統計数理研究所 75 周年記念事業（2019）

マルコフ連鎖モンテカルロ法の世界を拡げる
ー 高次元の多峰性分布からのサンプリングー

Expanding world of Markov chain Monte Carlo
– Methods for sampling high-dimensional multimodal distributions–

モデリング研究系 伊庭 幸人 （Yukito Iba）

要　　旨

多峰性分布に強いマルコフ連鎖モンテカルロ法とその応用について述べた．

キーワード：レプリカ交換モンテカルロ法，マルチカノニカル法，多峰性分布，レアイベントサンプリング

1.　はじめに

マルコフ連鎖モンテカルロ法 (MCMC) は高次元・多変量の確率分布からの乱数生成（サンプリング）の手法である．当初は統計物理の世界で使われていたのが，1990 年代以降に統計学や機械学習の世界で利用されるようになり，大きな影響をもたらしたことはよく知られている．著者は長年マルコフ連鎖モンテカルロ法の応用に興味を持っているが，その間，以下の 2 つの点にこだわりを持ち続けてきた．

1. 通常の MCMC ではサンプリングしにくい多峰性のある分布（確率の高い領域が複数に分かれている分布）に対して有効な手法の開発．
2. 統計データ解析でも統計物理でもない分野への応用．

MCMC というと，ベイズ統計やそのソフトウェアを思い浮かべる方が多いかもしれないが，それとはかなり違う問題意識で研究を進めてきたといえると思う．以下，その軌跡を紹介したい．

2.　拡張アンサンブル MCMC とその応用

多峰性の分布に対して MCMC を適用する手法として Simulated Annealing 法が知られている．Simulated Annealing 法では「温度」に対応するパラメータを最初は「高温」に設定し次第に「低温」にすることで，重要でない分布の山（局所的極大）にトラップされることを防ぐ．しかし，これはあくまで最適化の範疇であって，多峰性の分布を正しくサンプルすることができるとは限らない．統計物理での有限温度の分布，ベイズ統計での事後分布からのサンプリングには別のアイディアが必要である．1993 年ごろにこの問題に興味を持った著者は，今日「レプリカ交換モンテカルロ法」（パラレルテンパリング）と呼ばれている方法を独自に思いついたが，あまり知られていない先行研究（木村・瀧（1990）及び Geyer（1991））の存在に気づいたため，査読つき論文を出すには至らなかった（短報は「統計数理」の研究報告会要旨に掲載され，オンラインでも見られる． 伊庭（1993））．それから数年たってから考えたのは，温度に

限らず，収束を悪くしている特定の制約条件を緩めることでよりよい結果が得られるということで，この考え方をタンパク質の格子モデルに適用して開発した手法（MSOE 法）は当該分野で高く評価されている（第 2 報の論文 Chikenji et al.（1999)は web of science で引用数 74）．また，それまでに考えたことを踏まえて，関連する一連の手法を「拡張アンサンブル法」としてまとめたレビュー論文 Iba（2001)は 2018 年現在 web of science で 142 件引用されている．

3. さまざまな分野での高次元分布からのサンプリング

MCMC が有用な分野としては統計物理とベイズ統計が代表格であるが，これらの分野では，それぞれの主役である「ギブス分布（カノニカル分布)」「事後分布」が高次元の確率分布であるという点が共通である．そう考えると，統計物理やベイズ統計以外でも MCMC によって面白いことができる分野はもっとあるのではないか．「最適化からサンプリングへ」というスローガンは，単に MCMC の応用ということを超えて，いろいろな学問の世界で「確率構造」「測度構造」を考える糸口になるかもしれない．過去 10 年ほど行ってきた研究は，こうした発想に基づくものである．

具体的な応用としては，ランダム行列やランダムグラフの大偏差の数値計算，力学系の珍しい軌道のサンプリング，ランダム系の不純物に関する平均（クエンチ平均）の効率的計算，ラテン方陣や分割表の数の計算，複雑な帰無仮説下での検定統計量の分布の計算（サロゲーション法への応用）などがある．これらの例の多くでは，与えられた分布の極端な裾からのサンプリングが必要とされるが，そこで多峰性分布からのサンプリングの手法が重要となる．マルチカノニカル法，レプリカ交換モンテカルロ法などを利用することで，与えられた確率分布のもとでの生起確率が 10^{-200} といった極端なレアイベント（大偏差事象）のサンプリング及び確率計算ができることが示されている．興味を持たれた方は，解説 Iba et al.（2014)の中の事例と引用文献を参照されたい．

4. 今後の展望

マルチカノニカル法やレプリカ交換モンテカルロ法（パラレル・テンパリング）など多峰性分布に対応した MCMC 手法は，ベイズ統計にもとづくデータ解析でも有用なはずであり，研究レベルではレプリカ交換モンテカルロ法を中心にいろいろな適用例がある．しかし，現場での応用が MCMC ソフトウェア中心に展開しているため，十分活用されていないのが現状である．これらの手法に対応する統計ソフトウェアが登場することで，より自由な統計モデリングが行えるようになることを望みたい．

参 考 文 献

Chikenji, G., Kikuchi, M. and Iba, Y. (1999). Multi-self-overlap ensemble for protein folding: ground state search and thermodynamics, *Physical Review Letters*, **83** (9), 1886–1889.

Geyer, C. J. (1991). *Computing science and statistics: Proceedings of 23rd Symposium on the Interface* (ed. E. Keramidas), , 156–163, Interface Foundation, Fairfax Station.

Iba, Y. (2001). Extended ensemble Monte Carlo, *International Journal of Modern Physics C*, **12** (05), 623–656.

Iba, Y., Saito, N. and Kitajima, A. (2014). Multicanonical MCMC for sampling rare events: an illustrative review, *Annals of the Institute of Statistical Mathematics*, **66** (3), 611–645.

伊庭幸人 (1993). メトロポリス的モンテカルロ法の緩和について (統計数理研究所研究活動 (平成 4 年度 研究報告会要旨)), 統計数理, **41** (1), 65–67.

木村宏一, 瀧 和男 (1990). 時間的一様な並列アニーリングアルゴリズム, 電子情報通信学会 NC-90-1, 1–8.