# 水産資源管理談話会報

## 第 34 号

日本鯨類研究所　資源管理研究センター

２００４年　９月

翻訳・公表希望者は以下の手続きとり、著者の許可を得た上で翻訳・公表する。

1. 翻訳・公表希望者は文章（FAX、手紙）で著者、表題および会報の号を明記し、資源管理談話会事務局を通じて要請し、著者の許可を得て翻訳・公表する。

2. 翻訳公表物を資源管理談話会事務局に送付する。

# 目次

（投稿）

# MCMC 入門

平松一彦（遠洋水研）

## 1．はじめに

水産資源評価において MCMC という言葉が聞かれるようになった。MCMC はマルコフ連鎖モンテカルロ法（Markov chain Monte Carlo algorithms）の略であるが、これだけ聞いても何の意味かさっぱりわからない。統計学の専門書や論文をみるとそこには数学的な定義が示されており、さらにわけがわからなくなる。

本稿は MCMC の基本的なイメージを示すことを目的とする。数学的な厳密性は問わないので、厳密には正しくない説明もあるかもしれない。いろいろと奥が深そうであるが、とりあえず MCMC とは任意の確率分布から乱数発生させるひとつの方法と思っていればよい。一様分布や正規分布からの乱数であれば、多くのソフトで組み込み関数として用意されている。しかし、それ以外の乱数を効率よく発生させるのは結構難しい。MCMC はそのひとつの方法である。

それにしてもたかが乱数の発生方法ごときがなぜそんなに注目されているのか？　水産資源学といったい何の関係があるのか？

## 2．乱数の利用法

乱数はこれまでにも水産資源学に利用されている。例えば、再生産関係を用いて資源の将来予測をするような場合、再生産関係式からのばらつきとして対数正規分布などが使われる。こういった計算には正規分布や一様乱数ぐらいが利用できれば十分である。しかし、最近組み込み関数で対応できないような分布からの乱数発生が必要となる状況が生じてきている。

その大きな要因はベイズ統計の利用の広がりである。ベイズ統計では事前分布と尤度関数から事後分布を計算し、事後分布から様々な推測を行う。事後分布の計算には複雑な積分が必要となりまた必ずしも正規分布などの扱いやすい分布になるとは限らず、このためベイズ統計の適用は計算可能なごく一部の問題に限られてきた。MCMC などの乱数発生アルゴリズムを用いることにより、複雑な事後分布に対応する乱数を発生させることが可能となった。

例えばベイズ統計を用いてプロダクションモデルにより、$B_0$、r、K、q を推定する場合を考える。これらのパラメータを $\theta$ として事前分布を $p(\theta)$、データ X の下での尤度関数を $L(X|\theta)$ とすると事後分布 $p(\theta|X)$ は

$$p(\theta|X)=\frac{L(X|\theta)p(\theta)}{\int L(X|\theta)p(\theta)d\theta} \tag{1}$$

である。事後分布 $p(\theta|X)$を得るためには分母の積分が必要であるが、これが解析的に出来るのはごく限られた場合のみである。また、例えば $B_0$ の分布に興味がある場合には、(1)式をその他のパラメータ r、K、q について積分する必要がある。

分布が $L(X|\theta)\ p(\theta)$に比例する乱数が発生できれば、実質的には $p(\theta|X)$が得られたのと同様に様々な推測が可能である。例えば、r の平均値が知りたいのであれば

$$E[r]=\int r\cdot p(\theta|X)d\theta \tag{2}$$

の計算が必要となるが、これは発生した乱数の r の値を足し合わせればよい。また $B_0$の分布を見るのであれば、単純に $B_0$の値にのみ注目すればよい。

$(B_0, r, K, q)=(B_{01}, r_1, K_1, q_1),\ldots,(B_{0n}, r_n, K_n, q_n)$

という乱数列が得られているとすると、(2)式に相当する r の平均値は

$$E[r]=\frac{1}{n}\sum_{i=1}^{n} r_i \tag{3}$$

で得られる。また $B_0$の分布は$(B_{01},\ldots, B_{0n})$に注目すればよい。

以下に具体的な乱数発生の方法を示す。最初はマルコフ連鎖ではない、ただのモンテカルロ法による乱数の発生方法から紹介する。

## ３．モンテカルロ法による乱数発生法

### ３．１．棄却サンプリング

乱数発生が容易な確率分布 g(x)からの乱数を利用して、目的とする確率分布$\pi$(x)からの乱数発生を行う方法である。

$\pi(x)\leqq cg(x)$とする（c は定数）。この時 g(x)からの乱数 $x^*$と 0～1 区間の一様乱数 u を用いて

$$u\leqq \pi(x^*)/cg(x^*) \tag{4}$$

なら $x^*$を$\pi$からのサンプルとして受容する。

g(x)からの乱数のうち、余分なものは除去して$\pi$(x)からの乱数とするという方法であり、$\pi$(x)と cg(x)がずれていると、捨てられる乱数が多くなり非常に非効率となる。

例えば一様乱数 r からある関数 f(x)に従う乱数を作りたいとき、まず直感的に思いつくのは（区間が同じ）一様乱数 r にその関数 f を掛けることである。しかし、乱数に関数を直接掛けても、分布 r に従って rf(r)の値が発生するだけで、分布 f に従う乱数が作られるわけではない。r から f の分布を作るためには r の発生する頻度分布に f を掛ける必要がある。この直感的な方法に対応するのが、上記の棄却サンプリングである。r の分布と f の分布を比較して r に従う乱数から余分な部分を除いて f に従う分布を作り出す。分布 f の裾が長いと、発生させた一様分布のほとんどを捨ててしまうことになり、非効率となる。

### ３．２．SIR法(Sampling-importance resampling algorithm)

これも棄却サンプリング同様、乱数発生が容易な確率分布 g(x)からの乱数を利用して、目的とする確率分布 π(x)からの乱数発生を行う方法である。

1)求めたい π(x)の近似関数 g(x)を考える。

2)g(x)から M 個の標本 $x_1,\ldots,x_M$ を発生させる。

3)重み $w_j$ を $w_j \propto \pi(x_j)/g(x_j)$ とし、$x_1,\ldots,x_M$ の中から m(<M)個の X を重み $w_j$ で復元抽出により発生させる。

棄却サンプリングのように g(x)からの乱数を捨てたりはしないが、やはりこれも g の近似が悪いと非効率となりまた精度も悪くなる。

このように比較的単純な乱数発生方法では、目的とする分布関数に近い乱数発生が容易な関数が得られていないと非効率となる。次にマルコフ連鎖を用いた乱数発生方法の簡単な紹介をする。

## ４．マルコフ連鎖と MCMC

マルコフ連鎖とはランダムウォークみたいなものをイメージすればよい。例えば、ε を標準正規分布として

$$x_{i+1} = 0.5x_i + \varepsilon_i \qquad (5)$$

はマルコフ連鎖のひとつである。過去の履歴（例えば $x_{i-1}$ など）によらず次のステップの値が決まるような過程のことをマルコフ連鎖という。

ところで、(5)式は少し特殊なマルコフ連鎖になっていて、適当な初期値 $x_0$ から上式を用いて $x_i$ を発生させていくと、その頻度分布は平均 0 分散 4/3 の正規分布となる。すなわち(5)式は平均 0 分散 4/3 の正規乱数発生アルゴリズムになっているわけである。MCMC はこのような性質を利用して乱数を発生させる方法である。

あるいは以下のような考え方も可能である。最尤法の使用と共に最適化プログラムの使用も一般的になってきたが、最適化プログラムでは尤度関数の最大となるところを、初期値から始めてあっちへ行ったりこっちへ行ったりしながら探していく。MCMC もそれとよく似ているが、最適値のところで止まってしまうのではなく、関数の値の大きいところでは長くとどまり、小さなところではあまりとどまらないようにしながら永遠にさまよい続ける。そしてその頻度分布が関数に対応しているわけである。

なおマルコフ連鎖であればある分布の乱数発生に使えるかというとそうではなく、例えば(5)式を通常のランダムウォークである

$$x_{i+1} = x_i + \varepsilon_i \qquad (6)$$

とすると、これから得られる x の値は定常的な分布とはならない。また(5)式において初期値が分布の中心である 0 から遠く離れていると、しばらくの間は分布の裾野の値しか出てこない。このような初期状態による特異性を避けるために、得られるサンプル列の最初の部分は使わないのが普通である。

MCMC を実際に使うには、目的とする関数の乱数を効率よく発生させる(5)式に相当する式を見出す必要がある。これには幾つもの方法が考えられており、次にその代表的なものを紹介する。

## 5．種々の MCMC による乱数発生法

### 5．1．メトロポリス法 (Metropolis algorithm)

作成したい分布関数を f(x)とする。メトロポリス法は、以下のような手順を繰り返して f(x)に従う乱数 x を発生させる。

1)x に一様乱数 $\varepsilon$ (−a, a)を加え $x^*=x+\varepsilon$ とする。

2) $\gamma=f(x^*)/f(x)$を計算し、(0,1)の一様乱数 r を発生させ、

3) $r<\gamma$ なら x を $x^*$で置き換え、それ以外は x のままとする。

まず 1)で移動先の x の値 $x^*$を選び、そこの $f(x^*)$と現在の f(x)を比較し、$f(x^*)>f(x)$なら x を $x^*$で置き換え、$f(x^*)\leqq f(x)$の場合は確率 $\gamma$ で $x^*$に置き換える。それ以外の場合はそのままとする。この操作を繰り返す。

さきほどの最適化プログラムとのアナロジーでいうと、最適化プログラムでは常に目的関数の値の大きいほうへ移動していくが、メトロポリス法の場合は移動先 $f(x^*)$と現地点 f(x)の値を比較して移動先が大きければ無条件でそちらに移動し、大きくない場合でも確率 $f(x^*)/f(x)$でそちらに移る。

ここでは一変数の場合を示したが、多変数の場合はまず変数を選択し（順番あるいはランダム)、上記の操作を繰り返すことにより適用可能である。

### 5．2．ギブスサンプラー (Gibbs sampler)

２変量 f(x,y)の分布を考える。ギブスサンプラーは以下のような手順で $x_i$、$y_i$ から $x_{i+1}$、$y_{i+1}$ を発生させる。

1)$x_{i+1}$ を $f(x,y_i)$から発生させる。ここで $f(x,y_i)$は y に値 $y_i$ を代入した x のみの関数（条件付確率に相当）である。

2)$y_{i+1}$ を $f(x_{i+1},y)$から発生させる。

これを繰り返し、サンプルを発生させる。ここで $f(x,y_i)$や $f(x_{i+1},y)$は標本を簡単に発生できる必要がある。

### 5．3．Metropolis-Hasting 法

メトロポリス法を拡張したものとなっており以下の手順に従う。

1)条件付分布（サンプラーという）$q(y|x)$を設定する。

2)$q(x|x_i)$より $x^*$を発生させる。

3) $\alpha = \dfrac{f(x^*)q(x_i|x^*)}{f(x_i)q(x^*|x_i)}$ を計算し、(0,1)の一様乱数 r を発生させ

4)r＜ $\alpha$ なら x を x*で置き換え、それ以外は x のままとする。

q(y | x)は x の下で y が発生する確率を示す。q(y | x)の形が重要であるが

$$q(y|x)=q(x|y) \tag{7}$$

となるのがメトロポリス法で、特に

$$q=x+(\text{平均 0 の一様乱数}) \tag{8}$$

とすれば5．1で例示したメトロポリス法のアルゴリズムになる。他に正規分布や、f の条件付分布が使われる。後者の場合は5．2のギブスサンプラーであり、$\alpha$=1 となるため常に置き換わりが生じ効率がよい。

## 6．おわりに

様々な分野でベイズ統計の使用は急速に広がっているが、これには MCMC 等の計算法の開発による部分が大きい。水産分野でのベイズ統計の使用も増えており、たかが乱数発生の方法ではあるが MCMC に関する最低限の知識は必要になりつつある。しかし残念ながら日本語の文献、特に統計学の非専門家向けの解説はまだほとんどない。

CCSBT（みなみまぐろ保存委員会）における管理方式（MP）の開発において、MCMC がほとんど常識として用いられていたことが、本稿をとりまとめるきっかけとなった。CCSBT の MP 開発ではベイズ統計を用いて自然死亡係数 M 等のパラメータ推定を行い、その事後分布を用いて MP 候補をテストするためのシミュレーションが行われた。例えば、事後分布から MCMC により発生した乱数を用いて M の値を変えたシミュレーションなどが行われた。今後、こういった作業が様々な状況で必要となってくるのではないだろうか。

本稿を取りまとめるにあたり、伊庭(1996)、丹後(2000)、大森(2001)を参考にした。より正確な定義や MCMC の方法論、問題点についてはこれらを参考にしていただきたい。

## 参考文献

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, 1:49-84.

大森裕浩(2001). マルコフ連鎖モンテカルロ法の最近の展開. 日本統計学会誌, 31 :305-344 (http://www.e.u-tokyo.ac.jp/~omori/2003/MCMC/mcmc.pdf)

丹後俊郎(2000). 統計モデル入門. 朝倉書店, 東京, 246pp.