



HAL
open science

Chaîne et champ de markov : quelques récurrences et calcul d'une constante de normalisation

Xavier Guyon, Cécile Hardouin

► **To cite this version:**

Xavier Guyon, Cécile Hardouin. Chaîne et champ de markov : quelques récurrences et calcul d'une constante de normalisation. 2009. hal-00397201

HAL Id: hal-00397201

<https://hal.science/hal-00397201v1>

Preprint submitted on 19 Jun 2009

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.

Chaîne et champ de Markov : quelques récurrences et calcul d'une constante de normalisation

X. Guyon et C. Hardouin

SAMOS-MATISSE, Université Paris 1

June 19, 2009

Abstract

M. Khaled [KM1, KM2, 2008] établit une récurrence originale permettant un calcul récursif des lois marginales d'une loi temporelle jointe π de $Z = (Z_1, Z_2, \dots, Z_T) \in E^T$, loi utilisée en économétrie pour modéliser un régime caché [G. Lovinson, 2006]. Une conséquence intéressante est de faciliter le calcul de la constante de normalisation de π .

Nous généralisons ces résultats en observant que la loi π envisagée étant une loi de Gibbs sur $\mathcal{T} = \{1, 2, \dots, T\}$, π est aussi un champ de Markov sur \mathcal{T} dont il est aisé de manipuler les lois conditionnelles. Cette approche permet de généraliser la récurrence fondamentale de M. Khaled au cas d'un espace d'état E fini quelconque et d'une loi π de Gibbs générale, temporelle ou spatiale. Cette récurrence permet un calcul simplifié de la constante de normalisation de π et la performance numérique de cet algorithme est évaluée pour π un modèle d'Ising spatial.

Mots clés : modèle de Gibbs, chaîne de Markov, champ de Markov, loi marginale, constante de normalisation, calcul récursif, modèle d'Ising.

Classification AMS : 60G60, 62M40

1 Introduction

Soit T un entier > 0 fixé, $E = \{e_1, e_2, \dots, e_N\}$ un espace d'états fini à N éléments, $Z = (Z_1, Z_2, \dots, Z_T)$ une suite temporelle de loi jointe π sur E^T . On suppose que π est une *loi de Gibbs d'énergie* U_T dérivant de *potentiels* de singletons $(\theta_s)_{s=1,T}$ et de paires $(\Psi_s)_{s=2,T}$, c'est à dire, notant $z(t) = (z_1, z_2, \dots, z_t)$,

$$\pi(z(T)) = C \exp U_T(z(T)) \text{ avec } C^{-1} = \sum_{z(T) \in E^T} \exp U_T(z(T)) \text{ où}$$
$$U_t(z(t)) = \sum_{s=1,t} \theta_s(z_s) + \sum_{s=2,t} \Psi_s(z_{s-1}, z_s) \text{ si } 2 \leq t \leq T,$$

et $U_1(z_1) = \theta_1(z_1)$. Généralement, le calcul de la *constante de normalisation* C est impossible du fait de la grande cardinalité de E^T . L'objectif ici est d'obtenir des formules explicites permettant le calcul des lois marginales $\pi_t(z(t)) = \pi_t(z_1, z_2, \dots, z_t)$, $1 \leq t \leq T$ (avec $\pi = \pi_T$) ainsi que le calcul de la constante de normalisation C .

Le graphe de voisinage sur $\mathcal{T} = \{1, 2, \dots, T\}$ associé aux potentiels de π étant celui aux 2-plus proches voisins (2-ppv), π est aussi un *champ de Markov bilatéral* aux 2-ppv ([Kindermann et Snell, 1980]; [Guyon, 1995]; [Lauritzen, 1996]) : si $1 < t < T$,

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

En effet :

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

La loi conditionnelle bilatérale $\pi(z_t \mid z_{t-1}, z_{t+1})$ s'explique donc facilement dès que N , le cardinal de E , n'est pas trop grand.

2 Propriété de chaîne de Markov

Proposition 1 Z est une chaîne de Markov : $\pi(z_t \mid z_s, s \leq t-1) = \pi(z_t \mid z_{t-1})$ si $1 < t \leq T$.

Preuve : $\pi(z_t \mid z_s, s \leq t-1) = \frac{\pi_t(z_1, z_2, \dots, z_t)}{\pi_{t-1}(z_1, z_2, \dots, z_{t-1})}$. Identifions $\pi_t(z_1, z_2, \dots, z_t)$. Pour $1 \leq s < t \leq T$, notant $u_s^t = (u_s, u_{s+1}, \dots, u_t)$, on a :

$$\begin{aligned} \pi_t(z_1, z_2, \dots, z_t) &= \sum_{u_{t+1}^T \in E^{T-t}} \pi(z_1, z_2, \dots, z_t, u_{t+1}^T) \\ &= C \exp\left\{ \sum_{s=1, t} \theta_s(z_s) + \sum_{s=2, t} \Psi_s(z_{s-1}, z_s) \right\} \times \exp\{\theta_t^*(z_t)\} \text{ où} \\ \exp\{\theta_t^*(z_t)\} &= \sum_{u_{t+1}^T} \exp\left\{ \sum_{s=t+1}^T \theta_s(u_s) + \Psi_{t+1}(z_t, u_{t+1}) + \sum_{t+2}^T \Psi_s(u_{s-1}, u_s) \right\}. \end{aligned}$$

On en déduit que la loi de z_t conditionnelle au passé vaut :

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

■

Remarques :

1. Dans la preuve, on peut encore écrire la loi marginale $\pi_t(z_1, z_2, \dots, z_t)$ sous la forme

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

Ainsi, le champ marginal (Z_1, \dots, Z_t) est encore un champ de Gibbs aux 2-ppv pour les mêmes potentiels à l'exception du potentiel de singleton terminal qui vaut $\theta_t(z_t) = \theta_t(z_t) + \theta_t^*(z_t)$.

2. Une différence importante apparaît entre les formules (2) et (1) : (1) est calculable alors que (2) ne l'est pas du fait de la sommation en u_{t+1}^T de complexité N^{T-t} .

3 Récursions sur les lois marginales

3.1 Contribution conditionnelle au futur $\Gamma_t(z(t))$

Pour $t \leq T-1$, la loi conditionnelle au futur, $\pi(z_1, z_2, \dots, z_t \mid z_{t+1}, z_{t+2}, \dots, z_T)$, ne dépend que de z_{t+1} . En effet,

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

On peut encore écrire cette loi conditionnelle sous la forme suivante :

$$\begin{aligned}\pi(z_1, z_2, \dots, z_t \mid z_{t+1}) &= \frac{\exp\{U_t(z_1, \dots, z_t) + \Psi_{t+1}(z_t, z_{t+1})\}}{\sum_{u_t^i \in E^t} \exp\{U_t(u_1, \dots, u_t) + \Psi_{t+1}(u_t, z_{t+1})\}}, \text{ c-à-d} \\ \pi(z_1, z_2, \dots, z_t \mid z_{t+1}) &= C_t(z_{t+1}) \exp U_t^*(z_1, z_2, \dots, z_t; z_{t+1})\end{aligned}$$

où U_t^* est l'énergie conditionnelle au futur définie par :

$$U_t^*(z_1, z_2, \dots, z_t; z_{t+1}) = U_t(z_1, z_2, \dots, z_t) + \Psi_{t+1}(z_t, z_{t+1}). \quad (3)$$

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

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

Ainsi, $\gamma_t(z(t); e_i)$ est la contribution à la loi π_t de $Z(t)$ conditionnelle au futur $z_{t+1} = e_i$.

Definition 1 Pour $t \leq T - 1$, le vecteur $\Gamma_t(z(t))$ des **contributions conditionnelles au futur** est le vecteur de \mathbb{R}^N de i -ième coordonnée, $1 \leq i \leq N$:

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

Pour $t = T$, il n'y a pas de conditionnement et $\Gamma_T(z(T))$ est le vecteur constant de coordonnées $\gamma_T(z(T)) = \exp U_T(z(T))$. La définition de $\Gamma_T(z(T))$ est analogue à celle de Γ_t pour $t \leq T - 1$ sous la convention que $\Psi_{T+1} \equiv 0$. Avec cette même convention, définissons pour $1 \leq t \leq T$ la matrice A_t de taille $N \times N$ de terme général :

$$A_t(i, j) = \exp\{\theta_t(e_j) + \Psi_{t+1}(e_j, e_i)\}, \text{ et } i, j = 1, N. \quad (4)$$

Notons que A_T est une matrice à colonne constante égale à $(\exp\{\theta_T(e_j)\}, j = 1, N)$. On a alors la récurrence fondamentale suivante :

Proposition 2 Pour tout $1 \leq t \leq T$, $z(t) = (z_1, z_2, \dots, z_t) \in E^t$ et $e_i \in E$, on a :

$$\gamma_t(z(t-1), e_j; e_i) = A_t(i, j) \times \gamma_{t-1}(z(t-1); e_j). \quad (5)$$

et

$$\sum_{z_t \in E} \Gamma_t(z(t-1), z_t) = A_t \Gamma_{t-1}(z(t-1)) \quad (6)$$

Preuve : (i) Examinons d'abord le cas où $2 \leq t \leq T - 1$. L'énergie U_t vérifiant $U_t(z(t-1), z_t) = U_{t-1}(z(t-1)) + \theta_t(z_t) + \Psi_t(z_{t-1}, z_t)$, (3) donne pour tout $(z_t, z_{t+1}) = (a, b) \in E^2$:

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

La récurrence (5) en découle en posant $(e_j, e_i) = (a, b) = (z_t, z_{t+1})$. La sommation en $z_t = e_j$ donne alors la coordonnée i du terme de gauche de (6). D'où le résultat.

(ii) Pour $t = T$, puisque $\Psi_{T+1} \equiv 0$, $U_T^*(z(T-1), a) = U_{T-1}^*(z(T-1); a) + \theta_T(a)$. On obtient alors (6) pour $A_T(i, j) = \exp \theta_T(e_j)$. ■

3.2 Calcul récursif des lois marginales et de la constante de normalisation

Commençons par donner quelques notations sur les vecteurs lignes $1 \times N$ suivants : soient $E_1 = B_T = (1, 0, \dots, 0)$ et la suite $(B_t)_{t=T}$ définie par la récursion $B_{t-1} = B_t A_t$ si $t \leq T$; on note également $K_1 = \sum_{z_1 \in E} \Gamma_1(z_1) \in \mathbb{R}^N$. Le calcul de K_1 est facile si N n'est pas trop grand puisque sa i -ème coordonnée vaut $K_{1i} = \sum_{z_1 \in E} \exp\{\theta_1(z_1) + \Psi_2(z_1, e_i)\}$.

Proposition 3 *Calcul des lois marginales π_t et de la constante de normalisation C .*

(1) Pour $1 \leq t \leq T$:

$$\pi_t(z(t)) = C \times B_t \Gamma_t(z(t)). \quad (7)$$

(2) La constante de normalisation C de la loi globale π vérifie :

$$C^{-1} = E_1 A_T A_{T-1} \cdots A_2 K_1. \quad (8)$$

Preuve :

(1) Vérifions (7) par récurrence. Pour $t = T$, la relation est bien vérifiée puisque,

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

Supposons (7) vérifiée à l'ordre t , $2 \leq t \leq T$. Utilisant (6) :

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

(2) D'après (7), $\pi_1(z_1) = C \times B_1 \Gamma_1(z_1)$. Sommant en z_1 , on obtient $C \times B_1 K_1 = 1$. Le résultat résulte de l'égalité $B_1 = E_1 A_T A_{T-1} \cdots A_2$. \blacksquare

La formule (8) permettant le calcul de C se simplifie si les potentiels de π sont invariants dans le temps. En effet, dans ce cas, $A_t \equiv A$ pour $1 \leq t \leq T-1$, avec $A(i, j) = \exp\{\theta(e_j) + \Psi(e_j, e_i)\}$, $A_T(i, j) = \exp\theta(e_j)$ et $C^{-1} = E_1 A_T A^{T-2} K_1$. Si la taille N de E permet la diagonalisation de A , le calcul de C est possible indépendamment de la dimension temporelle T .

Exemples

1. J'ai fait la vérification de la formule donnant C^{-1} pour trois situations, les plus simples :

(i) un processus à , ceci pour $T = 2$ et $T = 3$. Ca marche.

(ii) pour $E = \{0, 1\}^2$ ($N = 4$ états), les vecteurs E_1, K_1 de dimension 4 et les matrices A sont 4×4 avec $T = 2$ et le modèle, posant $z = (x, y)$, et le modèle type Ising non isotropique :

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

C'est mieux de ne pas mettre d'isotropie, on controle mieux les calculs avec les 4 paramètres. Ici il suffit de calculer A_2 , en fait sa première ligne $A_2(1, \cdot) = E_1 A_2$, et K_1 . La formule n'est autre que $C^{-1} = \langle A_2(1, \cdot), K_1 \rangle$, soit 16 termes. Et on retrouve bien $C^{-1} = \sum_{z_1, z_2} \exp U(z_1, z_2)$!

2. Après quoi, il faudra s'assurer de la forme des potentiels θ_t et Ψ_t associés au modèle d'Ising et voir le calcul effectif. Dans ce calcul effectif, par exemple sur une image 10×10 , il n'y a pas lieu de diagonaliser. Il est plus facile de faire le produit de matrices $A_{10} \times A_9 \times \cdots \times A_2$ avec une matrice de "bord" A_{10} et les autres égales à A_2 . K_1 ne pose pas de problème.

3. La diagonalisation vaudra la peine si T est grand, et ce serait bien de valoriser un tel exemple. Une chose envisageable, qui nécessite du travail numérique, est de faire de l'estimation par MV pour une observation Ising sur un rectangle $S = \{1, M\} \times \{1, T\}$ d'un modèle d'Ising. On simule par échantillonneur de Gibbs un tel Ising et on examine les résultats asymptotiques pour T grand de comparaison du PMV, Codage et MV. Là, on peut explicitement programmer la maximisation de la vraie vraisemblance.

J'ai suivi ton exple 1 mais je reprendrai l'exple du ising classique avec $\{-1, +1\}$

La programmation (en matlab) donne un résultat immédiat pour les calculs de C

- sous la forme explicite en fonction des valeurs propres λ_1 et λ_2 , voir ci-dessous *Meth1*

- sous la forme $C^{-1} = E_1 A_T A^{T-2} K_1$ *Meth3*

- sous la forme $C^{-1} = E_1 A_T P D^{T-2} P^{-1} K_1$ *Meth2* et *Meth2bis* (selon qu'on utilise les matrices de passage explicites ci-dessous (*Meth2*) ou celles automatiquement calculées par matlab (*Meth2bis*))

et un calcul un peu plus long sous la forme de sommation. *Meth4*

Voici le tabelau des temps de caclul, en secondes

$\alpha = 1, \beta = 0.7$	<i>Meth1</i>	<i>Meth2</i>	<i>Meth2b</i>	<i>Meth3</i>	<i>Meth4</i>	Valeur C^{-1}
$T = 10$	0	0	0	0	0.0150	$3.6420e + 007$
$T = 20$	0	0	0	0	41.22	$2.2611e + 015$

Pour un T plus grand, j'ai des pbs de mémoire avec matlab, il n'est pas content d'avoir une matrice de taille 2 puissance T ! il va falloir que je fasse autrement pour stocker.

1. *****

Exemple 1

Prenons le modèle à $N = 2$ états avec $E = \{0, 1\} = \{e_1, e_2\}$, de potentiels de singletons $\theta_t(z_t) = \alpha z_t$ et de paires $\Psi_{t+1}(z_t, z_{t+1}) = \beta z_t z_{t+1}$ pour $t \leq T - 1$.

Dans ce cas, on obtient $A_t = A = \begin{pmatrix} 1 & e^\alpha \\ 1 & e^{\alpha+\beta} \end{pmatrix}$ pour $t = 1, T - 1$, $A_T = \begin{pmatrix} 1 & e^\alpha \\ 1 & e^\alpha \end{pmatrix}$, $E_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}$ et $K_1 = \begin{pmatrix} 1 + e^\alpha \\ 1 + e^{\alpha+\beta} \end{pmatrix}$.

Diagonalisation de A;

$$\det(A - \lambda I) = \lambda^2 - \lambda(1 + e^{\alpha+\beta}) + e^{\alpha+\beta} - e^\alpha$$

$$\Delta = 1 + e^{2(\alpha+\beta)} - 2e^{\alpha+\beta} + 4e^\alpha = (1 - e^{\alpha+\beta})^2 + 4e^\alpha > 0$$

$$\text{Deux valeurs propres } \lambda_1 = \frac{1+e^{\alpha+\beta}-\sqrt{\Delta}}{2} \text{ et } \lambda_2 = \frac{1+e^{\alpha+\beta}+\sqrt{\Delta}}{2}$$

$$\text{Vecteurs propres associés } u_1 = \begin{pmatrix} e^\alpha \\ \lambda_1 - 1 \end{pmatrix} \text{ et } u_2 = \begin{pmatrix} e^\alpha \\ \lambda_2 - 1 \end{pmatrix} \text{ d'où } P = \begin{pmatrix} e^\alpha & e^\alpha \\ \lambda_1 - 1 & \lambda_2 - 1 \end{pmatrix}$$

$$\text{et } P^{-1} = \frac{1}{\sqrt{\Delta}} e^{-\alpha} \begin{pmatrix} \lambda_2 - 1 & -e^\alpha \\ 1 - \lambda_1 & e^\alpha \end{pmatrix}$$

$$C^{-1} = E_1 A_T P D^{T-2} P^{-1} K_1$$

$$C^{-1} = \frac{1}{\sqrt{\Delta}} \left[(\lambda_1^{T-1}(\lambda_2 - 1) + \lambda_2^{T-1}(1 - \lambda_1))(1 + e^\alpha) + e^\alpha(1 + e^{\alpha+\beta})(\lambda_2^{T-1} - \lambda_1^{T-1}) \right]$$

Références

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

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

Lauritzen, *Graphical modelling* (1996), Clarendon Press Oxford.

Lovinson G. (2006), *A matrix-valued Bernoulli distribution*, Journal of Multivariate Analysis, 97:1573-1585.

Khaled M. (2008), A multivariate generalization of a Markov switching model, working paper, C.E.S., Université Paris 1.

Khaled M. (2008), *Estimation bayésienne de modèles espace-état non linéaires*, Thèse de l'Université Paris1.