マルコフ連鎖モンテカルロ法とその統計学への応用

伊庭幸人　統計数理研究所　予測制御系

統計物理では 40 年以上にわたって、動的モンテカルロ法とよばれるタイプのアルゴリズムが使われてきた。この方法は多変数の非ガウス分布からのサンプリングに適しており、物理では、液体・磁性体・結晶・素粒子などの各分野で大きな役割を果たしている。ここ 10 年ほどの間に、この方法が統計学の分野に導入され、盛んに使われるようになった。統計学者はこれをマルコフ連鎖モンテカルロ法と呼んでいる。はじめは空間統計や画像処理、また、“ボルツマンマシン学習” の名で神経回路網の分野での応用が行なわれたが、その後、統計学の他の計算法 (data augmentation など) との関連が認識され、“Gibbs Sampler” タイプの方法を医学統計・経済時系列などで現れるミクスチャ・モデルに応用することが盛んになった。現在ではさらに広く、たとえば、離散変数を含んだ推定やモデル選択などにも応用されている。

マルコフ連鎖モンテカルロ法の骨子は、与えられた分布を不変分布とするようなマルコフ連鎖をシミュレートしそこからのサンプルを利用することにある。与えられた分布を不変分布とするようなマルコフ連鎖を設計するためには、“詳細釣合の条件” を満たすようにするのがポイントである。典型的な方法としてはメトロポリス法と Gibbs Sampler(熱浴法) がある。両者は離散変数についてはそれほど違わないが、連続変数の場合にはかなり違う。メトロポリス法では、少数の変数を少しだけ動かす “試行” を乱数を用いて reject/accept するような方法をとるのに対し、 Gibbs Sampler ではある変数 (群) を他の変数を固定した条件つき分布からサンプルする方法をとる。メトロポリス法は適用範囲が広いが、その代償として状態空間内の乱歩による遅い緩和が目立つことがある。 Gibbs Sampler は条件つき分布が安価に計算できる場合、とくに多くの変数を同時にとりなおせる場合にとくに有効である。両者を混ぜ合わせたような方法も可能である。これは Metropolis-Hastings 法と呼ばれることがある。

マルコフ連鎖モンテカルロ法は分布からのサンプリング法であるから、ベイズ統計における事後分布の扱い (あるいは missing data や潜在変数の扱い) にもっとも適している。もちろん、ベイズ関係でなくても、サンプル生成が重要な問題、たとえば、尤度関数の規格化定数が解析的に求まらない場合の最尤法の計算なども有力な応用例である。ただし、マルコフ連鎖モンテカルロ法は、本来、分布からのサンプリング法 (あるいはそれを利用した期待値の計算法) であって、積分の計算法ではないことに注意すべきである。したがって、尤度関数の規格化定数や周辺尤度そのものを直接計算することはできない。積分そのものがどうしても必要な場合は、求める積分の適当なパラメータによる対数微分を計算し、その情報を利用して 1 次元積分に帰着させるのが最も簡単である。

マルコフ連鎖モンテカルロ法より少し前に話題となった方法として simulated annealing 法がある。与えられた確率分布 $P(\alpha)$ に対して変形した分布 $P_T(\alpha) = \frac{P(\alpha)^{1/T}}{\sum_\alpha P(\alpha)^{1/T}}$ を考える。この分布は “温度”T が小さいときにもとの分布 P の最大確率の状態の上に集中する。これを利用すると、T の大きいところから出発してマルコフ連鎖モンテカルロ法の計算を行ないながらゆっくりと $T \to 0$ とする (anneal する) ことで、効果的に大域的な最適化を行なうことができる。これが simulated annealing 法と呼ばれる手法である。マルコフ連鎖モンテカルロ法の特定の種類として simulated annealing 法があるというより、“マルコフ連鎖モンテカルロ法を最適化に利用すること” を simulated annealing 法というのである。最適化でなく、分布からのサンブリング

の場合には、単純に anneal の操作を行なっても、必ずしも収束は良くならない。これは、2 つの大きな峰のある分布の場合を考えてみれば容易に理解できる。anneal してもいずれか一方に行きつくだけで、両者を正しい割合で混ぜ合わせた期待値は得られない。なお、単純な anneal では駄目でも、うまく工夫をすれば annealing の考え方を分布からのサンプリングに生かすことは可能であり、ここ数年そうした方法が開発されている (simulated tempering 法、Metropolis-coupled chain 法など)。

マルコフ連鎖モンテカルロ法の最大の問題点は、生成されたサンプルが独立でないということである。ただし、各サンプルについて定義された量の期待値の計算が目的の場合は、すべてのサンプルを用いて差し支えない。個々のサンプルから量を計算する手間を無視すれば、サンプル列の最初の部分を十分長く捨てたあとでは、利用するサンプルの間隔が短いほど計算結果の分散は小さくなる。(伊庭 (1996) の説明は "統計量" という用語の濫用もあってやや強過ぎる表現となってしまったが、欲しいものが期待値より複雑な場合には全部のサンプルを使った方がいいとは (少なくともすぐには) いえない。相関のあることを忘れると不偏推定量が得られないこともある。ただし、途中のサンプルをむやみに捨てずに、有効に利用することが重要であることには変わりがない。)

真に問題なのは、サンプル間の相関のために、結果の分散がきわめて大きくなりうることである。また、最初の状態の影響がいつまでも残ってしまい、バイアスの原因となることもある。これらの評価法がいろいろ提案されているようであるが、極端な状態を含んださまざまな初期状態から長さと乱数を変えて複数回計算を行なうのが基本である。サンプルパスの様子をいろいろな方法を使って視覚化することも重要である。筆者はこの目的のために統計ソフトウエア "**S**" を使用している。

高次元のマルコフ連鎖の挙動はきわめて多彩であり、マルコフ連鎖の一般論から学べることには限界がある。むしろ、類似のモデルについて計算を行なっている研究者同士の (否定的結果を含めた) 情報交換が重要である。物理での応用においては、マルコフ連鎖モンテカルロ法の収束の悪い点で興味ある物理現象がおきているような例がみられる。データ解析への応用にあたっても (ハイパー) パラメータを変えていったときに、収束の遅くなる点があれば、何か興味あることが起きていないかどうか調べる必要がある。収束の悪さをデータやモデルの特徴を見い出す機会として利用できる場合もあると思われる。

マルコフ連鎖モンテカルロ法の統計への応用については、よく知られているように、早い時期 (少なくとも 1980 年ごろ) からの尾形と種村の仕事がある。simulated annealing 法についても、Tsuda,T and Kiyono,T (1964) をはじめとして、わが国の研究者による先駆的な研究が存在する。国内で出された新しい考えが海外で流行してはじめて広く利用されるようになることが多いのは残念なことである。本研究会のような試みが新しい考えを育てる糸口となることを願いたい。今回は概説的な内容の講演になったが、筆者の研究及びより具体的な解説については以下の文献を参照されたい。

Iba,Y.(1989) Bayesian statistics and statistical mechanics, Cooperative Dynamics in Complex Physical Systems, Ed. H.Takayama, Springer-Verlag, Berlin.

伊庭幸人 **(1991a)** メトロポリス的モンテカルロ法の巨視的パラメータ推定への応用 – 2 次元イジング模型の場合 –, 統計数理 39 No.1 1-21.

伊庭幸人 **(1991b)** メトロポリス的モンテカルロ法の擬ベイズ法への応用 – 変化点問題を例として –, 統計数理 39 No.2 225-244.

Iba,Y.(1992) An application of Metropolis-type algorithm to a complex classification problem, ISM Research Memorandum No.440.

伊庭幸人 **(1993)** 統計数理研究所 年度末発表会 (1992 年度) 要旨、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 41 No.1 65-67.

伊庭幸人 **(1996)** マルコフ連鎖モンテカルロ法とその統計学への応用, 統計数理 44 No.1 49-84.