

# Markov Chain Monte Carlo Simulation Methods

浅井 学

## 1 はじめに

本稿の目的は、近年、統計学の分野で注目を集め、また計量経済学への応用が増え続けているマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo, 以下 MCMC) を紹介することにある。第2章では MCMC の統計的性質について述べる。第3章では計量経済学への応用に使われる Gibbs サンプラーと Metropolis-Hastings を紹介する。第4章と第5章では、MCMC のページアン計量経済学への応用として ARMA(p, q) モデルと Stochastic volatility モデルを扱う。

## 2 マルコフ連鎖モンテカルロ法

### 2.1 マルコフ連鎖の概要

集合  $\Omega \subseteq R^p$  について部分集合の  $\sigma$ -algebra を  $\mathcal{F}$  とする。特に  $\mathcal{F}$  は  $E$  の部分集合の可算和 (countable collection) によって生成されるものとする。可測空間  $(\Omega, \mathcal{F})$  上の推移核 (transition kernel)  $P: \Omega \times \mathcal{F} \rightarrow [0, 1]$  は次の性質をもつとする。

1. 任意の固定された  $A \in \mathcal{F}$  について関数  $P(\cdot, A)$  は可測である。
2. 任意の固定された  $x \in \Omega$  について関数  $P(x, \cdot)$  は  $(\Omega, \mathcal{F})$  上の確率測度である。

また  $T = \{0, 1, 2, \dots\}$  とするとき確率変数ベクトル  $\{\Phi_i: i \in T\}$  についてマルコフ性は次のように定義される。

$$P(x, A) \equiv P_r(\Phi_{i+1} \in A \mid \Phi_i = x, \Phi_j, j < i) \quad x \in \Omega, A \subset \Omega$$

これは  $\Phi_{i+1}$  の確率分布は1時点前の値にのみ依存することを意味する。このマルコフ性をもつ確率過程をマルコフ連鎖と呼ぶ。 $x$  からスタートした推移核  $P$  をもつマルコフ連鎖の確率測度を  $P_x$  で表すことにする。

$\mathcal{F}_+$  を  $\mathcal{F}$  の非負の要素のクラスとする。非負関数  $h \in \mathcal{F}_+$  について、

$$\nu P(A) = \int_{\Omega} P(x, A) \nu(dx)$$

$$Ph(x) = \int_{\Omega} h(y)P(x, dy)$$

$$\nu h = \int_{\Omega} h(y)\nu(dy)$$

と定義する。 $h=Ph$  を満たすとき  $h$  は  $P$  について調和的 (harmonic) であるという。

ある関数  $p(x, y) : \Omega \times \Omega \rightarrow R^+$  を使って,  $x$  から  $y$  への推移核が

$$P(x, dy) = p(x, y)\nu(dy) + r(x)\delta_x(dy)$$

と表されたとする。ただし

$$p(x, x) = 0$$

$$\delta_x(dy) = \begin{cases} 1 & \text{if } x \in dy \\ 0 & \text{otherwise} \end{cases}$$

$$r(x) = 1 - \int_{\Omega} p(x, y)\nu(dy)$$

$\nu \in \mathcal{M}^+$  は  $(\Omega, \mathcal{F})$  上の測度,  $\mathcal{M}^+$  は正の測度のクラスで  $\mathcal{M}^+ = \{\lambda \in \mathcal{M}^+ : \lambda(\mathcal{F}) > 0\}$  である。このとき

$$\begin{aligned} P(x, A) &= \int_A P(x, dy) \\ &= \int_A p(x, y)\nu(dy) + r(x) \int_A \delta_x(dy) \end{aligned} \tag{2}$$

であるから  $p(x, x) = 0$  と  $\delta_x(dy)$  の定義に注意すると,  $x$  から  $x$  への推移は確率  $r(x)$  で発生し,  $x$  から ( $x$  以外の)  $A$  上の点へ推移する確率は上の式の最初の項によって表される。また  $r(x) = 0$  のとき,

$$\int_{\Omega} p(x, y)\nu(dy) = 1$$

であり,  $p(x, y)$  はマルコフ連鎖の推移密度関数となる。以上の議論から推移核は  $\Phi_i = x$  が与えられたときの  $\Phi_{i+1}$  の分布関数であることがわかる。

マルコフ性により  $n$  時点先の推移核は

$$P^{(n)}(x, A) = \int_A P(x, dy)P^{(n-1)}(y, A)$$

で与えられる。ただし  $P^{(1)}(x, dy) = P(x, dy)$  である。以下の議論により示される条件のもとで推移核の  $n$  回目の反復計算結果は  $n \rightarrow \infty$  のとき定常分布 (invariant distribution)  $\pi^*$  に収束することが示される。詳しい説明は Tierney (1994), Nummelin (1984), Meyn and Tweedie (1983) を参照されたい。

## 2.2 理論的背景

$\pi$  を  $\sigma$ -finite measure  $\nu$  の下での  $\pi^*$  の密度関数 (density) とする。このとき

$$\pi^*(dy) = \pi(y) \nu(dy)$$

である。これは通常定義における  $dF(x) = f(x) dx$  を考えればイメージしやすい。 $\Phi_i$  が  $\pi^*$  に従って分布するとき、マルコフ連鎖の他の部分列の要素も  $\pi^*$  に従うとすると

$$\pi^*(dy) = \int_{\Omega} P(x, dy) \pi^*(dx)$$

であるので

$$\pi^*(dy) = \int_{\Omega} P(x, dy) \pi(x) \nu(dx) \quad (3)$$

と書ける。(3)式を平衡条件 (invariant condition) とよぶ。(1)式の  $p(x, y)$  が反転可能性条件 (reversibility condition)

$$\pi(x, y) p(x, y) = \pi(y) p(y, x) \quad (4)$$

を満たすとき、マルコフ連鎖は反転可能であるという。反転可能なマルコフ連鎖は  $P(x, \cdot)$  の定常分布として  $\pi^*(x)$  をもつ。これは平衡条件を満たしているかどうかをみればすぐわかる。

$$\begin{aligned} & \int_{\Omega} P(x, A) \pi(x) \nu(dx) \\ &= \int_{\Omega} \left[ \int_A p(x, y) \nu(dy) \right] \pi(x) \nu(dx) + \int_{\Omega} r(x) \delta_x(A) \pi(x) \nu(dx) \\ &= \int_A \left[ \int_{\Omega} p(x, y) \pi(x) \nu(dx) \right] \nu(dy) + \int_A r(x) \pi(x) \nu(dx) \\ & \quad (\text{反転可能性条件により}) \\ &= \int_A \left[ \int_{\Omega} p(y, x) \pi(y) \nu(dy) \right] \nu(dx) + \int_A r(x) \pi(x) \nu(dx) \\ &= \int_A [1 - r(y)] \pi(y) \nu(dy) + \int_A r(x) \pi(x) \nu(dx) \\ &= \int_A \pi(y) \nu(dy) \end{aligned}$$

であるから(3)式は満たされる。

次に推移核の  $m$  回目の反復計算結果の収束を保証するために重要な定義をいくつか述べる。

**Definition 2.1** ( $\pi^*$ -既約性) 条件

$$\forall x \in \Omega, \pi^*(A) > 0 \Rightarrow P_r(\Phi_i \in A \mid \Phi_0 = x) > 0 \text{ for some } i$$

が成り立つとき、マルコフ連鎖を  $\pi^*$ -既約 (irreducible) という。

この条件は、 $\pi^*$  のもとで正の確率をもつすべての集合について、 $\Omega$  上のどの点からスタートしてもその集合に到達することを保証するためのものである。

**Definition 2.2** (非周期性) すべての  $i$  について

$$P_r(\Phi_i \in D_{i \bmod(p)} \mid \Phi_0 \in D_0) = 1$$

となるような  $\Omega$  の分割

$$\Omega = (D_0, D_1, \dots, D_{p-1}) \text{ for some } p \geq 2$$

は存在しないとき、マルコフ連鎖は非周期的 (aperiodic) であるという。このような  $\Omega$  の分割が存在するときは周期的 (periodic) であるという。ただし  $i \bmod(p)$  は  $i$  を  $p$  で割ったときの剰余項を意味する。

例えば  $p=2$  のとき

$$P_r(\Phi_1 \in D_1 \mid \Phi_0 \in D_0) = 1$$

$$P_r(\Phi_2 \in D_0 \mid \Phi_0 \in D_0) = 1$$

$$P_r(\Phi_3 \in D_1 \mid \Phi_0 \in D_0) = 1$$

$$P_r(\Phi_4 \in D_0 \mid \Phi_0 \in D_0) = 1$$

.....

となるような  $\Omega$  の分割  $\Omega = (D_0, D_1)$  が存在すれば、このマルコフ連鎖は周期的である。非周期的であるかどうかを示すのはかなりやっかいであるので、強非周期性 (strong aperiodicity) を示す方が簡単である。定常確率分布  $\pi^*$  をもつ既約マルコフ連鎖は、 $\Omega$  上の確率分布  $\lambda$ , 定数  $\beta > 0$ , 集合  $C \subset \Omega$  について

$$\lambda(C) > 0 \text{ and } P(x, A) \geq \beta \lambda(A) \text{ for all } x \text{ and all } A \subset \Omega$$

を満たすような  $\lambda, \beta, C$  が存在するとき、強非周期的であるという。強非周期的であれば、非周期的である。

既約性条件は、どの点からスタートしても興味のある集合に到達することを意味する。これに対し、次に述べる再帰的 (recurrent) 性質は、ほとんどすべての初期時点について、そのような集合に無限に何度も到達することを保証する。記号  $\{A_n \text{ i. o.}\}$  は、数列  $A_n$  が無限の頻度で (infinitely often) 発生することを意味する。すなわち  $\sum 1_{A_n} = \infty$  である。

**Definition 2.3** (再帰性) 定常分布  $\pi^*$  をもつ  $\pi^*$ -既約マルコフ連鎖  $\Phi_n$  は、 $\pi^*(B) > 0$  となるそれぞれの  $B$  について

$$P_x\{\Phi_n \in B \text{ i. o.}\} > 0 \text{ for all } x$$

$P_x\{\Phi_n \in B \text{ i. o.}\} = 0$  for  $\pi^*$ -almost all  $x$  となるとき再帰的 (recurrent) であるという。さらに

$$P_x\{\Phi_n \in B \text{ i. o.}\} = 1 \text{ for all } x$$

のとき Harris 再帰的 (または  $\pi^*$  再帰的) であるという。また  $\pi^*$ -既約再帰的マルコフ連鎖は、定常確率分布関数をもつとき、正再帰的 (positive recurrent) であるといい、もたないときは零再帰的 (null recurrent) であるという。

**Proposition 2.1** [Nummelin (1984, corollary 5.2)]  $P(\cdot, \cdot)$  が  $\pi^*$ -既約であり、定常分布  $\pi^*$  を

もつとする。このとき  $P(\cdot, \cdot)$  は正再帰的であり、 $\pi^*$  はその唯一の定常分布である。

### 2.3 分布の定常状態への収束について

2つの確率分布  $\lambda_1, \lambda_2$  について総変動距離 (total variation distance) を

$$\|\lambda_1 - \lambda_2\| = 2 \sup_{A \subset \Omega} |\lambda_1(A) - \lambda_2(A)|$$

で定義する。

再帰性は推移核の平均が収束するための十分条件である。

**Proposition 2.4** [Chib and Greenberg (1996)]  $\{\Phi_n\}$  を推移核  $P$  をもつ既約マルコフ連鎖であるとする。推移核の平均  $\bar{P}^{(n)}$  を

$$\bar{P}^{(n)}(x, A) = \frac{1}{n+1} \sum_{i=0}^n P^{(i)}(x, A) \text{ for all } x \in \Omega \text{ and } A \subset \Omega$$

と定義する。このとき

$$|\bar{P}^{(m)}(x, \cdot) - \pi^*(\cdot)| \rightarrow 0 \text{ for } \pi^*\text{-almost all } x$$

である。

さらに再帰性から SLLN (strong law of large numbers) が示される。

**Proposition 2.5** [Chib and Greenberg (1996)]  $\{\Phi_n\}$  を推移核  $P$ , 定常分布  $\pi^*$  をもつ既約マルコフ連鎖であるとする。また  $\Omega$  上の実数値関数  $f$  について  $\pi^*(|f|) < \infty$  とし、サンプルパスの平均を

$$\bar{f}_n = \frac{1}{n+1} \sum_{i=0}^n f(\Phi_i)$$

と定義する。このとき

$$P_x(\bar{f}_m \rightarrow \pi^*f) = 1 \text{ for } \pi^*\text{-almost all } x$$

である。

Proposition 2.5 は  $m \rightarrow \infty$  のとき、抽出したサンプルによる関数値の平均が求める密度関数 (target density) の下での期待値に強収束することを意味している。さらに強力な収束を示すためには、非周期性を仮定しなければならない。

**Proposition 2.6** [Chib and Greenberg (1996)]  $\{\Phi_n\}$  を推移核  $P$ , 定常分布  $\pi^*$  をもつ既約非周期的マルコフ連鎖であるとする。このとき

$$\|P^{(m)}(x, \cdot) - \pi^*(\cdot)\| \rightarrow 0 \text{ for } \pi^*\text{-almost all } x$$

である。

Proposition 2.6 は、マルコフ連鎖の  $m$  回反復計算したときの密度関数は、十分大きな  $m$  につい

て唯一の定常密度関数に近づくことをいっている。これは  $m$  が十分に大きいとき、 $P^{(m)}(x, A)$  からの抽出は定常分布からの抽出となることを意味する。もしマルコフ連鎖が Harris 再帰的であれば、Proposition 2.1, 2.4, 2.5, および 2.6 はすべての初期値  $x$  について成り立つ。Harris 再帰性は既約性からそのまま導き出されるわけではない。本稿では紹介しないが、Gibbs サンプラーや MH について Harris 再帰的であるための十分条件がある。

### 3 Gibbs サンプラーと Metropolis-Hastings アルゴリズム

#### 3.1 Gibbs サンプラー

MCMC シミュレーションの1つとして Gibbs サンプリング・アルゴリズムがある。Gibbs サンプリング・アルゴリズムでは、確率変数ベクトルをいくつかのブロックに分割し、完全な条件付き密度関数 (full conditional density<sup>1)</sup>) の積として推移密度関数が定義される。

$\pi(x)$ ,  $x \in S \subseteq R^d$  を (基準化されていない) 求める密度関数とする。 $x$  を  $x_1, \dots, x_d$  に分割し、第  $K$  ブロックの完全な条件付き密度関数を  $\pi(x_k | x_{-k}) \equiv \pi(x_k | x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_d)$  と書くことにする<sup>2)</sup>。このとき Gibbs サンプリング・アルゴリズムは次の反復計算により定義される。

1. Specify starting values  $x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})$  and set  $i=0$ .

2. Simulate

$$\begin{aligned} x_1^{(i+1)} & \text{ from } \pi(x_1 | x_2^{(i)}, x_3^{(i)}, \dots, x_d^{(i)}) \\ x_2^{(i+1)} & \text{ from } \pi(x_2 | x_1^{(i+1)}, x_3^{(i)}, \dots, x_d^{(i)}) \\ x_3^{(i+1)} & \text{ from } \pi(x_3 | x_1^{(i+1)}, x_2^{(i+1)}, x_4^{(i)}, \dots, x_d^{(i)}) \\ & \vdots \\ x_d^{(i+1)} & \text{ from } \pi(x_d | x_1^{(i+1)}, x_2^{(i+1)}, \dots, x_{d-1}^{(i+1)}). \end{aligned}$$

3. Set  $i=i+1$  and go to step 2.

故にこのアルゴリズムでは、条件付けする要素をサイクルの中で生成し、それぞれの完全な条件付き密度関数によりマルコフ連鎖の次の項  $x^{(i+1)}$  をシミュレートする。後で見るように MH では同じ点への推移を考えるのであるが、Gibbs サンプラーではこれを考慮しない。ゆえに  $r(x) = 0$  であり、 $x \equiv x^{(i)}$  から  $y \equiv x^{(i+1)}$  への推移は推移密度関数

$$p_G(x, y) = \prod_{k=1}^d \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_d) \quad (5)$$

に従って発生する。この推移密度関数が平衡条件(3)を満たしていることを確認する。簡単のため  $\nu$  は Lebesgue 測度であるとする。

1) あるブロックについて、データと残りのパラメータを与えられたときの条件つき分布。

2) 完全な条件付き密度関数  $\pi(x_k | x_{-k})$  は同時密度関数  $\pi(x)$  に比例していることに注意されたい。

$$\int_{\mathcal{Q}} p_G(x, y) \pi(x) dx = \int_{\mathcal{Q}} \prod_{k=1}^d \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_d) \pi(x) dx$$

ここで

$$\begin{aligned} & \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_d) \\ &= \frac{\pi(y_1, \dots, y_{k-1}, y_k, x_{k+1}, \dots, x_d)}{\pi(y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_d)} \\ &= \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1}, y_k) \pi(y_k | y_1, \dots, y_{k-1}) \pi(y_1, \dots, y_{k-1})}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1}) \pi(y_1, \dots, y_{k-1})} \\ &= \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1}, y_k) \pi(y_k | y_1, \dots, y_{k-1})}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \end{aligned}$$

であることと,  $\pi(x) = \pi(x_1 | x_2, \dots, x_d) \pi(x_2, \dots, x_d)$  と書けるから, 結局上の式は

$$\int_{\mathcal{Q}} \prod_{k=1}^d \frac{\pi(y_k | y_1, \dots, y_{k-1}) \pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_1 | x_2, \dots, x_d) \pi(x_2, \dots, x_d) dx$$

と書ける。 $\pi(y_k | y_1, \dots, y_k)$  は  $x$  と独立であり,  $\pi(y) = \prod_{k=1}^d \pi(y_k | y_1, \dots, y_k)$  であるから

$$\pi(y) \int_{\mathcal{Q}} \prod_{k=1}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_1 | x_2, \dots, x_d) \pi(x_2, \dots, x_d) dx$$

となる。 $x$  の各ブロックについて変換を施しながら逐次積分していくと

$$\begin{aligned} & \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=1}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \\ & \quad \times \left[ \int_{\mathcal{Q}} \pi(x_1 | x_2, \dots, x_d) dx_1 \right] \pi(x_2, \dots, x_d) dx_2 \dots dx_d \\ &= \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=1}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_2, \dots, x_d) dx_2 \dots dx_d \\ &= \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=2}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_2, \dots, x_d | y_1) dx_2 \dots dx_d \\ &= \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=3}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_3, \dots, x_d | y_1, y_2) \\ & \quad \times \frac{\pi(x_2, \dots, x_d | y_1)}{\pi(x_3, \dots, x_d | y_1)} dx_2 \dots dx_d \\ &= \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=3}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_3, \dots, x_d | y_1, y_2) \\ & \quad \times \pi(x_2 | x_3, \dots, x_d) dx_2 \dots dx_d \\ &= \pi(y) \int_{\mathcal{Q}} \dots \int_{\mathcal{Q}} \prod_{k=3}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_3, \dots, x_d | y_1, y_2) \\ & \quad \times \left[ \int_{\mathcal{Q}} \pi(x_2 | x_3, \dots, x_d) dx_2 \right] dx_3 \dots dx_d \end{aligned}$$

$$\begin{aligned}
&= \pi(y) \int_{\Omega} \cdots \int_{\Omega} \prod_{k=3}^d \frac{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_k)}{\pi(x_{k+1}, \dots, x_d | y_1, \dots, y_{k-1})} \pi(x_3, \dots, x_d | y_1, y_2) dx_3 \cdots dx_d \\
&\quad \vdots \\
&= \pi(y) \int_{\Omega} \pi(x_d | y_1, \dots, y_d) dx_d \\
&= \frac{\pi(y)}{\pi(y_1, \dots, y_d)} \int_{\Omega} \pi(x_d, y_1, \dots, y_d) dx_d \\
&= \pi(y) \frac{\pi(y_1, \dots, y_d)}{\pi(y_1, \dots, y_d)} = \pi(y)
\end{aligned}$$

となり平衡条件を満たすので、この推移密度関数は定常密度関数  $\pi^*$  をもつ。

次に実際に Gibbs サンプリング・アルゴリズムを使うときの問題点に触れる。

1. 確率変数ベクトルをブロックに分割するときには、高い相関を持つものどうしを一緒のグループに入れるべきである。そうしないとマルコフ連鎖はゆっくりと減衰する自己相関をもち、結果的に求める密度関数にゆっくりと収束することになる [Liu et al. (1994) and Chib and Greenberg (1996, Section 3.4) を参照]。
2.  $x$  の定義の中に潜在データ (latent data) やデータの欠落 (missing data) を組み込むことにより、扱いやすい完全な条件付き分布を構成できる場合がある。「data augmentation」として知られる、サンプラーに変数を加えるアイデアは Tanner and Wong (1987) によって紹介された。Chib and Greenberg (1996, Section 3) ではいくつかの例を使って説明している。
3. 伝統的な方法 [例えば棄却サンプリング法 (rejection sampling) や既知の乱数生成法による方法] を使っても完全な条件付き分布からサンプルを生成するのが困難な場合は、MH アルゴリズムから生成する [Müller (1991)] か、独立なサンプルを生成する方法を使う [Gilks and Wild (1992)]。

### 3.2 Metropolis-Hastings アルゴリズム

MH アルゴリズムは、定常分布  $\pi^*$  が扱いにくいときでも使える強力な MCMC 法である。MH アルゴリズムでは次のようにサンプルを生成する。潜在分布 (latent distribution) からの抽出により  $x$  という値をとるとき、この数列 (sequence) の次の値は候補を生成する密度関数 (candidate generating density)  $q(x, y)$  (もしくは proposal density) から抽出された値  $y$  によって生成される。このとき候補となる点 (candidate point)  $y$  は確率  $\alpha(x, y)$  で採択される。ただし

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

である。候補が棄却されたときは次の値は現在の値をそのままとる。

ここで次の点に注意されたい。

1.  $\alpha(x, y)$  の計算においては基準化定数がわからなくてもよい。
2. 候補を生成する密度関数が対称すなわち  $q(x, y) = q(y, x)$  のときは採択率 (acceptance probability) は  $\pi(y) / \pi(x)$  となる。これは Metropolis et al. (1953) が提案したオリジナル

のアルゴリズムである。

MH アルゴリズムではマルコフ連鎖の推移核は次の式で与えられる。

$$P_{MH}(x, dy) = q(x, y)\alpha(x, y)dy + \left[1 - \int_{\Omega} q(x, y)\alpha(x, y)dy\right]\delta_x(dy) \quad (6)$$

これにより  $x$  から  $y$  への推移 ( $x \neq y$ ) は

$$p_{MH}(x, y) \equiv q(x, y)\alpha(x, y), \quad x \neq y$$

に従って発生する。

関数  $p_{MH}(x, y)$  が反転可能性条件(4)を満たしていることを確認する。2.2 節で示したように反転可能性条件が満たされればそのマルコフ連鎖は定常分布をもつ。 $\alpha(x, y) < 1$  のケースについて考える。このとき

$$\alpha(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} < 1$$

であるので、

$$\frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} > 1$$

であり、定義により  $\alpha(y, x) = 1$  である。ここで

$$\pi(x)p_{MH}(x, y) \equiv \pi(x)q(x, y)\alpha(x, y) = \pi(y)q(y, x)$$

最後の等式は  $\pi(y)p_{MH}(y, x)$  に等しいので、 $P_{MH}(x, dy)$  は定常分布  $\pi^*$  をもつ。

### 3.3 Proposal density $q(x, y)$ の選択

以下では、どのように MH アルゴリズムの候補を生成する密度関数  $q(x, y)$  を選ぶのか、という問題に触れる。詳しくは Tierney (1994) を参照されたい。

#### 1. ランダム・ウォーク MH アルゴリズム: $q(x, y) = q_1(y-x)$

ただし  $q_1(\cdot)$  は多変量分布である。このとき候補  $y$  は確率過程  $y = x + z$  によって生成される。ここで  $z$  は  $q_1$  に従って分布する。候補は現在の値とノイズの和に等しいのでランダム・ウォークを基にした MH 連鎖と呼ばれる。関数  $q_1$  は多変量正規分布や多変量  $t$  分布、また Geweke (1989) の Split- $t$  分布であってもよい。例えば、 $q_1$  が対称であるとする  $q_1(z) = q_1(-z)$  である。このとき

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

となる。

#### 2. 独立連鎖 (Independence chain): $q(x, y) = q_2(y)$

ランダム・ウォーク連鎖とは対照的に、候補  $y$  は現在の位置  $x$  とは独立に抽出される。

ここでも関数  $q_2$  は多変量正規分布や多変量  $t$  分布であってもよいが、採択率  $\alpha(\cdot, \cdot)$  が尤もらしい値をとるように密度関数の位置母数 (location parameter) や尺度母数 (scale parameter) を調節しなければならない。一般的に独立連鎖による方法は非常にうまくいくか、非常に悪くなるかのどちらかである [Roberts (1995) を参照]。

3. 3番目の方法として、候補を生成する密度関数を特定するのに  $\pi(\cdot)$  の形を利用する方法がある [Chib and Greenberg (1995)]。例えば  $\pi(t)$  が

$$\pi(t) \propto \phi(t) h(t)$$

と書けるとする。但し、 $h(t)$  は密度関数で、 $\phi(t)$  は一様分布である。このとき候補を生成する密度関数を  $q(x, y) = h(y)$  とおけば、採択率  $\alpha(x, y)$  は

$$\alpha(x, y) = \min \left\{ \frac{\phi(y)}{\phi(x)}, 1 \right\}$$

で与えられ、関数  $\phi(\cdot)$  を計算するだけでよい。

4. 棄却サンプリング連鎖 (Rejection sampling chain) [Tierney (1994)]  
 5. 自己回帰連鎖 (Autoregressive chain) [Tierney (1994)]

この方法は1次の自己回帰過程によって表される。

$$y = a + B(x - a) + z$$

ただし  $a$  はベクトル、 $B$  は行列、 $z$  の密度関数は  $q$  である。このとき

$$q(x, y) = q(y - a - B(x - a))$$

である。 $B = I$  のときはランダム・ウォーク連鎖になり、 $a = 0$ 、 $B = 0$  のときは独立連鎖となる。

### 3.4 単要素 MH アルゴリズム

Gibbs サンプラーによる抽出を考える。完全な条件付き分布の近似的な関数からより簡単に抽出できるときは、この近似的な関数を使ってよいのだろうか？ この近似関数を使った Gibbs サンプラーのサンプルの分布は正確には定常分布  $\pi^*$  に対応していないため、どれほど長い連鎖を作っても  $\pi^*$  には収束しない。Tierney (1994) や Gelman (1992) ではエルゴード性を保ったまま、このような近似関数を使う方法を提案している。実はこの方法は Metropolis *et al.* (1953) による単要素 (single-component) MH アルゴリズムを使っている。

単要素 MH アルゴリズムでは次のように反復計算を行なう。ベクトル  $x$  を  $\{x_1, x_2, \dots, x_d\}$  に分割する。このとき  $x_{-i} = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d\}$  とおく。 $t$  回の反復計算を行なったときの  $x_i$  の状態を  $x_{t,i}$  と書くことにする。 $t+1$  回目の反復計算の第  $i$  ステップについて、 $x_i$  は MH アルゴリズムによって生成される。つまり候補  $y_i$  は候補を生成する分布  $q_i(y_i | x_{t,i}, x_{t,-i})$  によって生成される。ただし  $x_{t,-i}$  は  $t+1$  回目の反復計算の第  $i-1$  ステップまで終えたときの  $x_i$  の状態である。すなわち

$$x_{t,-i} = \{x_{t+1,1}, \dots, x_{t+1,i-1}, x_{t+1,i+1}, \dots, x_{t,d}\}$$

このとき  $i$  番目の候補を生成する分布  $q_i(\cdot | \cdot, \cdot)$  からは  $x$  の第  $i$  要素のみが生成されることになる。候補は確率

$$\alpha_i(x_{-i}, x_i, y_i) = \min \left\{ \frac{\pi(y_i | x_{-i}) q_i(x_i | y_i, x_{-i})}{\pi(x_i | x_{-i}) q_i(y_i | x_i, x_{-i})}, 1 \right\}$$

で採択される。ここで  $\pi(x_i | x_{-i})$  は  $x_i$  の完全な条件付き分布であり

$$\pi(x_i | x_{-i}) = \frac{\pi(x)}{\int \pi(x) dx_i}$$

で定義される。 $y_i$  が採択されたときは  $x_{t+1,i} = y_i$  とおき、棄却されたときは  $x_{t+1,i} = x_{t,i}$  とおく。ステップ  $i$  において他の要素は変化しない。

### 3.5 実用上の注意点

まず単連鎖 (single chain) による方法と多重連鎖 (multiple chain) による方法とどちらを使うべきかという問題がある。多重連鎖では、まず初期値を選択し、関数  $p(x^{(i-1)}, x^{(i)})$  を使って数列を生成する。 $N_0$  回抽出した後で、定常密度関数  $\pi(\cdot)$  に収束したとみなされれば、 $N_0+1$  回目のサンプルは  $\pi(\cdot)$  から抽出したものとみなされる。そして新たな初期値が選択され、この過程を繰り返していく。この方法では独立なサンプルを抽出できるが、何度も  $N_0$  個のサンプルを生成し、捨てなければならない。

単連鎖では  $\{x^{N_0+1}, x^{N_0+2}, \dots, x^{N_0+M}\}$  を  $\pi(\cdot)$  から抽出した  $M$  個のサンプルとみなす。マルコフ性によりこれらのサンプルはそれぞれ 1 時点前のサンプルと相関を持つにも拘わらず、この数列は定常分布に収束するので、実は使える。通常、サンプル間の自己相関は急速にゼロに近づくので、時点  $t$  と  $t+n_1$  の相関がゼロとみなせるような、ある値  $n_1$  が存在する。この場合、収束した後で  $n_1$  個ごとにサンプルを抽出すれば、近似的に独立なサンプルとなる。注意しなければならない点として、多重連鎖であれば一見収束したように見える場合でもそれぞれの連鎖を比較して確かめることができるが、単連鎖ではそれはできない。どちらの連鎖を使うべきかという論争はまだ決着がついていない。

次に連鎖の長さをどのように決めるか、すなわちどこまで連鎖を続ければ分布は定常状態にあるとみなせるのかという問題がある。この点については、Cowles and Carlin (1996) が、診断のための検定統計量をいくつか比較している。

## 4 計量経済学への応用例 1

本節では、回帰モデルの誤差項が ARMA(p, q) 過程に従うモデルをベイジアン計量経済学の枠組みで推定する方法を説明した Chib and Greenberg (1994) の研究を紹介する。

#### 4.1 モデルと事前分布の仮定

次の回帰モデルを考える。

$$y_t = x_t \beta + \varepsilon_t, \quad t=1, \dots, n \quad (7)$$

但し  $y_t$  は scalar である。このとき  $\varepsilon_t$  が ARMA(p, q) 過程に従うとする。

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}, \quad u_t \sim \text{iid}N(0, \sigma^2) \quad (8)$$

またはラグ・オペレーター  $L$  を使って

$$\phi(L)\varepsilon_t = \theta(L)u_t \quad (9)$$

但し  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ ,  $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$  である。(7)式と(8)式は状態空間モデル [Harvey (1993, Section 4) を参照] として次のように表現される。

$$y_t = x_t \beta + z \alpha_t \quad (10)$$

$$\alpha_t = G \alpha_{t-1} + f u_t \quad (11)$$

但し  $z = (1, 0, \dots, 0) : 1 \times m$ ,  $m = \max(p, q+1)$

$$\alpha_t = \begin{bmatrix} \alpha_{1t} \\ \vdots \\ \alpha_{mt} \end{bmatrix} : m \times 1, \quad f = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{bmatrix} : m \times 1,$$

$$G = \begin{bmatrix} \phi_1 & \vdots & & & \\ \phi_2 & \vdots & & & \\ \phi_3 & \vdots & & I_{m-1} & \\ \vdots & \vdots & & & \\ \dots & \dots & \dots & \dots & \dots \\ \phi_m & \vdots & 0 & \dots & 0 \end{bmatrix} : m \times m$$

ここで  $G, f$  についてそれぞれ  $s > p$  のとき  $\phi_s = 0$ ,  $r > q$  のとき  $\theta_r = 0$  とする。次の仮定をおく。

**Assumption M** (モデル) :  $p, q$  は既知であるとし, (7)式と(8)式によりデータ  $y = (y_1, \dots, y_n)'$  は生成される。

**Assumption S** (定常性) :  $\phi(L)$  の全ての根は単位円の外側にある。

**Assumption I** (反転可能性) :  $\theta(L)$  の全ての根は単位円の外側にある。

**Assumption P** (事前分布) :

$$\begin{aligned} & [\beta, \phi, \theta, \sigma_2, \alpha_0] \\ & = [\beta][\phi][\theta][\sigma^2][\alpha_0 | \beta, \phi, \theta, \sigma^2] \\ & = N_k(\beta | \beta_0, B_0^{-1}) \times N_p(\phi | \phi_0, \Phi_0^{-1}) I_s \times N_q(\theta | \theta_0, \Theta_0^{-1}) I_s \\ & \quad \times IG(\sigma_2 | \nu_0/2, \delta_0/2) \times [\alpha_0 | \beta, \phi, \theta, \sigma_2] \end{aligned} \quad (12)$$

ただし,  $\phi = (\phi_1, \dots, \phi_p)'$ ,  $\theta = (\theta_1, \dots, \theta_q)'$ ,  $IG$  は逆ガンマ分布,  $I_A$  は集合  $A$  についての指示関数,  $S_\phi$  は Assumption S を満たす  $\phi$  の集合,  $S_\theta$  は Assumption I を満たす  $\theta$  の集合, ハイパーパラメータ  $\beta_0, B_0, \phi_0, \Phi_0, \theta_0, \Theta_0, \nu_0, \delta_0$  は既知であるとする。

初期状態  $\alpha_0$  について考察する。定常性の仮定により(11)式は VAR(1) になるので

$$E(\alpha_t | \beta, \phi, \theta, \sigma^2) = (I_m - GL)^{-1} f E(u_t | \beta, \phi, \theta, \sigma^2) = 0$$

であり,  $u_t$  が正規分布に従うので,  $\beta, \phi, \theta, \sigma^2$  が与えられたとき  $\alpha_t$  も正規分布に従う。また  $u_t$  と  $\alpha_{t-1}$  には相関がないので

$$\begin{aligned} E(\alpha_t \alpha_t' | \beta, \phi, \theta, \sigma^2) &= G E(\alpha_{t-1} \alpha_{t-1}' | \beta, \phi, \theta, \sigma^2) G' + \sigma^2 f f' \\ &= G E(\alpha_t \alpha_t' | \beta, \phi, \theta, \sigma^2) G' + \sigma^2 f f' \end{aligned}$$

となる。つまり

$$\text{vec}(E(\alpha_t \alpha_t' | \beta, \phi, \theta, \sigma^2)) = \sigma^2 (I_{m^2} - G \otimes G)^{-1} \text{vec}(f f') \quad (13)$$

これらの性質は初期状態  $\alpha_0$  についても満たされている。すなわち

$$\begin{aligned} E(\alpha_0 | \beta, \phi, \theta, \sigma^2) &= 0, \quad E(\alpha_0 \alpha_0' | \beta, \phi, \theta, \sigma^2) \equiv \Omega \\ \text{vec}(\Omega) &= \sigma^2 (I_{m^2} - G \otimes G)^{-1} \text{vec}(f f') \end{aligned}$$

## 4.2 主な結果

### 4.2.1 準備

未知のパラメータをまとめて  $\psi = (\beta, \phi, \theta, \sigma^2)$  とおく。事前分布の密度関数を  $\pi(\psi)$ , 尤度関数を  $f(y | \psi)$  とおくと Bayes の定理により事後分布の密度関数は

$$f(\psi | y) \propto \pi(\psi) f(y | \psi)$$

となる。厳密な尤度関数は扱いにくいので, 以下では初期値を固定されたものとみなし条件付きの尤度関数  $f(y | \psi, \lambda)$  を使う。ただし

$$\lambda = (\varepsilon_0, \dots, \varepsilon_{-p+1}, u_0, \dots, u_{-q+1})$$

である。このとき条件付き尤度関数は次のように表される。

$$\begin{aligned} f(y | \psi, \lambda) &= \prod_{t=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} u_t^2\right] \\ &= \prod_{t=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (y_t - \hat{y}_{t|t-1})^2\right] \end{aligned} \quad (14)$$

ただし,

$$\hat{y}_{t|t-1} = x_t \beta + (\phi(L) - 1)(y_t - x_t \beta) + (\theta(L) - 1)u_t$$

であり、 $\hat{y}_{t|t-1} - x_t\beta$  は時点  $t-1$  までの情報が与えられたときの  $\varepsilon_t$  の 1 期先の予測値である。

実は ARMA モデルの状態空間表現を使うと、 $\lambda$  の  $p+q$  個の全ての要素を使う必要はなく  $m$  個の要素だけでよい。このことを示す。まず時点  $t=1$  では上の式により

$$\hat{y}_{1|0} = x_1\beta + \phi_1\varepsilon_0 + \cdots + \phi_p\varepsilon_{-p+1} + \theta_1u_0 + \cdots + \theta_qu_{-q+1}$$

である。また(10), (11)式から

$$\begin{aligned}\hat{y}_{1|0} &= x_1\beta + z\alpha_{1|0} \\ &= x_1\beta + \alpha_{1|0} \\ &= x_1\beta + \phi_1\alpha_{10} + \alpha_{20}\end{aligned}$$

となる。よって

$$\phi_1\varepsilon_0 + \cdots + \phi_p\varepsilon_{-p+1} + \theta_1u_0 + \cdots + \theta_qu_{-q+1} = \phi_1\alpha_{10} + \alpha_{20}$$

となる。すなわち  $\lambda$  の要素は  $\alpha_0$  の行要素を通じて予測式に入ってくる。次に  $1 \leq t \leq p$  について

$$\phi_t(L) = 1 - \phi_1L - \cdots - \phi_{t-1}L^{t-1}$$

と定義する。このとき

$$\begin{aligned}\phi_t(L)(y_t - x_t\beta) &= \phi_t(L)z\alpha_t = z\phi_t(L)\alpha_t \\ &= z(\alpha_t - \phi_1\alpha_{t-1} - \cdots - \phi_{t-1}\alpha_1) \\ &= \alpha_{1t} - \phi_1\alpha_{1,t-1} - \cdots - \phi_{t-1}\alpha_{11}\end{aligned}\tag{15}$$

である。(15)式に

$$\alpha_{rt} = \phi_r\alpha_{1,t-1} + \alpha_{r+1,t-1} + \theta_{r-1}u_t$$

を逐次代入していくと

$$\begin{aligned}\phi_t(L)(y_t - x_t\beta) &= (\phi_1\alpha_{1,t-1} + \alpha_{2,t-1} + u_t) - \phi_1\alpha_{1,t-1} - \cdots - \phi_{t-1}\alpha_{11} \\ &= \alpha_{2,t-1} - \phi_2\alpha_{1,t-2} - \cdots - \phi_{t-1}\alpha_{11} + u_t \\ &= (\phi_2\alpha_{1,t-2} + \alpha_{3,t-2} + \theta_1u_{t-1}) - \phi_2\alpha_{1,t-2} - \cdots - \phi_{t-1}\alpha_{11} + u_t \\ &\vdots \\ &= (\phi_{t-1}\alpha_{11} + \alpha_{t+1} + \theta_{t-2}u_2) - \phi_{t-1}\alpha_{11} + u_t + \theta_1u_{t-1} + \cdots + \theta_{t-3}u_3 \\ &= \phi_t\alpha_{10} + \alpha_{t+1,0} + u_t + \theta_1u_{t-1} + \cdots + \theta_{t-1}u_1\end{aligned}$$

となる。(10), (11)式とこの関係により

$$\begin{aligned}\hat{y}_{t|t-1} &= x_t\beta + \phi_1\alpha_{1,t-1} + \alpha_{2,t-1} \\ &= x_t\beta + \phi_t(L)(y_t - x_t\beta) - u_t + \phi_1\alpha_{1,t-1} + \cdots + \phi_{t-1}\alpha_{11} \\ &= x_t\beta + \phi_t\alpha_{10} + \alpha_{t+1,0} + \theta_1u_{t-1} + \cdots + \theta_{t-1}u_1\end{aligned}$$

$$\begin{aligned}
& +\phi_1\alpha_{1,t-1}+\cdots+\phi_{t-1}\alpha_{11} \\
= & x_t\beta+\phi_t\alpha_{10}+\alpha_{t+1,0}+\theta_1u_{t-1}+\cdots+\theta_{t-1}u_1 \\
& +\phi_1(y_{t-1}-x_{t-1}\beta)+\cdots+\phi_{t-1}(y_1-x_1\beta)
\end{aligned} \tag{16}$$

が得られる。この場合も  $y_t$  は  $\alpha_0$  のみを通じて  $\lambda$  に依存している。同様に

$$\begin{aligned}
\bar{y}_{t|t-1} = & x_t\beta + \sum_{i=1}^p \phi_i (y_{t-i} - x_{t-i}\beta) \\
& + \theta_1 u_{t-1} + \cdots + \theta_{t-1} u_1 + \alpha_{t+1,0}, \quad t=p+1, \dots, n
\end{aligned} \tag{17}$$

ただし  $\theta_j=0 (j>q)$ ,  $\alpha_{r0}=0 (r>m)$  である。以上の結果から

$$f(y|\psi, \lambda) = f(y|\psi, \alpha_0)$$

である。

以下では MCMC アルゴリズムを使って、次の条件つき密度関数をシミュレートする。

$$\begin{aligned}
& \pi(\beta|y, \psi_{-\beta}, \alpha_0), \quad \pi(\phi|y, \psi_{-\phi}, \alpha_0), \quad \pi(\theta|y, \psi_{-\theta}, \alpha_0), \\
& \pi(\sigma^2|y, \psi_{-\sigma^2}, \alpha_0), \quad \pi(\alpha_0|y, \psi)
\end{aligned}$$

これらの条件つき密度関数は同時事後密度関数

$$\pi(\psi, \alpha_0|y) \propto \pi(\psi)\pi(\alpha_0|\psi), \quad (y|\psi, \alpha_0) \tag{18}$$

に比例していることに注意されたい。

#### 4.2.2. 完全な条件付き分布

ここでは(18)式の中身を完全な条件付き分布で表現することを目的とする。以下では完全な条件付き分布を簡単に求めるために2つの変換を定義する。

**Definition 4.1**  $s \leq 0$  のとき、スカラー  $y_s = y_s^* = 0$ , ベクトル  $x_s = x_s^* = 0$  とし、 $r > m$  について  $\alpha_{r0} = 0$  とおく。 $t=1, \dots, n$  について

$$\begin{aligned}
y_t^* &= y_t - \sum_{s=1}^p \phi_s y_{t-s} - \sum_{i=1}^q \theta_i y_{t-i}^* - \phi_t \alpha_{10} - \alpha_{t+1,0} \\
x_t^* &= x_t - \sum_{s=1}^p \phi_s x_{t-s} - \sum_{i=1}^q \theta_i x_{t-i}^*
\end{aligned}$$

と定義する。

#### Lemma 4.1

$$y^- = \begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix} \sim n+1, \quad X^* = \begin{bmatrix} x_1^* \\ \vdots \\ x_n^* \end{bmatrix} \sim n \times k$$

とする。このとき

$$f(y^*|\psi, \alpha_0) = (2\pi\sigma^2)^{-n} \exp\left[-\frac{1}{2\sigma^2} (y^* - X^*\beta)' (y - X^*\beta)\right]$$

である。

**Proof.** 定義により

$$y_1^* - x_1^* \beta = y_1 - x_1 \beta - \phi_1 \alpha_{10} - \alpha_{20} = u_1$$

が成り立つ。同様に  $1 \leq t \leq n$  のとき

$$\begin{aligned} y_t^* &= x_t^* \beta \\ &= y_t - x_t \beta - \phi_t \alpha_{10} - \alpha_{t+1,0} - \sum_{s=1}^p \phi_s (y_{t-s} - x_{t-s} \beta) - \sum_{i=1}^q \theta_i (y_{t-s}^* - x_{t-s}^* \beta) \\ &\quad (y_1^* - x_1^* \beta = u_1 \text{ を与え, 逐次計算すると)} \\ &= y_t - x_t \beta - \phi_t \alpha_{10} - \alpha_{t+1,0} - \sum_{s=1}^p \phi_s (y_{t-s} - x_{t-s} \beta) - \sum_{i=1}^q \theta_i u_{t-i} \\ &= y_t - \hat{y}_{t|t-1} = u_t \end{aligned}$$

である。///

Definition 4.1 の変換は線形回帰モデル

$$y^* = X^* \beta + u, \quad u \sim N_n(0, \sigma^2 I_n)$$

を直接扱っていることになる。この関係からすぐに  $\beta$ ,  $\sigma^2$  の完全な条件付き分布が得られる。さらに  $\phi$  の完全な条件付き分布を得るために次の変換を行なう。

**Definition 4.2**  $s \leq 0$  について、スカラー  $y_s = \bar{y}_s = \bar{x}_s = 0$ , ベクトル  $x_s = 0$  とし,  $r > m$  について  $\alpha_{r0} = 0$  とする。  $t = 1, \dots, n$  について次のように定義する。

$$\begin{aligned} \bar{y}_t &= y_t - x_t \beta - \sum_{i=1}^q \theta_i \bar{y}_{t-i} - \alpha_{t+1,0} \\ \bar{x}_t &= y_t - x_t \beta - \sum_{i=1}^q \theta_i \bar{x}_{t-i} \end{aligned}$$

**Lemma 4.2**

$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} \sim n \times 1, \quad \bar{X} = \begin{bmatrix} \alpha_{10} & 0 & \cdots & \cdots & 0 \\ \bar{x}_1 & \alpha_{10} & 0 & \cdots & 0 \\ \bar{x}_2 & \bar{x}_1 & \alpha_{10} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{p-1} & \bar{x}_{p-2} & \cdots & \cdots & \alpha_{10} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{n-1} & \bar{x}_{n-2} & \bar{x}_{n-3} & \cdots & \bar{x}_{n-p} \end{bmatrix} \sim n \times p$$

とおく。このとき

$$f(\bar{y} | \phi, \alpha_0) = (2\pi\sigma^2)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} (\bar{y} - \bar{X}\phi)' (\bar{y} - \bar{X}\phi) \right]$$

である。

**Proof.** 行列  $\bar{X}$  の行要素を  $\bar{X}_t \sim 1 \times p$  と書くことにする。定義により

$$\bar{y}_1 - \bar{X}_1 \phi = y_1 - x_1 \beta - \phi_1 \alpha_{10} - \alpha_{t+1,0} = y_1 - \hat{y}_{1|0} = u_1$$

が成り立つ。同様に  $1 \leq t \leq p$  について逐次、次の計算を行なうと

$$\begin{aligned} \bar{y}_t - \bar{X}_t \phi &= y_t - x_t \beta - \sum_{i=1}^{t-1} \theta_i \bar{y}_{t-i} - \sum_{s=1}^{t-1} \phi_s \bar{x}_{t-s} - \phi_t \alpha_{10} - \alpha_{t+1,0} \\ &= y_t - x_t \beta - \sum_{i=1}^{t-1} \theta_i \left( \bar{y}_{t-i} - \sum_{s=1}^{t-i-1} \phi_s \bar{x}_{s-i} - \phi_{t-i} \alpha_{10} \right) \\ &\quad - \sum_{i=1}^{t-1} \theta_i \left( \sum_{s=1}^{t-i-1} \phi_s \bar{x}_{s-i} + \phi_{t-i} \alpha_{10} \right) - \sum_{s=1}^{t-1} \phi_s \bar{x}_{t-s} - \phi_t \alpha_{10} - \alpha_{t+1,0} \\ &= y_t - x_t \beta - \sum_{i=1}^{t-1} \theta_i u_{t-i} - \sum_{i=1}^{t-1} \theta_i \left( \sum_{s=1}^{t-i-1} \phi_s \bar{x}_{s-i} + \phi_{t-i} \alpha_{10} \right) \\ &\quad - \sum_{s=1}^{t-1} \phi_s (y_{t-s} - x_{t-s} \beta) - \sum_{s=1}^{t-1} \phi_s \sum_{i=1}^{t-s-1} \phi_s \sum_{i=1}^{t-s-1} \theta_i \bar{x}_{s-s-i} - \phi_t \alpha_{10} - \alpha_{t+1,0} \\ &= y_t - x_t \beta - \sum_{i=1}^{t-1} \theta_i u_{t-i} - \sum_{s=1}^{t-1} \phi_s (y_{t-s} - x_{t-s} \beta) \\ &\quad - \sum_{i=1}^{t-1} \theta_i \phi_{t-i} \alpha_{10} - \phi_t \alpha_{10} - \alpha_{t+1,0} \end{aligned}$$

(定義により  $\alpha_{10} = y_0 - x_0 \beta = 0$  であるから)

$$\begin{aligned} &= y_t - x_t \beta - \sum_{i=1}^{t-1} \theta_i u_{t-i} - \sum_{s=1}^{t-1} \phi_s (y_{t-s} - x_{t-s} \beta) - \phi_t \alpha_{10} - \alpha_{t+1,0} \\ &= y_t - \hat{y}_{t|t-1} = u_t \end{aligned}$$

となる。また  $p < t \leq n$  についても同様の結果が成り立つので

$$u_t = y_t - \bar{X}_t \phi, \quad t=1, \dots, n$$

である。///

Definition 4.2 の変換の結果

$$\bar{y} = \bar{X} \phi + u, \quad u \sim N_n(0, \sigma^2 I_n)$$

を直接扱うことになるので、この関係からすぐに  $\phi$  の完全な条件付き分布が求められる。

次にいくつかの表記を定義し、完全な条件付き分布について述べる。

表記

$$B_n \equiv B_0 + \sigma^{-2} X^* X^*, \quad \Phi_n \equiv \Phi_0 + \sigma^{-2} \bar{X}' \bar{X}$$

Prior density について

$$\pi(\alpha_0 | \beta, \phi, \theta, \sigma^2) \propto p(\phi, \theta, \sigma^2) = (\sigma^2)^{-m/2} |\Omega(\phi, \theta)|^{1/2} \exp\left[-\frac{1}{2\sigma^2} \alpha_0' \Omega(\phi, \theta)^{-1} \alpha_0\right],$$

$$p_1(\phi) \equiv p(\phi | \theta, \sigma^2), \quad p_2(\theta) \equiv p(\theta | \phi, \sigma^2).$$

$$d_1 \equiv \|y^* - X^* \beta\|^2, \quad d_2 \equiv \alpha_0' \Omega(\phi, \theta)^{-1} \alpha_0.$$

まず  $\alpha_0$  の完全な条件付き分布の平均と共分散行列を求める。これは全時点の情報及び他のパラメータの情報を与えられたときの  $\alpha_0$  について考えればよいので、Kalman フィルターによる平滑化を考えればよい。一般に、 $\hat{\alpha}_{t|n}$  と  $R_{t|n}$  をそれぞれ時点  $t$  における平滑化推定量とその共分散行列とすると平滑化方程式は

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t} + B_t^* (\hat{\alpha}_{t+1|n} - G \hat{\alpha}_{t|t}),$$

$$R_{t|n} = R_{t|t} + B_t^* (R_{t+1|n} - R_{t+1|t}) B_t^{*'},$$

$$t = n-1, n-2, \dots, 0$$

と表現される。ただし

$$B_t^* = R_{t|t} G R_{t+1|t}^{-1}, \quad 0 \leq t \leq n-1,$$

$R_{t+1|t}^{-1}$  は Moore-Penrose の逆行列で

$$R_{t+1|t}^{-1} = (R'_{t+1|t} R_{t+1|t})^{-1} R'_{t+1|t}$$

である。この平滑化方程式を時点  $n$  から逆方向に作用させていけば、 $\alpha_0$  の完全な条件付き分布の平均  $\alpha_{0|n}$  と共分散行列  $R_{0|n}$  が求められる。以上の表記を使って、誤差項が ARMA( $p, q$ ) に従う回帰モデルのシミュレーションに使う完全な条件付き分布を提示する。

**Proposition 4.1** Assumptions M, S, I, および P の下で、 $\beta, \phi, \sigma^2, \theta$  の完全な条件付き分布は次のように与えられる。

$$(i) \quad \beta | y, \phi_{-\beta}, \alpha_0 \sim N_k(B_n^{-1}(B_0 \beta_0 + \sigma^{-2} X^* y^*), B_n^{-1}),$$

$$(ii) \quad \phi | y, \phi_{-\phi}, \alpha_0 \propto p_1(\phi) \times N_p(\Phi_n^{-1}(\Phi_0 \phi_0 + \sigma^{-2} \bar{X}' y), \Phi_n^{-1}) I_{S_\phi},$$

$$(iii) \quad \sigma^2 | y, \phi_{-\sigma^2}, \alpha_0 \sim IG((\nu_0 + n + m) / 2, (\delta_0 + d_1 + d_2) / 2),$$

$$(iv) \quad \alpha_0 | y, \phi \sim N_m(\hat{\alpha}_{0|n}, R_{0|n}),$$

$$(v) \quad \pi(\theta | y, \phi_{-\theta}, \alpha_0) \propto p_2(\theta) \times \prod_{t=1}^n \exp\left[-\frac{1}{2\sigma^2} u_t(\theta)^2\right]$$

$$\times \exp\left[-\frac{1}{2} (\theta - \theta_0)' \Theta_0 (\theta - \theta_0)\right] I_{S_\theta}.$$

**Proof** (i) Assumption P と Lemma 1 より明らか。

(ii) Assumption P と Lemma 2 より明らか。

(iii) Assumption P と Lemma 1 より

$$\begin{aligned} \sigma^2 | \mathbf{y}, \phi_{-\sigma^2}, \alpha_0 \\ &\sim IG(\nu_0/2, \delta_0/2) \times IG(m/2, d_2/2) \times IG(n/2, d_1/2) \\ &= IG((\nu_0+n+m)/2, (\delta_0+d_1+d_2)/2) \end{aligned}$$

(iv) 表記に関する説明より明らか。

(v) 完全な条件付き分布の定義より明らか。///

#### 4.2.3 実用のためのノート

Proposition 4.1 に明らかなように,  $\beta$ ,  $\sigma^2$ ,  $\alpha_0$  は完全な条件付き分布の形から Gibbs サンプラーにより直接生成できるのに対し,  $\phi$  と  $\theta$  については MH アルゴリズムを使う必要がある。

$\phi$  については, 候補を生成する密度関数  $q(\phi^{(i)}, \phi)$  を  $N_p(\Phi_n^{-1}(\Phi_0\phi_0 + \sigma^{-2}\bar{X}'\bar{y}), \Phi_n^{-1})I_{S_p}$  の密度関数とし (独立連鎖),  $\phi'$  をこの分布から抽出したサンプルとする<sup>3)</sup>。このとき MH アルゴリズムでは, 候補  $\phi'$  は確率

$$\alpha(\phi^{(i)}, \phi') = \min\left\{\frac{p_1(\phi')}{p_1(\phi^{(i)})}, 1\right\}$$

で採択される。

$\theta$  については, 候補を生成する密度関数として  $\pi(\theta | \mathbf{y}, \phi_\theta, \alpha_0)$  を切断された正規分布で近似したものを使う。

$$\begin{aligned} q(\theta | \mathbf{y}, \beta, \phi, \sigma^2, \theta) \\ \equiv q(\theta) \propto \exp\left[-\frac{1}{2}[\theta - \mathbf{m}(\theta)]'V(\theta)^{-1}[\theta - \mathbf{m}(\theta)]\right]I_{S_\theta} \end{aligned} \quad (19)$$

ただし  $\theta^*$  は  $\theta$  の非線形最小 2 乗推定値であり,

$$\begin{aligned} \mathbf{m}(\theta^*) &= V(\theta^*)[\Theta_0\theta_0 + \sigma^{-2}W(\theta^*)'(u(\theta^*) + W(\theta^*)\theta^*)] \sim q \times 1, \\ V(\theta^*) &= [\Theta_0 + \sigma^{-2}W(\theta^*)'W(\theta^*)]^{-1} \sim q \times q, \\ u(\theta^*) &= \mathbf{y}^*(\theta^*) - X^*(\theta^*)\beta \sim n \times 1, \\ W(\theta^*) &= \left.\frac{\partial u(\theta)}{\partial \theta'}\right|_{\theta=\theta^*} \sim n \times q, \end{aligned}$$

である。また  $W(\theta^*)$  の各要素  $W_{ii}$  は次の漸化式により計算される。

$$W_{ii} = \begin{cases} 0, & t \leq 0, \\ u_{t-i}(\theta^*) - \sum_{j=1}^i \theta_j^* W_{t-j,i}(\theta^*), & t=1, \dots, n, \quad i=1, \dots, q. \end{cases}$$

次にどのような近似計算をして(19)式を得たのかを述べる。まず  $u_i(\theta)$  を  $\theta^*$  のまわりで Taylor 展開し, 2 次以上の項を無視する。

$$u_i(\theta) \simeq u_i(\theta^*) - w_i(\theta - \theta^*)$$

3) 切断されていない正規分布からサンプルを抽出し,  $S_p$  に入るものを使えばよい。

ただし  $w_t$  は  $W$  の第  $t$  行である。Proposition 4.1(v) に代入すると

$$\begin{aligned}
 & \prod_{t=1}^n \exp\left[-\frac{1}{2\sigma^2} u_t(\theta)^2\right] \times \exp\left[-\frac{1}{2}(\theta - \theta_0)' \Theta_0(\theta - \theta_0)\right] I_S, \\
 & \simeq \prod_{t=1}^n \exp\left[-\frac{1}{2\sigma^2} (u_t(\theta^t) - w_t(\theta - \theta^t))^2\right] \\
 & \quad \times \exp\left[-\frac{1}{2}(\theta - \theta_0)' \Theta_0(\theta - \theta_0)\right] I_S, \\
 & = \exp\left[-\frac{1}{2\sigma^2} (u(\theta^*) - W(\theta^*)(\theta - \theta^*))'(u(\theta^*) - W(\theta^*)(\theta - \theta^*))\right] \\
 & \quad - \frac{1}{2}(\theta - \theta_0)'(\Theta_0(\theta - \theta_0)) I_S, \\
 & \propto \exp\left[-\frac{1}{2}[\theta - m(\theta^*)]' V(\theta^*)^{-1}[\theta - m(\theta^*)]\right] I_S,
 \end{aligned}$$

となる。候補を生成する密度関数  $q(\theta)$  から抽出された候補  $\theta'$  は確率

$$\alpha(\theta^{(i)}, \theta') = \min\left\{\frac{p_2(\theta')}{p_2(\theta^{(i)})}, 1\right\}$$

で採択される。以上のようにして回帰モデルの誤差項が ARMA(p, q) に従う場合でも、ベイジアン計量経済学の枠組みで推定することができる。

## 5 計量経済学への応用例 2

本節では Jacquier, E., N. G. Polson, and P. E. Rossi (以下 JPR) (1994) の研究を概説し、ベイジアン計量経済学の枠組みにおける Stochastic Volatility (SV) model の分析方法を紹介する。

### 5.1 モデルと MCMC によるアプローチ

一般的に SV model では、時系列データのベクトル  $y$  を確率モデル  $p(y|h)$  から生成したものとみなす。ただし  $h$  はボラティリティのベクトルであり、各時点における  $y_t$  の分散は  $h_t$  である。ボラティリティ  $h$  は観察されないため、確率メカニズム  $p(h|w)$  によって生成されると仮定する。このときデータの密度関数は

$$\begin{aligned}
 p(y|w) &= \int p(y, h|w) dh \\
 &= \int \frac{p(y, h, w)}{p(h, w)} \frac{p(h, w)}{p(w)} dh \\
 &= \int p(y|h, w) p(h|w) dh \\
 &= \int p(y|h) p(h|w) dh
 \end{aligned}$$

で与えられる。

まず簡単な SV モデルから考える。時系列データ  $y_t$  の分散が log-AR(1) 過程に従うとする。

$$y_t = \sqrt{h_t} u_t, \quad \ln h_t = \alpha + \delta \ln h_{t-1} + \sigma_v \nu_t,$$

ただし  $u_t$  と  $\nu_t$  は互いに独立で、ともに  $NID(0, 1)$  に従う。このとき

$$w = \begin{bmatrix} \alpha \\ \delta \\ \sigma_v \end{bmatrix}$$

である。ベイジアン・シミュレーションの枠組みで、平均の式に外生的な説明変数を加えたり、分散の対数が AR(p) 過程に従うように拡張するのは簡単である。

最も簡単な SV モデルにおいても尤度関数は  $T$  次元の積分

$$l(w) = \int p(y|h)p(h|w)dh$$

として表される。ただし

$$w = \begin{bmatrix} \alpha \\ \delta \\ \sigma_v \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ \vdots \\ h_T \end{bmatrix}$$

である。この積分を解析的に解くことは容易ではない (Appendix を参照) ので、数値的に計算するしかない。単純な方法として Monte Carlo 積分がある。例えば、「事前分布」 $p(h|w)$  からシミュレートしたサンプル  $h_i (i=1, \dots, H)$  を使って条件付き尤度関数の平均値

$$\hat{l}(w) = \frac{1}{H} \sum_{i=1}^H p(y|h_i) / H$$

を計算する。つまり、尤度の分布の平均がわからないので、条件付き尤度関数の期待値を求めている。よく知られているように、 $h_i$  の次元が増えるるとこのような方法はうまくいかない。Danielsson and Richard (1993) はインポートランス・サンプリング (importance sampling) の方法を使って、より正確な積分計算の方法を提案している。ここでは詳しく述べないが、Danielsson and Richard (1993) の方法は、積分計算に最適なインポートランス関数 (importance function) を設定するのに膨大な数の尤度関数を計算しなければならないため、実際的な方法ではない。

SV モデルの特定化は、階層的構造 (hierarchical structure) をもつ条件付き分布を用いて行なわれる。この階層は、 $y|h$  の条件付き分布、 $h|w$  の条件付き分布、 $w$  の事前分布の 3 階層に分れる。ある意味で  $p(h|w)$  に対応する SV の式は、ハイパーパラメータ  $w$  をもつ事前分布と解釈することも可能である。また  $h$  と  $w$  の同時事後分布はこれら 3 つの分布の積に比例している。この意味において JPR (1994) では、 $h$  を使ってパラメータベクトルを  $w$  から  $(w, h) \in \Omega \times H$  に拡張している。Bayes の定理から  $h$  と  $w$  の同時事後分布は

$$\pi(h, w|y) \propto p(y|h)p(h|w)p(w)$$

で与えられる。この同時事後分布のうち、SV モデルのパラメーターの推定には周辺事後分布  $\pi(w|y)$  を使い、周辺事後分布  $\pi(h|y)$  は観察されないボラティリティの平滑化問題の解を与える。これらの周辺分布を計算するために、定常分布  $\pi$  をもつ Markov 連鎖を構築する。

適切な Markov 連鎖サンプラーを構築するには同時事後分布をどのような条件付き分布に分解するかが問題となる。例えば、 $h$  が与えられたときの  $w$  の事後密度関数  $p(w|y, h)$  は、通常のベイズ計量経済学における線形モデルの問題として簡単に計算される。もしまだ  $p(h|w, y)$  から直接サンプルを抽出することが可能ならば、 $p(w|y, h)$  と  $p(h|w, y)$  から交互にサンプルを生成し、Markov 連鎖をつくることができる。

SV モデルでは  $p(h|w, y)$  から直接サンプルを生成するには大変にコストがかかるので、その代わりに  $p(h|w, y)$  をさらに  $p(h_t| h_{-t}, w, y)$  の集合に分割する。ただし  $h_{-t}$  はベクトル  $h$  のうち  $h_t$  を除いたものである。この 1 変量分布から直接サンプルを生成することはできないので、JPR (1994) では Metropolis 採択/棄却サンプリング法を使っている。この方法では条件付き密度関数  $p(h_t| h_{-t}, w, y)$  から直接サンプルを生成しなくても、Markov 連鎖の定常分布として事後分布が得られる。この方法は「cyclic independence Metropolis chain」として知られている。

## 5.2 MCMC の構築

Markov 連鎖を構築するために、まず  $p(w|h, y)$  から抽出する方法を説明し、次に  $p(h|w, y)$  から間接的に抽出する方法を説明する。簡単のため  $\beta = (\alpha, \delta)'$  とおき、 $w = (\beta', \sigma_v)'$  の事前分布として通常 of 自然共役分布を仮定する。

$$p(\beta, \sigma_v) = p(\beta|\sigma_v)p(\sigma_v)$$

$$\beta|\sigma_v \sim N(\bar{\beta}, \sigma_v^2 A^{-1}), \sigma_v^2 \sim IG(\nu_0, s_0^2)$$

ただし  $\bar{\beta}$ ,  $A$ ,  $\nu_0$ ,  $s_0$  はハイパーパラメーターである。このとき  $h$  をデータとしてみたときの  $w$  の事後分布は

$$p(w|h) = p(w, h)/p(h) = p(w)p(h|w)/p(h)$$

$$\propto p(w)p(h|w)$$

$$\propto p(w) \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left[-\frac{1}{2\sigma_v^2} (\ln h_t - \alpha - \delta \ln h_{t-1})^2\right]$$

であり、 $y$  をデータとしたときの  $w|h$  の事後分布は

$$p(w|h, y) = p(w, h, y)/p(h, y) = p(y|h, w)p(h, w)/p(h, y)$$

$$= p(y|h, w)p(w|h)p(h)/p(h, y)$$

$$\propto p(w|h)p(y|h, w)$$

$$(p(y|h, w) = p(y|h) \text{ であるから})$$

$$\propto p(w|h) \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{y_t^2}{2h_t}\right]$$

で与えられる。これにより、多変量正規分布やガンマ分布を使って簡単に  $w$  を抽出できる。また同様にして AR(p) モデルに拡張することもできる。過剰にサンプルを抽出するのを避けるには、自己回帰項の係数に「damping prior」を仮定すればよい。

ここで  $\beta | \sigma_v, h, y$  の分布と  $\sigma_v | \beta, h, y$  の分布を具体的に示しておく。

$$L = \begin{bmatrix} \ln h_1 \\ \vdots \\ \ln h_T \end{bmatrix} : T \times 1, \quad X = \begin{bmatrix} 1 & \ln h_0 \\ \vdots & \vdots \\ 1 & \ln h_{T-1} \end{bmatrix} : T \times 2, \quad H = \begin{bmatrix} \ln h_1 & 0 \\ \vdots & \vdots \\ 0 & \ln h_T \end{bmatrix} : T \times T,$$

と定義すると、 $w | h$  の事後分布は

$$\begin{aligned} p(w|h, y) &\propto p(\beta | \sigma_v) p(\sigma_v) p(h|w) p(y|h) \\ &\propto |A| \sigma_v^{-2} \exp\left[-\frac{1}{2\sigma_v^2} (\beta - \bar{\beta})' A (\beta - \bar{\beta})\right] \sigma_v^{-2(\nu_0 A + 1)} \exp[-s_0^2 / \sigma_v^2] \\ &\quad \times \sigma_v^{-T} \exp\left[-\frac{1}{2\sigma_v^2} (L - X\beta)' (L - X\beta)\right] |H|^{-1/2} \exp\left[-\frac{1}{2} y' H^{-1} y\right] \end{aligned}$$

となる。これにより

$$\begin{aligned} p(\sigma_v | h, y, \beta) &\propto \sigma_v^{-2(\nu_0 + T/2 + 2)} \exp\left[-\frac{1}{2\sigma_v^2} [(\beta - \bar{\beta})' A (\beta - \bar{\beta}) + (L - X\beta)' (L - X\beta) + 2s_0^2]\right] \end{aligned}$$

となる。これにより

$$\sigma_v | h, y, \beta \sim IG(T/2 + \nu_0 + 1, B/2)$$

ただし、

$$B \equiv (\beta - \bar{\beta})' A (\beta - \bar{\beta}) + (L - X\beta)' (L - X\beta) + 2s_0^2$$

である。また

$$\begin{aligned} p(\beta | h, y, \sigma_v) &\propto \exp\left[-\frac{1}{2\sigma_v^2} [(\beta - \bar{\beta})' A (\beta - \bar{\beta}) + (L - X\beta)' (L - X\beta)]\right] \\ &\propto \exp\left[-\frac{1}{2\sigma_v^2} [\beta - (X'X + A)^{-1}(X'L + A\bar{\beta})]' (X'X + A) [\cdot]]\right] \end{aligned}$$

であるから、

$$\beta | h, y, \sigma_v \sim N(C, \sigma_v^2 (X'X + A)^{-1})$$

である。ただし

$$C \equiv (X'X + A)^{-1}(X'L + A\bar{\beta})$$

である。

次に  $h | w, y$  から直接抽出せずに、1変量条件付き密度関数  $p(h_t | h_{t-1}, h_{t+1}, w, y)$ ,  $t=1, \dots, T$  について考える。もしこの条件付き分布から直接サンプルを抽出できるのならば Gibbs サンプラーを使えばよいのだが、これはできない。この条件付き分布は通使の扱いやすい分布には従わないからである。 $h_t$  の Markov 性と Bayes の定理より、 $h_t | h_{t-1}, h_{t+1}, w, y$  の密度関数は

$$\begin{aligned} & p(h_t | h_{t-1}, h_{t+1}, w, y_t) \\ &= p(h_{t-1}, h_t, h_{t+1}, w, y_t) / p(h_{t-1}, h_{t+1}, w, y_t) \\ &= p(y_t | h_{t-1}, h_t, h_{t+1}, w) p(h_{t-1}, h_t, h_{t+1}, w) / p(h_{t-1}, h_{t+1}, w, y_t) \\ & \quad (y_t \text{ は } h_t \text{ にのみ依存するので}) \\ &= p(y_t | h_t) p(h_{t+1} | h_t, h_{t-1}, w) p(h_t, h_{t-1}, w) / p(h_{t-1}, h_{t+1}, w, y_t) \\ & \quad (h_{t+1} \text{ は } h_{t-1} \text{ には依存しないので}) \\ &= p(y_t | h_t) p(h_{t+1} | h_t, w) p(h_t | h_{t-1}, w) p(h_{t-1}, w) / p(h_{t-1}, h_{t+1}, w, y_t) \\ & \propto p(y_t | h_t) p(h_{t+1} | h_t, w) p(h_t | h_{t-1}, w) \\ & \propto h_t^{-1/2} \exp\left[-\frac{y_t^2}{2h_t}\right] h_{t+1}^{-1} \exp\left[-\frac{(\ln h_{t+1} - \alpha - \delta \ln h_t)^2}{2\sigma_v^2}\right] \\ & \quad \times h_t^{-1} \exp\left[-\frac{(\ln h_t - \alpha - \delta \ln h_{t-1})^2}{2\sigma_v^2}\right] \\ & \propto h_t^{-1/2} \exp\left[-\frac{y_t^2}{2h_t}\right] h_t^{-1} \exp\left[-\frac{(\ln h_t - \mu_t)^2}{2\sigma^2}\right] \tag{20} \end{aligned}$$

となる。ただし

$$\begin{aligned} \mu_t &= (\alpha(1-\delta) + \delta(\ln h_{t+1} + \ln h_{t-1})) / (1+\delta^2) \\ \sigma^2 &= \sigma_v^2 / (1+\delta^2) \end{aligned}$$

である。この密度関数から直接サンプルを抽出できないので、採択／棄却サンプリング法の使用を考えてみる。確率変数  $x$  を密度関数  $p(x)$  から生成できないとき、

$$p(x) \leq cq(x) \text{ for all } x$$

となるような定数  $c$  が存在するように採択／棄却密度関数  $q(x)$  を設定する。問題となっている条件付き分布に採択／棄却法を適用するには、条件付き分布の基準化定数が未知であり、時間と共に変化することに注意しなければならない。すなわち、 $p$  よりも大きい値をとるような採択／棄却密度関数が使えらるとしても各時点においてサンプル抽出を効率的にするような定数  $c$  を求め

なければならない。この場合、乱数発生法の文献 [例えば Ripley (1987, Chapter 3)] にあるような境界となる関数からの抽出法は使えない。

この問題に対し JPR (1994) では Metropolis 連鎖を使っている。Metropolis 連鎖では、単変量条件付き密度関数  $p$  に形が似ていて、サンプルを簡単に抽出できるような  $q$  を設定する。このとき  $cq$  は必ずしも  $p$  より大きい値をとる関数である必要はない。JPR (1994) では候補を生成する密度関数として条件付き分布と少し似た分布を設定し、独立連鎖を構成している。まず(20)式の最初の部分はガンマ分布から生成された変数の逆関数の密度関数である。ここで逆ガンマは

$$X \sim 1/Z, Z \sim Ga(\phi, 1/\lambda)$$

$$P_x(x) = \lambda^\phi / \Gamma(\phi) x^{-(\phi+1)} \exp(-\lambda/x)$$

で定義される。(20)式の後半部分是对数正規密度関数であるが、この2次までのモーメント

$$E[h_t | \cdot] = \exp(\sigma^2/2), E[h_t^2 | \cdot] = \exp(2\sigma^2)$$

を逆ガンマ密度関数の2次までのモーメントに対応するように近似する。このとき(20)式の近似式はパラメータ

$$\phi = (1 - 2\exp(\sigma^2)) / (1 - \exp(\sigma^2)) + 1/2, \lambda = (\phi - 1)(\exp(\mu_t + \sigma^2/2)) + y_t^2/2$$

をもつ逆ガンマ密度関数として表される。

完全な状態空間  $(h, w) \in H \times \Omega$  において Markov 連鎖を構築ふるために、 $T$  個の Metropolis 独立採択/棄却連鎖をつなぎ合わせて  $h$  ベクトルの各要素のサンプルや  $p(w|h, y)$  からのサンプルを生成できるようにしている。すなわち、各ステップにおける採択/棄却密動関数から各連鎖のサンプルを生成し、 $h_t$  を修正する。この  $T$  個の Metropolis 連鎖を循環させて  $h$  を  $h'$  に更新し、 $w|h', y$  からサンプルを抽出して  $w$  を  $w'$  に更新する。結局この Markov 連鎖では  $(h, w)$  から  $(h', w')$  に更新されることになる。あとは Markov 連鎖が定常分布  $\pi$  として  $(h, w)$  の同時事後分布をもつことを示せばよい。

MCMC の背景となる考え方は、ergodic (既約および非周期的) で定常分布  $\pi$  をもつ推移核  $P$  を特定し、Markov 連鎖を構築することにある。よく使われる Markov 連鎖として Gibbs サンプラーと Metropolis アルゴリズムがある。いずれも任意の初期状態  $X_0$  に Markov 連鎖を適用していくと定常分布  $\pi$  から抽出したサンプルに収束する。一般的に Metropolis アルゴリズムは次のように使われる。まず  $Q(x, y)$  を候補を生成する密度関数とし、時点  $t$  において Markov 連鎖は  $X_t = x$  にあるとする。このとき時点  $t+1$  の候補  $y$  は  $Q(x, y)$  により生成され、確率  $\alpha(x, y)$  で採択される。ただし

$$\alpha(x, y) = \min \left\{ \frac{p(y)Q(x, y)}{p(x)Q(y, x)}, 1 \right\}$$

である。また棄却されたときは現時点の値をそのまま次の時点の値とする。 $\pi$  については定数項を省略した  $\pi(x) \propto p(x)$  さえわかればよい。

Metropolis の独立連鎖では現時点の値とは無関係に次の時点の候補が生成される。すなわち

独立連鎖では推移核は  $Q(x, y) = f(y)$  となり、採択率は  $\alpha(x, y) = \min\{w(y)/w(xy), 1\}$  で与えられる。ただし  $w(z) = p(z)/f(z)$  である。さらに棄却サンプリング法を使うときは採択/棄却密度関数を  $q$  として、 $f(x) \propto \min\{p(x), cq(x)\}$  となる。仮に  $cq(x)$  がすべての  $x$  について  $p(x)$  よりも大きい値をとるならば、採択率は常に  $\alpha(x, y) = 1$  となりこの Markov 連鎖は単なる iid の採択/棄却サンプリングに退化する。ところがこの問題では  $cq$  は必ずしも  $p$  よりも大きい値をとるわけではないので、 $w(y)/w(x)$  が 1 より小さい値をとる場合がでてくる。この場合 Markov 連鎖はある点に 1 回以上留まることになる。

完全な状態空間 ( $H \times \Omega$ ) の下で Metropolis 独立連鎖を実際に使う際には、 $h$  と  $w$  の同時事後分布の計算はできるけれども、高次元の採択/棄却サンプリングを行なうため非常に非効率的である。この問題を避けるため JPR (1994) では、ベクトル  $h$  の  $T$  個の要素のそれぞれの核と  $w|h$  の条件付き分布の核の積を推移核

$$P : P_1 P_2 \cdots P_T P_{w|h}$$

として定義している。ただし  $P_i$  は  $w$  が与えられたときに  $h_i$  を更新し、 $P_{w|h}$  は  $w$  を更新する。

さて初期状態  $X_0$  の影響がなくなるには Markov 連鎖の長さをどれほどとればよいのか。実証分析では通常、最初の  $T^*$  個のサンプル以降は定常分布に収束したものとみなし、最初の  $T$  個は捨ててしまう。実際には、シミュレートされたデータから経験的に  $T^*$  を選ぶことになる。また Cowles and Carlin (1996) で使われている方法で診断を行う。

### 5.3 平滑化と予測

SV モデルの用途として、観察されない条件付きボラティリティの値を内挿的に推定 (smoothing) することや外挿的に推定 (prediction) することはいずれも重要である。

平滑化の問題には、 $h_t | y$  の条件付き分布を計算しなければならない。ただし  $y' = (y_1, \dots, y_T)$  である。この分布が使えるのなら  $h_t$  の平滑化の推定値として  $E[h_t | y]$  を使えばよい。またこの密度関数は観測できない  $h_t$  についての不確実性を集約している。ところがこの密度関数  $p(h_t | y)$  を解析的に求めることは容易ではない。しかしながら同時事後分布  $\pi(h, w)$  から抽出したサンプルを使ってこの密度関数の Monte Carlo 推定値を使うことができる。

JPR (1994) では、Monte Carlo シミュレーションの副産物として平滑化問題を解決している。ここでの Monte Carlo サンプラーは定常分布として同時事後分布  $\pi(h, w)$  に従っている。平滑化問題の解は、 $h_t$  の周辺事後分布  $p(h_t | y)$  で与えられる。すなわちベクトル  $h$  について抽出したサンプルをそのまま周辺密度関数の推定値に使えばよい。さらに周辺分布は

$$p(h | y) = \int p(h | w, y) p(w | y) dw$$

であるから、パラメータの不確実性も直接説明している。非線形フィルタリングの問題の解は  $E[h_t | y]$  で与えられる。

ボラティリティの予測問題の目的は、将来のボラティリティで構成されるベクトルについてデータが与えられたときの予測密度関数  $p(h_T | y)$  を計算することにある。ただし  $h'_j = (h_{T+1}, \dots,$

$h_{T+k}$ ) である。JPR (1994) のアプローチでは、サンプル内と将来の両方の条件付き分散を含めた  $h$  の同時事後分布を計算し、将来の  $h$  について周辺化すれば予測分布が得られる。ここで

$$y_f = \begin{bmatrix} y_{T+1} \\ \vdots \\ y_{T+k+1} \end{bmatrix}, \quad h^* = \begin{bmatrix} h \\ h_f \end{bmatrix}, \quad y^* = \begin{bmatrix} y \\ y_f \end{bmatrix}$$

と定義する。次の3つの条件付き分布をつなぎ合わせて Markov 連鎖を構築する。

1.  $y_f | y \oplus h^*, w$
2.  $h^* | y^*, w$
3.  $w | h^*$

最初の条件付き分布については iid の正規変数を使ってサンプルを抽出できる。残りの条件付き分布については 5.2 節説明した cyclic Metropolis アルゴリズムを使って抽出すればよい。

## 6 結びに代えて

本稿では、MCMC シミュレーション法とその計量経済学への応用について紹介した。特に計量経済学への応用として ARMA(p, q) モデルの推定や SV モデルの推定方法を説明した。これらのモデルを組み合わせ、さらに誤差項に正規分布ではなく一般化指数分布を仮定したときの推定法については Asai (1999) を参照してほしい。本稿が読者諸氏の MCMC 理解の手助けになれば幸いである。

### Appendix

#### A. 1 SV モデルの尤度関数について

ここでは尤度関数が解析的には求められないことを示す。潜在変数である  $h$  を積分する代わりに、 $h$  をモデルから消去し変数変換を行なう。モデルより

$$\ln y_t^2 = \ln h_t + \ln u_t^2$$

である。 $\ln y_{t-1}^2$  について、両辺に  $\delta$  を掛けて  $\ln y_t^2$  から引くと

$$\begin{aligned} \ln y_t^2 &= \delta \ln y_{t-1}^2 + (\ln h_t - \delta \ln h_{t-1}) + \ln u_t^2 - \delta \ln u_{t-1}^2 \\ &= \delta \ln y_{t-1}^2 + \alpha + \sigma_v \nu_t + \eta_t \end{aligned}$$

ただし

$$\eta_t \equiv \ln u_t^2 - \delta \ln u_{t-1}^2$$

である。

$\eta_t$  の分布を求めるために変数変換

$$\begin{bmatrix} u_t^2 \\ u_{t-1}^2 \end{bmatrix} \rightarrow \begin{bmatrix} \eta_t \\ u_{t-1}^2 \end{bmatrix}$$

を考える。ここで

$$u_t^2 = u_{t-1}^2 e^{\eta_t}$$

であるから、変換の Jacobian は

$$J = \begin{vmatrix} \frac{\partial u_t^2}{\partial \eta_t} & \frac{\partial u_t^2}{\partial u_{t-1}^2} \\ \frac{\partial u_{t-1}^2}{\partial \eta_t} & \frac{\partial u_{t-1}^2}{\partial u_{t-1}^2} \end{vmatrix} = \begin{vmatrix} u_{t-1}^{2\delta} e^{\eta_t} & \delta u_{t-1}^{2(\delta-1)} e^{\eta_t} \\ -u_{t-1}^2 / \delta & 1 \end{vmatrix} = 2u_{t-1}^{2\delta} \exp \eta_t > 0$$

である。 $u_t^2 \sim \text{iid} \chi^2(1)$  であるから  $u_t^2$  と  $u_{t-1}^2$  の同時密度関数は

$$f(u_t^2, u_{t-1}^2) = [2^{1/2} \Gamma(1/2)]^{-2} u_t^{-1} \exp\left(-\frac{u_t^2}{2}\right) u_{t-1}^{-1} \exp\left(-\frac{u_{t-1}^2}{2}\right)$$

で与えられる。よって  $\eta_t$  と  $u_{t-1}^2$  の joint pdf は

$$\begin{aligned} f(\eta_t, u_{t-1}^2) &= J \times [2^{1/2} \Gamma(1/2)]^{-2} (u_{t-1}^{2\delta} e^{\eta_t})^{-1/2} \exp\left(-\frac{u_{t-1}^{2\delta} e^{\eta_t}}{2}\right) u_{t-1}^{-1} \exp\left(-\frac{u_{t-1}^2}{2}\right) \\ &= [2^{1/2} \Gamma(1/2)]^{-2} 2u_{t-1}^{\delta-1} e^{\eta_t/2} \exp\left[-\frac{1}{2}(u_{t-1}^{2\delta} e^{\eta_t} + u_{t-1}^2)\right] \end{aligned}$$

で与えられる。よって  $\eta_t$  についての周辺分布は

$$f(\eta_t) = \int_0^\infty f(\eta_t, u_{t-1}^2) du_{t-1}^2$$

を計算すればよいのであるが、これは解析的には簡単に解けない。従って  $\ln y_t^2$  の分布も解析的に求めることは容易ではない。

## A. 2 新たな方法

現在では  $h$  を生成するのに JPR (1994) よりも簡単な方法がある。ここでは惨択/棄却アルゴリズムを使う方法を紹介する。このアルゴリズムについては 5.2 節で紹介したが、JPR (1996) は  $p(h_t | h_{t-1}, h_{t+1}, w, y_t)$  を上回るような関数の導出は困難であると考えて、MH アルゴリズムを使った。Pitt and Shephard (1995) が提案し、拡張したアイデアを使えばこの bounding function を導出できる。 $h_t | h_{t-1}, h_{t+1}, w, y_t$  の密度関数の対数は(20)式より

$$\ln(p(h_t | h_{t-1}, h_{t+1}, w, y_t)) = \text{const} + \ln(p_t^*)$$

ただし

$$\ln(p_t^*) = -\frac{3}{2} \ln h_t - \frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2} (\ln h_t - \mu_t)^2 \quad (21)$$

である。 $h_t = \exp v_t$  とおき、 $\exp(\pm v_t)$  を  $v_t^* (= \ln h_t^*)$  のまわりで Taylor 展開し 2 次以上の項を無視すると

$$\begin{aligned} \frac{1}{h_t} &= \exp(-v_t) \geq \exp(-v_t^*) - (v_t - v_t^*) \exp(-v_t^*) \\ &= \frac{1}{h_t^*} - (\ln h_t - \ln h_t^*) / h_t^* \end{aligned}$$

となる。この結果を使い(2)式を書き直すと

$$\begin{aligned} \ln(p_t^*) &\leq -\frac{3}{2} \ln h_t - \frac{y_t^2}{2} \left\{ \frac{1}{h_t^*} (1 + \ln h_t^* - \ln h_t) \right\} - \frac{1}{2\sigma^2} (\ln h_t - \mu_t)^2 \\ &= \text{const} - \frac{1}{2\sigma^2} \left[ \ln h_t - \left( \mu_t + \frac{y_t^2 \sigma^2}{2h_t^*} - \frac{3\sigma^2}{2} \right) \right]^2 \\ &= \ln(g_t^*). \end{aligned}$$

すなわち  $g_t^*$  は平均と分散がそれぞれ

$$m_t^* = \mu_t + \frac{y_t^2 \sigma^2}{2h_t^*} - \frac{3\sigma^2}{2}, \quad \sigma^2$$

の正規分布の密度関数で、 $p(h_t | \cdot)$  の bounding function である。従って、 $h_t | \cdot$  の密度関数からサンプルを生成するには、 $LM(m_t^*, \sigma^2)$  から抽出したサンプルを確率  $p_t^* / g_t^*$  で採択すればよい。

#### 参考文献

- [ 1 ] Asai, M. (1999), "Bayesian Analysis of Stochastic Volatility Models with Heavy-Tailed Distribution," *MTEC Journal* 12, 19-40.
- [ 2 ] Chib, S. and E. Greenberg (1994), "Bayes Inference in Regression Models with ARMA(p, q) errors," *Journal of Econometrics* 64, 183-206.
- [ 3 ] Chib, S. and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm" *The American Statistician* 49, 327-335.
- [ 4 ] Chib, S. and E. Greenberg (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory* 12, 409-431.
- [ 5 ] Cowles, M. K. and B. P. Carlin (1996), "Markov chain Monte Carlo convergence diagnostics: a comparison review," *Journal of the American Statistical Association* 91, 863-904.
- [ 6 ] Danielsson, J. and J.-F. Richard (1993), "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics* 8, S153-S173.
- [ 7 ] Gelman, A. (1992), "Iterative and non-iterative simulation algorithm," in H. J. Newton, ed., *Computing Science and Statistics*, 433-438, Interface Foundation of North America, Fairfax Station.
- [ 8 ] Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica* 57, 1317-1340.
- [ 9 ] Gilks, W. R. and P. Wild (1992), "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics* 41, 337-348.
- [ 10 ] Harvey, A. (1993), *Time Series Models*, 2nd. ed., Harvester Wheatsheaf, Hertfordshire.
- [ 11 ] Jacquier, E., N. G. Polson, and P. E. Rossi (1994), "Bayesian Analysis of Stochastic Volatilities,"

- Journal of Business and Economic Statistics* 12, 371-417.
- [12] Liu, J. S., W. W. Wong, and A. Kong (1994), "Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes," *Biometrika* 81, 27-40.
- [13] Meyn, S. P. and R. L. Tweedie (1993), *Markov chains and stochastic stability*, Springer-Verlag, London.
- [14] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics* 21, 1087-1092.
- [15] Müller, P. (1991), "A generic approach to posterior integration and Gibbs sampling," Technical Report 91-09, Department of Statistics, Purdue University.
- [16] Nummelin, E. (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge.
- [17] Ripley, B. (1987), *Stochastic Simulation*, John Wiley, New York.
- [18] Roberts, G. O. (1995), "Markov chain concepts related to sampling algorithms," in W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds., *Markov chain Monte Carlo in Practice*, 45-47, Chapman & Hall, London.
- [19] Tanner, M. A. and W. H. Wong (1987), "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association* 82, 528-549.
- [20] Tierney, L. (1994) "Markov chains for exploring posterior distributions," *Annals of Statistics* 22, 1701-1762.
- [21] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.