Exact marginals and normalizing constant for Gibbs distributions
Cécile Hardouin, Xavier Guyon

To cite this version:

HAL Id: hal-00795132
https://hal.archives-ouvertes.fr/hal-00795132
Submitted on 27 Feb 2013

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.
Abstract

We present a recursive algorithm for the calculation of the marginal of a Gibbs distribution \( \pi \). A direct consequence is the calculation of the normalizing constant of \( \pi \).

1. Introduction

Usually, obtaining the marginals and/or the normalizing constant \( C \) of a discrete probability distribution \( \pi \) involves high dimensional summation: for example, for the binary Ising model on a simple grid \( 10 \times 10 \), the calculation of \( C \) involves \( 2^{100} \) terms. One way to prevent this problem is to change distribution of interest for an alternative as, for example in spatial statistics, replacing the likelihood for the conditional pseudo likelihood ([1]). Another solution consists of estimating the normalizing constant; see for example Pettitt & al ([8]) and Moeller & al ([7]) for efficient Monte Carlo methods, Bartolucci and Besag ([2]) for a recursive algorithm computing the exact likelihood of a Markov random field, Reeves and Pettitt ([9]) for an efficient computation of the normalizing constant for a factorisable model.

We present specific results for a Gibbs distribution \( \pi \). We derive results of Khaled ([5,6]) who gives an original linear recursion on the marginals of \( \pi \), the law of \( Z = (Z_1, Z_2, \cdots, Z_T) \in E^T \); this result eases the calculation of \( \pi \)'s normalizing constant. We generalize Khaled results noticing that if \( \pi \) is a Gibbs
distribution on $T = \{1, 2, \cdots, T\}$, then $\pi$ is a Markov field on $T$, so it is easy to manipulate its conditional distributions that are the basic tools of our forward recursions.

2. Markov representations of a Gibbs field

Let $T > 0$ be a fix positive integer, $E = \{e_1, e_2, \cdots, e_N\}$ a finite state space, $Z = (Z_1, Z_2, \cdots, Z_T) \in E^T$ a temporal sequence with distribution $\pi$. Let us denote $z(t) = (z_1, z_2, \cdots, z_t)$. We assume that $\pi$ is a Gibbs distribution with energy and potentials:

$$\pi(z(T)) = C \exp U_T(z(T)) \text{ with } C^{-1} = \sum_{z(T) \in E^T} \exp U_T(z(T)) \text{ where}$$

$$U_t(z(t)) = \sum_{s=1,t} \theta_s(z_s) + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s) \text{ for } 2 \leq t \leq T, \text{ and } U_1(z_1) = \theta_1(z_1).$$

So, $\pi$ is a bilateral 2 nearest neighbours Markov field ([4,3])

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

but $Z$ is also a Markov chain:

$$\pi(z_t | z_s, s \leq t - 1) = \pi(z_t | z_{t-1}) \text{ if } 1 < t \leq T. \quad (3)$$

An important difference appears between formulas (3) and (2): indeed, (2) is computationally feasible, when (3) is not.

3. Recursion over marginal distributions

3.1. Future-conditional contribution $\Gamma_t(z(t))$

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

$$\pi(z_1, z_2, \cdots, z_t | z_{t+1}, z_{t+2}, \cdots, z_T) = \frac{\pi(z_1, z_2, \cdots, z_T)}{\sum_{u_{t+1} \in E} \pi(u_{t+1}, z_{t+1}, z_{t+2}, \cdots, z_T)} = \pi(z_1, z_2, \cdots, z_t | z_{t+1}).$$

We can also write $\pi(z_1, z_2, \cdots, z_t | z_{t+1}) = C_t(z_{t+1}) \exp U_t(z_1, z_2, \cdots, z_t | z_{t+1})$ where $U_t^*$ is the future-conditional energy:

$$U_t^*(z_1, z_2, \cdots, z_t | z_{t+1}) = U_t(z_1, z_2, \cdots, z_t) + \Psi_{t+1}(z_t, z_{t+1}), \quad (4)$$

and $C_t(z_{t+1}) = \sum_{u_{t+1} \in E} \exp \{U_t^*(u_{t+1}, z_{t+1})\}$. Then, for $i = 1, N$:

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

With the convention $\Psi_{T+1} \equiv 0$, we define for $t \leq T$, the vector $\Gamma_t(z(t)) \in \mathbb{R}^N$ of the future-conditional contributions as

$$(\Gamma_t(z(t)))_i = \gamma_t(z(t); e_i), 1 \leq i \leq N.$$

and the recursion matrix $A_t$ by

$$A_t(i, j) = \exp \{\theta_t(e_j) + \Psi_{t+1}(e_j, e_i)\}, i, j = 1, N. \quad (5)$$

Then we get the following fundamental recurrence.
Proposition 3.1 For all $2 \leq t \leq T$, $z(t) = (z_1, z_2, \ldots, z_t) \in E^t$ and $e_i \in E$, we have:

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

and

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

3.2. Forward recursions on marginals and normalization constant

Let us define the following $1 \times N$ row vectors: $E_1 = B_T = (1, 0, \ldots, 0)$, and the $(B_t)_{t=T-2}$ defined by the forward recursion $B_{t-1} = B_t A_t$ if $t \leq T$; we also denote $K_1 = \sum_{z_1 \in E} \Gamma_1(z_1) \in \mathbb{R}^N$. We give below the main result of this work.

Proposition 3.2 Marginal distributions $\pi_t$ and calculation of the normalization constant $C$.

1. For $1 \leq t \leq T$:

$$\pi_t(z(t)) = C \times B_t \Gamma_t(z(t)).$$

2. The normalization constant $C$ of the joint distribution $\pi$ verifies:

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

The formula (9) reduces to $C^{-1} = E_1 A_T A_{T-1} \cdots K_1$ for time invariant potentials.

As a basic example, let us consider $E = \{0, 1\}$, $\theta_t(z_t) = \alpha z_t$, and $\Psi_{t+1}(z_t, z_{t+1}) = \beta z_t z_{t+1}$; the analytic expressions of $A$, $K_1$ are trivially derived. We computed $C^{-1} = E_1 A_T A_{T-1} \cdots K_1$ for increasing values of $T$; the computing time is always negligible for $T \leq 700$, whereas computing $C^{-1}$ by direct summation needs 750 seconds for $T = 20$, 6 hours for $T = 25$, and the method becoming ineffectual for $T > 25$.

4. Extensions to general Gibbs fields

There are various generalisations of the preceeding results.

4.1. Temporal Gibbs model

Let us give the following example as an illustration to possible extensions. Coming back to the previous model (1), we add the interaction potentials $\Psi_{2,s}(z_{s-2}, z_s)$. Then $\pi$ is a 4 nearest neighbours Markov field but also a Markov chain of order 2. Conditionally to the future, we get

$$\pi(z_1, z_2, \ldots, z_t \mid z_{t+1}, z_{t+2}, \ldots, z_T) = \pi(z(t) \mid z_{t+1}, z_{t+2}) = C_t(z_{t+1}, z_{t+2}) \exp U_t(z(t); z_{t+1}, z_{t+2}),$$

with

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

Then, for $a, b$ and $c \in E$, $U_t(z(t-1), a; (b, c)) = U_{t-1}(z(t-1); (a, b)) + \theta_t(a) + \Psi_{1,t+1}(a, b) + \Psi_{2,t+2}(a, c)$; analogously to the previous example, we define the future-conditional contributions and the $N^2 \times N^2$ matrices $A_t$ by

$$\gamma_t(z(t); (z_{t+1}, z_{t+2})) = \exp U_t(z(t); (z_{t+1}, z_{t+2}))$$

$$A_t(i, j, (k, i)) = \exp\{\theta_t(k) + \Psi_{1,t+1}(c_k, e_i) + \Psi_{2,t+2}(e_k, e_j)\}$$
Similarly as 3.1, we get the following recursion:

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

We thus obtain a recurrence (7) on the contributions \( \Gamma_t(z(t)) \) and analogous results as (8) and (9) for the bivariate Markov chain \( (Z_{t-1}, Z_t), t = 1, T \).

4.2. Spatial Gibbs fields

For \( t \in T = \{1, 2, \ldots, T\} \), let us consider \( Z_t = (z(t, i), i \in I) \), where \( I = \{1, 2, \ldots, m\} \), \( Z(t, i) \in F \). Then \( Z = (Z_s, s = (t, i) \in S) \) is a spatial field on \( S = T \times I \). We note again \( z_t = (z(t, i), i \in I) \), \( z(t) = (z_1, \ldots, z_t) \), \( z = z(T) \) and we suppose that the distribution \( \pi \) 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 \( S \). For \( A \subseteq S \), let us define \( H(A) = \sup\{|u-v|, \exists (u, i) \text{ and } (v, j) \in A\} \), and \( H = \sup\{H(A_k), k = 1, K\} \). With this notation, we write the Gibbs-energy

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

where \( S_t(k) = \{s = (u, i) : A_k + s \subseteq S \text{ and } t - H(A_k) \leq u \leq t\} \). Then \( (Z_t) \) is a Markov process of order \( H \) and \( Y_t = (Z_{t-H}, Z_{t-H+1}, \ldots, Z_t), t > H \) a Markov chain on \( E^H \) for which we get the results (8) and (9).

We applied the result to the calculation of the normalization constant for an Ising model. For \( m = 10 \) and \( T = 100 \), the computing time is less than 20 seconds.

References