

メトロポリスのモンテカルロ法の 擬ベイズ法への応用

—— 変化点問題を例として ——

統計数理研究所 伊 庭 幸 人

(1991 年 12 月 受付)

1. 本論文の動機

ベイズ統計における MAP 解を求める手段及び周辺分布を求める手段として、メトロポリスのモンテカルロ法が注目されている (たとえば, Geman, S. and Geman, D. (1984), Geman and McClure (1987), Hinton and Sejnowski (1986), Gelfand et al. (1990), Ogata (1990), 伊庭 (1991)).

一方で Akaike (1973) による AIC の導入は、モデルのパラメータ推定だけでなく、モデル選択それ自身も AIC の最小化というひとつの変分原理で扱える可能性を示した。そこでまず考えられるのはパラメータ推定だけでなく、モデル選択に関する最適化も simulated annealing 法で行うことである。この場合、最小化すべき「エネルギー」として AIC を用いるわけである。

AIC の最も小さいモデルをひとつだけ選ぶという立場では、simulated annealing 法は適用可能だが、分布を生成する手段としてのメトロポリスのモンテカルロ法を用いることはできない。しかし、AIC によるモデル選択では、AIC が小さい側から何個かのモデルを持ってきて、それらを眺めて考える場合も多い。実際、AIC の差が微小なモデルが多くある場合でも、それらに共通な部分があればその部分は本当らしく思える。そこで、AIC について重み付けした推定値を計算するということが考えられ、分布を生成する手段としてのメトロポリスのモンテカルロ法を適用する余地が出てくる。

AIC の差が小さいモデルがいくつもある場合に、AIC について重み付けして推定値を計算する方法を、Akaike は擬ベイズ法 (quasi-Bayesian method) と呼んでいる (Akaike (1979, 1980), Kitagawa and Akaike (1982)). この方法では、モデル M の事前確率を $\pi(M)$ としたとき、

$$(1.1) \quad \frac{\exp\left(-\frac{1}{2} AIC(M)\right) \pi(M)}{\sum_M \exp\left(-\frac{1}{2} AIC(M)\right) \pi(M)}$$

をモデルに関する事後分布と見なし、さまざまな量の期待値や周辺分布を計算する。また、モデルを特定すると、AIC の代わりに漸近的に AIC に一致するもっと精密な表現が得られる場合があるが、それを使った場合も擬ベイズ法と呼ぶ。 $\frac{1}{2} AIC(M) - \log(\pi(M))$ を「エネルギー」としたメトロポリスのモンテカルロ法は擬ベイズ法における事後分布を生成することになる。

以上の議論は非常に一般的に見えるが、実際にメトロポリスのモンテカルロ法が有効なのは、状態空間内を（なんらかの意味で）微小な変化の積み重ねで移動できる場合に限られるので、モデル選択のような状況への適用性は限定される。とくに分布の生成に用いる場合は温度 $T=1$ で緩和しなければならないため、制限が厳しくなる（simulated annealing 法はもっと応用範囲が広いが、今度は他の最適化手法との効率の比較が問題となってくる）。

ここでは、変化点問題の擬ベイズ法での取扱い（Kashiwagi (1991)）に対してメトロポリスのモンテカルロ法を適用する。この問題では、変化点のラベルをイジングスピンで表現することにより、各モデルをうまく表現することができる。

ここで扱ったモデルは純ベイズ的立場から書かれた Smith の論文（Smith (1980)）がもとになっているが、Smith が来日したおりの談話（1991）によると、彼自身メトロポリスのモンテカルロ法を変化点問題に適用した研究を行っているらしい。結果は出版済みあるいは投稿中のようであったが、該当論文はまだ入手していない。また、Besag et al. (1991) に対する Raftery and Banfield による discussion（直後に収録）には、Carlin, Polson and Stoffer の研究が引用され、モデルの indicator を付加的なパラメータとして Gibbs sampler を使用することによりモデルの不確実性を取り入れる等と紹介されている。詳細な点（擬ベイズ法を用いるかどうかなど）は不明であるが、思想的には本稿と類似した研究と思われる。

情報量基準をホップフィールドの方法で最小化する研究として Sakai (1989) があるが、擬ベイズの見方は含まれていないと思われる。

2. 例題の説明

2.1 変化点問題と貼り合わせモデル

時系列データの変化点問題とは、時系列データの性質の変化する点を捜す問題の総称である。ここでは、そのうち、‘2 項分布の貼り合わせ’の問題を考える（Smith (1980), Kashiwagi (1991)）。以下、この問題を Smith (1980) の例で説明する。

表 1 はある古文書（Lindisfarne scribes）について、ある事象（特殊な綴字法の出現）の頻度を表にしたものである。章の数を n （上の例では $n=13$ ）、 t 章での試行数（事象が起こる可能性

表 1.

t	m_t	y_t	y_t/m_t
1	21	9	0.43
2	36	10	0.28
3	44	13	0.30
4	30	6	0.20
5	52	24	0.46
6	45	11	0.24
7	48	9	0.19
8	57	11	0.19
9	48	7	0.15
10	22	3	0.14
11	20	3	0.15
12	21	4	0.19
13	20	4	0.20

のある箇所の数)を m_t , 事象の生起数を y_t とする. 著者が変われば文体も変わると考えられる. これを利用して, この表からどこで著者が変わったかを推定せよというのが問題 (Lindisfarne scribes problem) である. 各章は年代順に書かれたと仮定しているので, 問題は本質的に 1 次元的である. また, 2 章を書いた人が, 3 章は休んで, 4 章からまた復活した, などということは考えない. また, 以下では $m_t=0$ となる t がない場合を考える.

ここで, 変化点の間の区間では事象が生起確率一定の 2 項分布に従うと仮定しよう (‘貼り合わせモデル’). $j-1$ 番目の変化点から j 番目の変化点までの事象の生起確率を θ_j とする. 最初から 1 番目の変化点までの事象の生起確率は θ_1 , 最後 (k 番目) の変化点のあとの事象の生起確率は θ_{k+1} とする.

すると, 変化点の配置 M を固定した場合の $\{\theta_j\}$ に関する尤度は,

$$(2.1) \quad L_M(\{m_t, y_t\} | \{\theta_j\}) = \prod_{t=1}^n \binom{m_t}{y_t} \theta_{j(t)}^{y_t} (1 - \theta_{j(t)})^{m_t - y_t}$$

と書ける. ただし, $\theta_{j(t)}$ は上記の貼り合わせモデルにおける t 章における事象の生起確率である.

変化点の配置を固定した場合の $\{\theta_j\}$ の最尤推定値は簡単に求められて,

$$(2.2) \quad \hat{\theta}_j = \frac{\sum_{g(j-1) < t \leq g(j)} y_t}{\sum_{g(j-1) < t \leq g(j)} m_t}$$

となる ($j=1, \dots, k+1$). ここで, j 番目の変化点の位置 $g(j)$ は変化点が t 章と $t+1$ 章の間にあるとき $g(j)=t$ と定義し, $g(0)=0, g(k+1)=n$ とおいた. 最大対数尤度はこの値を (2.1) 式に代入することにより簡単に求められる.

以上, 式で書くと複雑だが実際には簡単なことである. たとえば, 表 1 のデータで 4 章と 5 章の間で著者の交代があったとする. すると, 1 章から 4 章までの特殊な文体の出現確率の最尤推定値は

$$(2.3) \quad \hat{\theta}_1 = \frac{9+10+13+6}{21+36+44+30} = \frac{38}{131},$$

5 章から 13 章までについては,

$$(2.4) \quad \hat{\theta}_2 = \frac{24+11+\dots+4}{52+45+\dots+20} = \frac{76}{333}$$

となる. この「4 章と 5 章の間で 2 項分布を貼り合わせたモデル」の最大対数尤度は,

$$(2.5) \quad 38 \log \left(\frac{38}{131} \right) + 93 \log \left(\frac{93}{131} \right) + 76 \log \left(\frac{76}{333} \right) + 257 \log \left(\frac{257}{333} \right)$$

である. ただし, ここではどのモデルにも共通な定数は除いた.

可能な変化点の組合せ (表 1 の例では 2^{12} 通り) のうちどれがもっともらしいか, という問題を考えるひとつの方法は, 変化点の組合せを与えることでひとつのモデル M が与えられたと考えて, AIC 最小原理ないし尤度比検定のような方法でモデル選択を行うことである. この場合, 上記の方法で求めた最大対数尤度が選択の基準になる. 別の立場としては Smith (1980) のような純ベイズ的な解析がある. この場合, すべてが分布の形で取り扱われるので, パラメータを fit して θ_i の最尤推定値や最大対数尤度を求めることはない. 以下で考える擬ベイズ法は両者の中間的方法である.

なお, 変化点の組を表わすのに, 変数 $\{S_t\}$ ($t=1, \dots, n-1, S_t \in \{0, 1\}$) を導入して, 点 t (t

章と $t+1$ 章の間の点) が変化点であるとき $S_t=1$, そうでないとき $S_t=0$ と表わすと便利である. このような「イジングスピン」の間の相互作用は簡単な形に書けないが, このデータ構造はメトロポリスのモンテカルロ法をコーディングする際にそのまま利用できる. 違うモデルを離散変数を導入することでひとつにまとめるという考え方は必ずしも新しいものではないが, もっと採用されてもよいと思われる. これは2次元画像処理の場合の 'line process' に通じる考え方でもある (Geman, S. and Geman, D. (1984)).

2.2 擬ベイズ法の適用

擬ベイズ法では, 変化点の組合せを与えることでモデル M が定められたと考え, 各 M について最大対数尤度を計算する. さらに, $AIC(M)$ を

$$(2.6) \quad -\frac{1}{2} AIC(M) = (\text{最大対数尤度}) - (\text{変化点の個数} + 1)$$

により定義する (以下では変化点の個数を k とする).

ここまでは通常のモデル選択と同じであるが, このあと, モデル M 同士の混合を考える. これは, 'モデルの事後確率' $P(M)$ を

$$(2.7) \quad P(M) = \frac{\exp\left(-\frac{1}{2} AIC(M)\right) \pi(M)}{\sum_M \exp\left(-\frac{1}{2} AIC(M)\right) \pi(M)}$$

で定めることにより可能になる.

ここで問題になるのは $\pi(M)$ の選び方である. すべての M が等確率というのは一見もっともらしいが, この場合適切とはいえない. これではどの点に変化点である事前確率も $1/2$ ということになるが, 変化点とは元来少ないはずのものである. もっと良いものとして, 次のような事前分布が考えられる.

2.2.1 事前分布 (1)

まず考えられる事前分布は, 各点が独立に確率 α で変化点である可能性があるとするものである (Akaike (1979, 1980)). モデル M の変化点の個数が $(n-1)$ 個中 k 個とすると,

$$(2.8) \quad \pi_\alpha(M) = \alpha^k (1-\alpha)^{(n-1)-k}$$

が M の事前確率になる.

これは, さきに導入した $\{S_t\}$ を利用すると,

$$(2.9) \quad \pi_h(M) = \frac{\exp(-h \sum_{t=1}^{n-1} S_t)}{(1 + \exp(-h))^{n-1}}$$

と書ける. ここで,

$$(2.10) \quad h = \log\left(\frac{1-\alpha}{\alpha}\right)$$

とおいた (S_t を $\{\pm 1\}$ でなく $\{0, 1\}$ にとったので, 普通の「イジングスピン」の場合によく現われるのと少し違う形になったが, 本質的には同じ).

この場合, α (または h) を決めなくては意味がないが, これは ABIC 法 (TYPE II 最尤法) によって決めることができる. これについてはあとで述べる.

2.2.2 事前分布 (2)

未知のハイパーパラメータを含まない事前分布としては、たとえば次のようなものが考えられる (Akaike (1980), Kitagawa and Akaike (1982), Kashiwagi (1991)).

$$(2.11) \quad \pi(M) = \frac{1}{n} \frac{1}{\binom{n-1}{k}}$$

これは、変化点の数 k が $\sim(n-1)/2$ 個であるようなモデルの数が組合せ論的に多いので、その寄与を打ち消すという形になっている。

上の説明はあまり説得力がないが、より納得のいく正当化としては、(2.8) 式で不明な α を $[0, 1]$ で平均して、

$$(2.12) \quad \int_0^1 \alpha^k (1-\alpha)^{(n-1)-k} d\alpha = \frac{\Gamma(k+1)\Gamma(n-k)}{\Gamma(n+1)} = \frac{1}{n} \frac{1}{\binom{n-1}{k}}$$

としたとも考えられる (Akaike (1980), Kitagawa and Akaike (1982)).

この事前分布を $\{S_t\}$ を使って表現しようとする、 S_t の間の高次の相互作用を含む極めて複雑な形になる。 n と k がともに大きくて、スターリングの公式が使える場合はやや簡単になって、

$$(2.13) \quad \pi(M) \sim \prod_t \exp(S_t \log u + (1-S_t) \log(1-u))$$

となる。ただし、 u は $\{S_t\}$ の関数で、

$$(2.14) \quad u = \frac{1}{n-1} \sum S_t$$

である。

2.3 ABIC 法の適用 (事前分布 (1) の場合)

普通のベイズの場合の ABIC 法と同様に、

$$(2.15) \quad -\frac{1}{2} ABIC(h) = \log \sum_M \exp\left(-\frac{1}{2} AIC(M)\right) \pi_h(M)$$

を最大化する h を選ぶ (Akaike (1979)). ここで、 \sum_M は 2^{n-1} 個のモデルに関する和を意味する。これは、モデル M を $\{S_t\}$ で表現すれば、あらゆる可能な $\{S_t\}$ の組合せについて和をとるのと同じである。

(2.9) を (2.15) に代入して、 h で微分すると、

$$(2.16) \quad -\frac{1}{2} \frac{\partial ABIC(h)}{\partial h} = -\left(\langle \sum_t S_t \rangle_{\text{pos}} - (n-1) \frac{\exp(-h)}{1+\exp(-h)}\right)$$

となる。ここで、 $\langle \rangle_{\text{pos}}$ は擬ベイズ的な事後分布

$$(2.17) \quad P(M) = P(\{S_t\}) = \frac{1}{Z} \exp(\text{モデルの最大対数尤度} - \sum_t S_t - h \sum_t S_t)$$

での期待値である (Z は規格化定数)。以下では、(2.16) の右辺を実際にプロットして零点を求めているが、十分条件の良い場合は 'EM 法' 的な iteration で解いても差し支えないと思われる。

上の式で求めた h を (2.9) に代入すると、推定された事前分布は、形式的に、

$$(2.18) \quad \pi(M) = \prod_i \exp(S_i \log v + (1 - S_i) \log(1 - v))$$

と書ける。ただし、 v は定数で、

$$(2.19) \quad v = \frac{1}{n-1} \langle \sum S_i \rangle_{\text{pos}}$$

である。ここで、 $\langle \sum S_i \rangle_{\text{pos}}$ は推定された事前分布から作られた事後分布による期待値である。

系が十分大きくて独立な多くの部分からなると見なせ、かつ変化点の数がある程度多いモデルのみを考えればよいときは、事前分布 (1) で ABIC 法を用いた場合と事前分布 (2) を用いた場合で、各点に変化点である事後確率はほぼ等しい値をとる。これを理解するには、式 (2.12) を用いて、事前分布 (2) に関する和を

$$(2.20) \quad \sum_M \exp\left(-\frac{1}{2} AIC(M)\right) \pi(M) = \int \sum_M \exp\left(-\frac{1}{2} AIC(M)\right) \alpha^k (1-\alpha)^{(n-1)-k} d\alpha$$

と書いた際に、 α に関する積分を鞍点で評価することが、事前分布 (1) における ABIC 法に丁度相当することに注意すればよい。

擬ベイズ法による結果が事前分布について頑健であることは、Akaike (1979, 1980) でも論じられているが、変化点問題の場合、考えられるモデル数が組合せ論的に多いため、事前分布 (1) で α を (たとえば) $1/2$ としたのでは、変化点の個数が非常に多くなってしまふ。これに対して、事前分布 (1) で正確な ABIC 法を用いた場合と事前分布 (2) を用いた場合では、各点に変化点である周辺事後確率について、良い一致が期待できる。実際、あとで見る例では、必ずしも上記の条件が満たされているとはいえないにもかかわらず、両者はかなり良く一致した。

擬ベイズ法では、パラメータの一部を fit してしまうので、AIC を使用する必要があった。こ

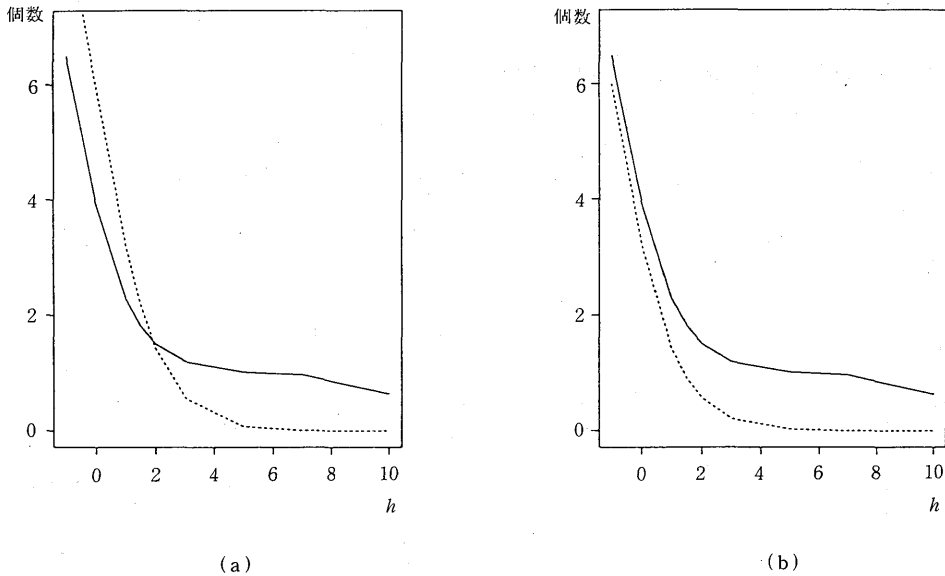


図1. AICを使う必然性。実線は $\langle \sum S_i \rangle_{\text{pos}}$ 。点線は (a) では $\frac{(n-1) \exp(-h)}{1 + \exp(-h)}$ ，
(b) では $\frac{(n-1) \exp(-(h+1))}{1 + \exp(-(h+1))}$ を表す。

表 2.

t	m_t	y_t	y_t/m_t
1	60	20	1/3
2	60	20	1/3
3	90	30	1/3
4	90	30	1/3
5	80	40	1/2
6	80	40	1/2
7	40	20	1/2
8	40	20	1/2
9	60	30	1/2
10	60	30	1/2
11	80	40	1/2
12	80	40	1/2
13	100	50	1/2

れは、最大対数尤度に対して変化点の数に比例した罰金を課することに当たる。(2.9) 式の h も変化点の数に対する罰金であり、どちらも S_t に対して 1 次の形（「磁場」の形）で入るので混同しやすいが、両者は別物である。(2.17) 式からわかるように、(2.16) の右辺のはじめの項の期待値を計算する場合には AIC の罰金項に起源を持つ罰金 1 と事前分布に起源を持つ罰金 h が両方課される。これに対し、あとの項を計算する際には h のみが登場する。AIC の代わりに罰金なしの最大対数尤度を用いると、(2.16) の代わりに、

$$(2.21) \quad -\frac{1}{2} \frac{\partial ABIC(h)}{\partial h} = -\left(\langle \sum_t S_t \rangle_{\text{pos}} - (n-1) \frac{\exp(-(h+1))}{1+\exp(-(h+1))} \right)$$

を使うのと同じことになるが、これではうまくいかない。

あとで示す方法で計算した $\langle \sum_t S_t \rangle_{\text{pos}}$ と $(n-1) \frac{\exp(-h)}{1+\exp(-h)}$ のグラフを図 1(a) に、 $\langle \sum_t S_t \rangle_{\text{pos}}$ と $(n-1) \frac{\exp(-(h+1))}{1+\exp(-(h+1))}$ のグラフを図 1(b) にそれぞれ示した。これは表 2 のような自明なデータに対するものであるが、後者では交点が存在せず、 $ABIC(h)$ の極値が存在しないことがわかる。

3. メトロポリスのモンテカルロ法の適用

3.1 アルゴリズム

擬ベイズ的な事後分布に関する期待値を計算するには、 2^{n-1} 個のモデルに関する平均を計算する必要がある。これにメトロポリスのモンテカルロ法を適用することを考える。

いまの問題では、 $\{S_i\}$ 間の相互作用は局所的でない。しかし、ひとつの S_{i^0} を除いて他を固定したときに、その S_{i^0} が 1 である (i^0 が変化点である) 確率は、それに最も近い左右 2 つの変化点の位置のみで決まるので、それを利用すると比較的簡単にメトロポリスのモンテカルロ法を構成することができる。以下、その要点を事前分布 (1) の場合について説明する。熱浴法 (Gibbs sampler) の場合を説明したが、狭義のメトロポリス法の場合も本質的な部分は変わら

ない。

1. 試行の対象となる S_{t^0} を選ぶ。

これは一定の法則で選んでもよいし、乱数を用いて選んでもよいが、初期条件を変えて緩和のチェックをする場合などを考えると、ランダムに選んだ方がより良いだろう。

2. ‘左側’の計算。

(a) $0 < t^L < t^0$ で、かつ、 $S_{t^L}=1$ であるような最大の t^L を探す。もし見つからないときは、 $t^L=0$ とする。

(b) $m_L = \sum_{t^L < t \leq t^0} m_t, \quad y_L = \sum_{t^L < t \leq t^0} y_t$

3. ‘右側’の計算。

(a) $t^0 < t^R < n$ で、かつ、 $S_{t^R}=1$ であるような最小の t^R を探す。もし見つからないときは、 $t^R=n$ とする。

(b) $m_R = \sum_{t^0 < t \leq t^R} m_t, \quad y_R = \sum_{t^0 < t \leq t^R} y_t$

4. これらを用いて条件付き確率を計算する。一般には、 $0 \log 0$ の処理に注意を要する。

(a) $-E_L = y_L \log \frac{y_L}{m_L} + (m_L - y_L) \log \frac{m_L - y_L}{m_L}$

(b) $-E_R = y_R \log \frac{y_R}{m_R} + (m_R - y_R) \log \frac{m_R - y_R}{m_R}$

(c) $-E_{sp} = -E_L - E_R$

(d) $-E_{mg} = (y_L + y_R) \log \frac{y_L + y_R}{m_L + m_R} + (m_L + m_R - y_L - y_R) \log \frac{m_L + m_R - y_L - y_R}{m_L + m_R}$

(e) $-\Delta E = -E_{sp} + E_{mg} - h - 1$

(f) $P_1 = \frac{1}{1 + \exp(\Delta E)}$

5. 熱浴法による updation.

$0 < rnd < 1$ の乱数 rnd を用いて、 $rnd < P_1$ なら $S_{t^0}=1$ 、さもなければ $S_{t^0}=0$ とする。

6. はじめに戻る。

上記の loop を $(n-1)$ 回まわることを 1 MCS と定義する。

事前分布 (2) は局所的ではないが、変化点の総数 h をつねに記録しておいて、それを使って条件付き確率を計算し、updation のたびに h を修正することにすれば、ほとんど同じ手間で計算を行うことができる。

メトロポリスのモンテカルロ法のプログラムは容易に作成できる。たとえば、実験のために作成したプログラムは、結果を印刷する部分を含めて 200~300 行たらず、このうちメトロポリス法の中核部分のサブルーチンは 80 行ほどであり、内容も非常にストレートである。

3.2 局所的極値と緩和

実験結果を述べる前に、上記のアルゴリズムの緩和の妨げになるような状況について簡単に考察しよう。最も簡単な自明でない状況としては、事象の生起確率（むしろ y_t/m_t というべきか）が段階的に変わる場合（図 2(a)）、山型もしくは谷型になっている場合がある（図 2(b)）。前者は‘OR 状況’でともいうべきものであり、全く変化点がないのに比べると一方に変化点がある可能性が高いが、2 つともあるとは考えにくい状況である。これは真の変化点の配置が図 2(a) のようになっていなくても、ある程度データが少なくなれば必ず生ずる状況であって、図 2(a) は極端な場合を示しただけである。逆に、後者は変化点があるとすれば 2 つあり、ないとすれ

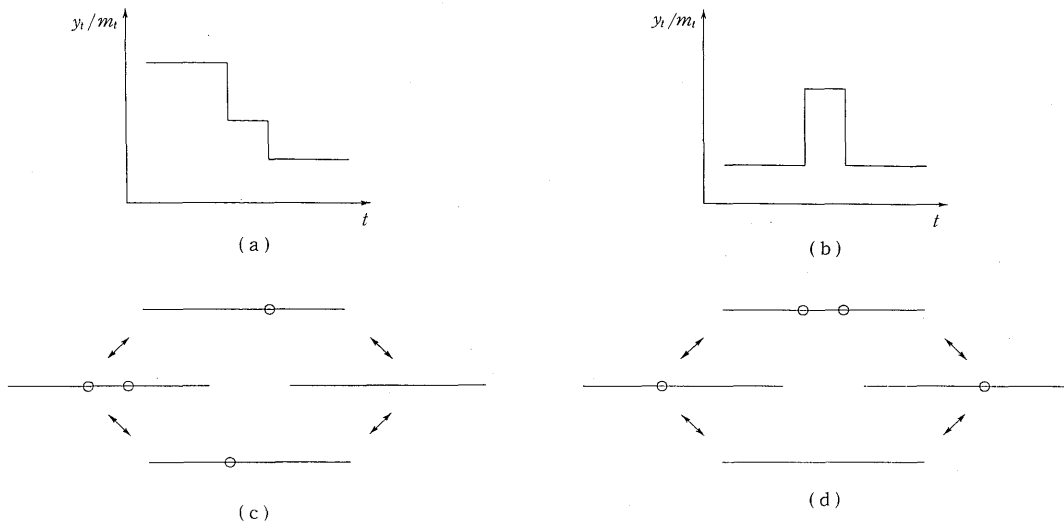


図2. アルゴリズムの緩和の妨げになるような状況の例. 説明は本文.

ばひとつもないと解釈される場合で, ‘AND 状況’ とでもいうべきものである. これらの場合, 図2(c) (‘OR 状況’ に対応), 図2(d) (‘AND 状況’ に対応)において, 真中の状態 (‘通過点’) の事後確率が上下の状態に比べて低くなり, マルコフ鎖の緩和が妨げられる.

事前分布 (1) の場合に, このようなことが問題になる可能性があるのは, ある点に変化点である事前確率 α が非常に小さい場合 (変化点に対する罰金が非常に大きい場合) である. あとで述べるように, α が大きいほど問題はやさしくなるので, α が大きい自明な場合と α が小さすぎて緩和しなくなる領域の間がメトロポリスのモンテカルロ法が有効な範囲といえる. 一般に, α が小さくなると, 変化点の数の期待値がよく収束しても, 各点に変化点である周辺事後確率の精度は悪いということが見られるが, これは上述の ‘OR 状況’ のためである.

事前分布 (2) の場合は, α に当たるものが $\{S_i\}$ で決まるので, 初期条件によっては非常に小さい α を与えたのと同じになる. この結果, 系の大きさ n がある程度以上大きいと変化点が全くないという状態が準安定状態になってしまう (このような正の feedback は磁性体とのアナロジーでいえば, 「強磁性」的な相互作用に相当する). それでも, n の大きさが 100 程度なら, 変化点が全くない場合と 1 個ある場合の事後確率の比は最大 $1/100e \sim 1/300$ 程度なので, 十分時間をかければ緩和すると思われる.

緩和の妨げになる状況が本質的に上記の 2 つのみだということは証明できないので, データごとに初期条件や乱数を変えて収束のテストを行うことが望ましい.

4. 実験結果

4.1 各点に変化点である周辺事後確率の計算

実験に用いたデータは $n=13$ の Lindisfarne scribes data (表1) 及び $n=100$ の模擬データ (表3) である. Lindisfarne scribes data では n が小さすぎて, すべての場合についての和をとることが容易にでき, メトロポリスのモンテカルロ法の利点が現われないが, 他との比較のために扱うことにした. 表3の模擬データは, 各章で事象の起きる数が各章独立の 2 項分布に従うとして作成した. 以下では, 狭義のメトロポリス法と熱浴法 (Gibbs sampler) を両方用い

表 3.

t	m_t	y_t	生起確率 (真の値)	t	m_t	y_t	生起確率 (真の値)
1	16	10	0.50	51	16	8	0.60
2	16	8	0.50	52	16	9	0.60
3	16	7	0.50	53	16	12	0.60
4	16	9	0.50	54	16	10	0.60
5	16	10	0.50	55	16	10	0.60
6	16	9	0.50	56	16	9	0.60
7	16	9	0.50	57	16	8	0.60
8	16	6	0.50	58	16	12	0.60
9	16	8	0.50	59	16	12	0.60
10	16	10	0.50	60	16	9	0.60
11	16	8	0.50	61	16	9	0.60
12	16	9	0.50	62	16	10	0.60
13	16	7	0.50	63	16	12	0.60
14	16	6	0.50	64	16	12	0.60
15	16	9	0.50	65	16	9	0.60
16	16	9	0.50	66	16	6	0.60
17	16	9	0.50	67	16	8	0.60
18	16	9	0.50	68	16	6	0.60
19	16	8	0.50	69	16	7	0.60
20	16	5	0.50	70	16	6	0.60
21	16	10	0.50	71	16	9	0.60
22	16	8	0.50	72	16	10	0.60
23	16	12	0.50	73	16	12	0.60
24	16	7	0.50	74	16	7	0.60
25	16	10	0.50	75	16	8	0.60
26	16	6	0.50	76	16	7	0.60
27	16	7	0.50	77	16	13	0.60
28	16	7	0.50	78	16	7	0.60
29	16	7	0.50	79	16	12	0.60
30	16	5	0.50	80	16	10	0.60
31	16	13	0.80	81	16	9	0.60
32	16	13	0.80	82	16	9	0.60
33	16	9	0.80	83	16	9	0.60
34	16	9	0.80	84	16	9	0.60
35	16	7	0.40	85	16	8	0.60
36	16	8	0.40	86	16	8	0.60
37	16	5	0.40	87	16	10	0.60
38	16	5	0.40	88	16	8	0.60
39	16	7	0.40	89	16	7	0.60
40	16	7	0.40	90	16	8	0.60
41	16	7	0.40	91	16	9	0.60
42	16	4	0.40	92	16	10	0.60
43	16	8	0.40	93	16	9	0.60
44	16	6	0.40	94	16	9	0.60
45	16	10	0.40	95	16	8	0.60
46	16	7	0.40	96	16	13	0.60
47	16	3	0.40	97	16	7	0.60
48	16	5	0.40	98	16	7	0.60
49	16	3	0.40	99	16	6	0.60
50	16	4	0.40	100	16	12	0.60

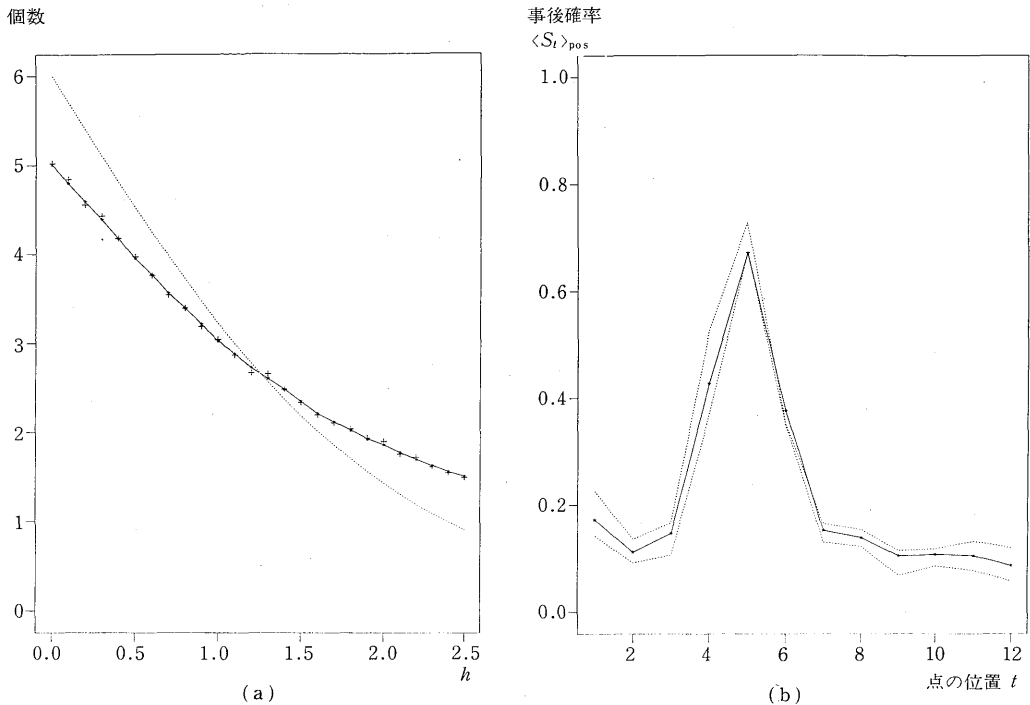


図3. Lindisfarne scribes data に関する事前分布 (1) による結果. (a) は h の推定に関する結果. 実線は $\langle \sum_t S_t \rangle_{\text{pos}}$, 点線は $\frac{(n-1) \exp(-h)}{1 + \exp(-h)}$. ★印と実線は確率 1/2 でランダムに変化点をおく初期条件からはじめて, 2000 MCS とばしたあとの 16000 MCS の平均, +印は 4000 MCS までの平均. (b) は各点に変化点である周辺事後確率. 実線は $h=1.3$, 上下の点線はそれぞれ $h=1.1$, $h=1.5$ とした場合. 確率 1/2 でランダムに変化点をおく初期条件からはじめて, 2000 MCS とばしたあとの 4000 MCS の平均.

たが, 結果に大きな差はなかった.

Lindisfarne scribes data に関する事前分布 (1) による結果は図 3(a)-(b) に示されている. アルゴリズムは熱浴法. 図 3(a) には $\langle \sum_t S_t \rangle_{\text{pos}}$ と $(n-1) \frac{\exp(-h)}{1 + \exp(-h)}$ のグラフを示した. 交点が h の推定値であって, およそ $h=1.3$ である. 図 3(b) は推定した h から求めた各点に変化点である周辺事後確率である. 結果は Kashiwagi (1991) のものに近いが, 変化点である確率が全体にやや押え気味になっている.

事前分布 (2) による結果を図 4 に示す. アルゴリズムは狭義のメトロポリス法である. 事前分布 (2) の場合は, ハイパーパラメータの推定のステップがないので, 図 3(b) に対応する周辺事後確率の図のみ示した. 図 4 では, 図 3(b) に比べて変化点である確率が全体にやや大きい結果になった. 対応する Kashiwagi (1991) の結果 (事前分布は事前分布 (2) と同じだが, 単なる AIC ではなく高次の補正項を持った式を用いている) により近い結果である.

表 3 の模擬データに関する事前分布 (1) による結果を図 5(a)-(b) に, 事前分布 (2) による結果を図 6 に示す. アルゴリズムは狭義のメトロポリス法である. 図 5(b) では, 3 つの変化点のうち 2 つが正しく検出され, ひとは位置がずれている. 位置のずれた変化点及びいくつかのピークの事後確率は正しいものに比べて相対的に低くなっている. 用いた模擬データがかなり難しいものであることを考えるとこれはまずまずの結果といえる. 図 6 は図 5(b) と

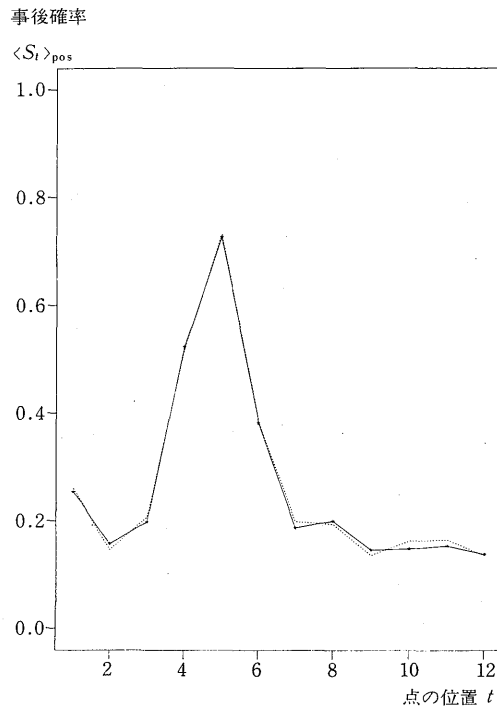


図4. Lindisfarne scribes dataに関する事前分布(2)による結果. 各点に変化点である周辺事後確率. 実線は変化点なしの初期条件からはじめて2000 MCS とばしたあとの8000 MCS の平均. 点線は2000 MCS までの平均.

ほとんど完全に一致する. 図6は各点が確率1/2で変化点であるようにランダムに変化点をばらまいた状態を初期状態としているが, ひとつも変化点がない状態から出発した場合との比較を図7に示す. よく一致していることがわかる.

図5(b), 図6の場合に変化点の数 k の揺らぎを調べたのが, それぞれ図8, 図9である. アルゴリズムは狭義のメトロポリス法である. 図8(a), 図9(a)がマルコフ鎖の「時系列」, 図8(b), 図9(b)がそれから求めた変化点の数 k のヒストグラムである. 図5(b), 図6がよく一致しているにもかかわらず, 変化点の数の揺らぎは事前分布(2)を用いた場合の方がかなり大きい.

図6($n=100$)では10000 MCS 計算しているが, これに要する時間は統計数理研究所のM-682Hで数十秒の程度であった. 図5の場合は, 個々の h に関するMCS当たりの計算時間はこれよりも短い, h を推定しなければならないのでその分時間がかかる.

以上の実験で, 提案した方法で一応望ましい結果が得られることが示された. また, ここで扱った例では事前分布(1)で α を推定した場合と事前分布(2)を用いた場合で, 各点に変化点である周辺事後確率にあまり違いがないことも示された.

4.2 2体相関の計算例

メトロポリスのモンテカルロ法では, さまざまな周辺分布について計算を行うことができる. 前節では, ある点に変化点であるような周辺事後確率を計算したが, ここでは2つの点に関する相関を計算してみよう. すなわち, s 点に変化点であると判断されることが, r 点に変化点である確率にどう影響してくるかを調べるわけである.

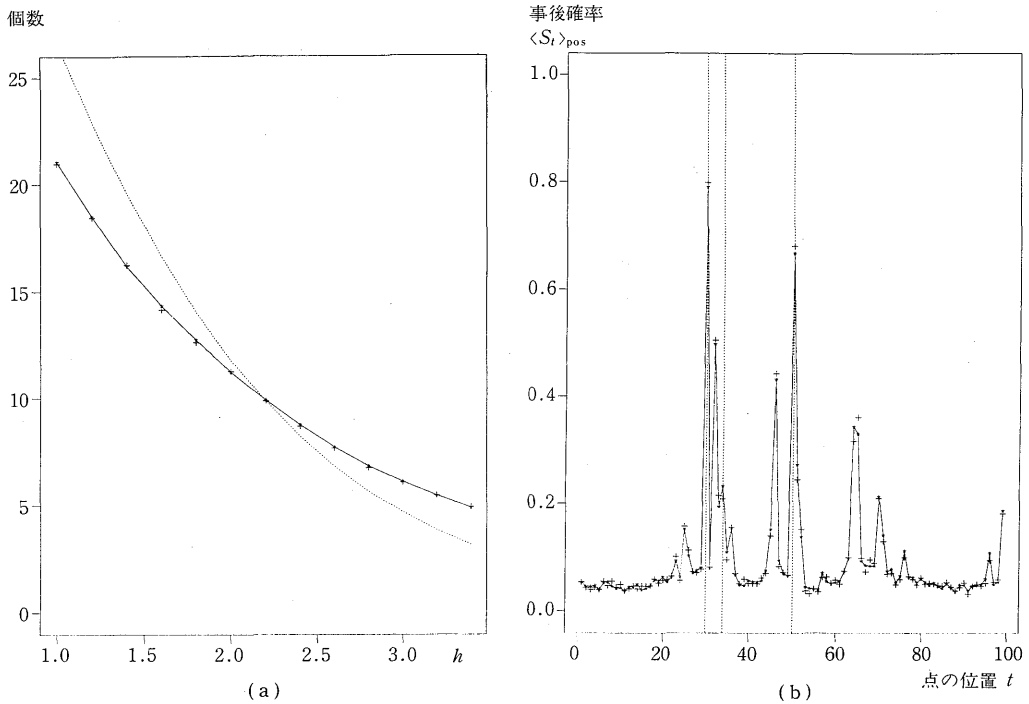


図5. 表3の模擬データに関する事前分布(1)による結果。(a)は h の推定に関する結果。実線は $\langle \sum_t S_t \rangle_{\text{pos}}$, 点線は $\frac{(n-1)\exp(-h)}{1+\exp(-h)}$. \star 印と実線は変化点なしの初期条件からはじめて1000 MCS とばしたあとの4000 MCS の平均, $+$ 印は2000 MCS までの平均。(b)は各点に変化点である周辺事後確率, $h=2.2$. \star 印と実線は変化点なしの初期条件からはじめて1000 MCS とばしたあとの4000 MCS の平均, $+$ 印は2000 MCS までの平均。縦の点線は変化点の真の位置。

相関を表わす量として, r 点に変化点である確率を $p(r)$, s 点に変化点である確率を $p(s)$, 両方がともに変化点である確率を $p(r, s)$ としたときに,

$$(4.1) \quad C(r, s) = \frac{p(r, s) - p(r)p(s)}{p(r)(1 - p(r))}$$

$$(4.2) \quad Q(r, s) = \frac{p(r, s)}{p(r)p(s)} - 1$$

の両者を計算した。 $C(r, s)$ は $r=s$ で1になるように規格化された相関関数であり, $Q(r, s)$ は系を点 s で切断した場合に, 点 r が変化点である確率が変化する割合を表わした量である。系を点 s で切断するのは, 点 s に変化点があると仮定するのと同じであることに注意されたい。とくに $Q(r, r) = (1 - p(r))/p(r)$ である。表3の模擬データについてこれらを計算したのが図10, 図11である。アルゴリズムは狭義のメトロポリス法である。

ここでは, $C(r, s)$, $Q(r, s)$ の両者とも全系に関してメトロポリスのモンテカルロ法を適用することによって計算した。 $Q(r, s)$ のような量は, 実際に系を切断して測定する方が良いと思われる。 s に変化点がある確率は低い, もし存在した場合には r への影響が大きいという場合に, 条件付き確率の測定精度が悪くなるからである。これに対して, $C(r, s)$ で $p(r)$ が0ま

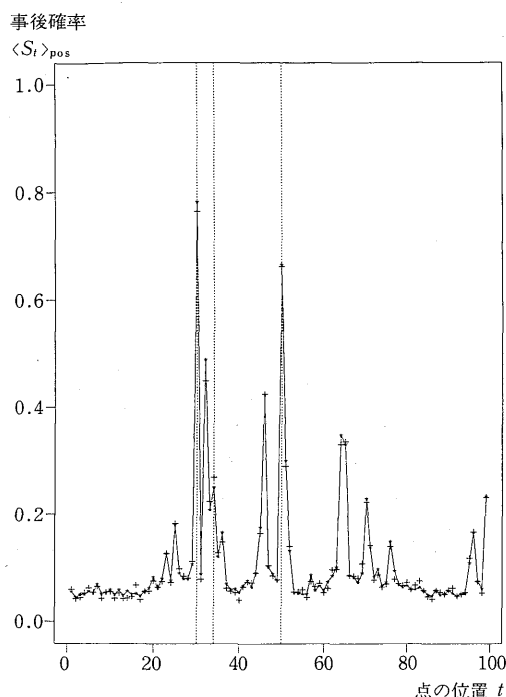


図6. 表3の模擬データに関する事前分布(2)による結果. 各点に変化点である周辺事後確率. ★印と実線は変化点なしの初期条件からはじめて2000 MCS とばしたあとの8000 MCS の平均. +印は2000 MCS までの平均. 縦の点線は変化点の真の位置.

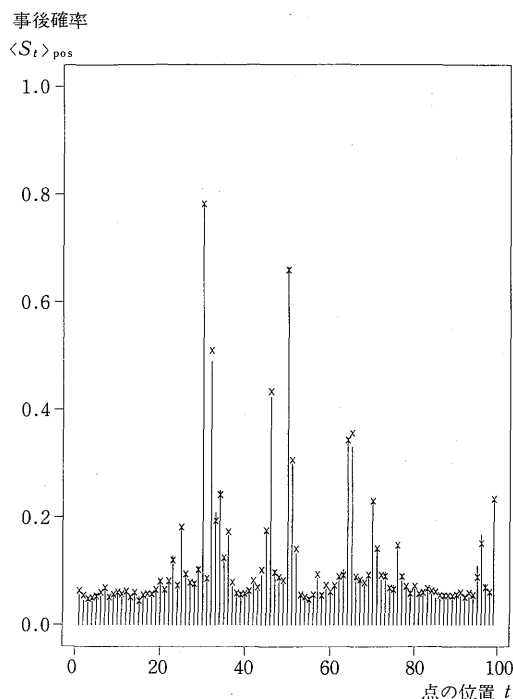


図7. 初期条件の影響(事前分布(2)の場合). 各点に変化点である周辺事後確率. 縦線は変化点なしの初期条件の場合, 'x'は確率1/2でランダムに変化点をおく初期条件の場合. とともに2000 MCS とばしたあとの8000 MCS の平均.

たは1にひどく近くない場合にはここで行った計算法で十分であろう.

5. 問題点

5.1 改良可能な点

柏木(私信)によれば, 問題のスケールが中程度の場合は, 事前に表を作っておくことにより, 変化点問題に関する計算を非常に高速化することができる. メトロポリスのモンテカルロ法の場合もこの手法は効果的と考えられる. また, マルコフ鎖のダイナミクスを改良することも考えられる. とくに, ある変化点を「他に移す」ような試行を取り入れることにより, 前に述べた「OR 状況」の場合の緩和を大幅に加速することが可能と思われる.

5.2 他のアルゴリズムとの比較

まず考えられるのは, 問題を分割して個々に処理する方法である. これは変化点の事前確率 α が大きい場合には効果的である. 表3の模擬データの場合, h の推定値は2.2であるから $\alpha \sim 0.1$, ある点に変化点である確率の下限はAICの罰金項を考慮して, $\alpha e^{-1} \sim 0.037$ である. 従って, $1/0.037 \sim 27$ の数倍以上離れた点の間には変化点が必ずひとつあることになる. ひとつでも変化点があれば, そこから先のことは関係ないので, この値が小さければ系を小さく分割して

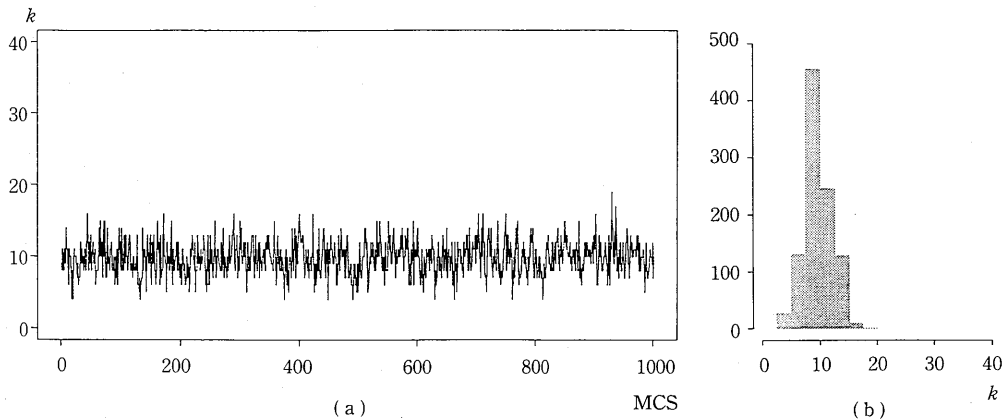


図8. 事前分布(1)の場合の変化点の数 k の揺らぎ. データは表3の模擬データ, $h=2.2$. (a)はメトロポリスのモンテカルロ法により生成されたマルコフ鎖のsample path上の k の値(最初の部分はとばしてある). (b)は k のヒストグラム(周辺分布).

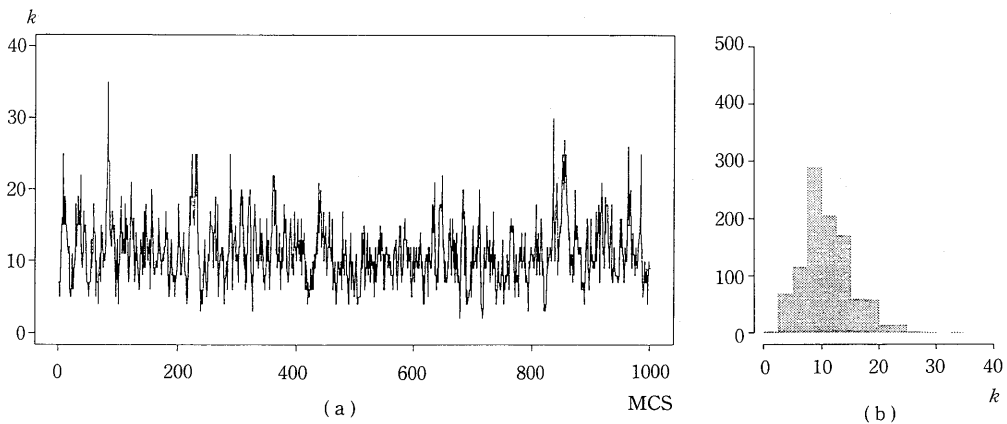


図9. 事前分布(2)の場合の変化点の数 k の揺らぎ. データ及び見方は図8に同じ.

論じてよいことになる. この例での27という値は小さくないが, 図11の結果を見ると, 結果論としては分割法で良い可能性もある.

さらによく考えてみると, この問題については非ガウスフィルターや隠れマルコフの方法と同じような原理で正確な周辺分布を求めるrecursiveなアルゴリズムが存在することがわかる. これは, 事前分布(1)の場合にとくに簡単に構成できるが, 事前分布(2)の場合にも構成可能である. この方法は, メトロポリスのモンテカルロ法より高能率であると考えられる. この点を考えると, メトロポリスのモンテカルロ法はこの問題を解くのに最良の方法とは必ずしもいえない.

5.3 アルゴリズム以外の問題

本稿の目的は, 主として, 擬ベイズ法のような複雑な状況にメトロポリスのモンテカルロ法を適用してみることにあった. 従って, 擬ベイズ法自体に関する問題やここで採用したモデルの欠点などについては, 以下で簡単にコメントするにとどめる.

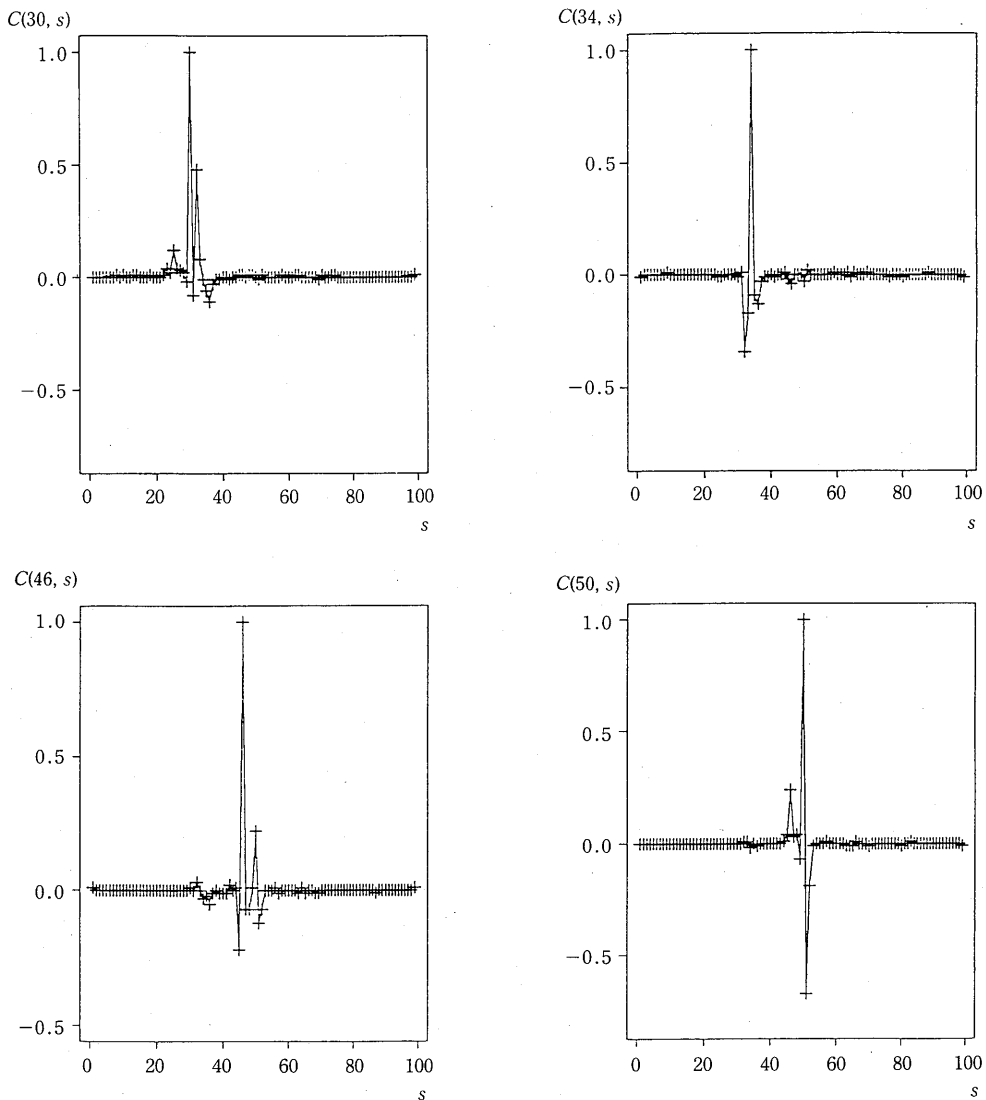


図 10. $C(r, s)$ の例. それぞれ $r=30, 34, 46, 50$ に関する結果. データは表 3 の模擬データ. 事前分布 (1) で $h=2.2$ とした場合 (図 5 (b) に対応). 実線は 5000 MCS とばしたあとの 80000 MCS の平均, + は 40000 MCS までの平均.

5.3.1 AIC に関する問題点

ここで考えた方法では, すべての考えられるモデルについて和をとることを基本としているが, これらの中には AIC で良さを測るのが不適当なほど変数の多いモデルが入っている (たとえば, ある区間の確率の推定値 θ_j が 0 や 1 になる場合). これらのモデルの AIC は一般には大きくなるが, 数が多いので, 結果的に変化点の数を多く評価しやすくなる可能性がある. Kashiwagi (1991) では AIC の罰金項を高次の項までとることにより状況を改善しようと試みている.

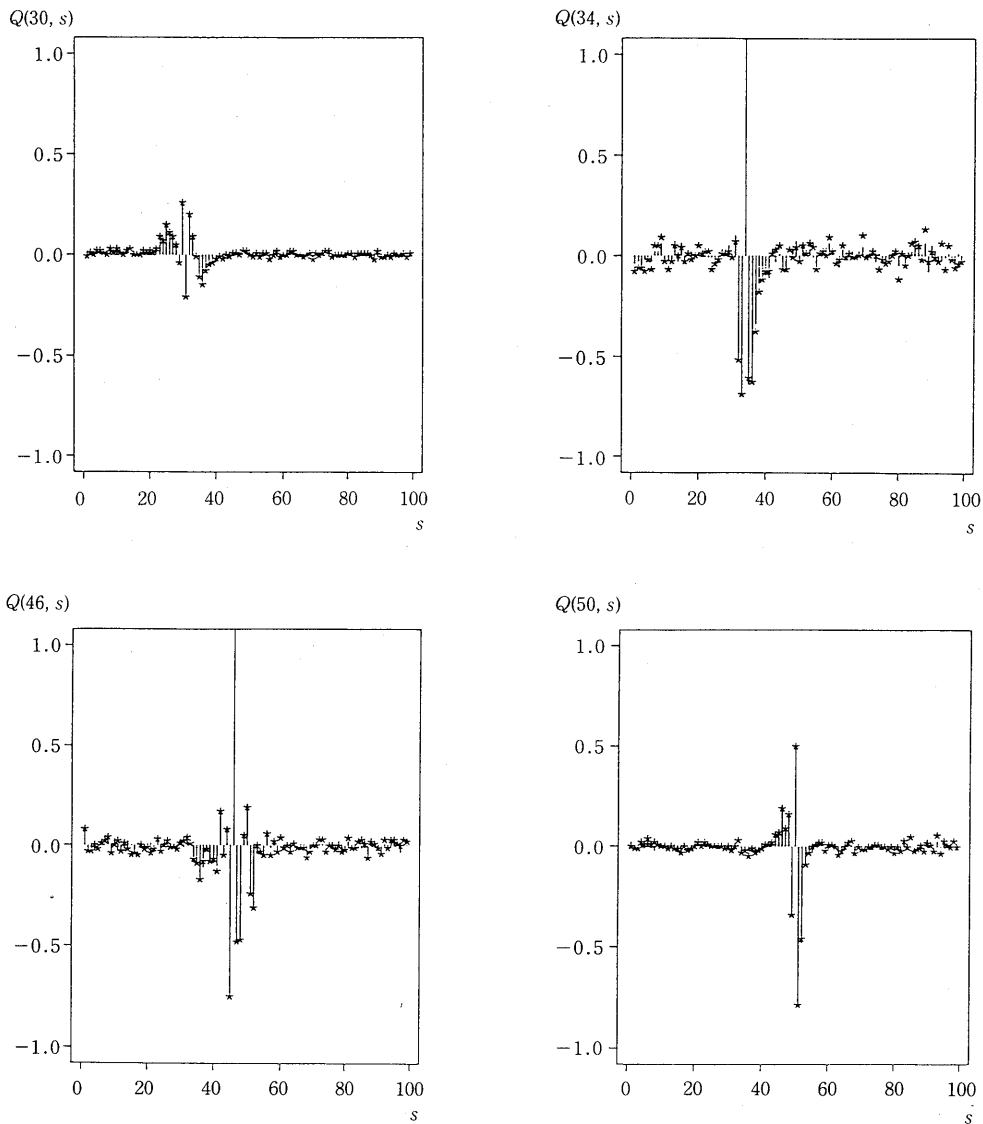


図 11. $Q(r, s)$ の例. r の値と実験条件は図 10 に同じ. 縦線は 5000 MCS とばして 80000 MCS の平均. \star は 40000 MCS までの平均. $r=s$ の場合については図からはみでたものがある.

5.3.2 適応的な方法の問題点

変化点と認める基準を適応的に定めているため, 真の変化点の数が増えてくると, 正しくない変化点にも高い事後確率が与えられる傾向が出てくる. h をデータから推定する代わりに事前分布 (2) を用いた場合も, 実質的には適応的に定めたことになっているため, 同様の傾向がある. これはやむをえないことのようにも思えるが, より良い理解のためには, 仮説検定の立場との比較や事前分布の再検討 (たとえば Kitagawa (1984)) が必要と思われる.

5.3.3 モデルの欠点

ここで論じた変化点問題の扱いは, 実際の問題, たとえば著者推定に適用するには重大な欠

点がある。それはトレンドが考慮されていないために、はっきりした変化点がなくだららと生起確率が変わる場合でも、変化点を検出してしまうことである。事後分布を考えることで、あいまいさを取り入れて考えているといっても、‘変化点の事後確率’というのはあくまでモデル(族)の内部で考えた確率なので、この欠点を救うには役にたたない。改良案としては、生起確率が(区分的に一定であるとする代わりに)区分的に直線であるとする方法やガウスの平滑化法(Ishiguro and Sakamoto (1983))に変化点を付け加える方法などがある。これらについても、Kashiwagi (1991)を参照されたい。

5.3.4 変化点の数の期待値

ここで述べた方法では推定が比較的良好な結果を与える場合にも、変化点の数の期待値が真の値より非常に大きくなる傾向がある。表3の模擬データでは、変化点の真の個数は3なのに、図8、図9からわかるように、推定された個数は10個近くになる。このことが上に述べたいくつかの問題点の影響なのか、もっと他に理由があるのかは、さらに検討する必要がある。

6. ま と め

擬ベイズ法での‘モデルに関する和’の部分にメトロポリスのモンテカルロ法を適用することを提案した。変化点問題を例として、実際のデータ及び模擬データに関して実験を行った。変化点に関する事前分布を学習する場合と与えておく場合の両者を扱い、両者の関係を論じた。

謝 辞

誠実に討論して下さった柏木宣久氏に感謝します。また、北川源四郎教授には擬ベイズ法についての質問に答えて頂き、レフェリーには英文要旨を直して頂きました。ここに謝意を表します。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B.N. Petrov and F. Csáki), 267-281, Akadémiai Kiadó, Budapest.
- Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, **66**, 53-59.
- Akaike, H. (1980). On the use of predictive likelihood of a Gaussian model, *Ann. Inst. Statist. Math.*, **32**, 311-324.
- Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Statist. Math.*, **43**, 1-59.
- Gelfand, A.E., Hills, S.E., Racine-poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Amer. Statist. Assoc.*, **85**, 972-985.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- Geman, S. and McClure, D.E. (1987). Statistical methods for tomographic image reconstruction, Proceedings of the 46th Session of the ISI, *Bull. Internat. Statist. Inst.*, **52**.
- Hinton, G.E. and Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines, *Parallel Distributed Processing*, Vol. 1 (eds. E. Rumelhart and J.L. McClelland), Cambridge MIT Press.
- 伊庭幸人 (1991). メトロポリスのモンテカルロ法の巨視的パラメータ推定への応用 —— 2次元イジング模型の場合 ——, *統計数理*, **39**, 1-21.

- Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115-137.
- Kashiwagi, N. (1991). Bayesian detection of structural changes, *Ann. Inst. Statist. Math.*, **43**, 77-93.
- Kitagawa, G. (1984). Bayesian analysis of outliers via Akaike's predictive likelihood of a model, *Comm. Statist. Simulation Comput.*, **13**, 107-126.
- Kitagawa, G. and Akaike, H. (1982). A quasi Bayesian approach to outlier detection, *Ann. Inst. Statist. Math.*, **34**, 389-398.
- Ogata, Y. (1990). A Monte Carlo method for objective Bayesian procedure, *Ann. Inst. Statist. Math.*, **42**, 403-433.
- Sakai, H. (1989). An application of BIC type method to harmonic analysis and a new criterion for order determination of an AR process, ISM Cooperative Research Report, **20**.
- Smith, A.F.M. (1980). Change-point problems: approaches and applications, *Bayesian Statistics* (eds. J.M. Bernardo, M.H. De Groot, D.U. Lindley and A.F.M. Smith), University Press, Valencia.

Metropolis-type Monte Carlo Algorithm and Quasi-Bayesian Estimation Procedure :
An Application to a Change-point Problem

Yukito Iba

(The Institute of Statistical Mathematics)

The Metropolis-type Monte Carlo algorithm is applied to the change-point problem for time series data, employing the quasi-Bayesian approach due to Akaike, where the posterior probability of a model is estimated to be proportional to $\exp(-AIC/2) * (\text{prior of the model})$. The posterior distribution over the space of the models is sampled by the Metropolis-type algorithm. The results of the numerical experiments for the Lindisfarne scribes data and an artificial data of larger size are reported and discussed.