Techniques d'élargissement et de grappes pour l'application de l'algorithme CFTP Monotone

M.Y. BOUNNITE and A. NASROALLAH

*Department of Mathematics, Cadi Ayyad University, Faculty of Sciences Semlalia
B. P. 2390, Marrakesh, Morocco*

Résumé. Le célèbre protocole du à Propp et Wilson (1996,1998) nommé couplage depuis le passé "Coupling From The Past (CFTP)", assure la simulation exacte de la distribution stationnaire d'une chaîne de Markov. Dans le cas où le cardinal de l'espace d'état est très grand, l'algorithme CFTP dans sa forme standard déploie un nombre important d'opérations ; réclamant ainsi, trop de place mémoire. Afin de remédier à cette complexité, nous proposons une approche d'élargissement de l'espace d'état, rendant la chaîne de Markov stochastiquement monotone et apériodique. En conséquent, le coût de la simulation est réduit en utilisant l'algorithme CFTP monotone. Un algorithme est également proposé dit "CFTP par Grappe" dédié a une structure donnée de l'espace d'état, permet de réduire le coût de la simulation.

Mots clés. Monte Carlo, simulation, chaîne de Markov, monotonie stochastique, couplage à partir du passé.

Abstract. The standard Coupling From The Past (CFTP) algorithm is an interesting tool to sample from exact stationary distribution of a Markov chain. But it's very expensive in time consuming for large chains. There exist a monotone version of CFTP, called MCFTP, that is less time consuming for monotone chains. In this work, we propose two techniques to get monotone chain allowing use of MCFTP : Widening technique based on adding two fictitious states and Clustering technique based on partitioning the state space in clusters. Usefulness and efficiency of our approaches are showed through a sample of Markov Chain Monte Carlo simulations.

Keywords. Monte Carlo, simulation, Markov chain, stochastic monotonicity, coupling from the past.

1 Algorithme CFTP Standard

Considérons une chaîne de Markov $\mathbf{X} = (X_n)_{n \geq 0}$ homogène à temps discret sur un ensemble d'état fini E , avec une matrice de transition $K(., .)$, et une distribution initiale

μ_0 . Supposons que X est irréductible, récurrente positive et apériodique. Il existe alors une distribution stationnaire unique $\pi = (\pi_x)_{x \in \mathbf{E}}$ de X ($\pi_x = \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = x)$). Afin d'estimer cette distribution stationnaire, on laisse évoluer cette chaîne pendant une période transitoire de longueur T_0 appelée 'burn-in' jusqu'au régime stationnaire. Ensuite, on génère un échantillon de l'évolution de la chaîne après T_0 , pour l'utiliser dans l'estimation de π ; cependant le problème est de prédire T_0 . L'algorithme de couplage depuis le passé, nécessite qu'un nombre presque sûrement fini de transitions garantit que la distribution obtenue est exactement la distribution stationnaire sans se soucier de T_0 .

Soit $\phi : \mathbf{E} \times [0, 1) \rightarrow \mathbf{E}$ une règle de transition X , appelée fonction de mise à jour, elle est définie par :

$$\mathbb{P}[\phi(x_1, U) = x_2] = K(x_1, x_2), \quad x_1, x_2 \in \mathbf{E}.$$

où u est une variable aléatoire uniforme sur $]0, 1[$ et $U \sim \mathcal{U}_{[0,1]}$. La fonction ϕ nous permet de déterminer à partir de l'état courant X_n , l'état suivant X_{n+1} par la relation de récurrence suivante :

$$\phi(X_n, u) = X_{n+1}$$

L'idée de l'algorithme CFTP consiste à lancer simultanément, depuis un instant $T < 0$ dans le passé jusqu'à l'instant 0, N copies de la chaîne de Markov étudiée X . Chaque copie démarre à l'instant T dans un état différent, en suivant la même fonction de mise à jour ϕ , et en utilisant la même suite de nombres aléatoires $u_1, u_2, \dots$ pour la mise à jour. On cherche le couplage de toutes ces chaînes, c'est-à-dire qu'elles aient atteint le même état à un instant T_c entre T et 0. Deux chaînes ou trajectoires qui se rencontrent, suivent exactement le même chemin après.

Considérons maintenant une suite de temps $(T_k)_{k \in \mathbb{N}}$ entiers telle que $\dots T_{k+1} < T_k < \dots < T_1 < T_0 < 0$: La transition composée de la chaîne X entre l'instant T_2 et l'instant T_1 est donnée en utilisant l'ensemble des variables aléatoires $u = (u_{T_2+1}, u_{T_2+2}, \dots, u_{T_1}) \in \mathcal{U}_{]0,1[}^{T_1-T_2}$ et la fonction de mise à jour composée Φ définie de $E \times \mathcal{U}_{]0,1[}^{T_1-T_2}$ dans E par :

$$\Phi_{T_1}^{T_2}(i, \mathbf{U}) = \phi(\phi(\dots \phi(i, U_{T_2+1}), \dots, U_{T_1-1}), U_{T_1})$$

où $X_{T_2} = i$.

Alors on dit qu'il y a couplage de toutes les trajectoires de la chaîne de Markov lancées à partir du temps $T < 0$ si et seulement si :

$$\exists i_0 \in E, \forall i \in E \quad i_0 = \Phi_0^T(i, \mathbf{u}) \quad \text{pour un } \mathbf{u} \in \mathcal{U}_{]0,1[}^{-T}$$

Maintenant en fixant un $T_0 < 0$ et une suite $u_{T_0+1}, \dots, u_0$, le couplage de toutes les trajectoires peut ne pas avoir lieu. Dans ce cas on choisit un nouveau $T_1 < T_0$ et on réitère la même procédure en utilisant la suite $u'_{T_1+1}, \dots, u'_{T_0-1}, u'_{T_0}, u_{T_0+1}, \dots, u_0$. Si le couplage a eu lieu on arrête, sinon on réitère le procédé pour un nouveau choix $T_2 < T_1$ et ainsi de suite jusqu'au couplage de toutes les trajectoires.

2 Élargissement de l'espace d'état

Sous les mêmes hypothèses et notations, l'idée de l'élargissement ou enrichissement de l'espace d'état est d'ajouter deux états à l'espace d'état E , de telle façon à rendre la chaîne de Markov stochastiquement monotone et apériodique, tout en sauvegardant sa propriété d'irréductibilité, sachant bien que Les deux états ajoutés ne sont que l'état minimal et l'état maximal selon l'ordre partiel défini. Le gain tiré est l'utilisation de l'algorithme CFTP monotone qui ne simule que deux trajectoires celles issues des états extrémaux enveloppant ainsi toutes les trajectoires issues du reste des états. Il en résulte donc une nouvelle chaîne de Markov $\mathbf{X}' = (X'_n)_{n \geq 0}$ dite chaîne de Markov élargie définie sur un nouvel espace d'état $E' = E \cup \{0, N + 1\}$ et un noyau de transition

$\mathbf{K}' = K'(i, j)_{0 \leq i, j \leq N+1}$ défini comme suit :

$$K'(i, j - 1) = \begin{cases} \frac{K(i, j)}{1 + \alpha} & \text{si } 1 \leq i \leq N \text{ et } 1 \leq j \leq N \\ \frac{\alpha}{2(1 + \alpha)} & \text{si } 1 \leq i \leq N \text{ et } N + 1 \leq j \leq N + 2 \end{cases}$$

Où

$$\alpha = \min\{K(i, j) \neq 0, i, j \in E\}$$

$$K'(0, j) = \begin{cases} \max_{1 \leq i \leq N} (K'(i, j)) & \text{si } j = 0 \\ \max_{1 \leq i \leq N} F_i(j) - \max_{1 \leq i \leq N} F_i(j - 1) & \text{si } j = 1, \dots, N + 1 \end{cases}$$

$$K'(N + 1, j) = \begin{cases} \min_{1 \leq i \leq N} (K'(i, j)) & \text{si } j = 0 \\ \min_{1 \leq i \leq N} F_i(j) - \min_{1 \leq i \leq N} F_i(j - 1) & \text{si } j = 1, \dots, N + 1 \end{cases}$$

Avec

$$F_i(t) = \sum_{l=0}^t K'(i, l) \quad (1)$$

Soit le fonction de mise à jour $\varphi : E' \times [0, 1) \rightarrow E'$, elle est définie par :

$$\mathbb{P}(\varphi(i, U) = j) = K'(i, j), \quad \forall i, j \in E'$$

Et vérifiant les inégalités d'ordre partiel défini sur E'

$$\varphi(0) \preceq \varphi(i) \preceq \varphi(N + 1), \quad \forall i \in E'$$

On peut réécrire les termes du noyau de transition K' de la chaîne X' en utilisant l'expression (1), on a donc :

$$K'(i, j - 1) = \begin{cases} F_i(j) - F_i(j - 1) & \text{si } 1 \leq i \leq N \text{ et } 1 \leq j \leq N \\ \frac{\alpha}{2(1 + \alpha)} & \text{si } 1 \leq i \leq N \text{ et } N + 1 \leq j \leq N + 2 \end{cases}$$

$$K'(0, j) = \begin{cases} \max_i (F_i(0)) & \text{si } j = 0 \\ \max_{1 \leq i \leq N} F_i(j) - \max_{1 \leq i \leq N} F_i(j-1) & \text{si } j = 1, \dots, N+1 \end{cases}$$

$$K'(N+1, j) = \begin{cases} \min_i (F_i(0)) & \text{si } j = 0 \\ \min_{1 \leq i \leq N} F_i(j) - \min_{1 \leq i \leq N} F_i(j-1) & \text{si } j = 1, \dots, N+1 \end{cases}$$

La chaîne X' ainsi définie, de matrice de transition K' sur le nouvel espace $E' = E \cup \{0, N+1\}$ est irréductible et apériodique.

Lemma 1 *Considérons la partition du noyau de transition $\tilde{K}$ suivante :*

$$\tilde{K} = \left[\begin{array}{c|c} T & W \\ \hline R & Q \end{array} \right]$$

où T une matrice carrée d'ordre N et Q est une matrice carrée d'ordre 2. Le noyau $\tilde{K}_N$ défini par :

$$\tilde{K}_N = T + W(I_{2 \times 2} - Q)^{-1}R,$$

gouverne l'évolution d'une chaîne de Markov irréductible sur $E = \{1, \dots, N\}$.

Theorem 1 *Soit $\tilde{\alpha} = [\tilde{\pi}_1, \dots, \tilde{\pi}_N, \tilde{\pi}_0, \tilde{\pi}_{N+1}]$ la probabilité stationnaire de $\tilde{K}$. Alors $\tilde{\alpha}_N = [\tilde{\pi}_{1N}, \dots, \tilde{\pi}_{NN}]$, où*

$$\tilde{\pi}_{iN} = \tilde{\pi}_i \left(\sum_{j=1}^N \tilde{\pi}_j \right)^{-1}$$

est la probabilité stationnaire de $\tilde{K}_N$.

Proof. Posons $x = [\tilde{\pi}_1, \dots, \tilde{\pi}_N]$ et $y = [\tilde{\pi}_0, \tilde{\pi}_{N+1}]$, $\tilde{\alpha}$ s'écrit donc $\tilde{\alpha} = [x, y]$, étant loi stationnaire de $\tilde{K}$ elle vérifie donc : $\tilde{\alpha} = \tilde{\alpha}\tilde{K}$, ce qui implique que

$$x = xT + yR \quad (2)$$

$$y = xW + yQ \quad (3)$$

L'égalité (3) est équivalente à $y = xW(I_{2 \times 2} - Q)^{-1}$, d'où

$$yR = xW(I_{2 \times 2} - Q)^{-1}R \quad (4)$$

En additionnant xT à (4), on obtient :

$$xT + yR = x(T + W(I_{2 \times 2} - Q)^{-1}R).$$

Maintenant, en remplaçant x dans la dernière équation par sa définition dans (2) et $\tilde{K}_N$ par sa définition également on a alors :

$$x = x\tilde{K}_N$$

D'où x et $\tilde{\alpha}_N$ sont proportionnels, et puisque $\tilde{\alpha}$ est une loi de probabilités, alors $x = \tilde{\alpha}_N$.

■

3 Algorithme CFTP par grappe

Dans cette section, nous présentons une version particulière de l'algorithme CFTP monotone, dite 'CFTP par grappe'. Pour cela, on suppose que l'espace d'état E est une partition de m sous-ensembles E_l ; i.e : $E = \sqcup_{l=1}^m E_l$, et $2m < |E|$. Chaque E_l est muni d'un ordre partiel $\preceq$ pour lequel il existe un élément maximal i_{max}^l et un élément minimal i_{min}^l . On suppose que la fonction de mise à jour est monotone sur chaque E_l , autrement dit pour tout $l = 1..m$:

$$\forall i^l, j^l \in E_l \text{ telque } i^l \preceq j^l \Rightarrow \phi(i^l, u) \Rightarrow \phi(j^l, u) \quad \forall u \in \mathcal{U}_{[0,1]}$$

. Et la fonction de mise à jour composée telle qu'elle est définie est comme suit :

$$\forall i^l \in E_l \quad \forall T_2 < T_1 < 0 \quad \Phi_{T_1}^{T_2}(i_{min}^l, \mathbf{u}) \preceq \Phi_{T_1}^{T_2}(i^l, \mathbf{u}) \preceq \Phi_{T_1}^{T_2}(i_{max}^l, \mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{U}_{[0,1]}^{T_1-T_2}$$

Dans ce cas, la restriction de la chaîne X sur chaque sous-ensemble E_l est stochastiquement monotone. Cependant, les trajectoires issues des états appartenant à chaque E_l sont comprises entre celles issues des états i_{max}^l et i_{min}^l pour tout $l = 1..m$; L'idée de cet algorithme consiste à lancer uniquement et simultanément, depuis un instant $T < 0$, les trajectoires issues des états extrémaux i_{max}^l et i_{min}^l pour chaque sous ensemble E_l , pour tout $l = 1, ..m$, au lieu de lancer tous les états de E , ceci réduit remarquablement le temps de simulation de la chaîne X en fonction du nombre des grappes (dans ce cas $2m$ trajectoires sont lancées et $2m < n$).

CFTP Clustering Algorithm

Étape 1. Initialiser le temps en $T := -2$ et poser $\mathbf{U}$ comme étant une séquence vide.

Étape 2. Générer des variables aléatoires $U_T, U_{T+1}, \dots, U_{T/2-1} \in [0, 1)$, et les insérer dans le vecteur $\mathbf{U}$ dans cet ordre (i.e. $\mathbf{U} := (U_T, U_{T+1}, \dots, U_{-1})$).

Étape 3. Lancer $2m$ copies de la chaîne $\mathbf{X}$ à partir de chaque état extrémal :

$i_{min}^l, i_{max}^l \in \mathbf{E}_l \quad \forall l = 1, ..m$ dans le temps T .

Laisser évoluer chaque chaîne depuis T jusqu'à l'instant 0 calculer la fonction de mise à jour ϕ en utilisant la séquence $\mathbf{U}$.

Calculer $\Phi_T^0(e_l, \mathbf{U})$, $\forall e_l \in \{i_{min}^l, i_{max}^l\} \quad \forall l = 1, ..m$.

Étape 4. (Temps de Coalescence)

Si $\exists x_0 \in \mathbf{E}$, $x_0 = \Phi_T^0(e_l, \mathbf{U})$, $\forall e_l \in \{i_{min}^l, i_{max}^l\} \quad \forall l = 1, ..m$, alors retourner x_0 et arrêter,

Sinon actualiser le temps de départ $T := 2T$ et aller à l'**Étape 2**.

REFERENCES

- [1] Fakhouri, H. and Nasroallah, A. (2009). On the simulation of Markov chain steady-state distribution using CFTP algorithm. *Monte Carlo Methods Appl.* Vol. 15 No. 2, pp. 91-105.
- [2] Connor, S. (2007). Coupling : Cutoffs, CFTP and Tameness. Ph. D. thesis, University of Warwick.
- [3] Connor, S. and Kendall, W. S. (2006). Perfect simulation for a class of positive recurrent Markov chains. Research Report 446, University of Warwick.
- [4] D.J. Daley, Stochastically monotone Markov chains, *Z. Wahrscheinlichkeitsth. Verw. Geb.* 10 (1968) 305-317.
- [5] Green, P. J. and Murdoch, D. J. (1998). Exact sampling for Bayesian inference : towards general purpose algorithms. In *Bayesian Statistics 6*. To appear.
- [6] Kendall, W. S. (1997). On some weighted Boolean models. In D. Jeulin, editor, *Advances in Theory and Applications of Random Sets*, pp. 105-120. World Scientific Publishing Company, Singapore.
- [7] Kendall, W. S. and Møller, J. (2000). Perfect Simulation Using Dominating Processes on Ordered Spaces, with Application to Locally Stable Point Processes. *Advances in Applied Probability*, Vol. 32, No. 3, pp. 844-865.
- [8] Mira, A., Møller, J. and Roberts, G. O. (2001). Perfect Slice Samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 63, No. 3, pp.593-606.
- [9] Møller, J. (1999). Perfect simulation of conditionally specified models. *J. R. Statist. Soc. B*, 61, Part 1, pp. 251-264.
- [10] Murdoch, D. J. and Green, P. J. (1997). Exact sampling from continuous state space. *Scandinavian Journal of Statistics*, 25 :483-502.
- [11] Popp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9 :223-252.