
マルコフ連鎖モンテカルロ法の

これまでとこれから

福島孝治 *Koji Hukushima*

所属：東京大学大学院総合文化研究科

○略歴

平成3年 筑波大学第一学群自然科学類 卒業

平成5年-8年 日本学術振興会 特別研究員 DC1

平成8年 筑波大学物理学研究科 博士課程修了、博士（物理学）

平成8年 東京大学物性研究所 助手

平成14年 東京大学大学院総合文化研究科 助教授

平成19年 同上 准教授

平成28年- 国立研究開発法人 物質・材料研究機構 派遣研究員

○講演概要

マルコフ連鎖モンテカルロ法 (MCMC) と現在呼ばれている方法は 1953 年の Metropolis らによる液体の統計力学的研究のために開発された数値計算手法に端をなす。ほぼ同時期に開発された分子動力学法 (MD) とともに、多粒子系の非摂動的な手法として現在でも多くの研究に用いられている。ちょうど計算機が世に現れ、その積極的な活用によって新たな研究の潮流を作り出そうとするまさに「今やるべき研究」であったと想像する。その後、並進対称性のある統計力学モデルへの応用が精力的になされ、熱力学極限への系統的な接近法が整備され、1980 年代ころには相転移や臨界現象の研究手法として確立されていく。計算機の進展と相まって、数値計算手法を手段とするシミュレーション科学が自然科学の方法として認知されていったのもその頃であった。また、同じころ、自然科学以外でも、統計科学の分野ではベイズ統計が MCMC を基礎として実践的に使えることが認識されるようになり、そのような研究が大きく発展する。

1990 年代になり、国内でも計算物理学が重点領域研究としてプロジェクト研究となり、大型計算機センターのスーパーコンピュータの利用が多くの研究者にとって身近になり、またワークステーションと呼ばれる中規模計算機が各大学に配備されることになる。その頃に私は大学院生として研究活動をスタートさせることになる。非一様で複雑な自由エネルギー構造を持つ系に MCMC を適用すると平衡状態の実現に問題があることが認識されるようになったころである。MCMC は原理的には使えるが、現実的にはうまくいかないというわけである。タンパク質研究に使うこともまだまだままならない状況であった。私が取り組んだ問題はスピングラスとよばれるランダムな磁性体の研究であるが、その平均場理論から複雑な自由エネルギーランドスケープの存在が明らかになり、その描像の正当性を確認することが重要な時期であった。そのためにも複雑な系にも「使える MCMC」を作ることが必至の状況であった。

博士課程に進学したときに、面白いと思って勉強して自分でプログラムを書いていたのが、遺伝アルゴリズムであった。しかし、これは研究室内で評判が非常に悪く、今でもトラウマになっている。おそらく、このままではまずい方向にしか進まないと思われたと推測するが、別の方向を考えるように紹介されたのが Berg らによるマルチカノニカル法の論文であった。この方法のよさが全く理解できずに、悶々としているところに、当時研究室の助手の根本さんにほぼ導かれるように議論したのが交換モンテカルロ法であった。議論した同じ週に根本さんが昼食会でアイデアを紹介したときには、私はプログラムの実装が終わって最初のデータが出ていたことを記憶している。その後の研究室で発表した資料が残っていた (図 1)。

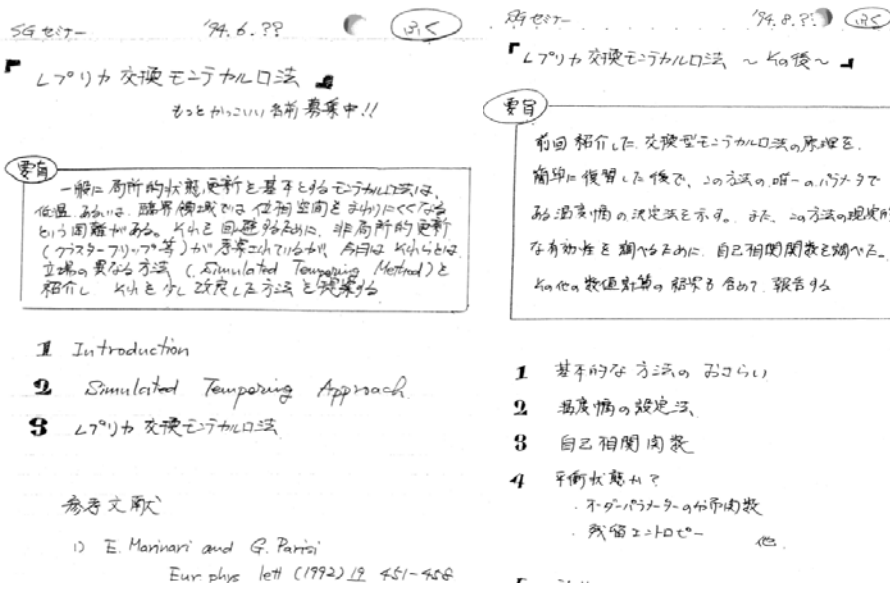


図1 研究室での発表のためのレジюме。第一報(左)と第二報(右)。この二か月の間におおよその原型は完成。その後、9月に日本物理学会で発表。翌年の3月の日本物理学会ではシンポジウムで発表することになる

これはスピングラス研究のために必要な手法として考案したが、同じように複雑な自由エネルギー構造をもつ系にも活用でき、Sugita-Okamotoによる分子動力学法への拡張とともにタンパク質のフォールディング問題への応用に繋がっていく。2000年代には拡張アンサンブル法と呼ばれる一連の改良が出揃い、とにかく「使えるMCMC&MD」ができたように思える。少し大げさな言い方をすれば風景が変わった気がする。少なくともスピングラス研究では、交換法「前」と「後」ではMCMCの計算結果の信頼度は大きく上がり、見えてくるものが変わった。MCMCは広い分野に应用されているが、このような大きな改良は自然科学者によることが多いのは興味深い観測事実である。元々、自然科学者は「問題の解き方」そのものにはさほど興味はなく、「問いた結果」に大きな関心が向く。その意味で、解法の改良に研究の矛先が向くことは少ないのだけど、結果を導くための方法がないとすれば方法から作ってしまおうと考える傾向はあるのかも

しれない。素材があるから包丁はよく研がれるのであって、素材を持たないことにはよい包丁は生まれないのである。

昔話はよいとして、「これから」について議論しよう。生物物理の諸問題の自由度を真面目に考えると、これまでのMCMCはまだそこそこ「使える」程度であって、生体系のシミュレーションツールとして真に使うには物足りない感じである。タンパク質の問題ではそれでも「使える」状況になったからこそ、次の段階としての力場の精度問題が表面化することになる。これも風景の変化の一例と考えることができる。力場の問題は第一原理電子状態計算を行うことが正攻法であるが、その精度でMCMCやMDを行うほどにまだ計算機の性能は足りないし、そもそも精度が必要かどうかはよくわからない。ただ、物質科学の現状を見てみると、第一原理計算の結果をもとに機械学習を用いて、古典ポテンシャルを評価する方法が精力的に試みられており、分子シミュレーションでも必要に応じて取り入れる必要があるように思われる。さらに、MCMCに関しては個人的には詳細つりあい条件の破れた方法が現状ではもっとも有力であると考えている。特にEvent-chain MCは有力で、分子シミュレーションのための壁を一つ一つ解消してきている。数年後には本当に使い物になるかどうかは明らかになっていることだろう。

MCMCの「これから」を議論するもう一つの方向性は、ベイズ統計に基づくデータ駆動科学的アプローチであろう。木川先生の講演にもあるように、実験結果と理論解析をうまく組み合わせることが新しい理解には重要であることは、これまでの自然科学の歴史を見ても明らかである。その中において、実験と数値計算によって得られた大量のデータから重要な特性を抽出する「操作」は容易ではなく、このデータ解析の中にMCMCが重要な役割を果たすことが予想される。機械学習が打ち出の小槌のように「何か」を出してくれることはないので、自然科学者がデータ解析の目的を明確にすることが重要であり、その解析の必要性が新たなブレークスルーを生むような予感がする。これが現代の「今やるべき研究」なのだと思う。