} \text{ and } t - H(A_k) \leq u \leq t\}$.

(Z_t) is a Markov field with the $2H$-nearest neighbours system but also a Markov process with memory H; $Y_t = (Z_{t-H}, Z_{t-H+1}, \cdots Z_t)$, $t > H$, is a Markov chain on E^H for which we get the results (3·5) and (3·6).

4·3 *Computing the normalization constant for the Ising model*

We specify here the calculus of C^{-1} in the case of a translation invariant potentials Ising model (Kindermann and Snell (1980), Guyon (1995)). Let $\mathcal{S} = \mathcal{T} \times \mathcal{I} = \{1, 2, \cdots, T\} \times \{1, 2, \cdots, m\}$ be the set of sites, and $F = \{-1, +1\}$ the state space. We consider $Z = (Z_{(t,i)},\ (t,i) \in \mathcal{S})$ a Markov field on $\mathcal{S}$ with the four nearest neighbours system, a site (t,i) being a neighbour of (s,j) if $\|(t,i) - (s,j)\|_1 = 1$. The joint distribution π of Z is characterized by the singletons and pairs potentials:

$$\begin{aligned}
\Phi_{t,i}(z) &= \alpha\ z_{(t,i)} \quad \text{for } (t,i) \in \mathcal{S}, \\
\Phi_{\{(t,i),(t,i+1)\}}(z) &= \beta\ z_{(t,i)} z_{(t,i+1)} \text{ for } 1 \leq i \leq m-1, \\
\text{and } \Phi_{\{(t-1,i),(t,i)\}}(z) &= \delta\ z_{(t-1,i)} z_{(t,i)} \text{ for } 2 \leq t \leq T.
\end{aligned}$$

Forgetting the spatial dimension, we consider the state $z_t = (z_{(t,i)},\ i = 1, m) \in E = \{-1, +1\}^m$. Then Z is a temporal Gibbs field with the following translation invariant potentials:

$$\begin{aligned} \theta_t(z_t) &= \theta(z_t) = \alpha \sum_{i=1,m} z_{(t,i)} + \beta \sum_{i=1,m-1} z_{(t,i)} z_{(t,i+1)}, \\ \Psi_t(z_{t-1}, z_t) &= \Psi(z_{t-1}, z_t) = \delta \sum_{i=1,m} z_{(t-1,i)} z_{(t,i)},\ 2 \leq t \leq T. \end{aligned}$$

Let us give some notations associated to $c = (c_i,\ i = 1, m)$ and $d = (d_i,\ i = 1, m)$, two states of $E = \{-1, +1\}^m$; we first introduce $n^+(c) = \sharp\{i \in \mathcal{I} : c_i = +1\}$ and $n^-(c) = \sharp\{i \in \mathcal{I} : c_i = -1\}$ $(n^+(c) + n^-(c) = m)$, then $v^+(c) = \sharp\{i = 1, m-1 : c_i = c_{i+1}\}$ and $v^-(c) = \sharp\{i = 1, m-1 : c_i \neq c_{i+1}\}$ $(v^+(c) + v^+(c) = m-1)$, and finally $n^+(c,d) = \sharp\{i \in \mathcal{I} : c_i = d_i\}$, $n^-(c,d) = \sharp\{i \in \mathcal{I} : c_i \neq d_i\}$ $(n^+(c,d) + n^+(c,d) = m)$.

Since the potentials are invariant, the matrix A_t given by (3·2) does not depend on t, we have $A_t = A$, $t \leq T-1$, the $2^m \times 2^m$ matrix whose general term is,

$$A(a,b) = \exp\{\alpha(n^+(b) - n^-(b)) + \beta(v^+(b) - v^-(b)) + \delta(n^+(a,b) - n^-(a,b))\}, a, b \in E.$$

Moreover, for $t = T$, $A_T(a,b) = \exp\{\alpha(n^+(b) - n^-(b)) + \beta(v^+(b) - v^-(b))\}$. Therefore, the normalization constant of π is $C^{-1} = E_1 A_T A^{T-2} K_1$.

Since K_1 is given by a summation over $E = \{-1, +1\}^m$, and the size A is $2^m \times 2^m$, this formula is practically useful for m not too big.

Example 3

Table 2 gives computing times for the normalizing constant for the Ising model above with parameters $\alpha = 0.15,\ \beta = 0.05,\ \delta = -0.08$. In the case $T = m = 2$, the model is the one given in example 2 with $F = \{0, 1\}$. We consider varying values of T and $m = 10$, that is we work with vectors and square matrices of size 2^{10} but without theoretical constraints on the size of $\mathcal{T} = \{1, 2, \cdots, T\}$. We compute C^{-1} using the powers of matrix A or its diagonalization.

$m = 10$	$C^{-1} = E_1 A_T A^{T-2} K_1$	$C^{-1} = E_1 A_T P D^{T-2} P^{-1} K_1$	Value C^{-1}
$T = 2$	0.3130	32.4850	1.3855e+006
$T = 10$	8.9220	40.8290	5.4083e+030
$T = 50$	15.4380	47.4060	4.8989e+153
$T = 100$	19.3600	51.0950	2.4344e+307

Table 2: Computing times of C^{-1} for an Ising model on a lattice $10 \times T$ for various T.

We observe that it's computationally more efficient to compute the powers A^{T-2} rather than to use the diagonalization of A. Indeed, the diagonalization procedure itself is expensive for large size matrices.

4·4 *Other generalizations*

First, we presented the forward recursion for Gibbs distributions with singletons and pairs potentials. The results can be extended to larger potentials (triple or more).

Another extension is to consider variable state spaces. The recurrence (3·4) and properties (3·5), (3·6) hold for different state spaces E_t of the components Z_t; in this case, the associated matrices A_t involved in (3·6)are not necessarly square.

Finally, we can extend the results to embedded sets T; for instance let us consider the decreasing sequence $\mathcal{T} = S_Q \supset S_{Q-1} \supset \cdots \supset S_1$ of parts of $\mathcal{T} = \{1, 2, \cdots, T\}$; similarly to the former future-conditional contributions, we define the contributions $\gamma_q(z(S_q); z(\mathcal{T} \setminus S_q))$, conditionally to the outer layout $z(\mathcal{T} \setminus S_q)$. Let us give the following example: we assume $S_q = S_{q-1} \cup \partial S_{q-1}$ for $q = 1, Q-1$. Then, we obtain the conditional energy

$$U_q^*(z(S_q); z(\partial S_q)) = U_{q-1}^*(z(S_{q-1}); z(\partial S_{q-1})) + \Delta_q(z(\partial S_{q-1})); z(\partial S_q))$$

with $\Delta_q(z(\partial S_{q-1})); z(\partial S_q)) = \sum_{u \in \partial S_{q-1}} \theta_u(z_u) + \sum_{u \in \partial S_{q-1}, v \in \partial S_q, <u,v>} \Psi_{\{u,v\}}(z_u, z_v)$. then we define the matrices A_q by

$$A_q(\partial S_q; \partial S_{q-1}) = \exp \Delta_q(z(\partial S_{q-1}); z(\partial S_q)).$$

As an illustration let us set the following decreasing sequence $\mathcal{T} = S_{T-1} = \{1, 2, \cdots, T\}$, $S_{T-2} = \{1, 2, \cdots, T-1\}$, $\cdots$, $S_2 = \{1, 2, 3\}$ and $S_1 = \{2\}$. For $q = T-1, \cdots, 3$, the conditional contributions and the matrices A are defined in the usual way (3·1) and (3·2), while for $q = 2$, A_2 is a $N \times N^2$ matrix with $A_2(z_4, (z_1, z_3)) = \exp\{\theta_1(z_1) + \theta_3(z_3) + \Psi_4(z_3, z_4)\}$.

5 Conclusion

The technique proposed in this paper to evaluate marginals and normalisation constant is applicable to Markov chains, Markov fields. It overcomes the need to resort to approximate alternatives when one wants to evaluate the likelihood or the normalizing constant of a Gibbs field. It makes feasible the exact evaluation of the normalizing constant for moderate set of sites, eliminating any Monte Carlo procedure, variational scheme, or (approximate)

Bayesian computations steps. As a statistical consequence, it is therefore possible to perform true maximum likelihood estimation for such Gibbs field. Another application is to allow exact simulation.

We gave several illustrations of computing times for the normalizing constant. For one dimensional two states Gibbs fields, we are able to compute instantaneously the normalizing constant for a sequence of length 700, as well as for a sequence with four states and of length 400. We could keep computing for bigger lengths using another software. However, our main goal here is to provide a new method and we let the users choose their way of programming.

For spatial processes, we have computed the normalizing constant for an Ising model on a lattice 10×100 in 20 seconds. Following the discussion above, we could increase T, one of the side of the lattice; the limitation of the procedure ensures from the manipulation of $N^m \times N^m$ matrices. So the method seems to fail for large square lattices. As a comparison, Pettitt *et al.* (2003) compute the normalizing constant for an autologistic model defined on a cylinder lattice for which the smallest row or column is not greater than 10. They suggest to split a large lattice into smaller sublattices along the smallest row or column. A similar idea could apply here.

References

F. Bartolucci, J. Besag, 2002. A recursive algorithm for Markov random fields. *Biometrika* **89** (3), 724-730.

J. Besag, 1974. Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.* **B** 148, 1-36.

R. Kindermann and J.L. Snell, 1980. *Markov random fields and their applications*. Contemp. Maths.

X. Guyon, 1995. *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer-Verlag, New York.

G. Lovinson, 2006. A matrix-valued Bernoulli distribution. *J. of Mult. Anal.* **97**, 1573-1585.

M. Khaled, 2008. Estimation bayésienne de modèles espace-état non linéaires. *Ph.D. Thesis, Université Paris 1* http://mkhaled.chez-alice.fr/khaled.html

M. Khaled, 2008. A multivariate generalization of a Markov switching model. *working paper, C.E.S., Université Paris 1* http://mkhaled.chez-alice.fr/khaled.html

J. Moeller, A.N. Pettitt, R. Reeves, K.K. Berthelsen, 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalizing constants. *Biometrika* **93**, 2, 451-458.

A.N. Pettitt, N. Friel, R. Reeves, 2003. Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice. *J. Roy. Statist. Soc.* **B**, 65 Part 1, 235-246.

R. Reeves, A.N. Pettitt, 1982. Efficient recursions for general factorisable models. *Biometrika* **91**, 3, 751-757.