

# マルコフ連鎖モンテカルロ法の最近の展開\*

大森裕浩<sup>†</sup>

2001年7月

\*本稿に対して浅井学助教授（立命館大学）、國友直人教授（東京大学）、繁榊算男教授（東京大学）、和合肇教授（名古屋大学）、渡部敏明教授（東京都立大学）及び本誌レフリーより貴重なコメントを頂いた。ここに記して感謝の意を表したい。

<sup>†</sup>東京都立大学経済学部. 〒192-0397 東京都八王子市南大沢 1-1. e-mail:omori@bcomp.metro-u.ac.jp

## 1 はじめに

1990年以降、マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo, MCMC) は統計学の分野で研究が進み現在では多くの実証分析で応用されている。この手法はベイズ統計学で事後分布や事後確率の計算を行うためのシミュレーションの方法である。ベイズ推論においてはもとの密度関数自体が複雑でなくても共役でない事前分布を用いると事後分布の形が複雑になったり、また未知母数の個数が非常に多いときにはひとつでも分布型が複雑なものがあれば周辺事後分布の積分の計算が困難になることが多い。このような場合、これまでは高速な大型の計算機を利用することができなければ数値積分により計算することができなかった。この問題を解決するために1990年代になって事後分布をシミュレーションによって求めるという方法が積極的にとりあげられ、その理論的な裏付けも急速に進められてきた。

その方法にはギブスサンプラー、M-H アルゴリズム、データ拡大法、代替サンプリングなどがあるが、現在はマルコフ連鎖モンテカルロ法と総称している。これらの方法は、Geman and Geman (1984) が画像復元のために使用したギブスサンプラーを Gelfand and Smith (1990) が離散分布から連続分布に拡張して事後分布を求める方法の主流となってきた。本稿ではまず第2章でMCMCの基本となるモンテカルロ法を説明し、第3章でマルコフ連鎖モンテカルロ法の概略について説明する。第4章、第5章でもっともよく使われている手法のギブスサンプラーとM-H アルゴリズムを紹介する。さらに第6章では収束に関する諸問題を取りあげ、収束の判定方法、判断基準、収束を加速する方法について紹介し、第7章では収束に関する理論的結果を説明する。最後に第8章で得られた標本に基づいてどのような推論を行うか、モデル選択のためのベイズ比の計算をどうすればよいかを説明する。

## 2 モンテカルロ法の基礎

### 2.1 よく使われる乱数の発生方法

モンテカルロ法ではさまざまな乱数を発生させることが必要になるが一様乱数や正規乱数など標準的な確率分布に従う乱数の発生は多くのソフトウェアで組み込み関数として用意されている<sup>1</sup>。そこで、ここでは、そのような関数の用意されていない確率分布に従う乱数を発生させる方法について紹介する。特に頻繁に用いられる方法は以下に述べる棄却サンプリングである。

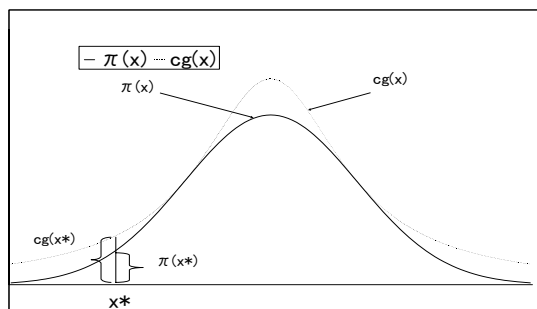
#### 2.1.1 棄却サンプリング

密度関数が  $\pi(x)$  であるような確率分布から乱数を発生させたいが、簡単にはできないということはしばしば起こる。このとき、正規分布など乱数発生が簡単な確率分布 (密度関数を  $g(x)$  とする) からの乱数を利用して  $\pi(x)$  からの乱数発生を行うのが棄却サンプリング (rejection sampling) である。Acceptance-Rejection sampling (Accept-Reject sampling, A-R 法, 受容-棄却サンプリング) ともいわれる。もし「 $g(x)$  が  $\pi(x) > 0$  であるようなすべての  $x$  について  $\pi(x) \leq cg(x)$  (ただし  $c$  は定数)」であるならば、

- (1)  $g(x)$  から乱数  $x^*$  を、 $(0,1)$  区間上の一様乱数  $u$  をそれぞれ発生させる。

<sup>1</sup>乱数発生のための基本的な参考文献としては Ripley (1987) がよい。

図 1:  $\pi(x)$  (実線) と  $cg(x)$  (点線).



(2) もし、 $u \leq \pi(x^*)/cg(x^*)$  ならば  $x^*$  を  $\pi(x)$  からの標本(サンプル)として受容し、そうでなければ棄却して(1)へ戻る。

とする。(1)(2)の手順で得られる確率変数  $x^*$  は  $\pi(x)$  に従う (Ripley, 1987)。このことはベイズの定理から

$$p(x|u \leq \pi(x)/cg(x)) = \frac{\text{pr}(u \leq \pi(x)/cg(x)|x)g(x)}{\int \text{pr}(u \leq \pi(x)/cg(x)|x)g(x)dx} = \frac{\{\pi(x)/cg(x)\}g(x)}{\int \{\pi(x)/cg(x)\}g(x)dx} = \frac{\pi(x)}{\int \pi(x)dx} = \pi(x)$$

であることから理解できるであろう。

また目標となる密度関数  $\pi(x)$  の基準化定数がわからなくてもよい。なぜなら  $\pi(x) = h(x)/\int h(t)dt$ ,  $c' = c/\int h(t)dt$  とすると「 $\pi(x) \leq cg(x)$ 」は「 $h(x) \leq c'g(x)$ 」、「 $\pi(x^*)/cg(x^*)$ 」は「 $h(x^*)/c'g(x^*)$ 」となるからである。

ただし、この手法には次のような問題点がある。第1は  $\pi(x)$  を上から覆うような  $cg(x)$  は必ずしも容易に見つけれないという問題である。もしそのような  $g(x)$  がみつけれない場合には、後の節で説明するようにさらに M-H 法を組み合わせることによってこの問題を解決することになる (5.3.3 節)。第2に、仮にそのような  $g(x)$  が見つかったとしても  $g(x)$  の  $\pi(x)$  に対する近似がよくなければ(2)における受容率が低くなり、効率の悪い乱数発生法(サンプリング法)になる。特に分布のすそにおいて条件  $\pi(x) \leq cg(x)$  が満たされるかどうかは判定が難しいからといって  $c$  の値を大きくしてしまうと受容率が低くなり効率が悪くなる。

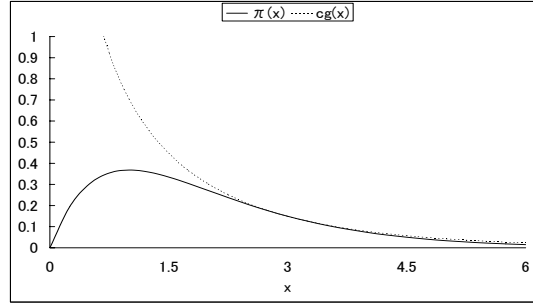
例.  $X \sim \text{Gamma}(2, 1)$  のとき自由度1の  $\chi^2$  乱数を用いて棄却法を行うことができる。 $X$  の密度関数を  $\pi(x)$  とし、自由度1の  $\chi^2$  分布の密度関数を  $g(x)$  とすると

$$\pi(x) = x \exp -x, \quad g(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} \exp -\frac{x}{2}, \quad x > 0$$

である。ここで

$$\frac{\pi(x)}{g(x)\sqrt{2\pi}} = x^{\frac{3}{2}} \exp -\frac{x}{2} \leq 3^{\frac{3}{2}} \exp -\frac{3}{2},$$

図 2: ガンマ分布 (Gamma (2,1)) の近似



であるから、 $c = \sqrt{2\pi}3^{\frac{3}{2}} \exp -\frac{3}{2}$  ととればよい。この例の場合、図 2 のように 0 付近での近似が悪いため  $n = 10000$  回の発生で受容率は低く約 34% である。

例. 確率的ボラティリティモデルでは  $X_t = Z_t \exp(\alpha_t/2)$ ,  $Z_t \sim N(0, 1)$  のようなモデルを考える。このとき  $X_t$  の密度関数は

$$f(x_t|\alpha_t) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{\alpha_t}{2} - \frac{x_t^2}{2} \exp(-\alpha_t) \right\}$$

である。 $\alpha_t$  の事前分布を  $\alpha_t \sim N(\mu_t, \sigma_t^2)$  として事後分布を求めると

$$\log \pi(\alpha_t|x_t, \mu_t, \sigma_t^2) = \text{定数} - \frac{\alpha_t}{2} - \frac{x_t^2}{2} \exp(-\alpha_t) - \frac{(\alpha_t - \mu_t)^2}{2\sigma_t^2}$$

である。ここで

$$\exp(-\alpha_t) \geq (1 + \mu_t) \exp(-\mu_t) - \alpha_t \exp(-\mu_t)$$

に注意すると

$$\begin{aligned} \log \pi(\alpha_t|x_t, \mu_t, \sigma_t^2) &\leq \text{定数} - \frac{\alpha_t}{2} + \frac{x_t^2}{2} \alpha_t \exp(-\mu_t) - \frac{(\alpha_t - \mu_t)^2}{2\sigma_t^2} \\ &= \text{定数} - \frac{(\alpha_t - \mu^*)^2}{2\sigma_t^2}, \quad \mu^* = \mu_t + \sigma_t^2 \{x_t^2 \exp(-\mu_t) - 1\}/2, \end{aligned}$$

であり、 $g(x)$  は平均  $\mu^*$ 、分散  $\sigma_t^2$  の正規分布の密度関数となる。

### 2.1.2 適応的棄却サンプリング

指数型分布族をはじめとして特に目標 (target) となる密度関数の対数が凹関数であるときには Gilks and Wild (1992) による適応的棄却サンプリング (adaptive rejection sampling) という方法がある。この手法でも基準化定数は必要ではないので  $\pi(x)$  を目標となる密度関数または、密度関数に比例する関数とする。そして  $\partial^2 \log \pi(x)/\partial x^2 < 0$  であると仮定する。

まず  $\pi(x)$  の台 (support) 上の  $n+2$  個の点  $x_0 < x_1 < \dots < x_{n+1}$  を選び、点集合  $\Omega_n$  を  $\{x_0, x_1, \dots, x_{n+1}\}$  とおく。次に  $l(x) = \log \pi(x)$  とし、 $i = 0, 1, \dots, n$  について点  $(x_i, l(x_i))$  と  $(x_{i+1}, l(x_{i+1}))$  を結ぶ直線  $m_i(x)$  を引く。するとこの直線  $m_i(x)$  は区間  $[x_i, x_{i+1}]$  では  $l(x)$  より下にあるが、その区間の外では  $l(x)$  より大きくなっている。そこで

$$l_{1,n}(x) = \begin{cases} m_i(x), & x \in [x_i, x_{i+1}], i = 0, 1, \dots, n \text{ のとき} \\ -\infty, & x \notin [x_0, x_{n+1}], \end{cases}$$

$$l_{2,n}(x) = \begin{cases} m_0(x), & x < x_0 \text{ のとき} \\ m_1(x), & x \in [x_0, x_1] \text{ のとき} \\ \min(m_{i-1}(x), m_{i+1}(x)), & x \in [x_i, x_{i+1}], i = 1, \dots, n-1 \text{ のとき} \\ m_{n-1}(x), & x \in [x_n, x_{n+1}] \text{ のとき} \\ m_n(x), & x > x_{n+1} \text{ のとき} \end{cases}$$

とおくと  $l_{1,n}(x) \leq l(x) \leq l_{2,n}(x)$  が成り立つ。従って  $g_{1,n}(x) = \exp l_{1,n}(x)$ ,  $g_{2,n}(x) = \exp l_{2,n}(x)$  とすれば  $g_{1,n}(x) \leq f(x) \leq g_{2,n}(x)$  が成り立つ。ここで  $g_n(x) = g_{2,n}(x) / \int g_{2,n}(x) dx$  とおくと

- (1)  $g_n(x)$  から乱数  $x^*$  を、 $(0,1)$  区間上の一様乱数  $u$  をそれぞれ発生させる。
- (2) もし、 $u \leq g_{1,n}(x^*)/g_{2,n}(x^*)$  ならば  $x^*$  を  $\pi(x)$  からの標本として受容し (1) へ戻る。そうでなければ
  - (a) もし、 $u \leq \pi(x^*)/g_{2,n}(x^*)$  ならば  $x^*$  を  $\pi(x)$  からの標本として受容し、点  $x^*$  を点集合  $\Omega_n$  に加えて  $\Omega_{n+1} = \Omega_n \cup x^*$  として (1) へ戻る。
  - (b) そうでなければ  $x^*$  を棄却して (1) に戻る。

$g_n(x)$  から乱数を発生させるためには、まずその基準化定数  $\int g_{2,n}(x) dx$  を求めて直線で区切られた区間 (図3では8個。  $x_1 \sim x_2$  と  $x_2 \sim x_3$  は区間がさらに2つに分けられることに注意) の確率を使ってまず区間を選び、それから選ばれた区間の中でどの点を選ぶかを一様乱数を使って分布関数から決定する。

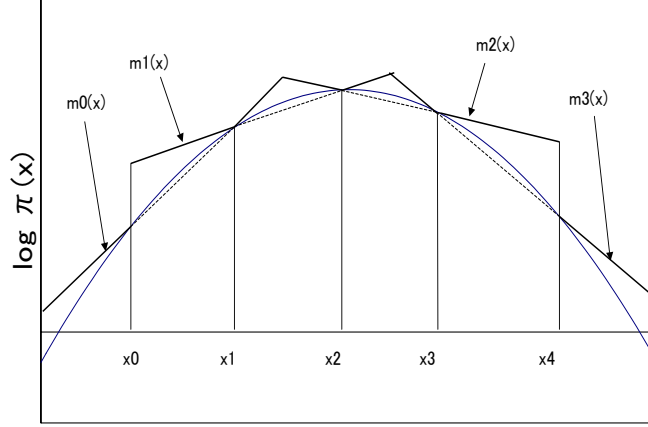
この方法は、 $\log \pi(x)$  が凹関数である場合にはいつでも使うことができ、 $n$  が大きくなるにつれて乱数のどんどん受容率も高くなっていくという長所がある。しかし、個別の密度関数では例えばモードのまわりでのテーラー展開による正規近似など別の方法が効率がよい場合もある。またプログラミングがやや複雑になることもあるので、その適用はケースバイケースで考える必要がある。

### 2.1.3 SIR 法 (Sampling/Importance Resampling)

Rubin (1987a) は一般に公開されているデータを用いて統計的推測を簡単に行うために、欠損値を埋める方法として SIR (Sampling/Importance Resampling) 法を提案した。この方法は、目標とする確率密度関数  $\pi(x)$  と近似密度関数  $g(x)$  との近さの度合いを有限個の点で求め、これを重みにして再び標本を抽出するため重みつき再サンプリング法 (weighted resampling method) とも呼ばれる。具体的には

- (1)  $g(x)$  から  $M$  個の標本  $X_1, \dots, X_M$  を発生させる。

図 3:  $l_1(x)$  (実線) と  $l_2(x)$  (点線).  $n = 3$ .



(2) 重み  $w_j$  を  $w_j \propto \pi(x_j)/g(x_j)$ ,  $j = 1, 2, \dots, M$ , とする。

(3)  $(x_1, \dots, x_M)$  の中から  $m (< M)$  個の  $X$  を復元抽出により確率  $Pr(X = x_j) = w_j / \sum_{i=1}^M w_i$  で発生させる。

という方法である。このことは  $M \rightarrow \infty$  のとき

$$\begin{aligned} \text{pr}(X \leq x) &= \frac{\sum_{\{i: x_i \leq x\}} w_i}{\sum_{i=1}^M w_i} = \frac{\frac{1}{M} \sum_{i=1}^M \pi(x_i)/g(x_i) I(x_i \leq x)}{\frac{1}{M} \sum_{i=1}^M \pi(x_i)/g(x_i)} \\ &\rightarrow \frac{\int_{-\infty}^x \{\pi(t)/g(t)\} g(t) dt}{\int \{\pi(t)/g(t)\} g(t) dt} = \frac{\int_{-\infty}^x \pi(t) dt}{\int \pi(t) dt} = \int_{-\infty}^x \pi(t) dt \end{aligned}$$

であることから理解できるであろう。

Rubin (1987a) では未知の母数を  $\theta$ 、観測されるデータを  $Y$ 、観測されないデータ (欠損値) を  $Z$  とし、事後分布から欠損値を発生させるためのアルゴリズムとして以下のように考えた。まず  $\theta$  の事前分布を  $\pi(\theta)$ 、 $\theta$  が与えられたときの  $(Y, Z)$  の同時確率密度関数を  $\pi(Y, Z|\theta)$ 、 $Y$  が与えられたときの  $(\theta, Z)$  の同時事後密度関数を  $\pi(\theta, Z|Y)$  とする。このとき  $\pi(\theta, Z|Y)$  をよく近似していて、かつ標本を発生させやすい密度関数  $g(\theta, Z|Y) = g(\theta|Y)g(Z|\theta, Y)$  を選ぶ。そして

(1)  $g(\theta, Z|Y)$  から  $M$  個の標本  $(\theta_j, Z_j)$  ( $j = 1, 2, \dots, M$ ) を発生させる。

(2) 重み  $w_j$  を次のように計算する。

$$w_j = w(\theta_j, Z_j|Y) \propto \frac{\pi(Y, Z_j|\theta_j)\pi(\theta_j)}{g(\theta_j, Z_j|Y)}, \quad j = 1, 2, \dots, M.$$

- (3)  $(z_1, z_2, \dots, z_M)$  の中から  $m (< M)$  個の  $Z$  を復元抽出により確率  $Pr(Z = z_j) = w_j / \sum_{i=1}^M w_i$  で発生させる。

もし  $(\theta, Y)$  が与えられたときの  $Z$  の条件付事後密度が  $\pi(Z|\theta, Y) = g(Z|\theta, Y)$  ならば、 $w_j$  は  $z_j$  に依存しなくなるから

- (i)  $g(\theta|Y)$  から  $M$  個の標本  $(\theta_j, j = 1, 2, \dots, M)$  を発生させる。  
(ii) 重み  $w_j$  を次のように計算する。

$$w_j = w(\theta_j|Y) \propto \frac{\pi(Y|\theta_j)\pi(\theta_j)}{g(\theta_j|Y)}, \quad \text{for } j = 1, 2, \dots, M.$$

ただし  $\pi(Y|\theta)$  は  $\theta$  が与えられたときの  $Y$  の周辺確率密度関数である。

- (iii)  $(\theta_1, \theta_2, \dots, \theta_M)$  の中から  $m (< M)$  個の  $\theta$  を復元抽出により確率  $Pr(\theta = \theta_j) = w_j / \sum_{i=1}^M w_i$  で発生させ、得られた  $\theta_j^*$  を用いて  $Z$  を  $\pi(Z|\theta_j^*, Y)$  から発生させる ( $j = 1, 2, \dots, m$ )。

となる。もし  $w_j$  の分布が非常に歪んでいるならば SIR 法による近似は非常に悪くなると考えられるので  $g$  のとり方を再検討する必要がある。また  $M/m \rightarrow \infty$  のとき、 $m$  個の  $(Z_j^*, \theta_j^*)$  の確率密度関数は

$$\frac{g(\theta, Z|Y)w(\theta, Z|Y)}{\int \int g(\theta, Z|Y)w(\theta, Z|Y)d\theta dz} = \frac{\pi(Y, Z|\theta)\pi(\theta)}{\int \int \pi(Y, Z|\theta)\pi(\theta)d\theta dz} = \pi(Z, \theta|Y)$$

になるが、Rubin (1987a) は  $M/m$  の値の決め方について、欠損している情報の割合に依存するが一般に 20 がよいであろうとしている。Gelfand and Smith (1990) はさらに事後分布  $\pi(Z|Y)$  と  $\pi(\theta|Y)$  を

$$\pi(Z|Y) = \frac{\sum_{j=1}^M w_j \pi(Z|Y, \theta_j)}{\sum_{j=1}^M w_j}, \quad \pi(\theta|Y) = \frac{\sum_{j=1}^M w_j \pi(\theta|Y, z_j)}{\sum_{j=1}^M w_j}$$

として求めた。SIR 法の精度は  $g$  が事後分布をよく近似しているかどうかによって左右されてしまい、後述する代替サンプリングやギブスサンプリングなどに比べて分布の近似という点で劣っている。

## 2.2 モンテカルロ積分と重点的サンプリング

積分  $\int k(x)dx$  の計算方法にはさまざまあるが、 $x$  がベクトルで高次の積分の場合には数値積分による近似方法はあまり正確ではなく、また正確な計算のためには大型計算機が必要であることが多い。これに対してシミュレーションを応用したモンテカルロ積分では安価な PC を用いて大型計算機と同等な精度の計算結果を得ることができる。

まず確率変数  $X$  の密度関数を  $f(x)$  とし、その台が  $k(x)$  の台を含むとしよう。このとき積分を

$$H = \int k(x)dx = \int \frac{k(x)}{f(x)}f(x)dx = \int h(x)f(x)dx = E[h(X)], \quad h(x) = \frac{k(x)}{f(x)}$$

と書き換える。すると  $X_1, \dots, X_n \sim \text{i.i.d. } f(x)$  のとき

$$\hat{H} = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

は  $H$  のモーメント推定量になり  $n \rightarrow \infty$  のとき大数法則により  $H$  に収束する。従ってもし  $X$  のある関数  $h(X)$  について、期待値  $E(h(X)) = \int h(x)f(x)dx$  を計算したい場合には上述のよう計算すればよい。

しかし  $f(x)$  から標本を発生させることが難しい場合にはどうすればよいだろうか。その場合、モンテカルロ積分と同様な操作で重みつきサンプリングを以下のように行う。まず乱数を発生させやすい近似分布  $g(x)$  を用いて

$$E_f(h(X)) = \int h(x)f(x)dx = \int \left( \frac{f(x)}{g(x)} h(x) \right) g(x)dx = E_g \left[ \frac{f(x)}{g(x)} h(x) \right]$$

とし、 $X_1, \dots, X_n \sim \text{i.i.d. } g(x)$  のとき

$$\hat{E}_f(h(X)) = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g(X_i)} h(X_i)$$

とすればよい。このような重みつきサンプリングの方法を重点的サンプリング (importance sampling) という。

ところで確率変数  $X, Y$  の同時確率密度関数を  $\pi(x, y)$  とし  $Y$  の周辺確率密度関数  $p(y)$ ,

$$p(y) = \int \pi(x, y)dx = \int \pi(y|x)\pi(x)dx$$

を求め、 $p(y)$  から標本を発生させたい場合はどうであろうか。もし  $\pi(x)$  から確率変数を簡単に発生することができるならば

- (1)  $X_1, \dots, X_n \sim \text{i.i.d. } \pi(x)$  を発生させる。
- (2)  $\hat{p}(y) = (1/n) \sum_{i=1}^n \pi(y|x_i)$  を計算する。

さらに  $p(y)$  から標本を発生させるには (2) で  $\hat{p}(y)$  の代わりに  $\pi(y|x_i)$  から発生させればよい。もし  $\pi(x)$  から標本を発生させることが難しい場合には Geweke (1989) は  $\pi(x)$  に近く、標本を発生させやすい分布  $g(x)$  ( $g$  の台は  $\pi$  の台を含む) を用いて、

- (i)  $X_1, \dots, X_n \sim \text{i.i.d. } g(x)$  を発生させる。
- (ii)  $\hat{p}(y) = \sum_{i=1}^n w_i \pi(y|x_i) / \sum_{i=1}^n w_i$ 、 $w_i = \pi(x_i)/g(x_i)$  を計算する。

とすると確率 1 で  $\hat{p}(y)$  が  $p(y)$  に収束することを証明した。 $p(y)$  からの乱数を発生させるには

- (ii)  $\pi(y|x_i)$  から  $Y_i$  を発生させる ( $i = 1, \dots, M$ )。
- (iii)  $(y_1, \dots, y_M)$  に確率  $(w_1, \dots, w_M)$  を与えて乱数を発生させる。

とすればよい。この方法は重みつきブーストラップとも呼ばれ、 $M$  が大きくなるに従い近似がよくなることが証明されている (Smith and Gelfand, 1992)。前述した Rubin (1987a) の SIR 法はこの方法がもとになっている。



### 3 マルコフ連鎖モンテカルロ (Markov Chain Monte Carlo) 法

#### 3.1 マルコフ連鎖とは

まず、マルコフ連鎖について Tierney (1994) に従って定義と記号を導入しよう。

**定義 3.1.1** 不変分布 (*invariant distribution*)  $\pi$  をもつ定常な (*time-homogeneous*) マルコフ連鎖とは、空間  $E$  に値をとる確率変数列  $X_n$  ( $n \geq 0$ ) で、推移核 (*transition kernel*) として次のような  $P$  をもつものである (通常は  $E$  は  $k$  次元ユークリッド空間で、 $\pi$  は  $\sigma$ -有限な測度  $\mu$  に関して密度をもつと仮定される)。すべての可測集合  $A$  について

$$P(X_n, A) = P\{X_{n+1} \in A | X_0, \dots, X_n\}, \quad \pi(A) = \int \pi(dx)P(x, A)$$

$X_0$  の分布は連鎖の初期分布であり、 $X_0$  が与えられたときの  $X_n$  の条件付き分布は、 $P^n$  を推移核  $P$  を  $n$  回繰り返すことと定義すれば

$$P\{X_n \in A | X_0\} = P^n(X_0, A)$$

となる。

不変分布は確率分布であるときに定常分布 (*stationary distribution*) とも呼ばれる。

**定義 3.1.2** 不変分布  $\pi$  は、すべての可測集合  $A$  について

$$\lim_{n \rightarrow \infty} P^n(x, A) = \pi(A) \quad \text{for } \pi\text{-almost all } x$$

を満たすとき、連鎖の均衡分布 (*equilibrium distribution*) と呼ばれる。

**定義 3.1.3** 不変分布  $\pi$  をもつマルコフ連鎖は初期状態にかかわらず  $\pi$  が正の確率を与える集合にはいる確率が正であるとき  $\pi$ -既約 ( $\pi$ -irreducible) であるという。

連鎖がどんなところから出発しても、何回か反復するうちにどこにでも到着することができる (ただし確率密度が正であるようなところ) という性質で、irreducible という単語は「これ以上は単純化することができない」という意味をもつ。

**定義 3.1.4** マルコフ連鎖は、一定の時間間隔で必ず訪れる状態空間があるとき周期的 (*periodic*) であるといい、そうでないとき非周期的 (*aperiodic*) であるという。

非周期的であるとは、例えば空間を 2 個以上の集合に分けて周期的にそれらの集合を訪れるということがないという性質で、事後分布からのサンプリングでは特定の場所ばかりサンプリングしてしまうということがないということを意味する。

**定理 3.1.1** (Tierney, 1994)

マルコフ連鎖  $P$  が (1)  $\pi$ -既約 (2)  $\pi P = \pi$  (3) 非周期的であるとき、 $\pi$  は唯一の不変分布であり、また連鎖の均衡分布である。

この  $\pi(x)$  が我々の求めたい事後分布であり、マルコフ連鎖を反復することにより均衡分布として求めるのがマルコフ連鎖モンテカルロ法である。マルコフ連鎖モンテカルロ法には大きく分けて代替サンプリング、ギブスサンプラー、M-H アルゴリズムの3種類とその混合型がある。特によく用いられるのがギブスサンプラー、M-H アルゴリズムであるが、これらについては後の章で説明することとしまし、代替サンプリングについて説明する。

### 3.2 データ拡大法 / 代替サンプリング

Tanner and Wong (1987) によって紹介されたデータ拡大 (data augmentation) 法は、代替サンプリング (substitution sampling) と基本的には同じもので、観測されるデータ  $Y$  を観測できないデータ  $Z$  によって拡大 (augment) させる手法である。  $Y$  と  $Z$  が与えられたとき事後分布  $\pi(\theta|Y, Z)$  が簡単に計算できるか、乱数の発生が簡単であることを仮定しており、そのアルゴリズムは、次のように与えられる。

- (a) 予測分布  $\pi(Z|Y)$  に対する現在の近似分布から  $m$  個の標本  $z_1, z_2, \dots, z_m$  を発生させる ( $m$  は固定された値)。
- (b) 得られた  $z_j$  ( $j = 1, \dots, m$ ) を用いて、現在の  $\pi(\theta|Y)$  に対する近似を以下のように事後分布の混合分布  $g_i(\theta|Y)$  として更新する。

$$g_i(\theta|Y) = \frac{1}{m} \sum_{j=1}^m \pi(\theta|z_j, Y).$$

- (a1) (b) で得られた  $g_i(\theta|Y)$  から標本  $\theta^*$  を発生させる。まず  $m$  個の密度関数  $\pi(\theta|z_j, Y)$ ,  $j = 1, \dots, m$  のなかから1つを確率  $1/m$  で選び、選んだ密度関数を用いて  $\theta^*$  を発生させる。
- (a2) (a1) で得られた  $\theta^*$  を用いて  $\pi(Z|\theta^*, Y)$  から標本  $z$  を発生させて (a1) に戻ることを繰り返す。標本が  $z_1, \dots, z_m$  と  $m$  個得られたら (b) に戻る。

Rubin (1987b) は  $z_1, z_2, \dots, z_m$  を multiple imputations と呼び (a), (a1), (a2) を imputation ステップ、(b) を事後ステップと呼んだ。このようにして得られる標本は、 $m$  の値に関わらず事後分布からの標本に収束するが、 $m$  の値が大きければ大きいほど  $g_i(\theta|Y)$  の  $\pi(\theta|Y)$  に対する近似はよくなる。従って現実の応用においては最初は小さい  $m$  の値からはじめて標本経路が安定的になってきたら大きい  $m$  の値に変更してやるとよい。一方、代替サンプリングのアルゴリズムは

- (a1) 予測分布  $\pi(Z|Y)$  に対する現在の近似分布から  $m$  個の標本  $z^*$  を発生させる。
- (a2) 得られた  $z^*$  を用いて  $\pi(\theta|z, Y)$  から標本  $\theta^*$  を発生させる。
- (a3) (a2) で得られた  $\theta^*$  を用いて  $\pi(Z|\theta^*, Y)$  から標本  $z^*$  を発生させて (a2) に戻る。(a2), (a3) を  $i$  回反復して最後に得られた標本を  $(\theta^{(i)}, Z^{(i)})$  とし (a1) に戻る。これを  $m$  回繰り返し  $m$  個の  $(\theta^{(i)}, Z^{(i)})$  が得られたら順に  $(\theta_j^{(i)}, Z_j^{(i)})$ , ( $j = 1, 2, \dots, m$ ) とする。

(b) 得られた  $(\theta_j^{(i)}, Z_j^{(i)})$ ,  $(j = 1, 2, \dots, m)$  を用いて次のように事後分布を推定する。

$$\pi(\hat{\theta}|Y) = \frac{1}{m} \sum_{j=1}^m \pi(\theta|Z_j^{(i)}, Y)$$

(a1) ~ (a3) を反復することは、 $m = 1$  のデータ拡大法に等しいので Tanner (1993) は連鎖データ拡大 (chained data augmentation) 法と呼んでいる。また次の節で見るようにこの 2 変数の代替サンプリングはギブスサンプリングの特別な場合になっている。

この例では  $\theta$  と  $Z$  の 2 変数であったが例えば  $(\theta_1, \theta_2, \theta_3)$  の 3 変数にも容易に拡張できる。Gelfand and Smith (1990) は、3 変数以上で条件付密度関数が  $\pi(\theta_i|\theta_j, j \neq i)$  だけしか知られていない場合には、代替サンプリングとギブスサンプリングは本質的に収束までの計算時間が同じであること、計算方法はギブスサンプリングのほうが簡単であるが、 $\pi(\theta_1|\theta_2)$ ,  $\pi(\theta_1|\theta_3)$ ,  $\pi(\theta_1|\theta_2, \theta_3)$  などのようにより多くの条件付密度関数が利用可能なときには代替サンプリングが速く収束するだろうことを指摘している。実際の問題への応用では変数の数が多く条件付密度関数の情報が豊富ではないことが多いことから、どちらかといえばギブスサンプラーのほうが有用であるといえるだろう。

## 4 ギブスサンプラー

### 4.1 はじめに

ギブスサンプラー (Gibbs sampler) はもともと Geman and Geman (1984) による画像復元のアルゴリズムとして知られている。例えば航空写真で撮影された農場の写真から、どこにどのような品種の作物が栽培されているのかを知りたいとしよう。カメラの性能にもよるが写真は細かいところがぼやけており、画像の劣化が起こっている。従って写真をもとに判断すると、ぼやけた部分の品種を誤ってしまう可能性がある。そこで写真から得られる品種はノイズを含んだものとしてモデル化し、真の品種を求めるのである。具体的には農場を 2 次元の格子点で表し、格子点  $i$  上の真の品種を  $\mu_i$ , 写真から判断される品種を  $y_i$  とし、 $\mu_i$  にノイズが加わったものが  $y_i$  であると考え、 $y_1, \dots, y_n$  を  $n$  個の格子点とすると

$$f(y_1, \dots, y_n|\mu) = \prod_{i=1}^n f(y_i|\mu), \quad \mu = (\mu_1, \dots, \mu_n)'$$

となる。さらに  $\mu$  の事前分布としてギブス (Gibbs) 分布

$$\pi(\mu) \propto \exp -\beta E(\mu)$$

を考える。  $E(\mu)$  は  $\mu$  の関数でエネルギー関数と呼ばれ  $E(\mu) = \sum_{i<j} I(\mu_i \neq \mu_j)$  と置いた場合、Potts モデルといわれる。このようなモデルにおいて  $y_i$  が与えられたときの  $\mu$  の事後分布は

$$\pi(\mu|y) \propto \{\exp -\beta E(\mu)\} \times \prod_{i=1}^n f(y_i|\mu),$$

となるが、基準化定数を求めるのが難しいなどの問題点をもつ。しかし、 $\mu_{\setminus i} = (\mu_1, \dots, \mu_{i-1}, \mu_{i+1}, \dots, \mu_n)'$  が与えられたときの  $\mu_i$  の条件付き分布  $\pi(\mu_i|\mu_{\setminus i}, y)$  は 1 次元離散分布であり簡単に求まることから条件付き分布を利用して、事後分布を計算する方法としてギブスサンプラーは考案された。

Besag (2000) によればギブスサンプラーの考え方自体は古く Suomela (1976) のマルコフ確率場に関する博士論文 (Jyväskylä 大学) に遡るといふ。この手法はいろいろな分野で独立に発見されており、統計物理学の分野では Creutz (1979) により (heat bath algorithm として知られている)、空間統計学の分野では Ripley (1979) により、またブラウン大学におけるベイズ画像分析の研究で Grenander (1983) と Geman and Geman (1984) により、発見されている。

Geman and Geman (1984) は離散的な分布であるギブス分布への応用として紹介したためギブスサンプラーと名付けたが、この手法を Gelfand and Smith (1990) が *Journal of the American Statistical Association* のなかで事後分布一般に応用を拡張したため、統計学の分野で広まるようになった。従って厳密な意味ではギブスサンプラーという言葉は正しくなく、Robert (1994) が提案したようにベイジアンサンプラー (Bayesian Sampler) とでもいう方が正しいのだが、現在でもギブスサンプラーという通称が一般的に用いられている。

## 4.2 定義

未知の母数を  $\theta = (\theta_1, \dots, \theta_m)'$ , 観測されたデータを  $Y$  とすると、事後分布  $\pi(\theta|Y)$  を求めることが難しいときに次のように反復的に  $\theta$  を発生させることにより、事後分布からの標本を得るのが (systematic scan) ギブスサンプリング (Gibbs sampling) である。

- (1) まず初期値として  $(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_m^{(0)})$  を適当な分布から発生させる。
- (2)  $\theta_1^{(1)}$  を  $\pi(\theta_1|\theta_2^{(0)}, \dots, \theta_m^{(0)}, Y)$  から発生させる。
- (3)  $\theta_2^{(1)}$  を  $\pi(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_m^{(0)}, Y)$  から発生させる。
- (4)  $\theta_3^{(1)}$  を  $\pi(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_m^{(0)}, Y)$  から発生させる。同様に  $\theta_4^{(1)}, \dots, \theta_m^{(1)}$  を順次発生させていく。
- (5) 一般に  $\theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_m^{(i)})$  が得られたら、
  - (i)  $\theta_1^{(i+1)}$  を  $\pi(\theta_1|\theta_2^{(i)}, \dots, \theta_m^{(i)}, Y)$  から発生させる。
  - (ii)  $\theta_2^{(i+1)}$  を  $\pi(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_m^{(i)}, Y)$  から発生させる。
  - (iii)  $\theta_3^{(i+1)}$  を  $\pi(\theta_3|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \dots, \theta_m^{(i)}, Y)$  から発生させる。同様に  $\theta_4^{(i+1)}, \dots, \theta_m^{(i+1)}$  を順次発生させていく。

というように  $\theta^{(i+1)} = (\theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_m^{(i+1)})$  を得るという作業を ( $i = 1, 2, 3, \dots$ ) と繰り返し返していく。

このとき  $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots$  は

$$K(\theta^{(i+1)}, \theta^{(i)}) = \pi(\theta_1|\theta_2^{(i)}, \dots, \theta_m^{(i)}, Y) \prod_{j=2}^{m-1} \pi(\theta_j|\theta_1^{(i+1)}, \dots, \theta_{j-1}^{(i+1)}, \theta_{j+1}^{(i)}, \dots, \theta_m^{(i)}, Y) \\ \times \pi(\theta_m|\theta_1^{(i+1)}, \dots, \theta_{m-1}^{(i+1)}, Y)$$

のように  $K(\theta^{(i+1)}, \theta^{(i)})$  を推移核とするマルコフ連鎖であり、 $N \rightarrow \infty$  のとき  $\theta^{(N)}$  が事後分布  $p(\theta|Y)$  の標本になる。このプロセスを独立に  $t$  回繰り返して  $\theta^{(N),j} (j = 1, 2, \dots, t)$  が得られたとき  $(1/t) \sum_{j=1}^t f(\theta^{(N),j})$  が  $t \rightarrow \infty$  のとき  $E_\theta[f(\theta)|Y]$  に収束することが知られている。

例.  $X_1, \dots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$  とし、 $\mu, \sigma$  の事前分布をそれぞれ  $\pi(\mu) \propto d\mu$ ,  $\pi(\sigma) \propto d\sigma/\sigma$  とすると  $(\mu, \sigma)$  の事後分布  $\pi(\mu, \sigma|x)$  は

$$\pi(\mu, \sigma|x) \propto \sigma^{-n-1} \exp - \frac{n(\mu - \bar{X})^2}{2\sigma^2}$$

である。  $\tau = \sigma^{-2}$  とおくと

$$\pi(\mu, \tau|x) \propto \tau^{\frac{n}{2}-1} \exp - \frac{\tau n(\mu - \bar{X})^2}{2}$$

この事後分布から  $(\mu, \tau)$  の標本を発生させるためには

$$\mu|\tau, x \sim N(\bar{X}, (n\tau)^{-1}), \quad \tau|\mu, x \sim \text{Gamma}\left(\frac{n}{2}, \frac{2}{n(\mu - \bar{X})^2}\right)$$

を交互に発生させればよい。

### 4.3 格子点ギブスサンプラー

ギブスサンプラーを行うためには条件付分布から標本を簡単に発生できる必要があるが、それが困難な場合はどうしたらよいであろうか。この問題に対して Ritter and Tanner (1992) は次のような格子点上で行う格子点ギブスサンプラー (Griddy Gibbs sampler) を提案した。  $\pi(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m, Y)$  から標本を抽出することが困難なとき、

- (1) 密度関数  $\pi(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m, Y)$  の値を  $n$  個の点  $\theta_{j1}, \theta_{j2}, \dots, \theta_{jn}$  で計算し、その値を  $w_k$  ( $k = 1, 2, \dots, n$ ) とする。
- (2)  $w_k$  ( $k = 1, 2, \dots, n$ ) を用いて  $\pi(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m, Y)$  の分布関数の近似を求める。
- (3)  $(0,1)$  区間上の一様乱数を発生させ (2) で得られた近似分布関数を用いて標本を発生させる。

つまり、標本の発生が困難な条件付分布については  $n$  個の適当な格子点を選んで分布関数を階段関数や線形関数で近似するのである。近似分布関数の求めかたは

- (i) 密度関数を確率関数  $g(\theta_{jk}) = w_k / \sum_{l=1}^n w_l$  を持つ離散分布で近似する。
- (ii) あるいは密度関数を区間ごとに値が一定の密度関数

$$g(\theta_{jk}) = \frac{w_k}{\sum_{i=1}^n w_i (a_{i+1} - a_i)}, \quad x_{jk} \in [a_k, a_{k+1}] \text{ のとき}$$

とする。

などが考えられる。この方法は密度関数の定数項の値を計算する必要がないが格子点の選び方、格子点の数により近似精度が大きく左右される。このため Ritter and Tanner (1992) では、近似分布関数から発生したデータの経験分布関数を用いて格子点の位置や数を柔軟に変化させていく適応的格子点 / 格子点成長 (Adaptive Grid/Grid-grower) 法を提案している。格子点を用いたギブスサンプラーの収束に関する理論的な結果は得られていない。

Damien, Wakefield and Walker (1999) は条件付分布からのサンプリングが困難であっても潜在変数を導入することにより単純なサンプリングに帰着できる場合があることを示しているが、それ以外の場合では後述する棄却サンプリング連鎖 M-H アルゴリズム (A-R M-H) や Griddy Multiple-Try Metropolis アルゴリズムを用いることになる。

## 5 M-H (Metropolis-Hastings, メトロポリス-ヘイスティングス) アルゴリズム

### 5.1 はじめに

メトロポリス (Metropolis) アルゴリズムは、もともと統計物理学の分野で相互に作用しあう分子からなるシステムの均衡状態をシミュレーションするために Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) によって導入されたものである。<sup>2</sup>統計物理学ではある温度  $T$  でシステムが均衡状態にあるとき、状態が  $x$  にある確率はギブス分布

$$\pi(x) = \frac{1}{Z(T)} \exp\left(-\frac{E(x)}{kT}\right)$$

(ただし  $Z(T) = \int \exp(-E(x)/kT) dx$ ,  $E(x)$  は  $x$  におけるエネルギー関数、 $k$  は既知の定数でボルツマン定数) に従うことが知られている。Metropolis *et al.* (1953) は、この均衡状態へシステムがたどりつくまでの過程を次のようにシミュレーションした。まず、温度は  $T$  で一定とし、システムがエネルギー  $E(x)$  の初期状態  $x$  にあるとする。

- (1) ランダムに分子を選んで新しい状態の候補  $y$  を選択する。
- (2) この新しい状態の候補  $y$  のエネルギー  $E(y)$  が  $E(x)$  より小さければ新しい状態として  $y$  に移動し、 $y$  を新しく  $x$  として (1) に戻る。
- (3) そうでなければ、確率

$$\frac{\pi(y)}{\pi(x)} = \exp\left\{-\frac{E(y) - E(x)}{kT}\right\}$$

で新しい状態として  $y$  に移動し、 $y$  を新しく  $x$  として (そうでなければもとの状態  $x$  にとどまり)、(1) へ戻る。

そして、この過程を反復していくとやがてある温度  $T$  におけるシステムの均衡状態へたどりつくことを示したのである。このアルゴリズムはメトロポリスアルゴリズムと呼ばれ、Hastings (1970) によ

<sup>2</sup>Metropolis は第二次大戦中のアメリカで秘密裏に進められた原子爆弾製造計画であるマンハッタンプロジェクトのメンバーであった。戦後はシカゴ大学で教鞭をとったり、Los Alamos 国立研究所の先駆者の一人として高速コンピュータの開発を行ったりして、モンテカルロ法の研究に貢献したことで知られている。2000年10月17日ニューメキシコ州で、84歳で亡くなった。

て M-H (Metropolis-Hastings, メトロポリス-ヘイスティングス) アルゴリズムに一般化され、Peskun (1973) によって対称な推移行列  $P$  が与えられたとき最適な性質をもつことが示された (定理 2.1.1)。

## 5.2 一般的な定義

均衡分布  $\pi$  が  $\mu$  に関して密度関数をもち、推移核  $Q$  を  $Q(x, dy) = q(x, y)\mu(dy)$  とする。また  $E^+ = \{x : \pi(x) > 0\}$  として  $x \notin E^+$  は  $Q(x, E^+) = 1$  を満たし  $\pi$  は退化した分布ではないと仮定する。すると M-H 核  $P_{MH}$  は次のように定義される。

$$P_{MH}(x, dy) = p(x, y)\mu(dy) + r(x)\delta_x(dy),$$

ただし、

$$p(x, y) = \begin{cases} q(x, y)\alpha(x, y), & x \neq y \text{ のとき} \\ 0, & x = y \text{ のとき} \end{cases}$$

$$r(x) = 1 - \int p(x, y)\mu(dy),$$

$$\alpha(x, y) = \begin{cases} \min\left(\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right), & \pi(x)q(x, y) > 0 \text{ のとき} \\ 1, & \pi(x)q(x, y) = 0 \text{ のとき} \end{cases}$$

$\delta_x(\cdot)$  は点  $x$  で 1 をとり、それ以外では 0 である関数である。この M-H アルゴリズムは、次のように進められる。

- (1) 現在の点が  $X_n = x$  であるとき提案された密度 (proposal density)  $q(x, y)$  を用いて  $y$  を発生させ  $X_{n+1}$  の候補とする ( $q(x, y)$  は  $x$  が与えられたときの  $y$  の条件付密度なので  $q(y|x)$  と書くこともある)。
- (2) (1) で得られた  $y$  を確率  $\alpha(x, y)$  で  $X_{n+1}$  として受容する。棄却した場合には  $X_{n+1} = x$  とする。
- (3) (1) に戻る。

(2) で用いられる  $\alpha(x, y)$  のなかの比を

$$\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} = \frac{\pi(y)/q(x, y)}{\pi(x)/q(y, x)}$$

と書き直してみると、 $x, y$  の 2 つの点のうち提案された密度  $q$  に比べて均衡分布の確率密度  $\pi$  が大きい方が選ばれる確率が高くなっている。

ここでは  $\pi$  が  $P_{MH}$  の不変分布になっているが、 $P_{MH}$  が既約かどうかは  $q$  や  $\pi$  に依存している。もし、 $P_{MH}$  が既約で  $\pi(\{x : r(x) > 0\}) > 0$  ならばアルゴリズムは非周期的となる。以下では  $E = R^k$ 、 $\mu$  はルベーク測度、 $f$  は  $E$  上の密度関数として具体例を見ていく。

### 5.3 いろいろな M-H アルゴリズム

#### 5.3.1 酔歩連鎖

$q(x, y) = f(y - x)$  のとき、 $Y = x + Z$ ,  $Z \sim f(z)$  となり、推移核  $Q$  は酔歩過程となる。Tierney (1994) は  $f$  の例として一様分布や正規分布、多変量  $t$  分布、Split- $t$  分布 (Geweke, 1989) をあげている。もし、 $f(z) > 0, z \in R^k$  (正符号条件という) ならば  $P_{MH}$  は既約で非周期的となり、収束に関する節で説明するようにマルコフ連鎖は収束する。正符号条件を満たしていなくても  $f$  が原点の近傍で正であり  $E^+$  が開集合で連結されていれば、同様に収束する。しかし、一般に収束速度が遅くなることが多いので、 $Z$  の分布の分散の最適な大きさについて研究がなされている (Gelman, Gilks and Roberts, 1996, Roberts, Gelman and Gilks, 1997)。また、代替的な方法として定常分布  $\pi$  をもつ拡散過程を離散化することでシミュレーションを行う Langevin アルゴリズムや Langevin-Hastings アルゴリズムがある (Grenander and Miller, 1994, Phillips and Smith, 1996, Stramer and Tweedie, 1997, Roberts and Rosenthal, 1998a,b) が、収束速度は必ずしも改善されないという指摘がある (Roberts and Tweedie, 1996)。

#### 5.3.2 独立連鎖

$q(x, y) = f(y)$  のとき、新しい点の候補  $y$  は密度関数  $f$  によって毎回独立に発生させられることになり、推移核  $Q$  は独立連鎖となる。この場合、新しい候補の点  $y$  を受け入れる確率  $\alpha(x, y)$  が

$$\alpha(x, y) = \min \left\{ \frac{w_y}{w_x}, 1 \right\}, \quad w_x = \frac{\pi(x)}{f(x)}$$

であるが、 $w_x$  は重点的サンプリングで  $g(x) = f(x)$  としたときの重みであることから、 $x$  より  $y$  の重みのほうが大きいときには必ず  $y$  を選択し、そうでないときには重みの比に応じて選択する形になっている。このように独立連鎖は重点的サンプリングと関係が深いので Tierney (1994) は  $f$  の選択も自由度の低い多変量  $t$  分布か Split- $t$  分布がよいのではないかと指摘している。独立連鎖では  $E^+$  上で  $f$  が  $\mu$  に関してほとんど正であるならば  $P_{MH}$  は既約で非周期的となりその収束が証明される。

例. 1 階の定常な自己回帰過程  $y_t = \phi y_{t-1} + \epsilon_t$ ,  $\epsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$ ,  $t = 1, 2, \dots, n$ ,  $y_0 \sim N(0, \sigma^2/(1 - \phi^2))$ ,  $|\phi| < 1$  を考える。 $\sigma^2$  を既知とし、 $\phi$  に関する事前分布を  $\pi(\phi)$  とおくと、その事後分布は

$$\pi(\phi|y, \sigma^2) \propto \pi(\phi)(1 - \phi^2)^{\frac{1}{2}} \exp -\frac{(\phi - \hat{\phi})^2}{2\hat{\sigma}_\phi^2}, \quad \hat{\phi} = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^{n-1} y_t^2}, \quad \hat{\sigma}_\phi^2 = \frac{\sigma^2}{\sum_{t=1}^{n-1} y_t^2},$$

現在の  $\phi$  を  $\phi_x$  とし、候補  $\phi_y$  を  $N(\hat{\phi}, \hat{\sigma}_\phi^2)$  に従って発生させると、

$$\frac{w_y}{w_x} = \frac{\pi(\phi_y)(1 - \phi_y^2)^{\frac{1}{2}} \exp -\frac{(\phi_y - \hat{\phi})^2}{2\hat{\sigma}_\phi^2}}{\pi(\phi_x)(1 - \phi_x^2)^{\frac{1}{2}} \exp -\frac{(\phi_x - \hat{\phi})^2}{2\hat{\sigma}_\phi^2}} \times \frac{\exp -\frac{(\phi_x - \hat{\phi})^2}{2\hat{\sigma}_\phi^2}}{\exp -\frac{(\phi_y - \hat{\phi})^2}{2\hat{\sigma}_\phi^2}} = \frac{\pi(\phi_y)(1 - \phi_y^2)^{\frac{1}{2}}}{\pi(\phi_x)(1 - \phi_x^2)^{\frac{1}{2}}}$$

となるので確率

$$\min \left\{ \frac{\pi(\phi_y)(1 - \phi_y^2)^{\frac{1}{2}}}{\pi(\phi_x)(1 - \phi_x^2)^{\frac{1}{2}}}, 1 \right\}$$



で  $\phi_y$  を採択すればよい。

### 5.3.3 棄却サンプリング連鎖

独立連鎖の1つに棄却サンプリング連鎖がある。これは独立連鎖で  $f(x) = \pi(x)$  としたいが  $\pi(x)$  からの標本の発生が難しいためにすべての  $x$  について  $\pi(x) < cg(x)$ , ( $c$  は正の定数) を満たすような  $g(x)$  を使って棄却サンプリングによって  $\pi(x)$  からの標本の発生を行うものである。しかし、すでに述べたように棄却サンプリングには  $g(x)$  の近似精度の問題、 $c$  の選択の問題がある。そこで、これを修正して  $\pi(x) < cg(x)$  を仮定せず、 $f(x) = \min(\pi(x), cg(x))$  として独立連鎖を行えばよい。 $f(x)$  (基準化定数を考慮すれば  $f(x)/\int f(t)dt$ ) からの標本を発生するには棄却サンプリングを用いて

- (1)  $g(y)$  から標本  $y$  を発生させる。
- (2) (0,1) 区間上の一様乱数  $u$  を発生させる。
- (3) もし、 $u < f(y)/cg(y)$  ならば  $y$  を確率標本として受容し、そうでなければ棄却して(1)へ戻る。

この場合、すべての  $x$  について  $f(x) < cg(x)$  であるから、分布のすそに関する問題は生じない。最初に  $c$  を決めるときに  $\pi(x)$  の分布の中心で  $\pi(x) < cg(x)$  となるようにしてやればよいだけである。しかし標本発生効率率は、 $g(x)$  の  $\pi(x)$  への近似精度に依存していることに変わりはない。収束に関する結果は、 $f(x) = \min(\pi(x), cg(x))$  として独立連鎖についての結果がそのままあてはまる。

例.1 変数の確率密度関数  $\pi(x)$  を正規分布で近似することを考える。 $l(x)$  を  $\log \pi(x)$  の基準化定数を除く部分とし、これをある値  $\mu$  のまわりで次のようにテーラー展開する ( $l(x)$  は2階微分可能で  $l''(\mu) < 0$  と仮定する)。

$$l(x) \approx l(\mu) + (x - \mu)l'(\mu) + \frac{1}{2}(x - \mu)^2 l''(\mu) = h(x)$$

$h(x)$  を整理すると

$$h(x) = \text{constant} - \frac{1}{2v_\mu} \{x - (\mu + v_\mu l'(\mu))\}^2, \quad v_\mu = -1/l''(\mu),$$

という形に書くことができ、ちょうど正規分布の密度関数の形になっている。ここで  $\mu$  は何でもよいがサンプリングの効率をよくするためにモードないしモードに近い値であることが望ましい。このとき

- (1) A-R Step.
  - (i) 標本  $y \sim N(\mu + v_\mu l'(\mu), v_\mu)$  を発生させる。
  - (ii) (0,1) 区間上の一様乱数  $u$  を発生させる。
  - (iii) もし、 $u < \exp\{l(y) - h(y)\}$  ならば  $y$  を標本として受容し、そうでなければ棄却して(i)へ戻る。
- (2) M-H Step.  $y$  を確率

$$\min \left\{ \frac{e^{l(y)} \min(e^{l(x)}, e^{h(x)})}{e^{l(x)} \min(e^{l(y)}, e^{h(y)})}, 1 \right\},$$

で受容する。棄却された場合には、現在の  $x$  にとどまる。

### 5.3.4 自己回帰連鎖

自己回帰連鎖は酔歩連鎖と独立連鎖のちょうど中間にあたるもので、 $q(x, y) = f(y - \phi x)$ 、つまり、 $Y = \phi x + Z$ 、 $Z \sim f(z)$ 、(ただし  $\phi$  は定数か行列)とするものであり、 $\phi = 0$  であれば酔歩連鎖、 $\phi = 1$  ならば独立連鎖になっている。

### 5.3.5 格子点連鎖

ギブスサンプラーのところで紹介した格子点による方法では、格子点を十分細かくとらなければならなかったが、Tierney (1994) では M-H アルゴリズムが、その問題を次のように克服することができる と指摘している。簡単化のために

- (1) 適当な  $m$  個の格子点  $x_1, x_2, \dots, x_m$  を決める。
- (2)  $m$  個の格子点のなかから一つ  $x^*$  を確率関数  $Pr(X = x_i) = \pi(x_i) / \sum_{j=1}^m \pi(x_j)$  ( $i = 1, 2, \dots, m$ ) を用いて (SIR のときのように) 選択する。
- (3)  $Z$  を  $f(z)$  から標本を発生させ  $Y = x^* + Z$  とする。

このとき  $q(x, y)$ 、 $\alpha(x, y)$  はそれぞれ次のようになる。

$$q(x, y) = \sum_{i=1}^m \left( \frac{\pi(x_i)}{\sum_{j=1}^m \pi(x_j)} \right) f(y - x_i), \quad \alpha(x, y) = \min \left\{ \frac{\pi(y) \sum \pi(x_i) f(x - x_i)}{\pi(x) \sum \pi(x_i) f(y - x_i)}, 1 \right\}$$

この方法が効率的に機能するためには、格子点が分布の台に十分広がっていること、密度関数  $f$  のすそが十分広がっていることが必要である。格子点のとりかたについてよくわからない場合には Tierney (1994) は、また現在の標本の点  $x$  を中心にして両側に  $h$  ずつの増分で  $m$  個の格子点 (従って  $x - mh, x - (m-1)h, \dots, x - h, x, x + h, \dots, x + (m-1)h, x + mh$  の  $2m + 1$  個) の格子点をとることを提案している。

## 5.4 ヒットエンドラン・アルゴリズム

$\pi(x)$  が多峰分布であったり、状態空間が制約された空間であるときに有効な方法に Schmeiser and Chen (1991) によるヒットエンドラン (hit-and-run) アルゴリズムがある。  $x$  を  $k$  次元ベクトル、 $u_t$  を長さが 1 の  $k$  次元単位球面上の一様分布をするベクトルとする。最も単純なアルゴリズムでは、現在の  $x$  を  $x_t$  とすると次の標本  $x_{t+1}$  は

- (1)  $u_t$  を発生させる。
- (2) 実数  $\lambda_t$  を  $f(\lambda_t) \propto \pi(x_t + \lambda_t u_t)$  から発生させる。
- (3)  $x_{t+1} = x_t + \lambda_t u_t$  とする。

のようにして発生させる。(2) の  $f(\lambda_t)$  のとりかたは他にもあり、また (3) においても M-H ステップを含むものも提案されているが、詳しくは Chen, Shao and Ibrahim (2000) などを参照されたい。ヒットエンドラン・アルゴリズムは  $\pi(x)$  の分布が非常に狭い範囲で高い峰をもつような場合には、その部

分からうまく標本を発生させることができず、収束が遅くなる可能性がある。特に高次元の場合にはそういう危険性が高い。同様なアルゴリズムに Gilks, Roberts and George (1994), Roberts and Gilks (1994) の適応的方向サンプリング (adaptive direction sampling) や Gilks, Roberts and Sahu (1998) がある。

## 5.5 MTM(Multiple-Try Metropolis) アルゴリズム

通常の M-H アルゴリズムでは、現在の点  $x$  と候補となる点  $y$  の 2 点において提案された密度  $q$  と均衡分布  $\pi$  の確率密度の比を比較して標本を発生していたが、Liu, Liang and Wong (2000) では点の周囲の情報も集めるために、点  $x$  と  $y$  のそれぞれの周囲の点を  $k$  個使って確率密度の比を比較する方法として MTM(Multiple-Try Metropolis) アルゴリズムが提案された。

まず、 $q(x, y) > 0$  である時には必ず  $q(y, x) > 0$  であるような提案密度  $q$  を仮定する。また非負の対称な関数  $\lambda(x, y) = \lambda(y, x) > 0$  が  $q(x, y) > 0$  のとき必ず  $\lambda(x, y) > 0$  であるとする。このとき

$$w(x, y) = \pi(x)q(x, y)\lambda(x, y)$$

と定義する。MTM アルゴリズムは

- (1)  $k$  個の候補点  $y_1, \dots, y_k$  を  $q(x, \cdot)$  から発生させる。
- (2)  $k$  個の候補点  $y_1, \dots, y_k$  のなかから  $y$  を

$$Pr(y = y_j) = \frac{w(x, y_j)}{\sum_{i=1}^k w(x, y_i)}$$

に従って発生させる。さらに  $q(y, \cdot)$  から  $k - 1$  個の候補  $x_1^*, \dots, x_{k-1}^*$  を発生させ  $x_k^* = x$  とおく。

- (3)  $y$  を確率

$$\alpha(x, y) = \min \left( \frac{\sum_{i=1}^k w(y_i, x)}{\sum_{i=1}^k w(x_i^*, y)}, 1 \right),$$

で受容し、棄却した場合は  $x$  にとどまる。

この方法の応用の 1 つとして Griddy-Gibbs MTM (MTM-Gibbs) アルゴリズムがある。格子点ギブスサンプラーでは事後分布を近似することによって標本を発生させていたがこれを正確に均衡分布から発生させることを可能にするものである。  $x$  が  $p \times 1$  ベクトル  $x = (x(1), \dots, x(p))$  であるとし、  $x(i)$  を  $x(j), j \neq i$  が与えられたときに次のようにサンプルする。

- (1)  $T(x(i), y|x_{-(i)})$  を推移核とし、  $y_1, \dots, y_k \sim \text{i.i.d. } T(x(i), y|x_{-(i)})$  とする。例えば  $T =$  定数とすると  $x(i)$  の定義域が有界のときに  $y_j$  を定義域で一様に発生させることになり、  $T = x(i) + \epsilon$ ,  $\epsilon \sim N(0, \sigma^2)$  とすると現在の値  $x(i)$  を平均とする正規分布に従って発生させることになる。ここで

$$w(y_j, x(i)) = \pi(x(1), \dots, x(i-1), y_j, x(i+1), \dots, x(p))T(y_j, x(i))\lambda(y_j, x(i))$$

とおく。

(2)  $k$  個の候補点  $y_1, \dots, y_k$  のなかから  $y$  を

$$Pr(y = y_j) = \frac{w(x, y_j)}{\sum_{i=1}^k w(x, y_i)}$$

に従って発生させる。さらに  $T(y, \cdot)$  から  $k - 1$  個の候補  $s_1, \dots, s_{k-1}$  を発生させ  $s_k = x^{(i)}$  とおく。

(3)  $y$  を確率

$$\alpha(x, y) = \min \left( \frac{\sum_{i=1}^k w(y_i, x^{(i)})}{\sum_{i=1}^k w(s_i, y)}, 1 \right),$$

で受容し、棄却した場合は  $x^{(i)}$  にとどまる。

Liu, Liang and Wong (2000) は、MTM-Gibbs の計算時間は格子点ギブスサンプラーのただか2倍程度であるとしている。MTM にはこの他にもヒットエンドラン・アルゴリズムの修正である random-ray モンテカルロや適応的方向サンプリング (adaptive direction sampling) の修正である局所最適化 (local optimization-based) MTM などがある。

## 5.6 M-H アルゴリズムとギブスサンプラーの組み合わせ・関係

多変量の均衡分布  $\pi(x)$ ,  $x = (x_1, \dots, x_p)'$  に対して実際にマルコフ連鎖モンテカルロ法を応用する場合、M-H アルゴリズムを一度に行うのは難しい。つまり、現在の値  $p \times 1$  ベクトル  $x$  に対して、候補となるベクトル  $y$  を発生させても、提案された密度  $q(x, y)$  が  $\pi(y)$  をよりよく近似していないと  $y$  を棄却する確率が高くなってしまふからである。そこで  $x$  の成分 1 つ 1 つに対して順次 M-H アルゴリズムを行うという方法が通常とられる。つまり  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)'$  として条件付き分布  $\pi(x_i | x_{-i})$  に対して順次 M-H アルゴリズムを行っていく。つまり  $p$  個の M-H アルゴリズムを組み合わせて次の標本を発生させる方法である。この方法でも得られる標本は均衡分布に収束していくことが知られている。実はギブスサンプラーは成分ごとの M-H アルゴリズムの特別な場合である。つまり提案された密度が均衡分布であるような場合であり、M-H アルゴリズムでは発生させた候補  $y$  を確率 1 で受容することになる。

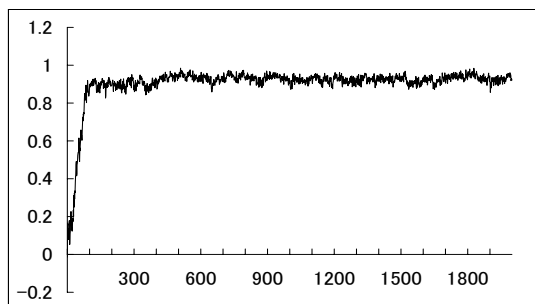
現実の応用においてはギブスサンプラーを行う場合に、条件付き分布から標本を発生させることが難しいため途中で M-H アルゴリズムで発生させることが多い。この方法は Müller (1991) により提唱された Metropolis within Gibbs として知られており、M-H アルゴリズムの部分では 1 ~ 5 回程度の連鎖を繰り返して条件付き分布から標本を発生させるというものである。しかし現実の応用では M-H アルゴリズムを 1 回だけ行って次のギブスサンプラーへ移っていくことが多い。その場合ギブスサンプラーが成分ごとの M-H アルゴリズムの特別な場合であることを考えれば、この手法は Metropolis within Gibbs というよりはむしろ Gibbs within Metropolis ともいうべき方法である。

ところで一般に、手法の組み合わせかたには混合型と循環型がある。  $P_1, P_2, \dots, P_m$  をマルコフ推移核とすると、混合型は正の確率  $\alpha_1, \alpha_2, \dots, \alpha_m$  に従って毎回マルコフ核を確率的に選択する方法で、循環型は  $P_1, P_2, \dots, P_m, P_1, P_2, \dots$  のように順番に核を選択して  $P_m$  の次に再び  $P_1$  に戻って繰り返すという方法である。混合型の場合、 $m$  個の核のひとつが既約で非周期的であれば混合核もまた既約で非周期的になり収束が保証されるが、循環型の場合には必ずしも既約で非周期的にならず、ケースバイケースで調べなくてはならない。

## 6 マルコフ連鎖の収束判定の方法

実際にマルコフ連鎖モンテカルロ法を使うときには標本が均衡分布に収束するまでは、初期値に依存する期間 (burn-in period, 稼働検査期間) であるとして棄て、それ以降の標本を用いて推論を行うことになる。その場合、反復を何回以上行えば初期値に依存せず、均衡分布に収束するのかという問題が生じる。収束の速度に関する理論の節で述べるように、収束に必要な反復回数を理論的に導く試みはなされているが、実用性に欠けているため現在の段階では、まず反復を行い得られた系列を用いてマルコフ連鎖が収束しているかどうかを検査するという方法がとられている。ここではそのなかで応用範囲のひろいと思われるものについてとりあげていく。もちろん、Cowles, Roberts and Rosenthal (1999) で指摘されているようにどの方法も完全ではなく、いくつかの方法を併用しながら、収束の判定をしなくてはならない。収束の判定手法の包括的なサーベイは Robert and Casella (1999) や Mengersen, Robert and Guihenneuc-Jouyaux (1999) を、実験に基づく比較については Cowles and Carlin (1996) や Brooks and Roberts (1999) を参照されたい。

図 4: 真の値は 0.9 であるようなパラメータの sample path. 初期値は 0.



### 6.1 標本経路は安定的かを判定する方法

#### 6.1.1 時系列プロットによる方法

収束を判定する方法で最も簡単な方法は得られた標本の時系列プロットをつくり、その変動が初期値に依存せず安定的な動きになっているかどうかで均衡分布に収束したかどうかを判定する方法である。図 4 では最初の 200 個は初期値に依存しているが、300 個を過ぎたあたりから均衡分布へ収束を始めたと考えられる。

#### 6.1.2 Liu and Liu (1993)

Liu and Liu (1993) では、ギブスサンプラーによるマルコフ連鎖の収束を調べるために、まず、ある分布に基づいて初期値を  $m$  個とりだしそこからそれぞれ別々に連鎖を開始して次のような統計量を計算する。二つの異なる連鎖、第  $i$  連鎖と第  $j$  連鎖の  $t$  回目の反復で得られた  $\theta$  とマルコフ推移核  $K$

を用いて

$$U^{(i,j,t)} = \frac{\pi(\theta^{(j,t)})K(\theta^{(j,t-1)}, \theta^{(i,t)})}{\pi(\theta^{(i,t)})K(\theta^{(j,t-1)}, \theta^{(j,t)})}$$

とする。このとき  $p_t(\theta)$  を反復時点  $t$  における  $\pi(\theta)$  への近似分布とすると

$$E_0(U^{(i,j,t)}) = \text{Var}_\pi \left( \frac{p_t(\theta)}{\pi(\theta)} \right) + 1$$

である。つまり  $U^{(i,j,t)}$  は  $\pi(\theta)$  とその近似分布  $p_t(\theta)$  との距離尺度の不偏推定量になっている。そこで Liu and Liu (1993) は、各反復時点  $t$  で  $m(m-1)$  個の  $U$  の対数の平均  $\overline{\log U}_t$  を計算して  $t$  に対してプロットして  $X$  軸に平行な直線に近づくならば  $p_t(\theta)$  が  $\pi(\theta)$  に収束すると考えた。ほかに  $U$  の系列を  $m/2$  組のペアに分けて Gelman and Rubin (1992) の方法を応用したり、各時点  $t$  での  $U$  の累積分布関数をプロットしたりするなどが提案されている。この方法は  $\pi$  と  $K$  が明示的に求まる場合に有効で、ギブスサンプラーによるモンテカルロ連鎖などへ応用できる。

### 6.1.3 ギブスストッパー (Gibbs Stopper)

Ritter and Tanner (1992) は、次のように現在の段階で予測された事後分布と、均衡分布である真の事後分布の比を用いてギブスサンプラーの収束を判定しようとした。つまり、

$$w_i = \frac{\pi(\theta^{(i)}|y)\pi(\theta^{(i)})}{g_i(\theta^{(i)})}$$

(ただし  $g_i$  は  $i$  回目の反復のあとの予測された事後分布) として、各反復時点  $i$  で  $w_i$  の系列についてヒストグラムを作り、分布がある定数に退化するかどうかで収束を判定しようというものである。ここで

$$g_i(\theta) = \int K(\theta^*, \theta)g_{i-1}(\theta^*)d\theta^*$$

であるが ( $K$  はマルコフ推移核)、積分が困難な場合には  $m$  個の独立なギブスサンプラーによる連鎖から、

$$\hat{g}_i(\theta) = \frac{1}{m} \sum_{j=1}^m K(\theta^j, \theta)$$

を用いてもよい。また、ひとつの長い連鎖を  $m$  個の部分に分割してこれを  $g$  の推定にあててもよい。その場合、 $K$  を計算するために条件付分布の基準化定数が必要になる。

### 6.1.4 Raftery and Lewis (1992a,b, 1996)

Raftery and Lewis (1992a,b, 1996) では、Gelman and Rubin (1992) とは異なり、マルコフ連鎖モンテカルロ法は1つの長いマルコフ連鎖を発生させてそこから  $k$  個おきに標本を抽出するものと想定している。そして以下で説明するような問題において指定した推定の精度を満たすような連鎖の長さ  $k$  の大きさ、初期値に依存している標本の系列の長さ (burn-in 期間) を理論的に導いてくれる。

まずマルコフ連鎖  $\theta$  の関数  $U = f(\theta)$  について累積確率  $Pr(U \leq u|Data) = q$  ( $u$  は固定され

た点)の値の 100s% 信頼区間を点推定量  $\pm r$  の形で求める問題を考える。  $Pr(U \leq u | Data)$  の推定量を作るために  $Z_t = I(U_t \leq u), (t = 1, 2, \dots)$  ( $I$  は定義関数) を定義する。  $Z_t$  自体はマルコフ連鎖ではないので、十分大きな  $k$  をとることによって  $Z_t^{(k)} = Z_{1+(t-1)k}$  を近似的にマルコフ連鎖とする (Raftery and Lewis, 1992a では厳密な証明は与えられていない)。すると  $Z_t^{(k)}$  は、次のように均衡分布  $\pi$ 、推移行列  $P$  で与えられるマルコフ連鎖と考えられる。

$$\pi = (\pi_0, \pi_1) = (q, 1 - q) = \left( \frac{\alpha}{\alpha + \beta}, \frac{\beta}{\alpha + \beta} \right), \quad P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

ここで

$$P = \begin{pmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{pmatrix} + \frac{(1 - \alpha - \beta)^m}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

であることから  $Z_t^{(k)}$  の系列も初期値に依存するので、最初の  $m$  個を除く必要がある。  $Z_t^{(k)}$  の確率関数の値と均衡分布の確率関数の値との差が  $\epsilon$  以下になるように  $m$  を定めると

$$m = \left\lceil \log \left( \frac{\epsilon(\alpha + \beta)}{\max(\alpha, \beta)} \right) \right\rceil / \log(1 - \alpha - \beta)$$

となるので  $Z_t^{(k)}$  を発生させて  $m + 1$  個目からの標本を使う。すると  $\bar{Z} = \sum_{t=m+1}^{m+n} Z_t^{(k)} / n$  に中心極限定理を用いて

$$Pr[\bar{Z} - r \leq q \leq \bar{Z} + r] = s, \quad r = z_{(1-s)/2} \sqrt{\frac{\alpha\beta(2 - \alpha - \beta)}{n(\alpha + \beta)^3}}$$

(ただし、 $z_\alpha$  は標準正規分布の  $100(1 - \alpha)\%$  点) で与えられる。  $r$  が予め与えられた値になるようにするには  $n$  を

$$n = \frac{z_{(1-s)/2}^2}{r^2} \frac{\alpha\beta(2 - \alpha - \beta)}{(\alpha + \beta)^3}$$

ととればよい。  $M = mk, N = nk$  とすると、全体として  $M + N$  個の標本をマルコフ連鎖モンテカルロ法で抽出して、最初の  $M$  個を初期値に依存するものとして棄て、残りの  $N$  個を均衡分布からの標本とすれば良いことになる。

$m, n$  の値を決めるためには、 $q, r, s, \epsilon$  については実験者が与えてやらなければならないが  $k, \alpha, \beta$  については観測値から推定しなくてはならない。そこで、Raftery and Lewis (1992) はまず試験的に与えられた  $q, r, s, \epsilon$  のもとで最小の  $N = N_{\min}$  でマルコフ連鎖モンテカルロ法を行うとした。  $N = N_{\min}$  となるのは  $Z_t$  が互いに独立な場合であると考えられるから、 $1 - \alpha = \beta = \pi_0 = q, M = 0, k = 1$  となり、 $N_{\min} = \frac{z_{(1-s)/2}^2 q(1-q)}{r^2}$  で与えられる。この  $N_{\min}$  個のデータから  $\alpha, \beta$  を推定する。  $k$  についても、得られた観測値を用いて各  $k$  の値毎に 1 次のマルコフ過程か 2 次のマルコフ過程かを BIC 基準によって選択し、1 次のマルコフ過程を選択するような最初の  $k$  を求める値とする (Raftery, 1986, Bishop, Fienberg and Holland, 1975)。この推定された  $\alpha, \beta, k$  を用いて  $M, N$  を計算し必要ならば最初に発生させた  $N_{\min}$  の系列に付加的にマルコフ連鎖モンテカルロ法を行い、再度  $\alpha, \beta, k$  を推定させて  $M, N$  を計算する、ということを繰り返して行けばよい。

Raftery and Lewis (1992a) では、主として  $q = 0.025, s = 0.95, \epsilon = 0.01$  をとり、 $r$  については  $r = 0.0125$  (精度に厳しい場合には  $r = 0.005$ ) を勧めている ( $N_{\min} = 600$ )。このうち  $q = 0.025$  は

恣意的であり、実際にはいくつかの異なる  $q$  に対して計算をしなくてはならないだろう。この手法は、すべてのマルコフ連鎖モンテカルロ法について有効であり、またプログラムが FORTRAN と S で書かれたものがカーネギーメロン大学統計学部の Statlib に公開されている。

ただし、Brooks and Roberts (1999) では、関心の対象が分位数でない限り、この方法は誤って収束が速いと判断しがちなので適用には注意が必要であるとしている。

## 6.2 定常分布への収束判定法

### 6.2.1 Geweke (1992)

Geweke (1992) では  $\theta_n$  をマルコフ連鎖、 $g$  をその関数として  $g$  のスペクトル密度  $S_g(\omega)$  が存在し、 $\omega = 0$  であるとき

$$\frac{g_n - E[g(\theta)]}{\sqrt{S_g(0)/n}} \Rightarrow N(0, 1), \quad g_n = \frac{1}{n} \sum_{i=1}^n g(\theta^{(i)})$$

であることを利用して、次のような手順で連鎖の収束を判定するものである。この仮定が満たされる限り、この方法はすべてのマルコフ連鎖に適用できる。

- (1)  $\theta^{(i)}$  ( $i = 1, 2, \dots, n$ ) を発生させ、 $g_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} g(\theta^{(i)})$ ,  $g_{n_2} = \frac{1}{n_2} \sum_{i=n-n_2+1}^n g(\theta^{(i)})$  を計算する (通常  $n_1 = 0.1n, n_2 = 0.5n$  とする)。
- (2) 標準正規分布の  $z_\alpha$  を上側  $100\% \alpha$  点、 $Z = (g_{n_1} - g_{n_2}) / \sqrt{S_g(0)/n_1 + S_g(0)/n_2}$  として  $|Z| \leq z_{\alpha/2}$  ならば収束をしていると判定する。

Cowles and Carlin (1996) は、この方法はスペクトルウィンドウのとりかたに左右されやすいことを指摘している。

### 6.2.2 Gelman and Rubin (1992)

Gelman and Rubin (1992) は、均衡分布が正規分布であるマルコフ連鎖を考えて、発生された標本の分布を  $t$  分布によって (従って分散の大きな分布で) 近似し、その分散が正規分布の分散に近いかどうかによって連鎖の分布が収束しているかどうかを判断しようとした。具体的には次のようなステップを踏んで行われる。

- (1) 1つの連鎖の長さが  $2n$  であるような  $m (\geq 2)$  個の独立のマルコフ連鎖を発生させる。第  $i$  の連鎖の第  $j$  番目の観測値を  $X_{ij}$  とおく。 ( $i = 1, \dots, m, j = 1, \dots, 2n$ ) とする。
- (2) (1) で得られたデータで、各連鎖について最初の  $n$  個の分を初期値に依存するものとして捨てて、残りの後半の  $n$  個を用いて次の  $\sqrt{R}$  を計算する。

$$\sqrt{R} = \sqrt{\left( \frac{n-1}{n} + \frac{m+1}{m} \frac{B/n}{W} \right) \frac{df}{df-2}},$$

$$\frac{B}{n} = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{x}_{..})^2}{m-1}, \quad \bar{x}_i = \frac{\sum_{j=n+1}^{2n} x_{ij}}{n}, \quad \bar{x}_{..} = \frac{\sum_{i=1}^m \bar{x}_i}{m}$$



$$\begin{aligned}
W &= \frac{\sum_{i=1}^m s_i^2}{m}, \quad s_i^2 = \frac{\sum_{j=n+1}^{2n} (x_{ij} - \bar{x}_i)^2}{n-1} \\
df &= \frac{2\hat{V}^2}{\hat{Var}(\hat{V})}, \quad \hat{V} = \frac{n-1}{n}W + \frac{m+1}{mn}B \\
\hat{Var}(\hat{V}) &= \left(\frac{n-1}{n}\right)^2 \frac{1}{m} \frac{\sum_{i=1}^m (s_i^2 - W)^2}{m-1} + \left(\frac{m+1}{mn}\right)^2 \frac{2B^2}{m-1} \\
&\quad + \frac{2(m+1)(n-1)}{mn^2} \frac{n}{m} \left\{ \frac{\sum_{i=1}^m (s_i^2 - W)(\bar{x}_i^2 - \bar{\bar{x}}^2)}{m-1} \right. \\
&\quad \left. - 2\bar{\bar{x}} \frac{\sum_{i=1}^m (s_i^2 - W)(\bar{x}_i - \bar{x}_{..})}{m-1} \right\}, \quad \bar{\bar{x}}^2 = \frac{\sum_{i=1}^m \bar{x}_i^2}{m}
\end{aligned}$$

ここで  $B/n$  は連鎖間での変動を、 $W$  は各連鎖内の変動を表わしている。

Gelman and Rubin (1992) は、この  $\sqrt{R}$  が 1 に近いかどうかで収束を調べ、1 に近い場合には収束していると考えて各連鎖の後半  $n$  個の系列をあわせて均衡分布の標本（最初の  $n$  個は初期値に依存していると考えられるため）とし、1 に近くない場合には収束が起こっておらずに標本が初期値に依存していると考え、 $n$  の値を大きくして再び連鎖を発生させてから  $\sqrt{R}$  を計算することを提案した。これは基本的に均衡分布が一次元分布の場合に考えられたものであるが多次元分布の場合には各変数に対してこの  $\sqrt{R}$  を計算してこれを用いて総合的に収束の判定を行うことになる。

この方法はどんなマルコフ連鎖モデルにも応用が可能だが、正規分布の理論を基礎に考えられているため、均衡分布が正規分布とかなり異なる場合には注意が必要である。たとえば Gelman and Rubin (1992) も指摘しているように均衡分布が 2 峰型であるときに  $m$  個の連鎖の初期値がすべて一方のモードの周辺であり  $n$  の値が小さいときにはそのモードの周辺ばかりで標本の発生が行われて収束したと誤認し、もう一方のモードを検出できないことが起こりうるからである。これを防ぐためには均衡分布の形状に関する情報をできるだけ得て、初期値も均衡分布の台に広く散らばるように発生させることが必要である。この Gelman and Rubin (1992) の方法も  $S$  のプログラムとしてカーネギーメロン大学統計学部の Statlib で公開されている。

### 6.2.3 再生時間 (regeneration time)

Mykland, Luke and Yu (1995) は確率過程の再生時間を利用して、観測されたマルコフ連鎖が収束しているかどうかを図を用いて判断しようとした。この方法は、すべてのマルコフ連鎖に応用可能であるが、応用が簡単なのは主として M-H アルゴリズムの独立連鎖である。

確率過程  $\{X_n : n = 0, 1, \dots\}$  は、時点  $T_i$  において ( $T_0 \leq T_1 \leq \dots$ ) その将来の過程が過去の過程とは独立であり、かつ過去の過程と同じ分布に従うとき再生的であるといわれる。Mykland, Luke and Yu (1995) はモンテカルロ法によって発生させられるマルコフ連鎖が収束をしているならば、その再生時間は安定しているはずであり、再生時間と次の再生時間の間の長さである tour の長さが安定していると考えた。そこで  $T_i/T_n$  を  $i/n$  ( $i = 1, 2, \dots, n$ ) に対してプロットし、それが傾き 1 の直線上にそっていれば連鎖は収束しており、直線から乖離して不規則な動きを示していれば連鎖は収束していないと判断すればよいとした。このプロットは SRQ (scaled regeneration quantile) プロットと呼ばれる。

ここでは独立連鎖の M-H アルゴリズムを例に考える。まず  $q(x, y) = f(y)$  はルベーク測度上で定義

されているものとし、 $w(y) = \pi(y)/f(y)$  として初期値  $X_0$  を  $f(y) \min(w(y)/c, 1)$  に比例する密度関数をもつ分布から棄却法によって発生させる。次にマルコフ連鎖  $X_n$  の再生時間  $T$  をみつけるために Splitting 変数  $S_n$  という過程を定義する。 $S_n$  は独立連鎖の次の候補点  $X_{n+1}$  を発生させて棄却されたら  $S_n = 0$  とし、受容されたら  $S_n$  を次のような  $r(x, y)$  を成功確率とするベルヌーイ確率変数とする。

$$r(x, y) = \begin{cases} \max\{\frac{c}{w(x)}, \frac{c}{w(y)}\} & w(x) > c \text{ かつ } w(y) > c \text{ のとき} \\ \max\{\frac{w(x)}{c}, \frac{w(y)}{c}\} & w(x) < c \text{ かつ } w(y) < c \text{ のとき} \\ 1 & \text{それ以外のとき} \end{cases}$$

Mykland, Luke and Yu (1995) は、 $S_n = 1$  となる時点が確率過程の再生時間  $T$  であることを証明した (Theorem 1, 2, 3)。もし  $S_n = 1$  で  $\sum_{j=1}^n S_j = i$  ならば、 $T_i = n$  となる。こうして得られた  $T_1, T_2, \dots, T_k$  を用いて SRQ プロットを作り、収束の判定を図により行う (初期値は  $T_0 = 0$  を満たすようにとるので直線は原点から始まり (1,1) で終わる)。

tour の長さの期待値は、 $\int (\pi(x)/w(x))dx / [\int \min\{\sqrt{c}/w(x), 1/\sqrt{c}\}\pi(x)dx]^2$  で与えられる。独立連鎖のなかでも棄却法を用いる場合には  $f(x) \propto \min(\pi(x), ag(x))$  として上の  $r$  の代わりに

$$r(x, y) = \begin{cases} 1 & \pi(x) \leq ag(x) \text{ あるいは } \pi(y) \leq ag(y) \text{ のとき} \\ \min\{\frac{ag(x)}{\pi(x)}, \frac{ag(y)}{\pi(y)}\} & \text{それ以外のとき} \end{cases}$$

とすればよい。 $q(y, x) = q(x, y)$  であるような M-H アルゴリズムやギブスサンプラーへの応用の仕方についても Mykland, Luke and Yu (1995) は言及しているが、すべての場合に簡単に適用できるとはいえない。

#### 6.2.4 Heidelberger and Welch (1983)

Heidelberger and Welch (1983) は「マルコフ連鎖が一様エルゴード性を持つ」という帰無仮説を検定する方法を与えている。もし、マルコフ連鎖  $\theta^{(i)}$  ( $i = 1, 2, \dots, n$ ) が一様エルゴード的であるならば、 $S(0)$  をスペクトル密度関数の推定値を 0 で評価したものとすると

$$B_n(t) = \frac{\sum_{i=1}^{[nt]} \theta^{(i)} - [nt]\bar{\theta}}{\sqrt{nS(0)}}, \quad 0 \leq t \leq 1$$

の分布が近似的に Brownian bridge であることから Cramer-von Mises 統計量

$$\int_0^1 B_n^2(t) dt$$

(統計量は数値積分によって計算される) を用いて次のようなプロセスで収束の判定は行われる。

- (1) まず、反復回数の上限として  $j_{\max}$ 、平均の信頼区間の 1/2 の幅の (平均に対する) 相対的な大きさを  $\epsilon$  と決めて、試験的に  $j_1 = 0.1j_{\max}$  回だけ反復を行う。
- (2) 反復したうちの後半 50% を使ってスペクトル密度の推定値  $S(0)$  を計算し、検定を行う。

- (3) もし帰無仮説が棄却されたら、 $\theta^{(i)}$  の最初の 10% を除いて検定を再び行う。それでも棄却されれば、さらにまた同じ 10% の長さだけ除いて検定を行うということを繰り返し、受容されれば (5) に行き、そうでなければ残りの長さがもとの 50% になるまで続け、(4) に進む。
- (4) 検定のための反復回数が十分でないため、さらに反復を行い合計でこれまでの 1.5 倍分の反復をし (ただし  $j_{\max}$  を超えない)、(2) に戻る。反復回数が  $j_{\max}$  でも棄却されたときには連鎖は収束していないと結論する。
- (5) 帰無仮説が受容されたときの検定に使ったすべての  $\theta^{(i)}$  ( $n^*$  個とする) を用いて平均、スペクトル密度を推定し直して  $\bar{\theta}^*$ ,  $S^*(0)$  とすると  $z_{\alpha/2}\sqrt{S^*(0)/n^*} < \epsilon\bar{\theta}^*$  ならば収束していると判定し作業を終了する。そうでなければ (4) のように反復をさらに行って (2) に戻る。

この方法はすべてのマルコフ連鎖に適用できるが、その検定は連鎖の初期値に依存する長さに対して  $j_{\max}$  が小さすぎると検出力が低いことが指摘されている。

### 6.3 単一連鎖と複数連鎖のどちらがよいのか

事後分布・均衡分布からのサンプリングを行う場合、(1) 初期値からはじめて長い 1 つの系列、単一連鎖 (single chain) を発生させ、そのサンプルを使って推論を行う方法と (2) 初期値から (1) より短い、複数の連鎖 (multiple chain) を発生させてそれらの最後のサンプルを用いて推論を行う方法の 2 つがある。複数連鎖の場合は初期値をうまくとらなければ、得られたサンプルが均衡分布に十分収束していない可能性があり間違いをおかしやすい。単一連鎖でもサンプルが均衡分布の一部だけをひろっていてもまだ他の部分をサンプリングしていないのにあたかも収束しているように勘違いしてしまう可能性がある (“you’ve only seen where you’ve been”) が、長い系列の方が全体を拾う可能性が高いので現在では単一連鎖のほうが複数連鎖よりもどちらかといえば望ましいと考えられている。

### 6.4 初期値の問題 - 完全シミュレーション

図 4 でも見たように初期値が均衡分布から離れたところから始めると、均衡分布に収束するまでに burn-in 期間をもうけなければならない。そこで初期値を均衡分布から発生させるためのアルゴリズムとして Propp and Wilson (1996) による完全シミュレーション (perfect simulation) アルゴリズムが提案された (もともとは exact sampling と呼んでいたが Kendall (1998) により perfect と改めて名付けられた)。まず  $x$  の状態空間が有限個 ( $k$  個) である場合を考えて、標本を  $-m$  時点から発生させはじめに初期値を 0 時点で発生させる。状態空間のどこから標本の発生を開始しても、サンプリングに用いる乱数を同じにとる限り、いったん同じ状態に入ればそれ以降は同じ標本経路をとることになる。従って十分  $m$  を大きくとれば異なる  $k$  個の状態から  $k$  個のマルコフ連鎖を走らせても (サンプリングに用いる乱数を同じにとる限り) 0 時点かそれ以前において、すべて同じ標本経路になるはずである (このようなサンプリングを coupling from the past, CFTP という)。そのようにして得られた 0 時点での標本は均衡分布からの標本となる。しかし  $k$  が小さくない限りそのようなサンプリング方法は計算負荷が大きい。そこで Propp and Wilson (1996) は単調性という仮定において計算負荷の小さい方法を提案した (また、その際の  $m$  の期待値に関する上限が Gibbs, 2000 で示されている)。その後、完全シミュ

レーションはさまざまに拡張されており、特に Murdoch and Green (1998) では状態空間が連続な空間である場合についていくつかの例を考慮している。

### 6.5 収束を加速させる方法

サンプリングするパラメータ  $\theta$  の次元が大きい場合に、モデルの構造によっては  $\theta$  の成分を1つずつ発生させていると得られた標本の自己相関が非常に高くなり、均衡分布への収束が遅くなることがある（例：確率的ボラティリティモデル、順序プロビットモデルなど）。このような場合、収束を速めるためにいくつかの方法が提案されている。

図 5: Single move と Multi-move での標本経路

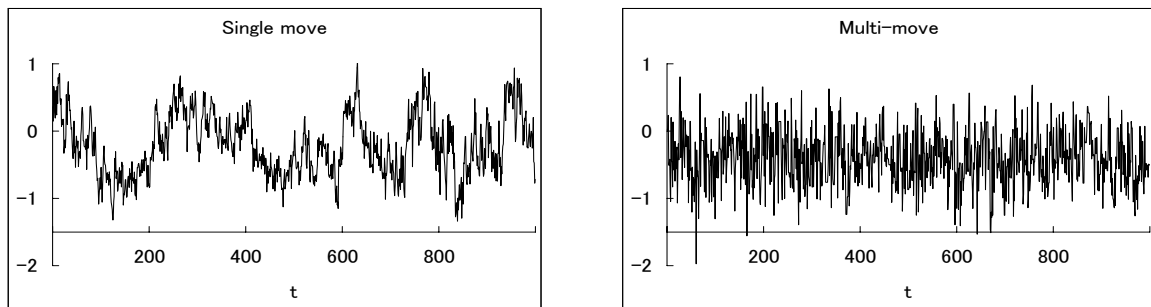
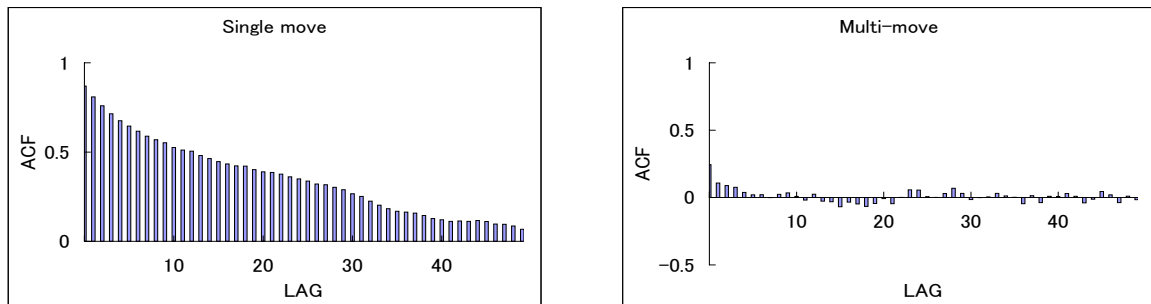


図 6: Single move と Multi-move での標本自己相関関数



第1はブロック化 (blocking) といひ、パラメータをいくつかのブロックに分けて、ブロックごとに標本を発生させる方法である。例えば  $\theta = (\theta_1, \dots, \theta_{10})'$  ならば  $\theta_i$  を一つずつ発生させる (single move としばしばいう) のではなく  $(\theta_1, \dots, \theta_3)'$ ,  $(\theta_4, \dots, \theta_6)'$ ,  $(\theta_7, \dots, \theta_{10})'$  という3つのブロックを発生させることを繰り返す (multi-move) のである。これにより標本の自己相関が飛躍的に小さくなることが知られている。図5では、パラメータが1000個あり  $\theta = (\theta_1, \dots, \theta_{1000})'$  であるようなモデルで  $\theta_{500}$  の標本経路を single move と multi-move (ブロックの個数は10個) についてプロットしたものである。

multi-move では事後分布からまんべんなくサンプリングしているのに対して single move では前回の標本の値に強く依存して偏ったサンプリングしかできず、均衡分布からのサンプリングがなかなか進まないことが見てとれる。図 6 は得られた標本の標本自己相関関数であるが、multi-move ではほとんど 0 であり直前の標本に依存していないのに対して、single move では標本自己相関が高く、以前の値に縛られていることがわかる。もっともブロックのサイズを大きくすると、条件付き分布からの発生が難しくなりやすいので適当なサイズを見つけるのには試行錯誤が少し必要である。確率的ボラティリティモデルのように非線形な時系列モデルにおける潜在変数のサンプリングについて Shephard and Pitt (1997) が multi-move の有効性を示している。<sup>3</sup>

第 2 はパラメータの変換 (reparameterization) である。パラメータの変換によってサンプリングする空間が変わり、分散分析モデルや順序プロビット、変量効果ポアソンモデルのパラメータのサンプリングに有効であることが知られている (例えば Chen, Shao and Ibrahim, 2000 を参照)。またパラメータの変換ではないが、Liu and Wu (1999) による parameter expanded data augmentation アルゴリズムや Liu and Sabatti (2000) による generalized multigrid モンテカルロ法では通常の MCMC のステップに加えてパラメータすべてにある確率変数を加えたり、かけたりするステップを加えることにより標本の自己相関を減少させることができるとしている。

その他、均衡分布が多峰型分布のような場合に局所的なサンプリングしてしまう危険を減らし、全体をサンプリングできるように重み変数をサンプリングの空間に新たに加えてサンプリング空間を拡大 (augment) する方法が Liu, Liang and Wong (2001) により dynamic weighting アルゴリズムとして提案されている。

## 7 収束に関する理論的結果

以下では収束に関する理論的結果についていくつか紹介する。より包括的で詳しい説明は Robert and Casella (1999) や Polson (1996) を参照されたい。

### 7.1 一般的な結果

マルコフ連鎖によって発生された系列の分布の収束に関する証明は Tierney (1994) にあり、その証明は Nummelin (1984) に大きく依存している。まず  $\mathcal{E}$  を空間  $E$  上の  $\sigma$ -field とし、 $(E, \mathcal{E})$  上のマルコフ推移核  $P : E \times \mathcal{E} \rightarrow [0, 1]$  を (i) すべての固定された集合  $A \in \mathcal{E}$  について  $P(\cdot, A)$  は可測である、(ii) すべての固定された  $x \in E$  について  $P(x, \cdot)$  は  $(E, \mathcal{E})$  上の確率測度である、とする。

**定理 7.1.1** (Tierney, 1994 の Theorem 1)

$P$  が  $\pi$  既約で  $\pi P = \pi$  を満たすとき、 $P$  は正再帰的 (positive recurrent) であり  $\pi$  は  $P$  の唯一の不変分布である。もし  $P$  がまた非周期的ならば  $\pi$  に関してほとんどすべての  $x$  について  $\|P^n(x, \cdot) - \pi\| \rightarrow 0$  (ただし、 $\|\cdot\|$  は total variation distance を表わす) が成り立つ。もし  $P$  が Harris 再帰的 (Harris recurrent) であるならば、この収束はすべての  $x$  について成り立つ。

<sup>3</sup>ただしアルゴリズムに一部誤りがある (Watanabe and Omori, 2001).

再帰的 (recurrent) であるとは、どんなところから (ただし確率が正であるところ。 “for almost every starting value”) 出発しても、確率 1 で再びもとの場所に無限回帰ってくるという性質で、確率的に連結 (connected) しているという意味でもある。さらに 正再帰的 (positive recurrent) であるということは定常分布が存在するということの意味する (また  $X$  が正再帰的ならば、 $X$  は再帰的である)。さらに Harris 再帰的とは状態空間のどんなところから出発しても (“for every starting value”), 確率 1 で再びもとの場所に無限回帰ってくるという性質を意味する。

同様な定理は Chan (1993) の Theorem 1.1 でも与えられている。Tierney (1994) の Theorem 1, Theorem 2, Corollary 1 から

系 7.1.1  $P$  が  $\pi$ -既約で  $\pi P = \pi$  を満たし非周期的であるとする。もし  $P(x, \cdot)$  が  $\pi$  に関してすべての  $x$  について絶対連続ならば、すべての  $x$  について  $\|P^n(x, \cdot) - \pi\| \rightarrow 0$  が成り立つ。

## 7.2 M-H アルゴリズムの収束

M-H (Metropolis-Hastings) 核  $P_{MH}$  の収束については Tierney (1994) の Theorem 1, Theorem 2, Corollary 2 から次が成り立つ。

系 7.2.1  $P_{MH}$  が  $\pi$ -既約で非周期的ならば、すべての  $x$  について  $\|P_{MH}^n(x, \cdot) - \pi\| \rightarrow 0$  が成り立つ。

$P_{MH}$  の既約性の条件には次がある。

補題 7.2.1 (1) (Roberts and Smith, 1994 の Theorem 3 (ii)). もし (i)  $q(x, y)$  が  $\pi$ -既約であり、(ii)  $q(x, y) = 0$  と  $q(y, x) = 0$  が同値であるとき (つまりすべての  $(x, y) \in E \times \mathcal{E}$  について  $\alpha(x, y) > 0$  であるとき)  $P_{MH}$  は  $\pi$ -既約である。

(2) (Mengersen and Tweedie, 1996 の Lemma 1.1). もしすべての  $x \in E$  について  $\pi(y) > 0$  が  $q(x, y) > 0$  を意味するならば、 $P_{MH}$  は  $\pi$ -既約である。

$P_{MH}$  の非周期性の十分条件としては次がある。

補題 7.2.2 (1) (Nummelin, 1984 の Section 2.4).  $P_{MH}$  は  $\pi$ -既約で  $\pi(\{x : r(x) > 0\}) > 0$  ならば非周期的である。

(2) (Mengersen and Tweedie, 1996 の Lemma 1.2). もし、 $\pi(x)$  と  $q(x, y)$  がすべての  $(x, y)$  について正で連続であるならば  $P_{MH}$  は非周期的であり、 $\mu(C) > 0$  ( $\mu$  はルベーグ測度) であるようなすべてのコンパクト集合  $C$  は *small set* である。

(3) (Roberts and Smith, 1994 の Theorem 3 (i)). もし  $q(x, y)$  が非周期的か、あるいはある数  $n \geq 1$  について  $Pr(X_n = X_{n-1}) > 0$  が成り立つならば  $P_{MH}$  は非周期的である。

また Smith and Roberts (1993) の Appendix にも収束に関する結果がコンパクトにまとめられている。これらをまとめると Mengersen and Tweedie (1996) の Lemma 1.1, Lemma 1.2 と Tierney (1994) の Theorem 1, Corollary 2 に基づいて

命題 7.2.1  $\pi(x)$  と  $q(x, y)$  がすべての  $(x, y)$  について正で連続で  $\pi P_{MH} = \pi$  であるならば  $P_{MH}$  はすべての  $x$  について  $\|P_{MH}^n(x, \cdot) - \pi\| \rightarrow 0$  を満たす。

ことがわかる。

### 7.3 ギブスサンプラーの収束

ギブスサンプラーについてはそれぞれのケースで系 7.1.1 を確かめることによって収束を導くことができる。Chan (1993) もギブスサンプラーの収束の十分条件を次のように与えている。

定理 7.3.1 (Chan, 1993 の Theorem 1.2).  $\pi(x)$  が状態空間  $E$  上のすべてにおいて正であるならば、ギブスサンプラーによるマルコフ連鎖はエルゴード的である (従って収束する)。

Chan (1993) は、さらに状態空間  $E$  上で  $\pi(x) > 0$  となるための十分条件を示している (Condition (C2))。Roberts and Smith (1994) も次のようなギブスサンプラーの収束の十分条件を与えている。

補題 7.3.1 (Roberts and Smith, 1994 の Theorem 1& 2).

- 状態空間の測度が離散測度であるとき、ギブスサンプラーの推移核  $P$  は定義可能で非周期的である。従って  $P$  が  $\pi$ -既約ならばギブスサンプラーは収束する。
- 状態空間の測度がルベグ測度であるとき、 $\pi$  が 0 で下に半連続で、 $\int \pi(x) dx_i$  がすべての  $i$  について局所有界であり、 $E$  が連結されているならば  $P$  は  $\pi$ -既約でギブスサンプラーは収束する。

### 7.4 組み合わせアルゴリズムの収束

変数がたくさんあり、乱数発生をすべて同時にではなく 1 回に 1 変数ずつ発生させていく M-H アルゴリズムの収束については次のような結果が得られている。

定理 7.4.1 (Chan and Geyer, 1994 の Theorem 1).

1 回に 1 変数ずつ乱数発生させていく  $R^d$  上の  $P_{MH}$  が、ルベグ測度に関して絶対連続である均衡分布  $\pi$  をもつとする。このとき  $P_{MH}$  が Harris 再帰的であるための十分条件は、すべての条件つき推移核 (無条件推移核  $P_{MH}$  を含む) が、条件つきの均衡分布と測度に関して固定された変数のいかなる値に対しても既約であることである。

従って、さらにこの  $P_{MH}$  が非周期的であればすべての  $x$  について  $\|P_{MH}^n(x, \cdot) - \pi\| \rightarrow 0$  を満たす。その他、データ拡大法の収束に関しては次の正規条件の下で重要な結果が Tanner and Wong (1987) に示されている。

定理 7.4.2  $K$  が正規条件「 $K(\theta, \phi) = \int \pi(\theta|Z, Y)\pi(Z|\phi, Y)dZ$  が一様に有界で  $\theta$  に関して equicontinuous である。また、いかなる  $\theta_0 \in \theta$  についても  $\theta_0$  の開近傍  $U$  が存在し、すべての  $\theta, \phi \in U$  について  $K(\theta, \phi) > 0$  である。」を満たすとき

1.  $g^*(\theta) = \pi(\theta|Z, Y)$  が唯一の  $g(\theta) = \int K(\theta, \phi)g(\phi)d\phi$  を満たす確率密度関数である。

2. 初期値  $g_0$  が  $\sup_{\theta} |g_0(\theta)/\pi(\theta|Z, Y)| < \infty$  を満たすならば

$$\int |g_i(\theta) - \pi(\theta|Z, Y)| d\theta \rightarrow 0 \text{ as } i \rightarrow \infty.$$

標本空間が有限な場合には、ある条件のもとで真の事後分布に  $O(\log n)$  回 (ただし  $n$  はデータの個数) の繰り返しののちに収束することが Rosenthal (1993) によって証明されている。

## 7.5 収束の速度

### 7.5.1 収束速度の種類

マルコフ連鎖が収束する速度に関する理論的な結果は、さらに強い仮定を加えることによって証明されている。正 Harris 再帰的で非周期的なマルコフ連鎖はエルゴード的であるといわれるが、次のようにエルゴード性には、通常のエルゴード性よりも強いものとして (1) 二次のエルゴード性、(2) 幾何エルゴード性、(3) 一様エルゴード性があり、(3) は (2) を、(2) は (1) を、(1) は通常のエルゴード性を意味するという関係になっている。

- (1) 二次のエルゴード性.  $S_B$  を集合  $B$  への到達時間とする。不変分布  $\pi$  をもつエルゴード的な連鎖は  $\pi(B) > 0$  であるすべての  $B \in \mathcal{E}$  について  $\int_B \pi(dx) E_x[S_B^2] < \infty$  を満たすとき、二次のエルゴード性をもつ。このとき  $\pi$  に関してほとんどすべての  $x$  について次が成り立つ。

$$n \|P^n(x, \cdot) - \pi\| \rightarrow 0.$$

- (2) 幾何エルゴード性、幾何収束. 不変分布  $\pi$  をもつエルゴード的なマルコフ連鎖は、非負の  $\pi$  可積分な拡張された実数値関数  $M(x)$  と正の定数  $r < 1$  が存在して、すべての  $x$  について次が成り立つとき幾何エルゴード性をもつ、幾何収束をするという。

$$\|P^n(x, \cdot) - \pi\| \leq M(x)r^n$$

- (3) 一様エルゴード性. 不変分布  $\pi$  をもつエルゴード的なマルコフ連鎖は、正の定数  $M$  と正の定数  $r < 1$  が存在して、すべての  $x$  について次が成り立つとき一様エルゴード性をもつという。

$$\sup_{x \in E} \|P^n(x, \cdot) - \pi\| \leq Mr^n$$

一般に二次のエルゴード性の条件を確かめるのは難しいので、通常、幾何エルゴード性か一様エルゴード性の条件について調べられている。

### 7.5.2 収束速度を証明する道具

収束の速度を証明するのに必ずといっていいほど用いられる、幾何エルゴード性や一様エルゴード性の十分条件が minorization 条件と drift 条件である。



- [Minorization 条件]  $\pi$ -既約な  $P$  は、整数  $m \geq 1$ , 定数  $\beta > 0$ , 集合  $C \in \mathcal{E}$ ,  $\mathcal{E}$  上の確率測度  $\nu$  について

$$\pi(C) > 0, \quad \beta\nu(\cdot) \leq P^m(x, \cdot) \text{ for all } x \in C$$

を満たすとき、minorization 条件  $M(m, \beta, C, \nu)$  を満たすという。また、もし  $P$  がある  $m, \beta, \nu$  について minorization 条件  $M(m, \beta, C, \nu)$  を満たすとき、集合  $C$  は  $P$  の small set であるという。

minorization 条件は初期値  $x_0 \in C$  からマルコフ連鎖を  $m$  回走らせると、 $X_{m+1}$  を確率  $\beta$  で  $\nu$  によって発生する (従って  $X_m$  には依存しない) ことができること、そしてその場合初期値に依存しなくなることを意味する (Athreya and Ney, 1976, Robert and Casella, 1999 の 149 頁)。

- [Drift 条件]  $X_n$  がエルゴード的であるとき、非負の実数値  $\mathcal{E}$ -可測関数  $g$ , small set  $C$ , 定数  $r > 1$ , 整数  $m \geq 1$  が存在して

$$\begin{aligned} \sup_{x \in C^c} E [rg(X_{n+m}) - g(X_n) | X_n = x] &= \sup_{C^c} (rP^m g - g) < 0 \\ \sup_{x \in C} E [g(X_{n+m}); X_{n+m} \in C^c | X_n = x] &= \sup_C P^m(1_{C^c} g) < \infty \end{aligned}$$

であるとき drift 条件を満たすという。

drift 条件では  $X_n$  が状態空間の中心部  $C$  から離れたときには (エネルギー関数)  $g$  を小さくするように動いていくと解釈できる。

drift 条件は  $g$  を  $g : E \rightarrow [1, \infty]$  である実数値関数、 $C$  を small set とし、 $E(g(X_1) | X_0 = x) \leq \lambda g(x) + bI[x \in C]$  ( $\lambda < 1, b < \infty$ ) というように表現することも多い (Meyn and Tweedie, 1993)。

命題 7.5.1 (Tierney, 1994 の Proposition 1. Chan, 1993 の Theorem 2.1).  $X_n$  がエルゴード的で drift 条件を満たすとき、 $X_n$  は幾何エルゴード性をもつ。もし  $g$  が有界であるならばすべての  $x$  に対して  $\|P^n(x, \cdot) - \pi\| \leq \rho^n(a + bg(x))$  となるような  $a, b, 0 < \rho < 1$  が存在し、 $X_n$  は一様エルゴード性をもつ。

ギブスサンプラーへの応用例は Chan (1993) を参照するとよい。幾何エルゴード性における収束速度の上限については、Meyn and Tweedie (1994), Mengersen and Tweedie (1996), Rosenthal (1995a, 1995b, 1996a) などが一般的な条件のもとで導いているが、実用の目的には十分とはいえない (例外的に Rosenthal, 1996b が James-Stein 推定量に関連するモデルの中で意味のある上限を示している)。一様エルゴード性における収束速度の上限については、次のような結果が得られている。

定理 7.5.1 (Tierney, 1994 の Proposition 2. Mengersen and Tweedie, 1996 の Theorem 1.3). マルコフ推移核  $P$  が一様エルゴード性をもつための必要十分条件は状態空間  $E$  全体が small set であることである。さらに  $P$  が minorization 条件  $M(m, \beta, E, \nu)$  を満たすならば、

$$\|P^n(x, \cdot) - \pi\| \leq (1 - \beta)^{[n/m]}$$

である (必要十分条件の別の表現は Mengersen and Tweedie, 1996 を参照)。

### 7.5.3 M-H アルゴリズムの収束速度の証明

独立連鎖の  $P_{MH}$  については

定理 7.5.2 (Tierney, 1994 の Corollary 4. Mengersen and Tweedie, 1996 の Theorem 2.1.) 独立連鎖の  $P_{MH}$  は、ある  $\beta > 0$  が存在して

$$\frac{f(y)}{\pi(y)} \geq \beta, \quad y \in E$$

ならば  $P_{MH}$  は一様エルゴード性をもち、

$$\|P_{MH}^n(x \cdot) - \pi\| \leq (1 - \beta)^{[n]}$$

である。

この他、Mengersen and Tweedie (1996) は  $q(x, y) = q(x - y) = q(y - x)$  であり状態空間が  $R$  の  $P_{MH}$  は一様エルゴード性をもたない (Theorem 3.1) が、ある条件のもとで幾何エルゴード性をもつことを証明し (Theorem 3.2)、状態空間が  $R^n$  の場合への拡張を  $\pi$  が指数型分布族に属する場合に Roberts and Tweedie (1996) で行っている (Theorem 2.1)。

### 7.5.4 ギブスサンプラーの収束速度の証明

Geman and Geman (1984) は有限な状態空間について、均衡分布が正符号条件 (Besag, 1974) を満たすときギブスサンプラーが幾何収束をすることを証明した。より一般的な場合については Chan (1993) において幾何エルゴード性のための条件 (Theorem 2.1) が導出されており、ギブスサンプラーへの応用例も豊富に与えられている。Roberts and Polson (1994) は定理 7.5.1 の結果や (Lemma2)、ギブスサンプラーの収束を調べやすくするためにコンパクト性と連続性による条件を示している (Theorem 1, Corollary 2, 3)。また Shervish and Carlin (1992) は測度  $\mu(A) = \int_A (1/\pi(y)) d\lambda^p(y)$ , (ただし  $\lambda^p$  は  $\lambda^p = \lambda_1 \times \dots \times \lambda_p$  である直積測度) を定義し、この測度に関して二乗可積分な関数のヒルベルト空間  $\mathcal{H}$  を用いて、幾何エルゴード性を導いた (この結果は Baxter and Rosenthal (1994) で一般の  $P$  のために拡張された)。同様な結果を Liu, Wong and Kong (1995) も証明している。

### 7.5.5 組み合わせアルゴリズムの収束速度

最後に、同じ不変分布をもつ推移核を組み合わせた場合、先に説明した混合型では推移核の一つが一様エルゴード性をもてば全体も一様エルゴード性をもつものに対して、循環型では推移核の一つが minorization 条件  $M(1, \beta, E, \nu)$  をある  $\beta$  と  $\nu$  について満たすとき全体も一様エルゴード性をもつことが知られている (Tierney, 1994 の Proposition 3&4)。たとえば、組み合わせの中に重み関数  $w = \pi/f$  が有界な独立連鎖の M-H 核があれば全体として一様エルゴード性をもつ。

## 7.6 関数の期待値の収束

マルコフ連鎖モンテカルロ法を用いて  $\pi$  可積分な関数  $f$  の期待値  $E[f(x)]$  を求める場合、その推定量は

$$\hat{f}_n = \frac{1}{n} \sum_{i=1}^n f(x_i)$$

となる。エルゴード性をもつ系列を使った推定量の漸近的な性質は大数の法則に基づいて次のように得られている。

定理 7.6.1 (Roberts and Smith, 1994 の Theorem 1). 推移核  $P$  が  $\pi$ -既約で非周期的であり実関数  $f$  が  $E|f(X)| < \infty$  を満たすならば、確率 1 で  $\hat{f}_n \rightarrow E[f(X)]$  である。

定理 7.6.2 (Tierney, 1994 の Theorem 3).  $X_n$  がエルゴード的で均衡分布  $\pi$  を持ち、実関数  $f$  が  $E|f(X)| < \infty$  を満たすならば、初期分布にかかわらず確率 1 で  $\hat{f}_n \rightarrow E[f(X)]$  である。

定理 7.6.3 (Tierney, 1994 の Theorem 4).  $X_n$  が 2 次エルゴード的で均衡分布  $\pi$  を持ち、実関数  $f$  が有界ならば実数  $\sigma(f)$  が存在して、初期分布にかかわらず  $\sqrt{n}(\hat{f}_n - E[f(X)]) \Rightarrow N(0, \sigma^2(f))$  である。

定理 7.6.4 (Tierney, 1994 の Theorem 5).  $X_n$  が一様エルゴード的で均衡分布  $\pi$  を持ち、実関数  $f$  が  $E|f(X)|^2 < \infty$  を満たすならば実数  $\sigma(f)$  が存在して、初期分布にかかわらず  $\sqrt{n}(\hat{f}_n - E[f(X)]) \Rightarrow N(0, \sigma^2(f))$  である。

## 8 マルコフ連鎖モンテカルロ法を用いたベイズ推論

### 8.1 事後分布

マルコフ連鎖モンテカルロ法によって得られた標本に基づいてベイズ推論を行うことができる。各パラメータの周辺事後分布は標本のヒストグラムに基づいて描くことができるし、事後分布の平均や標準偏差は、得られた標本の平均、標準偏差として求めることができる。また信用区間 (credible interval) も標本の順序統計量を用いて、95% 信用区間ならば (2.5% 点, 97.5% 点) とすればよい。

また事後平均の標準誤差を求めるには、得られた標本が時系列であることに注意して、ウィンドウを用いた時系列の標本平均の分散推定量を求める式で計算する。あるいはマルコフ連鎖を  $n$  回反復したときに  $X_i$  と  $X_{i+n}$  が近似的に独立であるように  $n$  を十分大きくとって  $N = nm$  回の反復を行い、 $X$  の関数  $f(X)$  の標本平均  $\bar{f} = \sum_{i=1}^N f(X_i)/N$  を計算し、その分散を  $\hat{\sigma}_f^2/N$ 、ただし

$$\hat{\sigma}_f^2 = \frac{n}{m-1} \sum_{j=1}^m (\bar{f}_j - \bar{f})^2, \quad \bar{f}_j = \frac{1}{n} \sum_{i=(j-1)n+1}^{jn} f(X_i)$$

とすればよい (batch mean による方法)。

### 8.2 モデルの選択

モデル選択の基準としては周辺尤度 (marginal likelihood) やその比であるベイズ比 (Bayes Factor) がよく使われる。もちろん、事前分布が非報知事前分布などで積分できないような場合には、それらの基準を定義することができないので他のモデル選択の基準を用いることになるが、それらについては詳しくは Kass and Raftery (1995), Berger and Pericchi (1996), Berger and Mortera (1999)などを参照

されたい。ここでは事前分布が積分可能な分布であることを仮定し、周辺尤度やベイズ比の計算方法について紹介する。

まず候補となるモデルが  $M_1, \dots, M_K$  あり、その事前分布を  $\pi(M_i)$  とおく。  $y$  を観測値とし、  $\theta_i, \pi(y|\theta, M_i), \pi(\theta|M_i)$  をそれぞれモデル  $M_i$  におけるパラメータ、  $y$  の確率密度関数、パラメータの事前分布とする。すると  $M_i$  における周辺尤度は

$$\pi(y|M_i) = \int \pi(y|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i$$

となる。またモデル  $M_i$  の  $M_j$  に対するベイズ比  $B_{ij}$  は

$$B_{ij} = \frac{\pi(y|M_i)}{\pi(y|M_j)} = \frac{\pi(M_i|y)/\pi(M_i)}{\pi(M_j|y)/\pi(M_j)}$$

と定義される。

### 8.2.1 重点的サンプリングによる推定

ここではモデル  $M_i$  で  $\theta_i$  のサンプリングを行うことだけを考えることとし、すべての確率密度関数は  $M_i$  の条件付けとする。また表記を簡単化するために  $M_i$  は条件に表記せず  $\theta_i$  も  $\theta$  とおく。周辺尤度  $m(y) = \pi(y|M)$  のもっとも単純な推定値はモンテカルロ積分を用いて

$$\hat{m}_{MC}(y) = \frac{1}{n} \sum_{j=1}^n \pi(y|\theta^{(j)}), \quad \theta^{(1)}, \dots, \theta^{(n)} \sim \text{i.i.d. } \pi(\theta)$$

である。しかし、事前分布  $\pi(\theta_i)$  と尤度関数  $\pi(y|\theta)$  の形状が異なる場合にはこの推定量は精度が悪いことが指摘されている (Raftery, 1996)。特に尤度関数は事前分布より狭い領域に集中していることが多いので、事前分布から発生した極端に大きな  $\theta$  や小さな  $\theta$  の値に影響されて不安定になりやすい。そこで重点的サンプリングの考え方を適用して

$$m(y) = \int \frac{\pi(y|\theta)\pi(\theta)}{g(\theta)}g(\theta)d\theta$$

より  $g(\theta)$  から  $\theta^{(1)}, \dots, \theta^{(n)}$  を発生させて

$$\hat{m}_{IS}(y) = \frac{1}{n} \sum_{j=1}^n \frac{\pi(y|\theta^{(j)})\pi(\theta^{(j)})}{g(\theta^{(j)})}$$

を得る。もし  $g(\theta) = \pi(\theta)$  のとき  $\hat{m}_{IS}(y) = \hat{m}_{MC}(y)$  となる。もし  $g(\theta) = \pi(\theta|y) = \pi(y|\theta)\pi(\theta)/m(y)$  とおくと基準化定数  $m(y)$  がわからないので結局元に戻ってしまう。そこで Gelfand and Dey (1994) は

$$\begin{aligned} \frac{1}{m(y)} &= \int \frac{g(\theta)}{m(y)}d\theta = \int \frac{g(\theta)}{m(y)\pi(\theta|y)}\pi(\theta|y)d\theta \\ &= \int \frac{g(\theta)}{\pi(y|\theta)\pi(\theta)}\pi(\theta|y)d\theta \end{aligned}$$

であることを利用して  $\theta^{(1)}, \dots, \theta^{(n)}$  を事後分布  $\pi(\theta|y)$  からのサンプルとすると

$$\frac{1}{\hat{m}_{GD}(y)} = \frac{1}{n} \sum_{j=1}^n \frac{g(\theta^{(j)})}{\pi(y|\theta^{(j)})\pi(\theta^{(j)})}$$

を提案した。もし  $g(\theta) = \pi(\theta)$  のとき

$$\hat{m}(y) = \left[ \frac{1}{n} \sum_{j=1}^n \frac{1}{\pi(y|\theta^{(j)})} \right]^{-1}$$

となり、Newton and Raftery (1994) によって提案された尤度関数の調和平均となる。しかし Raftery (1996) による上記の推定量の比較によれば  $\pi(y|\theta^{(j)})$  に 0 に近い値のものがあると推定値が大きく変動し不安定になってしまうという問題があるので、 $n$  を十分大きくとるように (少なくとも  $n \geq 5000$ ) しなければならない。また  $\hat{m}_{GD}(y)$  の精度は  $g$  のとりかたに依存して非常に悪くなることもあるので注意が必要である。

### 8.2.2 周辺尤度の恒等式に基づく推定

重点的サンプリングの方法では  $g$  を注意深くとらなければ精度の悪い推定値になってしまうため Chib (1995) はそういった問題のない、周辺尤度の推定方法を提案した。すべての  $\theta$  に対して周辺尤度の恒等式 (basic marginal likelihood identity)

$$m(y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(\theta|y)}$$

が成り立つ。従って

$$\log m(y) = \log \pi(y|\theta) + \log \pi(\theta) - \log \pi(\theta|y)$$

となる。もし  $\theta = \theta^*$  において  $\pi(\theta^*|y)$  の推定値  $\hat{\pi}(\theta^*|y)$  が得られれば

$$\log \hat{m}(y) = \log \pi(y|\theta^*) + \log \pi(\theta^*) - \log \hat{\pi}(\theta^*|y)$$

によって安定的な周辺尤度の推定値を得ることができる。 $\theta^*$  はどの値でもよいが通常、 $\theta$  の事後平均などを使う。では  $\hat{\pi}(\theta^*|y)$  はどのように求めるのだろうか。 $\theta$  が  $p$  個のベクトルごとにサンプルされるとして  $\theta' = (\theta'_1, \dots, \theta'_p)$ ,  $\psi_{i-1} = (\theta_1, \dots, \theta_{i-1})$ ,  $\psi^{i+1} = (\theta_{i+1}, \dots, \theta_p)$  とする。すると  $\pi(\theta^*|y) = \prod_{i=1}^p \pi(\theta_i^*|y, \psi_{i-1}^*)$  であることから

$$\begin{aligned} \log \hat{\pi}(\theta|y) &= \sum_{i=1}^p \log \hat{\pi}(\theta_i^*|y, \psi_{i-1}^*), \\ \hat{\pi}(\theta_i^*|y, \psi_{i-1}^*) &= \frac{1}{n} \sum_{m=1}^n \pi(\theta_i^*|y, \psi_{i-1}^*, \psi^{i+1,(m)}), \end{aligned}$$

ただし、

$$\psi_{i-1}^* = (\theta_1^*, \dots, \theta_{i-1}^*), \quad \psi^{i+1,(m)} = (\theta_{i+1}^{(m)}, \dots, \theta_p^{(m)}),$$

である。

ところでこの方法は条件付分布の  $\pi(\theta_i^*|y, \psi_{i-1}^*, \psi^{i+1,(m)})$  の基準化定数が知られていないと使うことができない。従ってギブスサンプラーでは使うことができるが、M-H アルゴリズムでは使うことができない。そこでこの問題を解決するために Chib and Jeliazkov (2001) は以下のような M-H アルゴリズムのための推定値を提案した。M-H アルゴリズムにおいて点  $\theta$  から新しい点  $\theta^*$  に移る確率を  $\alpha(\theta, \theta^*)$  とすると

$$\alpha(\theta, \theta^*|y) = \min \left\{ \frac{\pi(y|\theta^*)\pi(\theta^*)}{\pi(y|\theta)\pi(\theta)} \frac{q(\theta^*, \theta|y)}{q(\theta, \theta^*|y)}, 1 \right\}$$

である。ここで

$$\alpha(\theta, \theta^*|y)\pi(\theta|y)q(\theta, \theta^*|y) = \alpha(\theta^*, \theta|y)\pi(\theta^*|y)q(\theta^*, \theta|y)$$

であることに注意すると

$$\pi(\theta^*|y) = \frac{\int \alpha(\theta, \theta^*|y)\pi(\theta|y)q(\theta, \theta^*|y)d\theta}{\int \alpha(\theta^*, \theta|y)q(\theta^*, \theta|y)d\theta} = \frac{E_{\pi(\theta|y)}\{\alpha(\theta, \theta^*|y)q(\theta, \theta^*|y)\}}{E_q\{\alpha(\theta^*, \theta|y)\}}$$

となるから  $\theta^{(l)} \sim \text{i.i.d. } \pi(\theta|y)$ ,  $\tilde{\theta}^{(m)} \sim \text{i.i.d. } q(\theta^*, \theta|y)$  として

$$\hat{\pi}(\theta^*|y) = \frac{\frac{1}{L} \sum_{l=1}^L \alpha(\theta^{(l)}, \theta^*|y)q(\theta^{(l)}, \theta^*|y)}{\frac{1}{M} \sum_{m=1}^M \alpha(\theta^*, \tilde{\theta}^{(m)}|y)}$$

を得る。

この結果から  $\psi_{-k}^i$  を  $\psi^i$  から  $\theta_k$  を除いた集合として

$$\begin{aligned} \theta_k^{(l)} &\sim \pi(\theta_k|y, \psi_{i-1}^*, \psi_{-k}^i), \quad k = i, \dots, p, \\ \tilde{\theta}_i^{(m)} &\sim q(\theta_i^*, \theta_i|y, \psi_{i-1}^*, \tilde{\psi}^{i+1,(m)}), \quad \tilde{\psi}^{i+1,(m)} = (\tilde{\theta}_{i+1}^{(m)}, \dots, \tilde{\theta}_p^{(m)}), \\ \tilde{\theta}_k^{(m)} &\sim \pi(\theta_k|y, \psi_i^*, \psi_{-k}^{i+1}), \quad k = i+1, \dots, p, \end{aligned}$$

を用いることにより

$$\hat{\pi}(\theta_i^*|y, \psi_{i-1}^*) = \frac{\frac{1}{L} \sum_{l=1}^L \alpha(\theta_i^{(l)}, \theta_i^*|y, \psi_{i-1}^*, \psi^{i+1,(l)})q(\theta_i^{(l)}, \theta_i^*|y, \psi_{i-1}^*, \psi^{i+1,(l)})}{\frac{1}{M} \sum_{m=1}^M \alpha(\theta_i^*, \tilde{\theta}_i^{(m)}|y, \psi_{i-1}^*, \tilde{\psi}^{i+1,(m)})}$$

が得られる。Chib and Jeliazkov (2001) による例では周辺尤度の計算結果は  $L, M$  の大きさ (5000 ~ 20000), ブロック化の方法 (いくつかのブロックに分けるか等), 提案された分布などに影響されないが、事後分布からのサンプリングの効率が悪ければ計算精度は悪くなることが示されている。

### 8.2.3 モデルとパラメータの同時推定 - リバーシブルジャンプ法

これまでモデルが与えられたもとでの計算方法をみてきたが、ここではモデルのサンプリングも含める計算方法について紹介する。モデルは  $M_j$  ( $j = 1, \dots, K$ ) で対応するパラメータは  $\theta_j$  とおく。

- (1) Carlin and Chib (1995) はギブスサンプラーによるモデルとパラメータの同時サンプリングの方法を提案した。まず尤度関数や事前分布  $\theta$  がモデルの条件付きで独立であるという仮定

$$\pi(y|\theta, M_j) = \pi(y|\theta_j, M_j), \quad \pi(\theta|M_j) = \pi(\theta_j|M_j)$$

を置いた。周辺尤度の計算には事前分布  $\pi(\theta_i|M_j)$  ( $i \neq j$ ) は不要なのでどのようにとってもよく、擬似事前分布 (pseudo-prior) といわれる (Carlin and Chib, 1995 や Godsill, 1997 では、できるだけ  $\theta_i$  事後分布に近いものを選ぶことにより、計算精度がよくなると指摘している)。このときギブスサンプラー

$$\pi(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_K, M_i, y) \propto \begin{cases} \pi(y|\theta_j, M_j)\pi(\theta_j|M_j) & i = j \text{ のとき} \\ \pi(\theta_j|M_i) & i \neq j \text{ のとき} \end{cases}$$

$$\pi(M_j|\theta, y) \propto \pi(y|\theta_j, M_j)\pi(M_j) \prod_{i=1}^K \pi(\theta_i|M_j)$$

によってサンプリングを行う。得られた  $M^{(1)}, \dots, M^{(n)}$  を用いて

$$\hat{\pi}(M_j|y) = \frac{1}{n} \sum_{i=1}^n I(M^{(i)} = M_j)$$

と推定すればよい。この方法は毎回  $\theta_i$  ( $i \neq j$ ) を擬似事前分布から発生させなければいけないので計算負荷が高いため、候補となるモデルの数  $K$  が大きいときには適していない。

- (2) Metropolised Carlin and Chib アルゴリズム。Dellaportas, Forster and Ntzoufras (1998) は (1) の方法で生じた擬似事前分布からのサンプリング負荷を減らすために M-H アルゴリズムと組み合わせて (1) を修正した。現在の状態が  $(M_i, \theta_i)$  であるとき

- (i) まずモデル選択のために提案された分布  $q_1(M_i, M_j)$  に従ってモデル  $M_j$  を選択する。
- (ii)  $i = j$  ならば (1) と同様にギブスサンプラーによりモデル  $M_i$  の新しい  $\theta_i$  を発生させる。
- (iii)  $i \neq j$  ならば (1) の擬似事前分布  $\pi(\theta_j|M_i)$  に従って候補  $\theta_j$  を発生させ、確率

$$\min \left\{ 1, \frac{\pi(y|\theta_j, M_j)\pi(\theta_j|M_j)\pi(M_j)}{\pi(y|\theta_i, M_i)\pi(\theta_i|M_i)\pi(M_i)} \times \frac{q_1(M_j, M_i)\pi(\theta_i|M_j)}{q_1(M_i, M_j)\pi(\theta_j|M_i)} \right\}$$

で採択する。

ここでは (1) と異なり 1 つの  $\theta_j$  だけを発生すればよいので計算負荷を大幅に軽減できる。

- (3) Green (1995) によるリバーシブルジャンプ (reversible jump) アルゴリズム。すべての  $\theta_j, j = 1, \dots, K$  を毎回サンプリングするのではなく、候補となるモデルをまずサンプリングしてから、そのモデルにおけるパラメータをサンプリングする。現在の状態が  $(M_i, \theta_i)$  であるとき

- (i) まずモデル選択のために提案された密度  $q_1(M_i, M_j)$  に従ってモデル  $M_j$  を選択する。
- (ii)  $i = j$  ならば通常の M-H アルゴリズムかギブスサンプラーで新しい  $\theta_i$  を発生させる。

- (iii)  $i \neq j$  ならば、まず確率変数  $u_i$  を提案された分布  $q_2(u_i|\theta_i, M_i, M_j)$  に従って発生させる。また  $g_{ij}$  を  $(\theta_i, u_i)$  から  $(\theta_j, u_j)$  への単射 ( $u_i, u_j$  は  $\dim(\theta_i)+\dim(u_i) = \dim(\theta_j)+\dim(u_j)$  を満たす) としたとき、 $(\theta_j, u_j) = g_{ij}(\theta_i, u_i)$  を確率

$$\min \left\{ 1, \frac{\pi(y|\theta_j, M_j)\pi(\theta_j|M_j)\pi(M_j)}{\pi(y|\theta_i, M_i)\pi(\theta_i|M_i)\pi(M_i)} \times \frac{q_1(M_j, M_i)q_2(u_j|\theta_j, M_j, M_i)}{q_1(M_i, M_j)q_2(u_i|\theta_i, M_i, M_j)} \times \left| \frac{\partial g_{ij}(\theta_i, u_i)}{\partial(\theta_i, u_i)} \right| \right\}$$

で採択する。

Han and Carlin (2000) では (1) ~ (3) と前述の Chib の方法を含めたいくつかの方法によるベイズ比の計算結果を比較している。(1) が最も時間がかかる方法であるということ以外は、計算時間、計算精度 (Chib の方法については精度を調べていないが)、どの方法が最もすぐれているということはずしもいえない、としている。

Green (1995) のリバーシブルジャンプ法は回帰係数の選択、時系列モデルの次数選択や変化点問題の他、混合問題 (Richardson and Green, 1997,1998), ノンパラメトリック回帰 (Denison, Mallick and Smith, 1998), グラフィカル正規モデル (Giudici and Green, 1999) やグラフィカル対数線形モデル (Dellaportas and Foster, 1999), 対象識別 (Pievatolo and Green, 1998, Rue and Hurn, 1999), 要因実験 (Nobile and Green, 2000), switching diffusion model (Liechty and Roberts, 2001), 不均一分散の時系列モデルの選択 (Vrontos, Dellaportas and Politis, 2000) などにも用いられている。リバーシブルジャンプ MCMC のための収束の判定方法として Brooks and Giudici (1999) ではグラフィカル正規モデルを例に用いて Gelman and Rubin (1992) と Brooks and Gelman (1998) を拡張した方法を提案している。

またリバーシブルジャンプ MCMC の代替的な方法として Stephens (2001) による birth-death MCMC がある。異なるモデルへの移行を birth-death process によって行う方法で、計算負荷はリバーシブルジャンプ法とあまり変わらないが複雑なヤコビアン計算は不要である。

モデル選択の方法として他には、回帰係数の選択のための確率的探索変数選択法 (stochastic search variable selection, George and McCulloch, 1993) や Bayesian model averaging (Madigan and Raftery, 1994, Madigan and York, 1995) などがあるが、モデル候補が多い場合には実行が難しくなる。詳しくは Chen, Shao and Ibrahim (2000) 等を参照されたい。

## 9 おわりに

現在、マルコフ連鎖モンテカルロ法の応用や理論の進展にはめざましいものがあり、本稿では基本的な考え方と最近の展開について紹介した。さまざまな分野における応用の紹介は紙数の制約上できなかったのも、また別の機会としたい。

マルコフ連鎖モンテカルロ法の実際の計算では Fortran, C 言語や OX, GAUSS, MATLAB, S-Plus, R などの行列言語を用いることが多いが、専用のパッケージも開発されている。もっとも有名なソフトウェアは BUGS (Bayesian inference Using Gibbs Sampling) で、イギリスのケンブリッジにある Institute of Public Health の MRC Biostatistics Unit により開発されたシステムである (現在無料で提供されており <http://www.mrc-bsu.cam.ac.uk/bugs> からダウンロードすることができる)。BUGS は行列言語 (S-Plus 言語の文法に従っている) で書かれたプログラムを実行し、事後分布の平均とその標準誤



差、標準偏差、95% 信用区間のほか、標本経路、事後密度関数のカーネル推定値、標本自己相関関数などを出力してくれる。詳しい収束の判定は得られた標本を一旦ファイルに出力してから CODA (Convergence Diagnostic and Output Analysis Software) や BOA (Bayesian Output Analysis) というプログラムで行う。BUGS は現在 1 次元の条件付分布からしか乱数を発生できないので、標本の自己相関が高くしてブロック化を必要とするようなサンプリングはまだできない。

CODA は MRC Unit で Best, Cowles and Vines (1995) によって開発された、BUGS のアウトプットを用いて MCMC の収束を判定する S-Plus の関数群であり (S-Plus version 4.0 まで対応)、CODA という関数を起動するとさまざまなメニューが現れて収束の診断ができるようになっている。BUGS のアウトプットはテキスト形式なので同様なフォーマットのファイルを作れば CODA を用いることができる。CODA は具体的に Geweke, Gelman and Rubin, Raftery and Lewis, Heidelberger and Welch の診断方法の他、標本自己相関関数、標本経路、事後密度関数、事後平均、標準誤差、95% 信用区間を出力する。最近では R の関数群としても用意されている。また BOA は CODA をもとに作られており、boa.menu という関数を起動して CODA とほぼ同じ機能を提供するが、R と S-Plus5.0 にも対応している。CODA と BOA はともに上記の MRC のサイトからダウンロードすることができる。

このほか、ミュンヘン大学の Lang and Brezger によって開発された Bayes X というソフトウェアがある (無料で配布されており、<http://www.stat.uni-muenchen.de/~lang/bayesx/bayesx.html> からダウンロードすることができる)。regress などの簡単なコマンドで書かれたプログラムによって回帰分析、ノンパラメトリック回帰、プロビット・ロジットモデル、多項プロビットモデル、ポアソン回帰などの分析を行い、パラメータの事後分布の平均、標準偏差、90% 信用区間を出力するほか、計算結果をプロットするための S-Plus 関数が付属している。

最後に参考文献をいくつかあげておく。基礎的なテキストとしては Gamerman (1997) や Gilks, Richardson and Spiegelhalter (1996), Gelman, Carlin, Stern and Rubin (1995) がよい。Robert and Casella (1999) では収束の理論が詳細に解説されており、Chen, Shao and Ibrahim (2000) では最近開発された手法を中心に解説がなされている。時系列分析への応用を中心としたテキストには Bauwens, Lubrano and Richard (1999) があり、順序データへ応用したテキストには Johnson and Albert (1999) がある。

また日本語の文献は少ないが、繁樹 (1995) に基礎的な手法やギブスサンプラーについての説明があり、伊庭 (1996) はマルコフ連鎖モンテカルロ法の統計学への応用について紹介している。和合 (1998) では計量経済学における応用を中心としたマルコフ連鎖モンテカルロ法及びベイズ統計学の最近の発展について包括的なサーベイが行われていて、大変有用である。また確率的ボラティリティモデルを主としたファイナンスへの応用をまとめた渡部 (2000) は我が国ではまだ数少ない貴重なマルコフ連鎖モンテカルロ法の応用研究書である。今後、日本においても多くの研究が進んでいくことと思われる。

## 参考文献

- Athreya, K. B. and Ney, P. (1978), "A new approach to the limit theory of recurrent Markov chains," *Transactions of the American Mathematical Society*, 245, 493-501.
- Bauwens, L. Lubrano, M. and Richard, J.-F. (1999), *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.
- Baxter, J. R. and Rosenthal, J. S. (1994), "Rates of convergence for everywhere-positive Markov chains", *Statistics and Probability Letters*, 22, 333-338.

- Berger, J. O. and Mortera, J. (1999), "Default Bayes factors of nonnested hypothesis testing," *Journal of the American Statistical Association*, 94, 542-554.
- Berger, J. O. and Pericchi, L. R. (1996), "The intrinsic Bayes Factor for linear models," in *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.25-44. Oxford: Oxford University Press.
- Besag, J. (2000), "Markov chain Monte Carlo for statistical inference," Working paper, No.9, Center for Statistics and the Social Sciences, University of Washington.
- Best, N. G., Cowles, M. K. and Vines, S. K. (1995), *CODA: Convergence Diagnostics and Output Analysis for Gibbs Sampler Output: Version 0.3*. Technical report, Biostatistics Unit-MRC, Cambridge, UK.
- Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975), *Discrete Multivariate Analysis*. Cambridge, Mass.:MIT Press.
- Brooks, S. P. and Gelman, A. (1998), "Alternative methods for monitoring convergence of iterative simulations," in *Journal of Computational and Graphical Statistics*, 7, 434-455.
- Brooks, S. P. and Giudici, P. (1999), "Convergence assessment for reversible jump MCMC simulations," in *Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.733-742. Oxford: Oxford University Press.
- Brooks, S. P. and Roberts, G. O. (1999), "On quantile estimation and Markov chain Monte Carlo convergence," *Biometrika*, 86, 710-717.
- Brooks, S. P. and Roberts, G. O. (1999), "Assessing convergence of Markov chain Monte Carlo algorithms," *Statistics and Computing*, 8, 319-335.
- Carlin, B. P. and Chib, S. (1995), "Bayesian model choice via Markov chain Monte Carlo methods," *Journal of Royal Statistical Society, Ser. B*, 57, 473-484.
- Carter, C. K. and Kohn, R. (1994), "On Gibbs sampling for state space models", *Biometrika*, 81, 541-543.
- Chan, K.S. (1993), "Asymptotic behavior of the Gibbs sampler", *Journal of American Statistical Association*, 88, 320-326.
- Chan, K.S., and Geyer, C. J. (1994), Comment on "Markov Chains for exploring posterior distributions (with discussion)", by Tierney, L., *Annals of Statistics*, 22, 1747-1758.
- Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2000), *Monte Carlo Methods in Bayesian Computation*. Springer: New York.
- Chib, S. (1995), "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313-1321.
- Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270-281.
- Cowles, M. K. and Carlin, B. P. (1996), "Markov chain Monte Carlo convergence diagnostics: A comparative review," *Journal of the American Statistical Association*, 91, 883-904.
- Cowles, M. K., Roberts, G. O. and Rosenthal, J. S. (1999), "Possible bias induced by MCMC convergence diagnostics," *Journal of Statistical Computation and Simulation*, 64, 87-104.
- Creutz, M. (1979), "Confinement and the critical dimensionality of space-time, *Physics Review Letters*, 43, 553-556.
- Damien, P., Wakefield, J. and Walker, S. (1999), "Gibbs sampling for non-conjugate and hierarchical models by using auxiliary variables," *Journal of Royal Statistical Society, Ser. B*, 61, 331-344.

- Dellaportas, P. and Forster, J. J. (1999), "Markov chain Monte Carlo model determination for hierarchical and graphical log-linear model," *Biometrika*, 86, 615-634.
- Dellaportas, P., Forster, J. J. and Ntzoufras, I. (1998), "On Bayesian model and variable selection using MCMC," Technical report, Department of Statistics, Athens University of Economics and Business, Greece.
- Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998), "Automatic Bayesian curve fitting," *Journal of Royal Statistical Society, Ser. B*, 60, 333-350.
- Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall: London.
- Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398-409.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995), *Bayesian Data Analysis*. Chapman and Hall: London.
- Gelman, A., Gilks, W. R. and Roberts, G. O. (1996), "Efficient Metropolis jumping rules", in *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.599-608. Oxford: Oxford University Press.
- Gelman, A. and Rubin, D.B. (1992), "Inference from iterative simulation using multiple sequences (with discussion) ", *Statistical Science*, 7, 457-511.
- Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *Institute of Electrical and Electronics Engineers, Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- George, E.I. and McCulloch, R. E. (1993), "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88, 881-889.
- Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 57, 1317-1339.
- Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.169-193. Oxford: Oxford University Press.
- Geyer, C. J. (1994), "On the convergence of Monte Carlo maximum likelihood calculations", *Journal of Royal Statistical Society, Ser. B*, 56, 261-274.
- Gibbs, A. L. (2000), "Bounding the convergence time of the Gibbs sampler in Bayesian image restoration," *Biometrika*, 87, 749-766.
- Gilks, W. R., Roberts, G. O. and George, E. I. (1994), "Adaptive direction sampling," *Statistician*, 43, 179-189.
- Gilks, W. R., Roberts, G. O. and Sahu, S. K. (1998), "Adaptive Markov chain Monte Carlo through regeneration," *Journal of the American Statistical Association*, 93, 1045-1054.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London.
- Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling", *Applied Statistics*, 41, 337-348.
- Giudici, P. and Green, P. J. (1999), "Decomposable graphical Gaussian model determination," *Biometrika*, 86, 785-801.

- Godsill, S. J. (1997), "On the relationship between MCMC model uncertainty methods," Technical report CUED/F-INFENG/TR.305, Signal Processing Group, Cambridge University Engineering Department.
- Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711-732.
- Grenander, U. (1983), "Tutorial in pattern theory," Report, Division of Applied Mathematics, Brown University.
- Grenander, U. and Miller, M. I. (1994), "Representations of knowledge in complex systems," *Journal of Royal Statistical Society, Ser. B*, 56, 549-581.
- Han, C. and Carlin, P. C. (2000), "MCMC methods for computing Bayes factors: a comparative review," Technical report, Division of Biostatistics, School of Public Health, University of Minnesota.
- Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, 57, 97-109.
- Heidelberger, P. and Welch, P. D. (1983), "Simulation run length control in the presence of an initial transient", *Operations Research*, 31, 1109-1144.
- Johnson, V. E. and Albert, J. H. (1999), *Ordinal Data Modeling*. Springer: New York.
- Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.
- Kendall, W. S. (1998), "Perfect simulation for the area-interaction point process," in *Probability Towards 2000*, Ed. C. C. Heyde and L. Accardi, pp.238-234. Springer-Verlag: New York.
- Liechty, J. C. and Roberts, G. O. (2001), "Markov chain Monte Carlo methods for switching diffusion models," *Biometrika*, 88, 299-315.
- Liu, J. S., Liang, F. and Wong, W. H. (2000), "The multiple-try method and local optimization in Metropolis sampling," *Journal of the American Statistical Association*, 95, 121-134.
- Liu, C. and Liu, J. S. (1993), "Comment on Markov Chain Monte Carlo", *Journal of Royal Statistical Society, Ser. B*, 55, 82-83.
- Liu, J. S. and Sabatti, C. (2000), "Generalized Gibbs sampler and multi-grid Monte Carlo for Bayesian computation," *Biometrika*, 87, 353-370.
- Liu, J. S., Wong, W. H. and Kong, A. (1995), "Covariance structure and convergence rate of the Gibbs sampler with various scans", *Journal of Royal Statistical Society, Ser. B*, 57, 157-169.
- Liu, J. S. and Wu, Y. (1999), "Parameter expansion scheme for data augmentation," *Journal of the American Statistical Association*, 94, 1264-1274.
- Madigan, D. and Raftery, A. E. (1994), "Model selection and accounting for model uncertainty in graphical model using Occam's window," *Journal of the American Statistical Association*, 89, 1535-1546.
- Madigan, D. and York, J. (1995), "Bayesian graphical models for discrete data," *International Statistical Review*, 63, 215-232.
- Meyn, S. P. and Tweedie, R. L. (1994), "Computable bounds for convergence rates of Markov chains", *Annals of Applied Probability*, 4, 981-1011.
- Mengersen, K. L., and Tweedie, R. L. (1996), "Rates of convergence of the Hastings and Metropolis algorithms", *Annals of Statistics*, 24, 101-121.

- Mengersen, K. L., Robert, C. P. and Guihenneuc-Jouyaux, C. (1999), "MCMC convergence diagnostics: a review", in *Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.415-440. Oxford: Oxford University Press.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, 21, 1087-1091.
- Meyn, S. P. and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*. Springer-Verlag: New York.
- Müller, P. (1991), "A generic approach to posterior integration and Gibbs sampling," Technical report # 91-09, Institute of Statistics and Decision Sciences, Duke University.
- Murdoch, D. J. and Green, P. J. (1998), "Exact sampling from a continuous state space," *Scandinavian Journal of Statistics*, 25, 451-482.
- Mykland, P., Tierney, L. and Yu, B. (1995), "Regeneration in Markov chain samplers", *Journal of American Statistical Association*, 90, 233-241.
- Nobile, A. and Green, P. J. (2000), "Bayesian analysis of factorial experiments by mixture modelling," *Biometrika*, 87, 15-36.
- Newton, M. A. and Raftery, A. E. (1994), "Approximate Bayesian inference with the weighted likelihood bootstrap," *Journal of Royal Statistical Society, Ser. B*, 56, 3-26.
- Nummelin, E. (1984), *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge, U.K.:Cambridge University Press.
- Peskun, P. H. (1973), "Optimum Monte-Carlo sampling using Markov chains", *Biometrika*, 60, 607-612.
- Pievatolo, A. and Green, P. J. (1998), "Boundary detection through dynamic polygons," *Journal of Royal Statistical Society, Ser. B*, 60, 609-626.
- Phillips, D. B. and Smith, A. F. M. (1996), "Bayesian model comparison via jump diffusions," in *Markov chain Monte Carlo in Practice*, Ed. W. R. Gilks, S. T. Richardson and D. J. Spiegelhalter, pp.215-240. Chapman & Hall: London.
- Polson, N. G. (1996), "Convergence of Markov chain Monte Carlo algorithms (with discussion)," in *Bayesian Statistics 5*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.297-321. Oxford: Oxford University Press.
- Propp, J. G. and Wilson, B. M. (1996), "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms*, 9, 223-252.
- Raftery, A. E. (1986), "A Note on Bayes factors for log-linear contingency tables with vague prior information", *Journal of Royal Statistical Society, Ser. B*, 48, 249-250.
- Raftery, A. E. (1996), "Hypothesis testing and model selection", in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, Chapman & Hall: London, pp.165-187.
- Raftery, A. E. and Lewis, S. (1992a), "How many iterations in the Gibbs sampler?" in *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.763-773. Oxford: Oxford University Press.
- Raftery, A. E. and Lewis, S. (1992b), "The number of iterations, convergence diagnostics and generic Metropolis algorithms", Unpublished manuscript, Department of Statistics, University of Washington.
- Raftery, A. E. and Lewis, S. (1996), "Implementing MCMC", in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, Chapman & Hall: London, pp.165-187.

- Richardson, S. and Green, P. J. (1997), “ On Bayesian analysis of mixtures with an unknown number of components,” *Journal of the Royal Statistical Society, Ser. B*, 59, 731-792. (with discussion).
- Richardson, S. and Green, P. J. (1998), “Corrigendum: On Bayesian analysis of mixtures with an unknown number of components,” *Journal of the Royal Statistical Society, Ser. B*, 60, 661.
- Ripley, B. D. (1979), “Algorithm AS 137: simulating spatial patterns: dependent samples from a multivariate density,” *Applied Statistics*, 28, 109-112.
- Ripley, B. (1987), *Stochastic Simulation*, New York:Wiley.
- Ritter, C. and Tanner, M. A. (1992), “Facilitating the Gibbs sampler:The Gibbs stopper and the Griddy Gibbs sampler ”, *Journal of American Statistical Association*, 87, 861-868.
- Robert, C. P. (1994), *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer-Verlag.
- Robert, C. P. and Casella, G. (1999), *Monte Carlo Statistical Methods*. New York: Springer.
- Roberts, G. O., Gelman, A. and Gilks, W. R. (1997), “Weak convergence and optimal scaling of random walk Metropolis algorithms”, *Annals of Applied Probability*, 7, 110-120.
- Roberts, G. O. and Gilks, W. R. (1994), “Convergence of adaptive direction sampling,” *Journal of Multivariate Analysis*, 49, 287-298.
- Roberts, G. O. and Polson, N. G. (1994), “On the geometric convergence of the gibbs sampler”, *Journal of Royal Statistical Society, Ser. B*, 56, 377-384.
- Roberts, G. O. and Rosenthal, J. S. (1998a), “Markov-chain Monte Carlo: Some practical implications of theoretical results,” *Canadian Journal of Statistics*, 26, 5-20.
- Roberts, G. O. and Rosenthal, J. S. (1998b), “Optimal scaling of discrete approximation to langevin diffusions”, *Journal of Royal Statistical Society, Ser. B*, 60, 255-258.
- Roberts, G. O. and Smith, A. F. M. (1994), “Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms”, *Stochastic Processes and Their Applications*, 49, 207-216
- Roberts, G. O. and Tweedie, R. L. (1996), “Exponential convergence of Langevin distributions and their discrete approximations”, *Bernoulli*, 2, 341-363.
- Roberts, G. O. and Tweedie, R. L. (1996), “Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms”, *Biometrika*, 83, 95-110.
- Rosenthal, J. S. (1993), “Rates fo convergence for data augmentation on finite sample spaces”, *Annals of Applied Probability*, 3, 319-339.
- Rosenthal, J. S. (1995a), “Minorization conditions and convergence rates for Markov chain Monte Carlo”, *Journal of American Statistical Association*, 90, 558-566.
- Rosenthal, J. S. (1995b), “Rates of convergence for Gibbs sampling for variance component models”, *Annals of Statistics*, 23, 740-761.
- Rosenthal, J. S. (1996a), “Markov chain convergence: From finite to infinite”, *Stochastic Processes and their Applications*, 62, 55-72.
- Rosenthal, J. S. (1996b), “Analysis of the Gibbs sampler for a model related to James-Stein estimators”, *Statistics and Computing.*, 6, 269-275.
- Rubin, D.B. (1987a), Comment on “The Calculation of posterior distributions by data augmentation (with discussion) ”, by Tanner, M. A. and Wong, W. H., *Journal of American Statistical Association*, 82, 543-546.
- Rubin, D.B. (1987b), *Multiple Imputation for Non-response in Surveys*, New York:Wiley.

- Rubin, D.B. (1988), "Using the SIR Algorithm to simulate posterior distributions", in *Bayesian Statistics 3*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp.395-402. Oxford: Oxford University Press.
- Rue, H. and Hurn, M. (1999), "Bayesian object identification," *Biometrika*, 86, 649-660.
- Schervish, M. J. and Carlin, B. P. (1992), "On the convergence of successive substitution sampling", *Journal of Computational and Graphical Statistics*, 1, 111-127.
- Schmeiser, B. and Chen, M. H. (1991), "General hit-and-run Monte Carlo sampling for evaluating multidimensional integrals," Technical report, School of Industrial Engineering, Purdue University.
- Shephard, N. and Pitt, M. K. (1997), "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 84, 653-667.
- Smith, A. F. M. and Gelfand, A. E. (1992), "Bayesian statistics without tears", *American Statistician*, 46, 84-88.
- Smith, A. F. M. and Roberts, G. O. (1993), "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods", *Journal of Royal Statistical Society, Ser. B*, 55, 3-23.
- Stephens, M. (2001), "Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods," *Annals of Statistics*, 28, 40-74.
- Stramer, O. and Tweedie, R. L. (1997), "Geometric and subgeometric convergence of diffusions with given stationary distributions," Technical report, University of Iowa.
- Suomela, P. (1976), Unpublished Ph.D. thesis. University of Jyväskylä, Finland.
- Tanner, M. A. (1996), *Tools for Statistical Inference*, 3rd ed. Springer-Verlag.
- Tanner, M. A. and Wong, W. H. (1987), "The Calculation of posterior distributions by data augmentation (with discussion)", *Journal of American Statistical Association*, 82, 528-550.
- Tierney, L. (1994), "Markov chains for exploring posterior distributions (with discussion)", *Annals of Statistics*, 22, 1701-1762.
- Vrontos, I. D., Dellaportas, P. and Politis, D. N. (2000), "Full Bayesian inference for GARCH and EGARCH models," *Journal of Business and Economic Statistics*, 18, 187-198.
- Watanabe, T. and Omori, Y. (2001), "Multi-move sampler for estimating non-Gaussian times series models: Comments on Shephard and Pitt (1997)," Research Paper Series, No.25, Faculty of Economics, Tokyo Metropolitan University.
- 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, 49-84.
- 繁樹算男 (1995), 「意思決定の認知統計学」朝倉書店.
- 和合肇 (1998), 「ベイズ計量経済分析における最近の発展」日本統計学会誌, 28, 253-305.
- 渡部敏明 (2000), 「ポラティリティ変動モデル」朝倉書店.