

ベイズ的方法における MCMC の利用

統計数理研究所

下平英寿

1 はじめに

対象を表現するモデルは、適当な測度 $\mu(\cdot)$ に関する確率密度関数 $f(y, x|\phi)$ とする。ここで、 y は観測データ、 $x = (\xi, \theta)$ は観測できない確率変数で、 ξ は欠測データ、 θ はパラメタ、そして、 ϕ はハイパーパラメタである。ベイズ的なアプローチでは、 y を与えた時の x の事後分布 (posterior) $f(x|y, \phi) = f(y, x|\phi)/f(y|\phi)$ が中心的な役割を果たす。($f(y|\phi) = \int f(y, x|\phi)\mu(dx)$ は、 ϕ の尤度関数である。) すなわち、ベイズ推測は一般に、適当な関数 $a(x)$ の期待値

$$E(a(x)|y, \phi) = \int a(x)f(x|y, \phi)\mu(dx) \quad (1)$$

として定式化される。この計算は一般に困難であるが、もし M 個の $x^{(1)}, \dots, x^{(M)}$ がそれぞれ $f(x|y, \phi)$ に従う確率変数の実現値なら、

$$\hat{E}(a(x)|y, \phi) = \frac{1}{M} \sum_{t=1}^M a(x^{(t)}) \quad (2)$$

によって推定できる。Markov Chain Monte Carlo (MCMC) は、その定常分布 (stationary distribution) $\pi(x)$ が目的の分布 $f(x|y, \phi)$ になるようにマルコフ連鎖を設計し、サンプル $\{x^{(t)}\}$ をコンピュータで生成する技法である。

2 MCMC サンプラの作り方

2.1 Metropolis-Hastings アルゴリズム

Metropolis ら (1953) 以来、MCMC は統計物理で長い間使われて来た。Hastings (1970) は MCMC をより一般的な形で統計分野へ導入し、Hastings の与えた MCMC の中で以下に示す形のアルゴリズムがもっとも効率が良い事は Peskun (1973) が示した。収束の正則条件の議論は Tierney (1994) などにある。なお 3.3 節にあるように、Green (1995) がより一般の状態空間への拡張を構成的にしている。

$x \in E \subset \mathcal{R}^k$ とする。 $\pi(x)$ は、 σ 有限測度 $\mu(\cdot)$ に関する x の proper な確率密度 (density) とする。初期値 $x^{(0)}$ を適当に決め、 $x^{(0)}, \dots, x^{(t)}$ がすでに得られたとして、次のように $x^{(t)}$ を更新して $x^{(t+1)}$ を決める。適当に与えた “proposal density” $q(x'|x^{(t)})$ に従い、 $x^{(t+1)}$ の候補となる x' を生成する。そして

$$\alpha(x'; x^{(t)}) = \min \left\{ 1, \frac{q(x^{(t)}|x')\pi(x')}{q(x'|x^{(t)})\pi(x^{(t)})} \right\} \quad (3)$$

を計算する。確率 $\alpha(x'; x^{(t)})$ で、 x' を “accept” し、 $x^{(t+1)} = x'$ とする。確率 $1 - \alpha(x'; x^{(t)})$ で、 x' を “reject” し、 $x^{(t+1)} = x^{(t)}$ とする。

このマルコフ連鎖の状態遷移を決める transition density は、 $x^{(t+1)} \neq x^{(t)}$ に対して

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

であり, $x^{(t+1)} = x^{(t)}$ では $1 - \int_{E \setminus \{x^{(t)}\}} p(x'|x^{(t)})\mu(dx')$ の point mass である. この $p(x'|x)$ は詳細釣り合い (detailed-balance) $p(x'|x)\pi(x) = p(x|x')\pi(x')$ を満たして、マルコフ連鎖は時間に可逆 (reversible) である. このことから直ちに

$$\pi(x') = \int p(x'|x)\pi(x)\mu(dx) \quad (4)$$

が言えて, $\pi(x)$ はこのマルコフ連鎖の定常分布 (又は invariant distribution) となる.

仮に $x^{(0)} \sim \pi(x)$ なら, (4) より各 $x^{(t)}$, $t \geq 1$ も $\pi(x)$ に従うことになるが, $x^{(0)}$ をそのように選ぶ事は一般には出来ない. (もし出来るなら, MCMC が不要!) 実際には, $x^{(1)}, \dots, x^{(m+M)}$ まで発生させ, はじめの m 個のサンプルを捨てて,

$$\hat{E}(a(x)) = \frac{1}{M} \sum_{t=m+1}^{m+M} a(x^{(t)}) \quad (5)$$

より (1) を推定する. 初期のサンプルを捨てることは “burn-in” と呼ばれる. 一般にマルコフ連鎖が proper な定常分布 $\pi(x)$ を持ち, irreducible (到達可能条件) で aperiodic (非周期的) なら, 定常分布 $\pi(x)$ は unique で, $t \rightarrow \infty$ の時の $x^{(t)}$ の極限分布が $\pi(x)$ に収束する. このことから, $M \rightarrow \infty$ で $\hat{E}(a(x)) \rightarrow E(a(x))$ が言える.

このように, 適当な正則条件を満たす任意の $q(x'|x)$ から, 目的の $\pi(x)$ を定常分布とするマルコフ連鎖を構成するアルゴリズムは驚く程簡単である. MCMC サンプラの設計として問題なのは, ステップ毎の計算量, 必要な精度を得るのに必要なステップ数 M などを考慮して, 「良い」 $q(x'|x)$ を与える事である.

例 1 (色々なサンプラ) 以下のサンプラは Metropolis-Hastings サンプラの特殊形である. Metropolis サンプラは $q(x'|x) = q(x|x')$ とする. random walk サンプラは $q(x'|x) = q(x' - x)$ の形で書けるもの. さらに $q(z) = q(-z)$ なら random walk Metropolis サンプラと呼ばれる. また, independence サンプラは $q(x'|x) = q(x')$ の形に書けるものである.

例 2 (Metropolis-Hastings サンプラによる正規分布の生成) 実用的な意味は無いが簡単な例として, $x \in \mathcal{R}$ とし, $\pi(x)$ を平均 0, 分散 1 の正規分布 $N(0, 1)$ とする (図 1). random walk Metropolis サンプラで, $q(\cdot|x)$ として区間 $[x - d, x + d]$ の一様分布を使った ($d = 0.1, 1, 10$). d が小さすぎると accept される確率は上がるが移動幅は小さい. d が大きすぎると移動幅は大きい accept される確率が下がって, 結局あまり動かない. 一般に, $\{x^{(t)}\}$ がくまなく $\pi(x)$ を代表するような良い $q(x'|x)$ を選ぶのは難しい.

2.2 ハイブリッドサンプラ

Hastings (1970) や Tierney (1994) は, $\pi(x)$ を定常分布に持つ複数の transition density $p_i(x'|x)$, $i = 1, 2, \dots$ を組み合わせるハイブリッドサンプラを考察した. これらを任意の順序で $x^{(t)}$ に適用して状態遷移をしても, 全体の定常分布は再び $\pi(x)$ である. 特に, 各ステップ t で $x^{(t)}$ に依存する確率で $p_i(x'|x)$ を選んで適用するものを mixture hybrid サンプラと呼び, また, あらかじめ決めた順序で循環的に $p_i(x'|x)$ を選ぶものを cycle hybrid サンプラとよぶ.

各 $p_i(x'|x)$ がベクトル x の一つの要素しか動かさないようなハイブリッドサンプラを, single-component Metropolis-Hastings サンプラと呼ぶ. Metropolis ら (1953) が提案したのはこのタイプの MCMC だった. x の添字の集合を $K = \{1, \dots, k\}$ とし, $A \subset K$ に対して, $x_A = (x_i : i \in A)$ と書き, 残りの要素を $x_{-A} = (x_i : i \in K \setminus A)$ と書く. (特に, $x_{-\{i\}} = x_{-i}$ と書く.) 各 $p_i(x'|x)$ に対応する $q_i(x'|x)$ は, x_i 以外の要素を $x'_{-i} = x_{-i}$ と固定した x' を提案する. 一般的

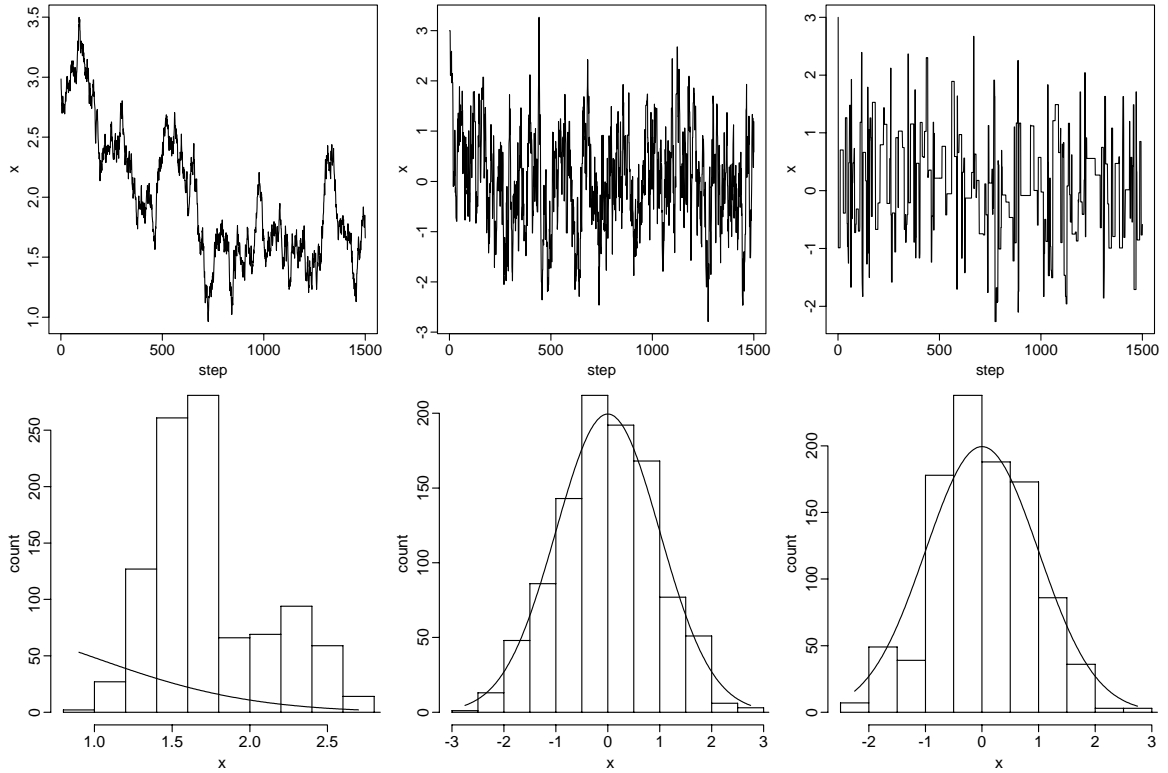


図 1: [上段] $x^{(t)}$ のプロット (左から $d = 0.1, 1, 10$) . $\{x^{(t)}\}_{t=501, \dots, 1500}$ のうち移動が accept された回数は, それぞれ 964, 819, 142 . [下段] $\{x^{(t)}\}_{t=501, \dots, 1500}$ のヒストグラムと $\pi(x)$.

にはいくつかの要素をまとめてブロックとした $A_i \subset K$ に対し, $x'_{-A_i} = x_{-A_i}$ と固定した x' の proposal density を $q_i(x'_{A_i}|x)$ と書くと, (3) は

$$\alpha_i(x'_{A_i}; x) = \min \left\{ 1, \frac{q_i(x_{A_i}|x'_{A_i}, x_{-A_i})\pi(x'_{A_i}|x_{-A_i})}{q_i(x'_{A_i}|x_{A_i}, x_{-A_i})\pi(x_{A_i}|x_{-A_i})} \right\}$$

となる. この場合, 各 $p_i(x'|x)$ は $\pi(x)$ に関して irreducible では無いが, ハイブリッドサンブラ全体として irreducible にすれば良い.

例 3 (Gibbs サンブラ) 各ブロック A_i について, x_{-A_i} を与えた時の x_{A_i} の条件付分布 (full conditional) $\pi(x_{A_i}|x_{-A_i})$ に従う確率変数が容易に生成できる場合には, Gibbs サンブラが便利である. $q_i(x'_{A_i}|x) = \pi(x'_{A_i}|x_{-A_i})$ とすると, $\alpha_i(x'_{A_i}; x) = 1$ となるので, full conditional に従って提案された x' は必ず accept される. この方法は heat bath 法として統計物理で以前から知られているが, Geman and Geman (1984) 以来 Gibbs サンブラという名称が統計分野で用いられるようになった.

3 MCMC の様々な利用

3.1 Unnormalized Density

$z = (x, y)$ とし, $z_K = x, z_L = y$ となるように, z の添字の集合を $K \cup L$ とする. z の規格化されていない密度関数 $h_\phi(z)$ は計算できるが, 規格化定数 $c(y, \phi) = \int h_\phi(x, y)\mu(dx)$ の計算が困難なもの考える. これから $f(x|y, \phi) = h_\phi(x, y)/c(y, \phi)$ を定常分布 $\pi(x)$ にもつ MCMC を構成するには, 規格化定数を計算する必要はない. $\alpha(x'; x)$ の計算で, $\pi(x')/\pi(x) = h_\phi(x', y)/h_\phi(x, y)$ となる.

特に $K \cup L$ の部分集合の族 $\Lambda \subset 2^{\{K \cup L\}}$ に関して

$$h_\phi(z) = \prod_{A \in \Lambda} H_A(z_A, \phi)$$

と factorization できる場合を考える． $|A|$ のほとんどが比較的小さい場合は，single-component Metropolis-Hastings サンプラは比較的少ない計算量で構成できる． $z_{-i} = z_{(K \cup L) \setminus \{i\}}$ と書くと，full conditional は，

$$f(z_i | z_{-i}, \phi) \propto \prod_{A \in \Lambda_i} H_A(z_A, \phi) \quad (6)$$

ただし， $\Lambda_i = \{A \in \Lambda | i \in A\}$ である．Markov random field (MRF) や，spatial statistics, time series など扱われる多くのモデルが，このクラスに入る．

例 4 (グラフィカルモデル) $z_i, i \in K \cup L$ をノードにもつ，サイクルの無い有向グラフ (directed acyclic graph, DAG) によって変数間の確率的依存関係を表現したモデルを考える (Lauritzen 1996)． z_i へ向かう枝をもつノードを $z_j, j \in \text{pa}(i)$ と書き，

$$f(z | \phi) = \prod_{i \in K \cup L} f(z_i | z_{\text{pa}(i)}, \phi)$$

によってモデルを定義する． $A_i = \{i\} \cup \text{pa}(i)$ とおくと， $\Lambda = \{A_i | i \in K \cup L\}$, $\Lambda_i = \{A_i\} \cup \{A_j | i \in \text{pa}(j)\}$ である．なお，conjugate prior をうまく利用した場合は容易に (6) に従う確率変数が生成できて Gibbs サンプラが利用できる．そうでない場合も，例 6 の ARS を (6) に適用して $z_i \sim f(z_i | z_{-i}, \phi)$ を容易に生成できる事が多い．

3.2 Importance Sampling

$f(x)$ と $g(x)$ は測度 $\mu(\cdot)$ 上の密度関数で， $w(x) = f(x)/g(x) < \infty$ a.e. とする．

$$\int a(x) f(x) \mu(dx) = \int a(x) w(x) g(x) \mu(dx) \quad (7)$$

なので，サンプル $x^{(1)}, \dots, x^{(M)} \sim g(x)$ があれば，

$$\frac{1}{M} \sum_t a(x^{(t)}) w(x^{(t)})$$

によって (7) が推定できる． $w(x^{(t)})$ の変動が大きいと推定が不安定になるので， $g(x) \approx f(x)$ の方が良い．importance sampling の考え方は， $\pi(x) = f(x|y, \phi)$ とするサンプラの構成が困難な場合やその収束性に問題がある場合など，いろいろな形で MCMC と組み合わせて利用できる．

例 5 (SIR アルゴリズム) $x^{(1)}, \dots, x^{(M)} \sim g(x)$ を計算した後で， $w(x^{(t)})$ で重みをつける代わりに，各々 $w(x^{(t)}) / \sum_t w(x^{(t)})$ の確率で $\{x^{(t)}\}$ から重複を許してリサンプリングする．得られたサンプルは $x^{*(1)}, \dots, x^{*(M^*)} \sim f(x)$ である．これは，Rubin (1987) の sampling/importance resampling (SIR) アルゴリズムと呼ばれる．時系列など次々とデータが入ってくるモデルでは，再帰的に SIR を用いる事で計算量を減らせる．Berzuini (1996), Berzuini ら (1996) の recursive updating や Kitagawa (1996) の Monte-Carlo filter はこの例である．

例 6 (Rejection Sampling) $f(x) \leq c g(x)$ をみたく x の密度関数 $g(x)$ と定数 c が既知とする． $x^{(1)}, \dots, x^{(M)} \sim g(x)$ の時，各 $x^{(t)}$ を確率 $f(x^{(t)}) / (c g(x^{(t)}))$ で accept し，それ以外は reject すると，accept された $x^{(t)}$ は $f(x)$ に従う．特に， $\log f(x)$ が凸関数なら，adaptive rejection sampling (ARS) が使えて， $c g(x)$ を事前に与える必要がない (Gilks and Wild 1992, Gilks 1996)．canonical link をもつ一般化線形モデル (GLIM) の full conditional (6) はすべてこの log-concave のクラスである (Dellaportas and Smith 1993)．

例 7 (ϕ の最尤推定) 経験ベイズのハイパーパラメタの推定 $\hat{\phi}$ は, 欠測データがある場合の最尤推定 (MLE) $f(y|\hat{\phi}) = \max_{\phi} f(y|\phi)$ とみなせる. Dempster ら (1977) の EM アルゴリズムは $\hat{\phi}$ を計算する一つの方法であり, $\hat{\phi}$ の仮の推定値 ϕ_0 から

$$Q(\phi, \phi_0) = E(\log f(x, y|\phi) | y, \phi_0)$$

を最大にする ϕ を計算し, これを ϕ_0 に置き換える事を反復して $\hat{\phi}$ を求める. $Q(\phi, \phi_0)$ の計算が困難な場合にこれを MCMC で行なうことも考えられるが, Geyer and Thompson (1992), Geyer (1993, 1994, 1996) は, importance sampling を利用した別のアプローチを示した.

$$\frac{f(y|\phi)}{f(y|\phi_0)} = E\left(\frac{f(x, y|\phi)}{f(x, y|\phi_0)} \middle| y, \phi_0\right) \approx \frac{1}{M} \sum_{t=1}^M \frac{f(x^{(t)}, y|\phi)}{f(x^{(t)}, y|\phi_0)} \quad (8)$$

ただし, $x^{(1)}, \dots, x^{(M)} \sim f(x|y, \phi_0)$ は MCMC で求め, その後で Newton-Raphson など (8) の最大化により $\hat{\phi}$ を求める. これは, Newton-Raphson の反復のたびに MCMC をする方法よりは, (原理的には) 効率が良い. 実際には, ϕ が ϕ_0 から遠いと (8) の近似が不安定になるので, ϕ_0 の適当な近傍で (8) を最大にする ϕ を求めてそれを ϕ_0 に置き換えて反復計算する. 反復のたびに MCMC サンプル $\{x^{(t)}\}$ を捨ててしまわないで, 過去のすべてのサンプルを合わせた物が $f(x|y, \phi_i), i = 0, 1, 2, \dots$ の mixture からのサンプルとみなす reweighting mixture (Geyer 1993) は, より収束が早いと期待される. なお, unnormalized density $h_{\phi}(z) \propto f(y, x|\phi)$ しか分からない時は, 規格化定数 $\int h_{\phi}(z)\mu(dz)$ も importance sampling で求める必要がある (Geyer 1994).

例 8 (周辺尤度の計算) モデルの評価などで $f(y|\phi) = \int f(y, x|\phi)\mu(dx)$ の値が計算したい場合がある. これは, 規格化定数 c が未知の密度関数 $h(x)/c$ に従うサンプル $x^{(1)}, \dots, x^{(M)}$ と, 適当な密度関数 $g(x)$ から importance sampling の式を 2 回使って,

$$f(y|\phi) \approx \sum_t \frac{f(y, x^{(t)}|\phi)}{h(x^{(t)})} \bigg/ \sum_t \frac{g(x^{(t)})}{h(x^{(t)})} \quad (9)$$

と推定できる (Gelfand and Dey 1994; Raftery 1996). 和の各項の変動を十分小さくすると良い推定が得られる. 特に $h(x) = f(y, x|\phi)$ とすれば分子の各項は 1 になり, 分母だけを計算すれば良い. Newton and Raftery (1994) は $g(x)$ を x の prior $f(x|\phi)$ にしたが, (9) は不安定になる. $g(x)$ が近似的に $h(x)/c$ であるものを探す必要がある.

別のアプローチとして, $c(0) = \int h_0(x)\mu(dx)$ が既知で $h_1(x) = f(y, x|\phi)$ となるような, s に関して滑らかな $h_s(x), s \in [0, 1]$ を定義し, s に沿って importance sampling を繰り返し行ない, $c(s) = \int h_s(x)\mu(dx)$ を次々と求める方法がある (Jerrum and Sinclair 1993). $0 = s_0 < s_1 < \dots < s_{m+1} = 1$ とする. $x^{(i,1)}, \dots, x^{(i,M)} \sim h_{s_i}/c(s_i)$ が MCMC サンプルとすると

$$\frac{c(s_{i+1})}{c(s_i)} \approx \frac{1}{M} \sum_t \frac{h_{s_{i+1}}(x^{(i,t)})}{h_{s_i}(x^{(i,t)})} \quad (10)$$

これを $i = 0, \dots, m$ について行ない, $c(0)$ から $f(y|\phi) = c(1)$ を得る. 一方, Ogata (1990) は, $\log(c(1)/c(0)) = \int_0^1 (d \log c(s)/ds) ds$ と書ける事を利用して,

$$\frac{d \log c(s)}{ds} \bigg|_{s_i} \approx \frac{1}{M} \sum_t \frac{\partial \log h_s(x^{(i,t)})}{\partial s} \bigg|_{s_i} \quad (11)$$

から s に関する 1 次元積分として $c(1)/c(0)$ を求めた. これは, (10) の対数が $\log c(s_{i+1}) - \log c(s_i) \approx (d \log c(s)/ds)|_{s_i}(s_{i+1} - s_i)$ と近似できることに対応している. Huang and Ogata (1997) では (10) と (11) の比較をしている. 小川と江口 (to appear) では $\{s_i\}$ の取り方を議論している. なお, これらとは別の方法を, $f(y|\phi) = \int f(y, x|\phi)/f(x|y, \phi)$ が任意の x で成り立つ事を利用して Chib (1995) が与えている.

3.3 x の次元が変わる場合

ここまで $x \in E \subset \mathcal{R}^k$ としたが, x の次元 $k = \dim(x)$ も確率変数とした場合にも, Metropolis-Hastings アルゴリズムは使える. ただし次元が変わる時の状態遷移を適切に表現するためには, 測度 $\mu(\cdot)$ 上の確率密度による議論は使えない. x は各 $E_i \subset \mathcal{R}^{k_i}$ のどれかの値を取るとし, $X = (i, x) \in E = \bigcup_i (\{i\} \times E_i)$ を状態空間とする. $q_m(dX'|X)$, $m = 1, 2, \dots$ を, X から m 番目の移動タイプを選び X' に遷移する mixture hybrid proposal kernel とすると, (3) と同じように

$$\alpha_m(X'; X) = \min \left\{ 1, \frac{q_m(dX|X')\pi(dX')}{q_m(dX'|X)\pi(dX)} \right\} \quad (12)$$

とすれば, 確率測度 $\pi(dX)$ に従うサンプルが得られる. Green (1995) は, この reversible-jump Metropolis-Hastings サンプラの一つの形式を構成的に与え, “template” と呼んだ.

ここで, 詳細釣合を保つためには, 次元を変える移動は “dimension-matching” が必要である. 簡単のため E_1 と E_2 だけを考え, 移動タイプ m は $E_1 \leftrightarrow E_2$ とする. 例えば, $k_1 = 1, k_2 = 2$ で $E_2 \rightarrow E_1$ は $X' = (1, (x_1 + x_2)/2)$ とすると, $E_1 \rightarrow E_2$ は $x'_1 + x'_2 = 2x_1$ を満たす必要がある. 具体的には, $x'_1 = x_1 + u_1, x'_2 = x_1 - u_1, u_1 \sim q_1(\cdot)$ などとすれば良い. 一般的には, 確率 $r_m(i, x)$ で移動タイプ m を選び, $u_i \sim q_i(u_i)$ を発生させ, $x' = x'(i, x, u_i)$ を提案するなら, $k_i + \dim(u_i) = k_{i'} + \dim(u_{i'})$ で $((i, x), u_i) \leftrightarrow ((i', x'), u_{i'})$ の bijection が存在する必要がある. このとき (12) は

$$\alpha_m(X'; X) = \min \left\{ 1, \frac{q_{i'}(u_{i'})r_m(i', x')\pi(i', x') \frac{\partial(x', u_{i'})}{\partial(x, u_i)}}{q_i(u_i)r_m(i, x)\pi(i, x) \frac{\partial(x, u_i)}{\partial(x', u_{i'})}} \right\}$$

となる. 一方, これより以前に提案された jump-diffusion サンプラ (Grenander and Miller 1994; Phillips and Smith 1996) では, x は Metropolis-Hastings サンプラで jump をして E_i 間を動き, 次の jump までの間は連続時間の Langevin 拡散過程で E_i 内を動く. しかし実際の計算では拡散過程は離散時間で近似するなど問題もある.

以下に示すベイズ的応用では, 各 E_i はデータ y を表現するひとつのモデルを表し, MCMC サンプルにおける E_i の出現頻度 (を prior $f(E_i)$ の比で調整したもの) から, 各モデル間の Bayes factor (式 (14)) が得られる. これはベイズ流のモデル選択を与える. また, MCMC サンプルを各 E_i に条件付けして取り出す事により, $f(x|y, \phi, E_i)$ に従うサンプルも得られる.

例 9 (変化点解析) $x(t)$ を rate とする区間 $[0, L]$ の Poisson 過程からのサンプルを $\{y_1, \dots, y_n\}$ とし, $x(t)$ が i 個の jump をもつ step 関数とする. jump の数, 位置, 高さに適当な prior を入れたベイズ解析を Green (1995) は示した. i が変わるとパラメタ数が変わるので, reversible-jump MCMC によって posterior にしたがうサンプルを生成している. ここでは同様に画像の segmentation も扱っている. Phillips and Smith (1996) では jump-diffusion サンプラでこれらを扱った.

例 10 (mixture モデル) i 個の正規分布 $N(\mu_j, \sigma_j^2)$, $j = 1, \dots, i$ を成分に持つ mixture からのサンプルを $\{y_1, \dots, y_n\}$ とする. 各成分の重み w_j や (μ_j, σ_j^2) , および成分の数 i に prior を入れたベイズ解析を Richardson and Green (1996) は与えた. Phillips and Smith (1996) でも扱われている.

例 11 (集団遺伝学の問題) ある生物種の n 個体を有効集団サイズ N_e の母集団からサンプルし, DNA データ $\{y_1, \dots, y_n\}$ を取る. これは過去に遡った n 個体の系図 X と, それに沿った頻度 μ の DNA 突然変異過程を表す確率モデル $f(y|X)$ で表現できる. 系図 X のトポロジが E_i に対応し, 各 E_i における x は枝の長さ (時間) に関するパラメタである. Kuhner ら (1995) は,

$\phi = N_e \mu$ をパラメタとする coalescent 過程 $f(dX|\phi)$ を simulate する形で系図のトポロジを部分的に変えていくアルゴリズムを用い, $f(dX|y, \phi_0)$ に従う MCMC サンプルを生成し, Geyer の importance sampling の手法を組み合わせ $\hat{\phi}$ を求めた. さらに, migration 過程も含めた解析も行なわれている¹. また, 同様の手法は系統樹推定にも用いられている (Li ら 1996).

3.4 モデルの良さ

仮定したモデル $f(y, x|\phi)$ が妥当かどうか, いくつかの候補の中でどれが良いのか, などを確かめる事は重要である. ひとつの方法は, “posterior predictive density” をデータと比較する事である (Gelfand 1996, Gelman and Meng 1996 など). x の条件付の y の適当な密度関数 $g(y|x, \phi)$ より, 将来の y' の予測分布

$$g(y'|y, \phi) = \int g(y'|x, \phi) f(x|y, \phi) \mu(dx) \quad (13)$$

を求め, この profile と実際の y を比較したり, $g(y|y, \phi)$ の値からモデルの良さを調べる. ここで, $g(y|x, \phi)$ としては $x = (\xi, \theta)$ から欠測データ ξ を除いたパラメタ θ に関する条件付分布 $f(y|\theta, \phi)$ が自然である. しばしば $f(y|\phi)$ を最大にする $\hat{\phi}$ で ϕ を置き換える. (13) の計算は, 積分を MCMC サンプルの和で推定し $(1/M) \sum_t g(\cdot|x^{(t)}, \phi)$ とするか, $g(\cdot|x, \phi)$ に相当する MCMC サンプラを x のサンプラと同時に走らせ, (13) に従うサンプル $\{y^{(t)}\}$ を直接得るなどする.

ベイズ流の別のアプローチとして Bayes factor (Kass and Raftery 1995; Raftery 1996 など)

$$\text{BF}_{12} = \frac{f(y|\phi, E_1)}{f(y|\phi, E_2)} \quad (14)$$

がある. ここで, モデル候補 i を $f(y, x|\phi, E_i)$ で表し, $f(y|\phi, E_i) = \int f(y, x|\phi, E_i) \mu(dx)$ である. (x や ϕ は E_i に依存している.) BF_{12} は y の周辺密度の比であり, この大きさがモデル E_1 のモデル E_2 に対する良さを与える. これはハイパーパラメタに関する二つのモデルの尤度比に他ならない. ϕ を MLE によって置き換える場合は, モデル $f(y|\phi)$ の良さは Akaike (1974) の AIC (ベイズモデルに使った場合は ABIC とも呼ばれる)

$$\text{AIC} = -2 \log f(y|\hat{\phi}) + 2 \dim(\phi)$$

で測る事もある. 二つのモデルの AIC の差は BF_{12} から直ちに得られる. E_1 と E_2 の x の空間が共通なら例 7 と同様に importance sampling を使って BF_{12} が計算できる. 一般の場合でも E_1 と E_2 を同時に含むモデルの posterior の MCMC が 3.3 節の方法で構成できれば, やはり BF_{12} が計算できる. このような方法が使えない時は例 8 の方法で $f(y|\phi, E_i)$, $i = 1, 2$ を計算し, これらの比として BF_{12} を得る.

4 MCMC の実装

4.1 誤差の評価と収束の診断

仮に $x^{(0)} \sim \pi(x)$ ならば (4) より $\{x^{(t)}\}$ は $\pi(x)$ に「収束」しているが, $\{x^{(t)}\}$ の自己相関が大きければ, 有限の M の (2) が (1) の良い推定だとは限らない. MCMC を使った $\hat{\mu}_M = (1/M) \sum_t a(x^{(t)})$ が, $\mu = E(a(x)|\pi)$ の良い推定量であるためには, μ に影響を与えるような $|a(x)|\pi(x)\mu(dx)$ の大きい $x \in E$ を $\{x^{(1)}, \dots, x^{(M)}\}$ がくまなく代表している必要がある. $\hat{\mu}_M$ の標準誤差を Geyer (1992) は議論している. 適当な正則条件下で, 中心極限定理 $\sqrt{M}(\hat{\mu}_M -$

¹Peter Beerli らの研究 (<http://evolution.genetics.washington.edu/PBhtmls/beerli.html>)

$\mu) \rightarrow N(0, \sigma^2)$ と言える． $\sigma^2 = \gamma_0 + 2 \sum_{i \geq 1} \gamma_i$, $\gamma_i = \text{cov}(a(x^{(t)}), a(x^{(t+i)}))$ であるが, γ_i の推定 $\hat{\gamma}_i = (1/M) \sum_{t=1}^{M-i} (a(x^{(t)}) - \hat{\mu}_M)(a(x^{(t+i)}) - \hat{\mu}_M)$ を単純に使うと σ^2 の推定の誤差が $O(1)$ になってうまく行かない． $\Gamma_m = \gamma_{2m} + \gamma_{2m+1}$ が, $\Gamma_1 > \Gamma_2 > \dots > 0$ であること等を利用した推定をする．これより, 所要の精度を得るのに必要な M を決める．現実には $x^{(0)} \sim \pi(x)$ とは出来ないから, $x^{(m)}$ がほぼ $\pi(x)$ に従うまでの最初の m サンプルを捨てて, (5) のように推定する． Raftery and Lewis (1992) は, MCMC サンプルから m や M を決める方法を与えている．

MCMC が理論的に irreducible でも, $\pi(x)$ が multimodal で一つの範囲から $\{x^{(t)}\}$ がなかなか抜け出せない場合, 上で述べた方法ではこれを検出できない．ひとつの現実的な方法は, 初期値 $x^{(i,0)}$ を変えながら $i = 1, \dots, l$ についてそれぞれ MCMC を実行し, $\{x^{(i,t)}\}_{t=1, \dots, M}$ の時系列としての l 個のプロットを眺めて (monitoring), m や M を判断する． Gelman and Rubin (1992) の potential scale reduction factor は, $\hat{\mu}_M$ の標準誤差が $M \rightarrow \infty$ でどれだけ小さくできるかを表していて, $\sqrt{\hat{R}} \approx \sqrt{1 + B/W}$ と近似できる．ただし W は各 MCMC サンプル $\{a(x^{(i,t)})\}_{t=1, \dots, M}$ の分散の平均, B は各 MCMC の $\hat{\mu}_M$ 間の分散である．

4.2 ダイナミクスの改良

$\{x^{(t)}\}$ の自己相関が極端に大きい場合, proposal density $q(x'|x)$ を工夫しマルコフ連鎖のダイナミクスを変えて, $\{x^{(t)}\}$ が良く「混ざる」(mixing) ようにする． Gilks and Roberts (1996) ではこのための技法をまとめている． Gibbs サンプラなどの single-component MCMC では, 関係の大きい成分を別々に update すると, mixing が悪くなる傾向がある．このような場合, 関係の深い x の成分をまとめて block として update するか, x の成分間がなるべく独立になるように reparametrization するのが, もっとも基本的な改良法である．ただし block をするとかえって mixing が悪くなる場合もあり注意が必要 (Roberts and Sahu 1997)．変数の関連に沿った updating order をとると mixing が良い． Gelfand ら (1995) の hierarchical centering など, 実際のモデルに即した reparametrize の方法も提案されている．

Besag and Green (1993), Higdon (1996) など議論されている方法は, auxiliary variable u の分布 $\pi(u|x)$ を工夫して与え, $(u, x) \sim \pi(u|x)\pi(x)$ となるように u も含めた MCMC を構成し mixing を改善する． Geyer (1991) の Metropolis-coupled MCMC では, $\pi_i(x) \propto \pi(x)^{1/i}$, $i = 1, \dots, l$ のように少しずつ定常分布のことなる MCMC を並列に走らせる．適宜ステップ t で状態 $x^{(i,t)}$ と $x^{(j,t)}$ を $\min \{1, \pi_i(x^{(j,t)})\pi_j(x^{(i,t)})/[\pi_i(x^{(i,t)})\pi_j(x^{(j,t)})]\}$ の確率で入れ換えても, 全体の定常分布は変わらない． i を大きくすると $\pi^{1/i}$ の multimodality が flat になって行くので, この状態交換により $\{x^{(1,t)}\}$ の mixing も改善すると期待できる． Geyer and Thompson (1995) の simulated tempering では, l 個の MCMC を並列に走らせる代わりに, i も状態に含めて一つの MCMC 系列となるように工夫したものである．

5 おわりに

MCMC の研究は現在盛んに進められており, technical report など最新の成果は WWW²を通して手にはいる．研究動向をサーベイした文献も多くある． Brooks (1997) は最新の動向まで含めて良くまとまっている． Gilks ら (1996) は, 25 編の論文からなり, 基礎から応用まで網羅している． Cowles and Carlin (1996) はそれまでに提案された 13 種類の収束の診断法を比較している． Kass ら (1997) は, この分野の研究者の本音がわかって面白い．日本語の文献として

²プレプリントサーバが <http://www.stats.bris.ac.uk/MCMC/> にある．

は伊庭 (1996)³が参考になる . Gibbs サンプラの汎用ソフト “BUGS”⁴は , グラフィカルモデルで $f(y, x|\phi)$ が記述でき , GLIM を含む多くの要素が実装されている . そのマニュアルと実例集 (Spiegelhalter ら 1996a,b) は , Gibbs サンプラの応用を理解するのに大変役立つ . BUGS の出力から収束の診断を行なうために S 言語で記述された “CODA” も便利である .

参考文献

- [1] AKAIKE, H. A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **19**, 6 (1974), 716–723.
- [2] BERZUINI, C. Medical monitoring, in Gilks, et al. [24], chapter 18, 321–337.
- [3] BERZUINI, C., BEST, N. G., GILKS, W. and LARIZZA, C. Dynamic Conditional Independence Models and Markov Chain Monte Carlo Methods, Technical report, Universita’ di Pavia (Sept. 1996).
- [4] BESAG, J. and GREEN, P. J. Spatial Statistics and Bayesian Computation, *J. Roy. Statist. Soc. Ser. B*, **55** (1993), 25–37, (Discussion: p53-102).
- [5] BROOKS, S. Markov Chain Monte Carlo and its Application, Technical report, University of Bristol (July 1997).
- [6] CHIB, S. Marginal Likelihood From the Gibbs Output, *J. Amer. Statist. Assoc.*, **90** (1995), 1313–1321.
- [7] COWLES, M. K. and CARLIN, B. P. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *J. Amer. Statist. Assoc.*, **91** (1996), 883–904.
- [8] DELLAPORTAS, P. and SMITH, A. F. M. Bayesian inference for generalised linear and proportional hazards models via Gibbs sampling, *Appl. Statist.*, **42** (1993), 443–460.
- [9] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. Maximum Likelihood From Incomplete Data Via the EM Algorithm, *J. Roy. Statist. Soc. Ser. B*, **39** (1977), 1–22, (C/R: p22-37).
- [10] GELFAND, A. E. Model determination using sampling-based methods, in Gilks, et al. [24], chapter 9, 145–161.
- [11] GELFAND, A. E. and DEY, D. K. Bayesian Model Choice: Asymptotics and Exact Calculations, *J. Roy. Statist. Soc. Ser. B*, **56** (1994), 501–514.
- [12] GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. Efficient Parametrisations for Normal Linear Mixed Models, *Biometrika*, **82** (1995), 479–488.
- [13] GELMAN, A. and MENG, X.-L. Model checking and model improvement, in Gilks, et al. [24], chapter 11, 189–201.
- [14] GELMAN, A. and RUBIN, D. B. Inference From Iterative Simulation Using Multiple Sequences, *Statist. Sci.*, **7** (1992), 457–472, (Discussion: p483-501, 503-511).
- [15] GEMAN, S. and GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intel.*, **6** (1984), 721–741.
- [16] GEYER, C. Estimating Normalizing Constants and Reweighting Markov Chain Monte Carlo, Technical report, University of Minnesota, USA (1993).
- [17] GEYER, C. J. Markov Chain Monte Carlo maximum likelihood, Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface (ed. Keramidas, E. M.), Fairfax Station (1991), Interface Foundation.
- [18] GEYER, C. J. Practical Markov Chain Monte Carlo, *Statist. Sci.*, **7** (1992), 473–483, (Discussion: p483-503).
- [19] GEYER, C. J. On the Convergence of Monte Carlo Maximum Likelihood Calculations, *J. Roy. Statist. Soc. Ser. B*, **56** (1994), 261–274.
- [20] GEYER, C. J. Estimation and optimization of functions, in Gilks, et al. [24], chapter 14, 241–258.
- [21] GEYER, C. J. and THOMPSON, E. A. Constrained Monte Carlo Maximum Likelihood for Dependent Data, *J. Roy. Statist. Soc. Ser. B*, **54** (1992), 657–683, (Discussion: p683-699).
- [22] GEYER, C. J. and THOMPSON, E. A. Annealing Markov Chain Monte Carlo With Applications to Ancestral Inference, *J. Amer. Statist. Assoc.*, **90** (1995), 909–920.
- [23] GILKS, W. R. Full conditional distributions, in Gilks, et al. [24], chapter 5, 75–88.
- [24] GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J. eds. *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London/New York (1996).

³関連文献が <http://www.ism.ac.jp/~iba/> にある .

⁴BUGS と CODA は <http://www.mrc-bsu.cam.ac.uk/bugs/> でマニュアルとともに手にはいる .

- [25] GILKS, W. R. and ROBERTS, G. O. Strategies for improving MCMC, in Gilks, et al. [24], chapter 6, 89–114.
- [26] GILKS, W. R. and WILD, P. Adaptive Rejection Sampling for Gibbs Sampling, *Appl. Statist.*, **41** (1992), 337–348.
- [27] GREEN, P. J. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, **82** (1995), 711–732.
- [28] GRENDER, U. and MILLER, M. I. Representations of Knowledge in Complex Systems, *J. Roy. Statist. Soc. Ser. B*, **56** (1994), 549–581, (Discussion: p581-603).
- [29] HASTINGS, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, **57** (1970), 97–109.
- [30] HIGDON, D. Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications, Technical report, ISDS, Duke University (July 1996).
- [31] HUANG, F. and OGATA, Y. Comparison of two methods for calculating the partition functions of various spatial statistical models, Research Memorandum 643, The Institute of Statistical Mathematics (July 1997).
- [32] 伊庭幸人 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** (1996), 49–84.
- [33] JERRUM, M. and SINCLAIR, A. Polynomial-time approximation algorithms for the Ising model, *SIAM J. Comput.*, **22** (1993), 1087–1116.
- [34] KASS, R. E. and RAFTERY, A. E. Bayes Factors, *J. Amer. Statist. Assoc.*, **90** (1995), 773–795.
- [35] KASS, R., CARLIN, B. P., GELMAN, A. and NEAL, R. Markov Chain Monte Carlo in Practice: A Roundtable Discussion, Technical report, Department of Statistics, Carnegie Mellon University (Feb. 1997).
- [36] KITAGAWA, G. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *J. Computat. Graph. Statist.*, **5** (1996), 1–25.
- [37] KUHNER, M. K., YAMATO, J. and FELSENSTEIN, J. Estimating Effective Population Size and Mutation Rate From Sequence Data Using Metropolis-Hastings Sampling, *Genetics*, **140** (August 1995), 1421–1430.
- [38] LAURITZEN, S. L. *Graphical models*, Oxford Univ. Pr., Oxford (1996).
- [39] LI, S., PEARL, D. K. and DOSS, H. Phylogenetic Tree Construction Using Markov Chain Monte Carlo, Technical report, Iowa State University (Sept. 1996).
- [40] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. Equations of state calculations by fast computing machine, *J. Chem. Phys.*, **21** (1953), 1087–1091.
- [41] NEWTON, M. A. and RAFTERY, A. E. Approximate Bayesian Inference With the Weighted Likelihood Bootstrap, *J. Roy. Statist. Soc. Ser. B*, **56** (1994), 3–26, (Discussion: p26-48).
- [42] OGATA, Y. A Monte Carlo Method for an Objective Bayesian Procedure, *Ann. Inst. Statist. Math.*, **42** (1990), 403–433.
- [43] 小川奈美子, 江口真透 マルコフ連鎖モンテカルロ法による高次元積分, *統計数理* (to appear).
- [44] PESKUN, P. H. Optimum Monte-Carlo Sampling Using Markov Chains, *Biometrika*, **60** (1973), 607–612.
- [45] PHILLIPS, D. B. and SMITH, A. F. M. Bayesian model comparison via jump diffusions, in Gilks, et al. [24], chapter 13, 215–239.
- [46] RAFTERY, A. E. Hypothesis testing and model selection, in Gilks, et al. [24], chapter 10, 163–187.
- [47] RAFTERY, A. E. and LEWIS, S. M. How many iterations in the Gibbs sampler?, *Bayesian Statistics 4* (eds. Bernardo, J. M., Berger, J. O., Dawaid, A. P. and Smith, A. F. M.), Oxford Univ. Pr., Oxford (1992), 765–776.
- [48] RICHARDSON, S. and GREEN, P. On Bayesian analysis of mixtures with an unknown number of components, Technical report, INSERM, France and University of Bristol (Feb. 1996).
- [49] ROBERTS, G. O. and SAHU, S. K. Updating Schemes, Correlation Structure, Blocking and Parametrization for the Gibbs Sampler, *J. Roy. Statist. Soc. Ser. B*, **59** (1997), 291–317.
- [50] RUBIN, D. B. Comments on “The Calculation of Posterior Distributions By Data Augmentation”, *J. Amer. Statist. Assoc.*, **82** (1987), 543–546.
- [51] SPIEGELHALTER, D., THOMAS, A., BEST, N. and GILKS, W. BUGS 0.5 Bayesian inference Using Gibbs Sampling, Manual, MRC Biostatistics Unit, Institute of Public Health (Aug. 1996).
- [52] SPIEGELHALTER, D., THOMAS, A., BEST, N. and GILKS, W. BUGS 0.5 Example, Volume 1 and 2, MRC Biostatistics Unit, Institute of Public Health (Aug. 1996).
- [53] TIERNEY, L. Markov Chains for Exploring Posterior Distributions, *Ann. Statist.*, **22** (1994), 1701–1728, (Discussion: p1728-1762).