HAL
open science

Recursion on the marginals and normalizing constant for Gibbs processes - Version 2

Cécile Hardouin, Xavier Guyon

To cite this version:

Cécile Hardouin, Xavier Guyon. Recursion on the marginals and normalizing constant for Gibbs processes - Version 2. 2010. hal-00452651

HAL Id: hal-00452651

https://hal.science/hal-00452651v1

Preprint submitted on 2 Feb 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recursions on the marginals and normalizing constant for Gibbs processes

X. Guyon[a], C. Hardouin[a,b,*]

[a]*SAMM, University Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex, France*
[b]*MODAL'X, University Paris Ouest, 200 avenue de la République, 92001 Nanterre Cedex, France*

Abstract

This paper presents recurrence formulas allowing the calculus of the marginals and the normalizing constant of a Gibbs distribution π.The numerical performances of different methods are evaluated on several examples, particularly for an Ising model on a lattice.

Keywords: Gibbs distribution, factorisable distribution, interaction potential, Markov Chain, Markov field, marginal law, normalizing constant, Ising model.

1. Introduction

Usually, the computation of the marginal distributions and/or the normalizing constant C of a discrete probability distribution π involves high dimensional summation, such that the direct evaluation of these sums becomes quickly infeasible in practice. For example, for an Ising model on a 10×10 grid, it involves summation over 2^{100} terms. This problem has a deep impact for many applications, for instance, maximum likelihood parameter estimation. In the same way, some significant efforts have been put to solve the problem and the litterature displays various alternatives for distributions involving such unreachable constants. For instance, in spatial statistics, we replace the likelihood by the

*corresponding author
Email addresses: guyon@univ-paris1.fr (X. Guyon), hardouin@univ-paris1.fr (C. Hardouin)

conditional pseudo likelihood (Besag, 1974; Gaetan and Guyon, 2010). Another common way to bypass the problem is to estimate C, see for example Moeller et al. (2006) for efficient Monte Carlo methods. Sometimes, it is possible to compute C with an efficient algorithm, see for example Liu (2001); of course this is an interesting preliminary for further statistical procedures such as simulation or estimation. We briefly recall hereafter some recent results on the subject. Bartolucci and Besag (2002) express the likelihood of a Markov field in terms of the product of conditional probabilities; Pettitt et al. (2003) obtain the exact normalizing constant for a general categorical $K-$ valued distribution on a $m \times n$ cylinder with a matrix method involving the computation of the eigenvalues of a $K^m \times K^m$ matrix, which makes it possible for $K^m \lesssim 1024$, that is for example $K = 2$ and $m = 10$. Then Reeves and Pettitt (2004) give recursions for π factorisable.

In this paper, we first summarize classical results on the calculus of C and of the marginal distributions of π, a general Gibbs distribution. Then, following an idea of Khaled (2008a, 2008b) (see also Lovinson, 2006), we propose a new algorithm based on π's conditional probabilities to compute the marginals.

The plan of the paper is the following. Section 2 sums up basic properties about π, a Gibbs process on a finite state space; then we recall some results which permit the calculus of the marginals and the normalization constant of π, these results being presented under a new writing. We also compare the numerical efficiency of various algorithms in terms of their computing times. Section 3 presents a new algorithm based on the conditionals of π to compute the marginals of π. Section 4 extends these results for general Gibbs fields, and evaluates the numerical performance for a spatial Ising model on a lattice $m \times T$. The paper ends with some generalisations.

2. Recursions for a temporal Gibbs distribution

2.1. Factoring joint distribution, Markov field and Markov chain properties

Let $T > 0$ be a positive integer, E a finite state space with N elements, $Z(T) = (Z_1, Z_2, \cdots, Z_T)$ a temporal random variable with a factorisable joint

distribution π on E^T (see Reeves and Pettitt, 2004): for $z(T) = (z_1, z_2, \cdots, z_T)$,

$$\pi(z(T)) = C^{-1} \exp \sum_{s=1,T-1} h_s(z_s, z_{s+1}) = C^{-1} \prod_{s=1,T-1} H_s(z_s, z_{s+1}) \qquad (1)$$

Let us give some notations; in all the following, we interprete H_s, $s = 1, T-1$, as $N \times N$ matrices of elements $H_s(u, v) = \exp h_s(u, v)$. For such two $N \times N$ matrices H and G, for a N-row vector $F = (B(v), v \in E)$, and for a N-column vector $B = (B(v), v \in E)$, we note, as in the classical way,

$$HG(u, v) = \sum_{w \in E} H(u, w)G(w, v)\ ,\ FH(v) = \sum_{u \in E} F(u)H(u, v) \text{ and } HB(u) = \sum_{v \in E} H(u, v)B(v).$$

HG is a $N \times N$ matrix and HB an N-column vector. We also denote π_S the marginal of π on the subset $S \subseteq \mathcal{T} = \{1, 2, \cdots, T\}$, and $\pi_s^t = \pi_{[s,t]}$ with $[s, t] = \{s, s+1, \cdots, t\}$ for $s < t$.

We can interpret π as a Gibbs distribution with energy $U_T(z(T)) = \sum_{s=1,T-1} h_s(z_s, z_{s+1})$ associated to saturated pairs potentials $(h_s)_{s=1,T-1}$. In particular, π is a bilateral Markov random field w.r.t. the 2-nearest neighbours system (see Kindermann et Snell, 1980; Lauritzen, 1996; Guyon, 1995):

$$\pi(z_t \mid z_s, s \neq t) = \frac{H_{t-1}(z_{t-1}, z_t)H_t(z_t, z_{t+1})}{H_{t-1}H_t(z_{t-1}, z_{t+1})} = \pi(z_t \mid z_{t-1}, z_{t+1}).$$

Let us note that this 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. On the other hand, the normalizing constant C as well as the marginals cannot be computed since they entail high dimensional summation.

$Z(T)$ is also a Markov chain; indeed, $\pi(z_t \mid z_s,\ s \leq t-1) = \frac{\pi_1^t(z_1, z_2, \cdots, z_t)}{\pi_1^{t-1}(z_1, z_2 \cdots, z_{t-1})}$; then, for $1 \leq s < t \leq T$, and using the notation $u_s^t = (u_s, u_{s+1}, \cdots, u_t)$, we write $\pi_1^t(z_1, z_2, \cdots, z_t) = \sum_{u_{t+1}^T \in E^{T-t}} \pi(z_1, z_2, \cdots, z_t, u_{t+1}^T)$. Finally, we have

$$\pi(z_t \mid z_s,\ s \leq t-1) = \frac{H_{t-1}(z_{t-1}, z_t) \sum_{u_{t+1}^T} H_t(z_t, u_{t+1}) \prod_{s=t+1}^{T-1} H_s(u_s, u_{s+1})}{\sum_{u_t^T} H_{t-1}(z_{t-1}, u_t) \prod_{s=t}^{T-1} H_s(u_s, u_{s+1})} = \pi(z_t \mid z_{t-1}).$$

But the analytic form of the transition cannot be explicited because it also requires high dimensional summation.

2.2. Recursions over marginal distributions

Let us set $B_T = \mathbf{1}$ the N-column vector with constant coordinates 1. By direct marginalization on z_T, we obtain :

$$\pi_1^{T-1}(z_1, z_2, \cdots, z_{T-1}) = C^{-1}\{\prod_{s=1,T-2} H_s(z_s, z_{s+1})\}(H_{T-1}B_T)(z_{T-1})$$

In the same way, defining $B_{t-1} = H_{t-1}B_t$, we have, for $t = T, 2$:

$$\begin{aligned} \pi_1^t(z_1, z_2, \cdots, z_t) &= C^{-1} \prod_{s=1,t-1} H_s(z_s, z_{s+1})(H_t B_{t+1})(z_t) \qquad (2) \\ &= C^{-1} \prod_{s=1,t-1} H_s(z_s, z_{s+1})(H_t \cdots H_{T-1}B_T)(z_t) \ . \end{aligned}$$

These equations, up to the constant C, gives the marginals π_1^t. Looking at $\pi_{\{1\}}$, and denoting $F_1 = {}^tB_T$ the $N-$row constant vector with components equal to 1, we express the normalizing constant

$$C = C \times \sum_{z_1 \in E} \pi_{\{1\}}(z_1) = F_1\{\prod_{s=1,T-1} H_s\}B_T. \qquad (3)$$

These results about the marginals and the constant can be found, eventually differently presented, in other works, see for instance paragraph 2.4 of Liu (2001), and Reeves and Pettitt (2004). However, our writing is interesting since it gives a very simple algorithm to compute efficiently the normalizing constant, in terms of matrices products.

Analogously, using forward recursions, and defining $F_t = F_{t-1}H_{t-1}$ for $t \geq 2$, we obtain :

$$\begin{aligned} \pi_t^T(z_t, z_{t+1}, \cdots, z_T) &= C^{-1}\{(F_{t-1}H_{t-1})(z_t) \prod_{s=t,T-1} H_s(z_s, z_{s+1})\} \\ &= C^{-1}\{(F_1 H_1 H_2 \cdots H_{t-1})(z_t) \prod_{s=t,T-1} H_s(z_s, z_{s+1})\}. \end{aligned}$$

More generally, let us consider the marginal of π on the subset $S = \{s_1, s_2, \cdots, s_q\} \subseteq \mathcal{T} = \{1, 2, \cdots, T\}$. For C given by (3), we have the following result:

Proposition 1. *1 - The marginal distribution of π on $S = \{s_1, s_2, \cdots, s_q\}$ with $1 = s_1 < s_2 < \cdots < s_{q-1} < s_q = T$ is*

$$\pi_S(z_1, z_{s_2}, \cdots, z_{s_{q-1}}, z_T) = C^{-1} \prod_{i=1,q-1} (\prod_{s=s_i}^{s_{i+1}-1} H_s)(z_{s_i}, z_{s_{i+1}}), \tag{4}$$

2 - The marginal on $S_{1,T} = S \setminus \{1, T\}$ is obtain by changing the first H-product $(\prod_1^{s_2-1} H_s)(z_1, z_{s_2})$ by ${}^t\mathbf{1}(\prod_1^{s_2-1} H_s)(z_{s_2})$ and the last product $(\prod_{s_{q-1}}^{s_q-1} H_s)(z_{s_{q-1}}, z_T)$ by $((\prod_{s_{q-1}}^{s_q-1} H_s)\mathbf{1})(z_{s_{q-1}})$.

2.3. Numerical performances

Formula (3) simplifies for time invariant potentials $h_s = h$ ($H_s = H$). In this case, $C = {}^t\mathbf{1}(H)^{T-1}\mathbf{1}$ (or $C = {}^t\mathbf{1}(H)^{T-2}H_{T-1}\mathbf{1}$ if $H_{T-1} \neq H$); if the size N of E allows the diagonalization of the matrix H, we can achieve the calculus of C independently of the temporal dimension T. Let us look at two examples for which we compare the computing times for different algorithms. For this study, we have used the software Matlab .

Example 1 : *binary temporal model*

Let us consider $E = \{0, 1\}$ and the autologistic Gibbs field π associated to time independent singletons and pairs potentials $\theta_t(z_t) = \alpha z_t$, $t = 1, T$ and $\Psi_t(z_t, z_{t+1}) = \beta z_t z_{t+1}$ for $t \leq T-1$. We have $h_t(z_t, z_{t+1}) = \theta_t(z_t) + \Psi_t(z_t, z_{t+1})$ for $t = 1, T-2$ and $h_{T-1}(z_{T-1}, z_T) = \theta_{T-1}(z_{T-1}) + \Psi_{T-1}(z_{T-1}, z_T) + \theta_T(z_T)$.

We present in Table 1 the times for the computation of C for increasing values of T and the following three methods: (1) $C = {}^t\mathbf{1}(H)^{T-2}H_{T-1}\mathbf{1}$; *(2) C* is computed by direct summation over E^T using a simple loop (each element is computed one by one and added to the previous calculus); (3) C is obtained by summation again, using a bitmap dodge which computes simultaneously the 2^T elements of E^T. We stopped computing C by summation (methods 2 and 3) for $T > 25$.

	Meth. 1	*Meth. 2*	*Meth. 3*	Value C
$T = 10$	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

Table 1 : Computing times (in seconds) of C for a binary temporal Gibbs distribution, $\alpha = 1$, $\beta = -0.8$.

We observe that the computing times of $C = {}^t\mathbf{1}(H)^{T-2}H_{T-1}\mathbf{1}$ are negligible for $T < 700$ while methods 2 and 3 become quickly unusable.

Example 2 : *bivariate binary temporal model*
We consider now $E = \{0,1\}^2$ and $Z(T)$ is the anisotropic Ising model with invariant saturated potentials $h_s = h$,

$$h((x_1,y_1),(x_2,y_2)) = \alpha x_1 + \beta y_1 + \gamma x_1 y_1 + \alpha x_2 + \beta y_2 + \gamma x_2 y_2 + \delta(x_1 x_2 + y_1 y_2)\ ,$$

with the convention that a pair potential equals 0 if a state is taken out of the time domain $\{1, 2, \cdots, T\}$. We computed the constant C in two ways, first calculating directly the power H^{T-2}, then making use of the diagonalization of H, i.e. calculating $C = {}^t\mathbf{1}PD^{T-2}P^{-1}H_{T-1}\mathbf{1}$. The parameter values are $\alpha = 1,\ \beta = -0.8,\ \gamma = -0.5,$ and $\delta = 0.04$. We were able to calculate C for $T \leq 430$ and then stopped since the software treats C as equals to infinity. The computing times for both methods are instantaneous, the size of H being too small to distinguish computations using power or diagonalization of H.

2.4. Results for r-range potentials

Reeves and Pettitt (2004) consider more general r-factorisable distribution $\pi(z(T)) = \prod_{s=1}^{T-r} H_s(z_s, z_{s+1}, \cdots, z_{s+r})$. There, the function H_s is defined on $E^* = E^{r+1}$. For H a real function defined on E^*, we set H^* defined on $E^* \times E^*$ by:

$$H^*(u,v) = H(u_2^{r+1}, v_{r+2}) \prod_{i=1}^{r} \mathbf{1}(u_{i+1} = v_i). \tag{5}$$

Then, respectively to the $(*)$-objects, and with the notations of §2.1, we obtain the same results, for example the normalizing constant $C = {}^t\mathbf{1}(\prod_{s=1}^{T-r} H_s^*)\mathbf{1}$. Recursive algorithms for the marginals of π follow in the same way as in (2) and (4).

3. A new recursive algorithm for marginals based on future conditionals

Let us clarify the Gibbs specification (1) of π in term of singletons and pairs potentials, and write :

$$h_s(z_s, z_{s+1}) = \theta_s(z_s) + \Psi_s(z_s, z_{s+1}) \text{ for } s = 1, T \ ,$$

with the convention $\Psi_T \equiv 0$. Then we get the energy's writing for $t = 1, T$, and $z_1^t = (z_1 z_2, \cdots, z_t)$:

$$U_t(z_1^t) = \sum_{s=1,t} \theta_s(z_s) + \sum_{s=1,t-1} \Psi_s(z_s, z_{s+1})$$

We use conditioning by the future in order to compute recursively the marginal π_1^t. First we define the *contribution* of π, conditionally to the future $(z_{t+1}, \cdots, z_T)$. For $t < T$, it is clear that $\pi(z_1, z_2, \cdots, z_t \mid z_{t+1}, \cdots, z_T) = \pi(z_1, z_2, \cdots, z_t \mid z_{t+1})$. Then, for $t < T$,

$$\pi(z_1^t \mid z_{t+1}) = C_t^{-1}(z_{t+1}) \exp U_t^*(z_1^t; z_{t+1}),$$

where $U_t^*(z_1^t; z_{t+1}) = U_t(z_1^t) + \Psi_t(z_t, z_{t+1})$ is the *future-conditional* energy, and $C_{t+1}(z_{t+1}) = \sum_{u_1^t \in E^t} \exp\{U_t^*(u_1^t; z_{t+1})\}$.

Definition 1. For $1 \leq t \leq T$, let $\gamma_t(z_1^t; u) = \exp U_t^*(z_1^t; u)$ be the future-conditional contribution of π_1^t, conditionally to the future $z_{t+1} = u$. Then we define $\Gamma_t(z_1^t)$ as the N-row vector of the future-conditional contributions of π at time t:

$$\Gamma_t(z_1^t)(u) = \gamma_t(z_1^t; u), \quad u \in E \ .$$

For $t = T$, $\Gamma_T(z(T))$ is the constant vector of components $\gamma_T(z(T)) = \exp U_T(z(T))$. Let us notice that $\Gamma_t(z_1^t)$ is analytically explicit.

For $1 \leq t \leq T$, let $H_t(u, v) = \exp\{\theta_t(u) + \Psi_t(u, v)\}, u, v \in E$ and let us define the sequence $(D_t, t = T, 2)$ of N-column vectors by $D_T = {}^t(1, 0, \cdots, 0)$,

and $D_{t-1} = H_t D_t$ for $t \leq T$. Then the following result gives a new recursion for the marginals, based on the future-conditional contributions.

Proposition 2. *Recursion for marginal distributions.*

1 - For $2 \leq t \leq T$ and $z_1^t = (z_1, z_2, \cdots, z_t) \in E^t$, we have:

$$\sum_{z_t \in E} \Gamma_t(z_1^{t-1}, z_t) = \Gamma_{t-1}(z_1^{t-1}) H_t. \tag{6}$$

2 - For $1 \leq t \leq T$,

$$\pi_1^t(z_1^t) = C^{-1} \times \Gamma_t(z_1^t) D_t. \tag{7}$$

Proof. 1 - For $2 \leq t \leq T$, $U_t(z_1^{t-1}, z_t) = U_{t-1}(z_1^{t-1}) + \theta_t(z_t) + \Psi_{t-1}(z_{t-1}, z_t)$; therefore,

$$\begin{aligned} U_t^*((z_1^{t-1}, z_t); z_{t+1}) &= U_{t-1}(z_1^{t-1}) + \theta_t(z_t) + \Psi_{t-1}(z_{t-1}, z_t) + \Psi_t(z_t, z_{t+1}) \\ &= U_{t-1}^*(z_1^{t-1}; z_t) + \{\theta_t(z_t) + \Psi_t(z_t, z_{t+1})\}. \end{aligned}$$

This implies $\gamma_t((z_1^{t-1}, z_t); u) = \gamma_{t-1}(z_1^{t-1}; z_t) \times H_t(z_t, u)$ and summation over z_t gives (6).

2 - We prove (7) by descending recurrence. For $t = T$, the equality is verified since

$$\pi_1^T(z_1^T) = \pi(z(T)) = C^{-1} \exp U_T(z(T)) = C^{-1} \times \Gamma_T(z(T)) D_T \ .$$

Let us assume that (7) is verified for some t, $2 \leq t \leq T$. We use (6) which gives:

$$\begin{aligned} \pi_1^{t-1}(z_1^{t-1}) &= \sum_{z_t} \pi_1^t(z_1^{t-1}, z_t) = C^{-1}\{\sum_{z_t} \Gamma_t(z_1^{t-1}, z_t)\} D_t \\ &= C^{-1} \Gamma_{t-1}(z_1^{t-1}) H_t D_t = C^{-1} \Gamma_{t-1}(z_1^{t-1}) D_{t-1}. \end{aligned}$$

■

Proposition (2) can be extended in a natural way to r-lag Gibbs processes. For example, let us consider the 2-lag factorisable distribution π, characterized by the following energy:

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

with the convention $\Psi_{1,T} \equiv \Psi_{2,T-1} \equiv \Psi_{2,T} = 0$. It is easy to see that π is a Markov field w.r.t. the 4-nearest neighbours system and:

$$\pi(z_1^t \mid z_{t+1}, z_{t+2}, \cdots, z_T) = \pi(z_1^t \mid z_{t+1}, z_{t+2}) = C_t(z_{t+1}, z_{t+2}) \exp U_t^*(z_1^t; z_{t+1}, z_{t+2}),$$

where

$$U_t^*(z_1^t; z_{t+1}, z_{t+2}) = U_t(z_1^t) + \Psi_{1,t}(z_t, z_{t+1}) + \Psi_{2,t-1}(z_{t-1}, z_{t+1}) + \Psi_{2,t}(z_t, z_{t+2}).$$

Then, we have:

$$U_t^*((z_1^{t-1}, z_t); (z_{t+1}, z_{t+2})) = U_{t-1}^*(z_1^{t-1}; (z_t, z_{t+1})) + \theta_t(z_t) + \Psi_{1,t}(z_t, z_{t+1}) + \Psi_{2,t}(z_t, z_{t+2}).$$

Following the previous scheme, we define for $t \leq T$, the N^2-row vector $\Gamma_t(z_1^t)$ by

$$\Gamma_t(z_1^t)(u, v) = \exp U_t^*(z_1^t; u, v), \quad u, v \in E \ .$$

Then, as in the proof of Proposition (2), but w.r.t the $E^* \times E^*$ matrices (H_s^*) (5) with in this case $E^* = E^2$, we obtain the recurrence (6) on the contributions $\Gamma_t(z_1^t)$ and the results (7) on the marginals.

4. The case of spatial Gibbs fields

4.1. A temporal multidimensional Gibbs process

The basic idea is to consider a spatial Gibbs field as a multidimensional Gibbs process.

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 take the notations $z_t = (z_{(t,i)},\ i \in \mathcal{I})$, $z(t) = (z_1, .., z_t)$ 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 $\mathcal{A} = \{A_k, k = 1, K\}$ of subsets 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 $r(A) = \sup\{|u - v|, \exists (u, i)$ and $(v, j) \in A\}$, and $r = r(\mathcal{A}) = \sup\{r(A_k),\ k = 1, K\}$ the biggest height of the potentials. With this notation, we write the energy U as the following**:**

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

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

(Z_t) is a Markov random field w.r.t. the $2r$-nearest neighbours system but also a Markov process with memory r: $Y_t = (Z_t, Z_{t+1}, \cdots Z_{t+r})$, $t = 1, T - r$, is a Markov chain on $E^* = E^r$ for which we get the results (7) and (3).

4.2. Computing the normalization constant for the Ising model

We specify here the calculus of C for a translation invariant Ising model. Let $\mathcal{S} = \mathcal{T} \times \mathcal{I} = \{1, 2, \cdots, T\} \times \{1, 2, \cdots, m\}$ be the set of sites, $F = \{-1, +1\}$ the state space, and $Z = (Z_{(t,i)},\ (t, i) \in \mathcal{S})$ the Markov random field on $\mathcal{S}$ with the 4-nearest neighbours system. The joint distribution π of Z is characterized by the potentials

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

Z is also a temporal Gibbs process with the following potentials:

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

We define the following counting statistics associated to c, $d \in E = \{-1, +1\}^m$: $n^+(c) = \sharp\{i \in \mathcal{I} : c_i = +1\}$, $n^-(c) = m - n^+(c)$, $v^+(c) = \sharp\{i = 1, m-1 : c_i = c_{i+1}\}$, $v^-(c) = (m-1) - v^+(c)$, and finally $n^+(c,d) = \sharp\{i \in \mathcal{I} : c_i = d_i\}$, $n^-(c,d) = m - n^+(c,d)$.

Then we apply the formula (3) w.r.t. the following $2^m \times 2^m$ matrices (H_t) defined for any $u, v \in E = \{-1, +1\}^m$ and $\delta(t) = \delta \times \mathbf{1}_{(t \leq T-1)}$ by :

$$H_t(u,v) = \exp\{\alpha(n^+(u) - n^-(u)) + \beta(v^+(u) - v^-(u)) + \delta(t)(n^+(u,v) - n^-(u,v))\}.$$

Example 3

Table 2 gives computing times for the normalizing constant C for this Ising model with parameters $\alpha = 0.15$, $\beta = 0.05$, $\delta = -0.08$. We fix $m = 10$ and consider increasing values of T. We compute C using (3) and two methods: (M1) calculates the power $H^{T-2}$of H, while (M2) uses diagonalization of H.

$m = 10$	M1	M2	C
$T = 2$	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 for an Ising model on a lattice $10 \times T$.

We observe that it's computationally more efficient to compute the powers H^{T-2} rather than to use diagonalization. Indeed, the diagonalization procedure itself is expensive for large size matrices, and we use here vectors and square matrices of size 2^{10}.

4.3. Some generalizations

The results can be extended to larger potentials, as triplet potentials and so on.

Another extension considers varying state spaces $E_t \ni z_t$; the recurrence (6) and properties (7), (3) still hold but in this case, the matrices H_t are no longer square.

Finally, we can extend results (2), (3) and (6) to a sequence of embedded subsets of $\mathcal{T} = \{1, 2, \cdots, T\}$. Let us consider a decreasing sequence $\mathcal{T} = S_Q \supset S_{Q-1} \supset \cdots \supset S_1$ of subsets of $\mathcal{T}$ and assume $S_q = S_{q-1} \cup \partial S_{q-1}$ for $q = 1, Q-1$.

Similarly to the former future-conditional contributions (7) presented in definition 1 and used in proposition 2, we define the contributions $\gamma_q(z(S_q); z(\mathcal{T} \setminus S_q))$ conditionally to the outer layout $z(\mathcal{T} \setminus S_q)$. Then, we obtain (7) respectively to the conditional energy U_q^*,

$$\begin{aligned} U_q^*(z(S_q); z(\partial S_q)) &= U_{q-1}^*(z(S_{q-1}); z(\partial S_{q-1})) + \Delta_q(z(\partial S_{q-1})); z(\partial S_q)), \text{ with} \\ \Delta_q(z(\partial S_{q-1})); z(\partial S_q)) &= \sum_{u \in \partial S_{q-1}} \theta_u(z_u) + \sum_{u \in \partial S_{q-1}, v \in \partial S_q, <u,v>} \Psi_{\{u,v\}}(z_u, z_v), \end{aligned}$$

and the matrices H_q :

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

5. Conclusion

This paper gives results permitting the evaluation of marginals and normalizing constant for a Gibbs processes π. Some of these results, known in the litterature, are presented with another formalism. A new result, based on the contributions of π conditionally to the future, permits another recursion to compute the marginals of π. In particular, these results are applicable to Markov chains and Markov fields. It overcomes the need to resort to approximate alternatives for the likelihood, and makes feasible the exact evaluation of the normalizing constant for moderate set of sites.

We gave several illustrations of the algorithm's efficiency in terms 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 able to manipulate large matrices.

Since a random field with states in F can be seen as a temporal m-multivariate Gibbs process in $E = F^m$, we give similar results on marginals and normalizing

constant for a Gibbs random field. We have computed the normalizing constant for an Ising model on a lattice 10×100 in 20 seconds. While we could increase the temporal parameter T, one of the side of the lattice, the limitation of the procedure ensures from the manipulation of $E^m \times E^m$ 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.

Références

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

Besag J., 1974. Spatial interaction and the statistical analysis of lattice systems, J.R.S.S. B, 36 : 192-236.

Gaetan C., Guyon X., 2010. *Spatial statistics and modeling,* Springer.

Guyon X., 1995. *Random fields on a network : modeling, statistics and applications,* Springer.

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

Khaled M., 2008b. *Estimation bayésienne de modèles espace-état non linéaires*, Ph D thesis, Université Paris1. http://mkhaled.chez-alice.fr/khaled.html

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

Lauritzen S. L., 1996. Graphical models, Oxford Science Publications.

Liu J., 2001. *Monte Carlo strategies in scientific computing,* Springer.

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

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

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

Reeves R., Pettitt A.N., 2004. Efficient recursions for general factorisable models, Biometrika 91, 3 : 751-757.