マルコフ連鎖モンテカルロ法

WIRAWAN DONY DAHANA

1. マルコフ連鎖モンテカルロ法（Markov Chain Monte Carlo）

統計的推論の問題において特定の分布の母数を推定する際に乱数を生成し，それを用いて母数を近似する方法をモンテカルロ法という．乱数を生成したい分布を目標分布（target distribution）というが，目標分布からのサンプリングは直接または間接的に行われる．目標分布から直接にサンプリングが困難な場合，サンプリングが容易な分布を提案分布（proposal distribution）として採用し，そこから間接的に乱数を生成する．例えば推定したい目標分布を $f(x)$ とすれば，この分布の母数を推定するために任意の関数 $h(\cdot)$ に関する積分を評価しなければならないが，モンテカルロ法では提案分布 $g(x)$ から T 個の乱数 $x^{(1)}, x^{(2)}, \cdots, x^{(T)}$ を生成し，$f(x)$ の母数を次のように近似する．

$$\int h(x)f(x)dx$$
$$\approx \frac{1}{T}(h(x^{(1)})+h(x^{(2)})+\cdots+h(x^{(T)}))$$

$h(x)=x$ の場合，上記の式は分布 $f(x)$ の平均の近似式を表している．大数の法則より $T\to\infty$ のとき平均による近似は対象の積分の値に確率収束することが保証される．モンテカルロ法の代表的な方法として棄却サンプリング（rejection sampling）と重点的サンプリング（importance sampling）がある[1]．

しかし，目標分布が高次元な場合や複雑な場合には従来のモンテカルロ法を適用することが困難であり，この問題を解決するためにマルコフ連鎖モンテカルロ法（Markov Chain Monte Carlo, MCMC）が用いられる．MCMC とは不変分布（invariant distribution）が目標分布であるようなエルゴード的なマルコフ連鎖を構成し，推移を繰り返すことにより目標分布を得るための推定法である[2]．

マルコフ連鎖が不変分布に収束するための十分条件が詳細釣り合い条件（detailed balance condition）と呼ばれる条件である．これは目標分布を $\pi(x)$ とすれば，状態空間に属するすべての x について，

$$\pi(x^{(t)})p(x^{(t+1)}|x^{(t)})=\pi(x^{(t+1)})p(x^{(t)}|x^{(t+1)})$$

が満たされることをいう．ここで $p(x^{(t+1)}|x(t))$ はマルコフ連鎖における推移核（transition kernel）であり，t の状態から $t+1$ の状態に推移する確率を表している．この条件が満たされるとき，十分に大きな t について推移核から生成される x^t は不変分布からの乱数と見なすことができる[3]．

2. MCMC のアルゴリズム

マルコフ連鎖モンテカルロ法の代表的な方法にはメトロポリス・ヘイスティングス法[4]とギブスサンプラー[5]がある．後者は前者の特殊なケースである．

2.1 メトロポリス・ヘイスティングス法（Metropolis-Hastings Algorithm）

多くの場合，詳細釣り合い条件は満たされていない．メトロポリス・ヘイスティングス法はこの条件が満たされるように目標分布と提案分布の違いを修正する操作を含めることで，目標分布からのサンプリングを可能とするアルゴリズムである．例えば以下のように収束条件が崩れたとき，

$$\pi(x^{(t)})p(x^{(t+1)}|x^{(t)})>\pi(x^{(t+1)})p(x^{(t)}|x^{(t+1)})$$

調節確率 $\alpha(x^{(t+1)}|x^{(t)})<1$ を導入して，次のように修正を行う．

$$\pi(x^{(t)})p(x^{(t+1)}|x^{(t)})\alpha(x^{(t+1)}|x^{(t)})$$
$$=\pi(x^{(t+1)})p(x^{(t)}|x^{(t+1)})$$

すなわち，

$$\alpha(x^{(t+1)}|x^{(t)})=\frac{\pi(x^{(t+1)})p(x^{(t)}|x^{(t+1)})}{\pi(x^{(t)})p(x^{(t+1)}|x^{(t)})}$$

である．$\alpha(x^{(t+1)}|x^{(t)})$ は現在の状態が $x^{(t)}$ であるとき，次に提案分布から取り出した標本 $x^{(t+1)}$ を採用するかどうかを表す確率を表していることから採択確率と呼

ウィラワン・ドニ・ダハナ
大阪大学 経済学研究科
〒560-0043 豊中市待兼山町 1-7

ぶ．提案分布から乱数を取り出す手続きは以下のとおりである．

- 初期値 $x^{(0)}$ を設定する．
- $t=0, 1, \cdots, T$ について $p(x^{(t+1)}|x^{(t)})$ から乱数 x を取り出し，一様分布 $U(0,1)$ から乱数 u を取り出す．
- $u \leq \alpha(x|x^{(t)})$ であれば $x^{(t+1)}=x$ とし，そうでなければ $x^{(t+1)}=x^{(t)}$ とする．
- $\{x^{(1)}, x^{(2)}, \cdots, x^{(T)}\}$ の値を帰す．

2.2 ギブスサンプラー（Gibbs sampler）

ギブスサンプラーは目標分布が多次元のとき，目標分布とその確率密度分布が条件付分布のかたちに分割できるとする．例えば目標分布が2次元の場合，$x=(x_1, x_2)$ であり目標分布を $\pi_1(x_1^{(t+1)}|x_2^{(t)})$ と $\pi_2(x_2^{(t+1)}|x_1^{(t)})$ に分割できる．これを全条件付分布（full conditional distribution）という．また，この条件付目標分布は条件付提案分布と一致していることが条件である．すなわち，

$$p_1(x_1^{(t+1)}|x_1^{(t)}, x_2^{(t)})=\pi_1(x_1^{(t+1)}|x_2^{(t)})$$
$$p_2(x_2^{(t+1)}|x_1^{(t)}, x_2^{(t)})=\pi_2(x_2^{(t+1)}|x_1^{(t)})$$

である．この際，採択確率 $\alpha_1(x_1^{(t+1)}|x_1^{(t)}, x_2^{(t)})$, $\alpha_2(x_2^{(t+1)}|x_1^{(t)}, x_2^{(t)})$ の値は1となり，候補は常に採択される．2次元分布について乱数の生成の手続きを以下に示す．

- 初期値 $x_1^{(0)}, x_2^{(0)}$ を設定する．
- $t=0$ のとき $p_1(x_1|x_1^{(t)}, x_2^{(t)})$ から乱数 x_1 を取り出し，$x_1^{(t+1)}=x_1$ とする．
- $t=0$ のとき $p_2(x_2|x_1^{(t+1)}, x_2^{(t)})$ から乱数 x_2 を取り出し，$x_2^{(t+1)}=x_2$ とする．
- $t=1, 2, \cdots, T$ まで繰り返す．
- $\{x^{(1)}, x^{(2)}, \cdots, x^{(T)}\}$ の値を帰す．

いずれのアルゴリズムも目標分布の母数を推定する際に初期値の影響を除去するために最初の数回の繰り返しの連鎖が取り除かれる．この期間をバーンイン期間（burn-in period）という．

3．収束の判定

MCMCにより生成された連鎖が目標分布に収束するかどうかを判定する必要がある．連鎖が収束していることを明確に判定できる方法はいまだに知られていないが，これまで提案されたいくつかの判定方法を以下に示す．

- 時系列プロットの目視による方法．マルコフ連鎖を構成する要素の値を時系列順に図示し，連鎖の変動に系統的な偏りがないか，不変分布の最頻値近辺にどのように変動するかを目視することで連鎖の収束を判定する方法である．
- Gewekeの方法．生成されたマルコフ連鎖を2つに分割し，両者に含まれる構成要素の値に差があるかどうかによって収束を判定する方法である．もし両者の間に統計的に有意な差がなければ連鎖が収束すると判定する．
- Gelman & Rubinの方法．同一の提案分布から生成された異なる複数の連鎖の間の分散と各連鎖の連鎖内の分散を比較することによって収束を判定する方法である．
- Heidelberger & Weltchの方法．連鎖の最初の何個の要素をバーンイン期間として破棄すれば収束が確認できるかを推定し，残りの要素を用いてその信頼区間の幅と母数の推定値を比較することによって収束の判定を行う方法である．
- Raftery & Lewisの方法．サンプリングされた連鎖が目標分布を近似したものになっているかどうかを標本百分位数をもとに，一定の近似精度を得るために必要なマルコフ連鎖の長さとバーンイン期間の長さを逆算する方法である．

参考文献

[1] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, Bayesian Data Analysis, Chapman & Hall/CRC, 2004.

[2] 豊田秀樹編著,『マルコフ連鎖モンテカルロ法』, 朝倉書店, 2008.

[3] L. Tierney, Markov Chains for Exploring Posterior Distributions, Annals of Statistics, 1994, Vol. 22, pp. 1701-1762.

[4] S. Chib and E. Greenberg, Understanding the Metropolis-Hastings Algorithm, The American Statistican, 1995, Vol. 49, No. 4, pp. 327-335.

[5] M. A. Tanner and W. H. Wong, The Calculation of Posterior Distributions by Data Augmentation, Journal of the American Statistical Association, 1987, Vol. 82, pp. 528-549.