

MCMC アルゴリズムによるロジットモデルの ベイズ推定に関する若干の考察

田 島 博 和

1. はじめに

確率的離散選択の代表的なモデルであるロジットモデルは、マーケティングや消費者行動の分野でも、消費者のブランド選択行動等を記述するために古くから用いられてきた [Malhotra, 1984; 片平・杉田, 1994; 土田, 2010 ほか]。最近では、MCMC (Markov-Chain Monte-Carlo) アルゴリズムを用いたベイズ推定により、選好に関する消費者の異質性を積極的に取り込んだ分析も行われるようになった [Rossi and Allenby, 2006; 稲田・室町, 2006 ほか]。そこで本稿では、MCMC アルゴリズムによるロジットモデルのベイズ推定に関して、特に MCMC アルゴリズムのひとつである酔歩 Metropolis-Hastings アルゴリズムを例にとり、そのサンプル系列の収束とチューニングについて、若干の考察を行いたい。

2. ロジットモデルの最尤推定とベイズ推定

選択肢がふたつの場合のロジットモデルは特に二項ロジットモデル (Binomial Logit Model) と呼ばれる。 n 期に選択肢 1 を選択する確率 p_{1n} は、説明変数の相対値を x_n 、パラメータのベクトルを θ とするとき、 $p_{1n} = e^{\theta x_n} / (1 + e^{\theta x_n})$ と書かれる。同様に選択肢 2 を選択する確率 p_{2n} は $p_{2n} = 1 / (1 + e^{\theta x_n})$ と書かれる。 n 期に選択肢 1 を選択したか否かを表す観測ダミー・データを δ_n 、観測データのベクトルを $\delta = \{\delta_n\}_{n=1}^N$ (ただし N は観測期間) とすると、モデルの尤度関数 (likelihood function) $L(\theta; \delta)$ は次のように書かれる。

$$L(\theta; \delta) \stackrel{\text{def}}{=} p(\delta|\theta) = \prod_{n=1}^N [p_{1n}^{\delta_n} (1 - p_{1n})^{1 - \delta_n}] = \prod_{n=1}^N \left[\left(\frac{e^{\theta x_n}}{1 + e^{\theta x_n}} \right)^{\delta_n} \left(\frac{1}{1 + e^{\theta x_n}} \right)^{1 - \delta_n} \right]$$

このとき (対数) 尤度関数を最大化するパラメータ $\hat{\theta}_{ML}$ が、最大尤度推定値 (最尤値, maximum likelihood estimator) である。

$$\log L(\hat{\theta}_{ML}; \delta) = \max_{\theta} \log L(\theta; \delta)$$

次にベイズ推定について説明する。条件付き確率の定義により $\Pr(\theta|\delta)\Pr(\delta) = \Pr(\theta, \delta) = \Pr(\delta|\theta)\Pr(\theta)$ だから $\Pr(\theta|\delta) = \Pr(\theta)\Pr(\delta|\theta) / \Pr(\delta)$ が成立する。

これはベイズ定理 (Bayes' theorem) と呼ばれる。したがって事後確率の期待値 $\hat{\theta}_B$ がベイズ推定値 (Bayesian estimator) である。しかし次式の最右辺の積分を解析的に解く事は、一般的には困難である。

$$\hat{\theta}_B \stackrel{\text{def}}{=} E_{\theta}[\Pr(\theta|\delta)] = \int_{\theta} \theta \Pr(\theta|\delta) d\theta = \frac{1}{\Pr(\delta)} \int_{\theta} \Pr(\theta) \Pr(\delta|\theta) d\theta$$

3. MCMC アルゴリズムによるベイズ推定

前節のベイズ定理により θ の事後分布 $\Pr(\theta|\delta)$ は事前分布と尤度関数の積 $\Pr(\theta)\Pr(\delta|\theta)$ に比例する。従って $\pi(\theta) \stackrel{\text{def}}{=} \Pr(\theta)\Pr(\delta|\theta)$ をターゲット分布 (target distribution) と呼ぶことにすると、容易に分かるように、ターゲット分布 $\pi(\theta)$ のランダムサンプル系列は、事後分布 $\Pr(\theta|\delta)$ のランダムサンプル系列になっている。従ってターゲット分布のランダムサンプル系列が得られるのであれば、モンテカルロ法によって事後分布の平均 (期待値)、すなわちベイズ推定値を数値的に求めることができる。ターゲット分布のランダムサンプル系列を得る方法はいくつか提案されている [照井, 2008] が、一般的には困難である。

MCMC アルゴリズム (Markov-Chain Monte-Carlo algorithm) は、ターゲット分布のランダムサンプリングを行うかわりに、ターゲット分布に収束するマルコフ連鎖によって生成されたサンプル系列を使い、モンテカルロ法によってターゲット分布の期待値、すなわちベイズ推定値を数値的に求める手法である。本稿では代表的な MCMC アルゴリズムのひとつである酔歩 Metropolis-Hastings アルゴリズム (random-walk Metropolis-Hastings algorithm) を紹介する。

酔歩 Metropolis-Hastings アルゴリズムの各ステップについて、その原理を表1で説明する。MCMC アルゴリズムがベイズ推定に有効であるのは、当該アルゴリズムによって生成されたサンプル系列 $\{\theta^{(n)}\}_n$ がターゲット分布のサンプル列とみなす事ができるからである。すなわ

図1 ターゲット分布によるサンプリング

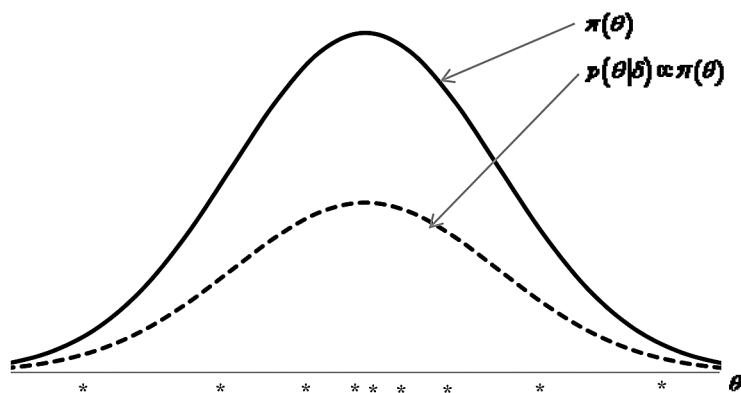


図 2 酔歩 Metropolis-Hastings アルゴリズム

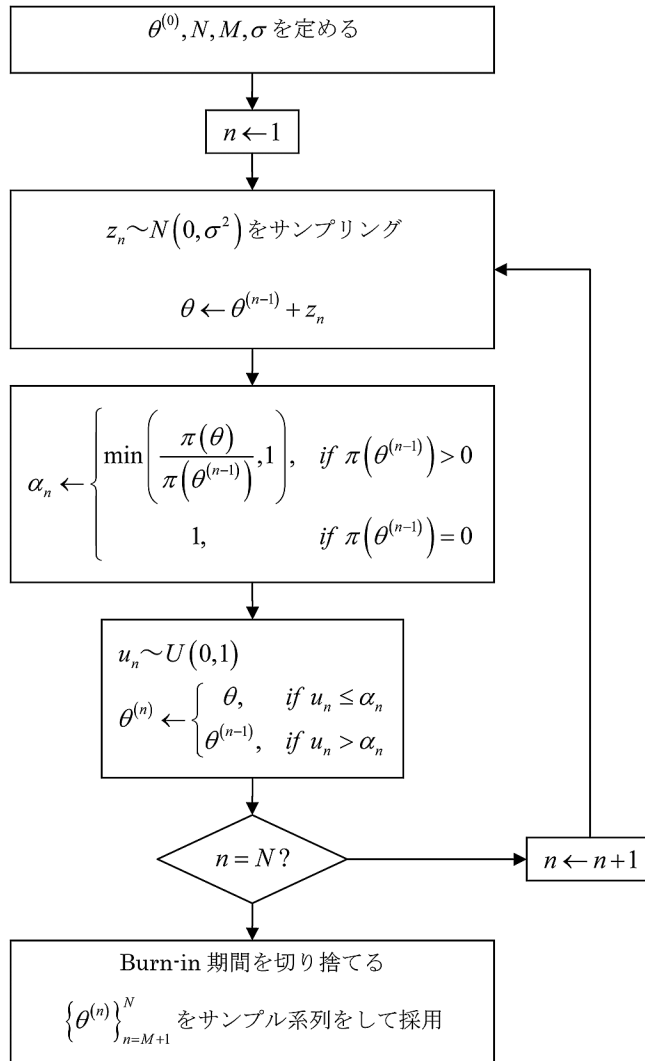
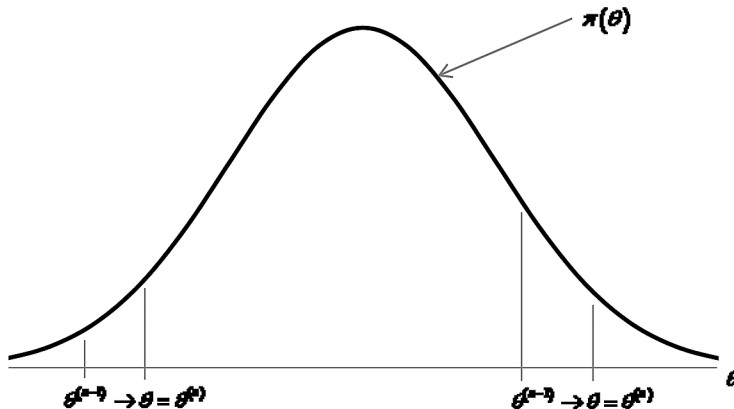


表 1 酔歩 Metropolis-Hastings アルゴリズムの各ステップの説明

ステップ①	パラメータの初期値 $\theta^{(0)}$ 、サンプル数 N 、burn-in（収束が安定しないために切り捨てられる部分）期間 M 、パラメータの更新幅 σ を定める。
ステップ②	反復計算回数を表す n に 1 を代入する。

ステップ③	$\theta^{(n)}$ の候補を酔歩によって定める。すなわち平均 0 分散 σ^2 の正規分布に従う乱数 z_n に対して、 $\theta \stackrel{def}{=} \theta^{(n-1)} + z_n$ とする。
ステップ④	<ul style="list-style-type: none"> ● $\pi(\theta^{(n-1)}) = \Pr(\theta^{(n-1)}) \Pr(\delta \theta^{(n-1)})$ を計算し、それが 0 か否かによって、α_n を上図の様に定める。 $0 < \pi(\theta^{(n-1)}) < \pi(\theta)$ ならば $\pi(\theta) / \pi(\theta^{(n-1)}) > 1$ だから $\alpha_n = 1$。 ● $\pi(\theta^{(n-1)}) > \pi(\theta)$ かつ $\pi(\theta^{(n-1)}) \neq 0$ ならば $\pi(\theta) / \pi(\theta^{(n-1)}) < 1$ だから $0 \leq \alpha_n < 1$。また明らかに $\alpha_n = 0 \Leftrightarrow \pi(\theta) = 0$ である。 ● $\theta = \theta^{(n-1)}$ ならば明らかに $\alpha_n = 1$。 ● $\pi(\theta^{(n-1)}) = 0$ の場合も明らかに $\alpha_n = 1$。
ステップ⑤	一様分布に従う乱数 $u_n, 0 \leq u_n \leq 1$ に対して、 $u_n \leq \alpha_n$ か否かによって上図の様に、 θ を $\theta^{(n)}$ として採用するか否かを定める。なお採用確率は $\Pr(u_n \leq \alpha_n) = \alpha_n$ に等しい。
ステップ⑥	反復回数 が所与の N に達していれば、burn-in 期間 $n \leq M$ を切り捨てて $\{\theta^{(n)}\}_{n=M+1}^N$ をベイズ推定のためのサンプル系列として採用する。

図 3 酔歩 Metropolis-Hastings アルゴリズムによるパラメータ更新



ち $N \rightarrow \infty$ のとき、サンプル系列の平均 $\sum_{n=1}^N \theta^{(n)} / N$ がターゲット分布の平均 $E[\pi(\theta)]$ 、従って事後分布の期待値 $E[\Pr(\theta|\delta)]$ 、すなわちベイズ推定値 $\hat{\theta}_B$ に収束する。詳細は次節に譲り、ここでは収束性と同等である次の事実を確認しておく。

定理 1 酔歩 Metropolis-Hastings アルゴリズムによって生成されたサンプル系列を $\{\theta^{(n)}\}_n$ 、そして $K(\theta, \varphi) = N(\theta - \varphi | 0, \sigma^2) \alpha(\theta, \varphi)$ とする (ただし $N(\cdot | 0, \sigma^2)$ は平均 0, 分散 σ^2 の正規分布の密度関数)。このとき $\pi(\theta^{(n)}) K(\theta^{(n)}, \theta) = \pi(\theta) K(\theta, \theta^{(n)})$ が成立する。

(注意) 上式は詳細釣り合条件 (detailed balance condition) と呼ばれるが、詳細は次節で説明する。

(証明) $N(\cdot | 0, \sigma^2)$ は偶関数だから $\pi(\theta^{(n)}) \alpha(\theta^{(n)}, \theta) = \pi(\theta) \alpha(\theta, \theta^{(n)})$ を証明すればよい。これは、 $\pi(\theta^{(n)})$, $\pi(\theta)$ に関して以下のように場合分けして考えれば自明である。

	$\pi(\theta^{(n)})$	$\pi(\theta)$	$\alpha(\theta^{(n)}, \theta)$	$\alpha(\theta, \theta^{(n)})$
$\pi(\theta^{(n)}) < \pi(\theta)$	=0	$\neq 0$	1	0
	$\neq 0$	$\neq 0$	1	$0 < \frac{\pi(\theta^{(n)})}{\pi(\theta)} < 1$
$\pi(\theta^{(n)}) = \pi(\theta)$	=0	=0	1	1
	$\neq 0$	$\neq 0$	1	1
$\pi(\theta^{(n)}) > \pi(\theta)$	$\neq 0$	=0	0	1
	$\neq 0$	$\neq 0$	$0 < \frac{\pi(\theta)}{\pi(\theta^{(n)})} < 1$	1

(証明終)

4. エルゴード・マルコフ系列の収束

本節では、MCMC アルゴリズムによって生成されたサンプル系列がターゲット分布 $\pi(\theta) = \Pr(\theta) \Pr(\delta|\theta)$ のランダムサンプル系列とみなせる事を確かめるために、エルゴード・マルコフ系列の収束に関する一般論について、とくに 2 値のマルコフ連鎖について詳しく解説する。

定義 1 離散的な時間パラメータを n に対して確率変数の列を $\{X^{(n)}\}_n$ 、また状態空間 $(X^{(n)})$ の実現値 $x^{(n)}$ の集合) を S とする。このとき状態空間 S が離散的で、しかも $X^{(n)}$ の確率分布が

MCMC アルゴリズムによるロジットモデルのベイズ推定に関する若干の考察

直近の実現値 $x^{(n-1)}$ のみに依存する、すなわち次式が成立するならば、 $\{X^{(n)}\}_n$ をマルコフ連鎖 (Markov chain) と呼ぶ¹⁾。

$$\Pr(X^{(n)} \in A | x^{(0)}, x^{(1)}, \dots, x^{(n-1)}) = \Pr(X^{(n)} \in A | x^{(n-1)})$$

マルコフ連鎖 $\{X^{(n)}\}_n$ に対して $p_{ij} = \Pr(X^{(n)} = j | X^{(n-1)} = i)$ を推移確率 (transition probability), その行列を推移行列 (transition matrix) と呼ぶ。

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{pmatrix}$$

なおここで p_{ij} が時点 n に依存しない。この性質を斉時的 (homogeneous) という。

定理 2 $\pi_j^{(n)} = \Pr(X^{(n)} = j)$, $\pi^{(n)} = (\pi_1^{(n)}, \pi_2^{(n)})$ に対して $\pi^{(n)} = \pi^{(n-1)} P = \dots = \pi^{(0)} P^n$ が成立する。

(証明)

$$\pi^{(0)} P = (\pi_1^{(0)}, \pi_2^{(0)}) \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = (p_{11}\pi_1^{(0)} + p_{21}\pi_2^{(0)}, p_{12}\pi_1^{(0)} + p_{22}\pi_2^{(0)}) \dots (*)$$

また $\pi_j^{(1)} = \Pr(X^{(1)} = j) = \sum_{i=1}^2 \Pr(X^{(1)} = j | X^{(0)} = i) \Pr(X^{(0)} = i) = \sum_{i=1}^2 p_{ij} \pi_i^{(0)}$ だから

$$(*) = (\pi_1^{(1)}, \pi_2^{(1)}) = \pi^{(1)}$$

(証明終)

定義 2 推移確率 P に対して $\pi = \pi P$ が成立する分布 π を P -不変 (invariant) な分布という。

また $\pi_1 p_{12} = \pi_2 p_{21}$ を詳細釣り合い条件 (condition of detailed balance) という。

数値例 $P = \begin{pmatrix} 1/4 & 3/4 \\ 3/8 & 5/8 \end{pmatrix}$, $\pi = (\frac{1}{3}, \frac{2}{3})$ に関して次の①~⑥が成立する。(証明は省略する)

- ① π は P -不変な分布である: $\pi = \pi P$ 。
- ② π, P は詳細釣り合い条件を満足する: $\pi_1 p_{12} = \pi_2 p_{21}$ 。
- ③ 任意の初期分布 $\pi^{(0)}$ に対して $\pi^{(N)} = \pi^{(0)} P^N \xrightarrow[N \rightarrow \infty]{\text{def}} \pi$ が成立する。
- ④ 任意の $i, j = 1, 2$ に対して P^N の (i, j) 成分を $p_{ij}^{(N)}$ とするとき, $p_{ij}^{(N)} \xrightarrow[N \rightarrow \infty]{} \pi_j$ が成立する。
- ⑤ 不変分布 π が時点 1 から N の間に状態 j を訪問する回数を $w(N, j)$ とするとき, $w(N, j)/N \xrightarrow[N \rightarrow \infty]{} \pi_j$ が成立する。
- ⑥ N 期の実現値 $\theta^{(N)} (= 1, 2)$ に対して $\sum_{n=1}^N \theta^{(n)} / N \xrightarrow[N \rightarrow \infty]{} E[\pi] = 5/3$ が成立する。

図 4 普遍性と詳細約合条件

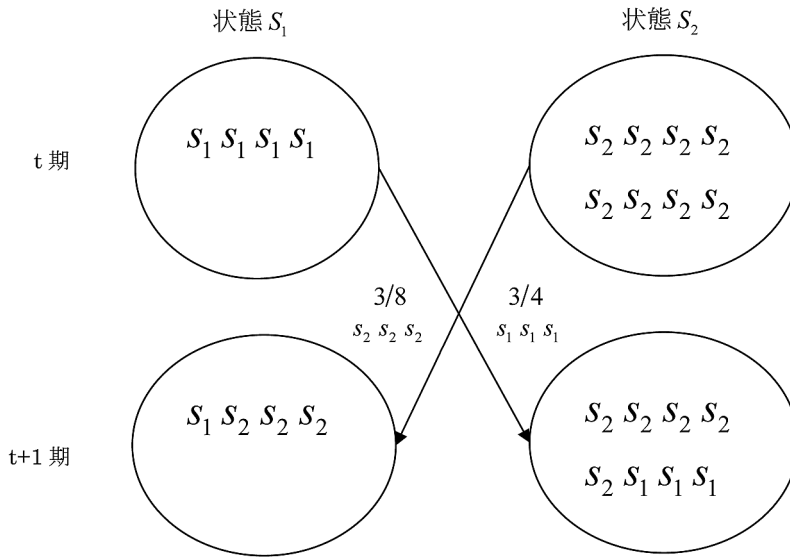
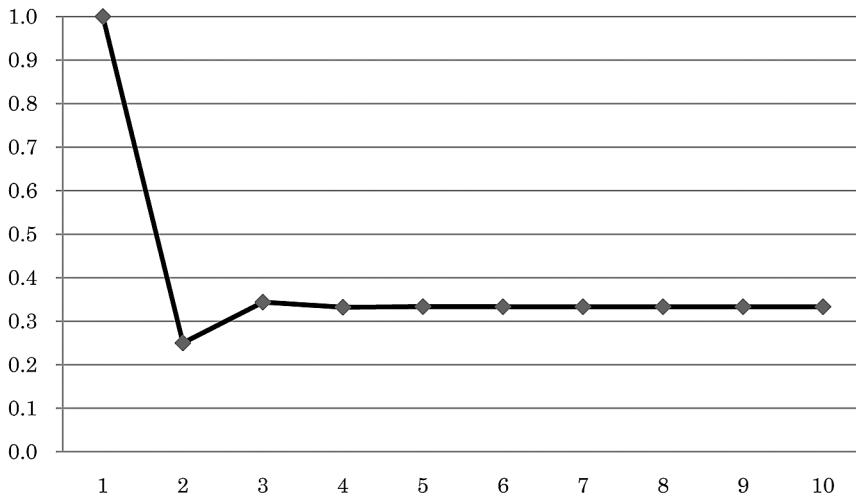


図 5 不変分布への収束



定義 3 マルコフ連鎖が既約 (irreducible) すなわち推移確率が全て正で、しかも非周期的 (aperiodic) すなわち $\{n | p_{ij}^{(n)} > 0, \forall i, j\}$ の最大公約数が 1 ならば、エルゴード的 (ergodic) であるという。

定理 3 推移確率が常に正ならば、或る π_j が存在して $p_{ij}^{(n)}$ は π_j に収束する。従って $\pi = (\pi_1 \pi_2)$ とすると任意の初期分布 $\pi^{(0)}$ に対して $\pi^{(N)} = \pi^{(0)} P^N \xrightarrow[N \rightarrow \infty]{def} \pi$ が成立する。

MCMC アルゴリズムによるロジットモデルのベイズ推定に関する若干の考察

(証明)

仮定より $p_{ij} > 0$ だから $\varepsilon = \min_{\text{def } i,j} p_{ij}$ とすると $\varepsilon > 0$ である。さらに $p_{ij}^{(n)}$ の i に関する最大値および最小値をそれぞれ $M(j, n), m(j, n)$ とすると、明らかに $m(j, n) \leq M(j, n)$ で、しかも任意の i に対して

$$p_{ij}^{(n)} = \sum_k p_{ik} p_{kj}^{(n-1)} \leq \sum_k p_{ik} M(j, n-1) = M(j, n-1)$$

が成立する。したがって $M(j, n) \leq M(j, n-1)$ 、同様に $m(j, n-1) \leq m(j, n)$ である。すなわち

$$m(j, 0) \leq m(j, 1) \leq \dots \leq m(j, n) \leq \dots \leq M(j, n) \leq \dots \leq M(j, 1) \leq M(j, 0)$$

となる。ところで

$$\begin{aligned} M(j, n+1) - m(j, n+1) &= \max_{\alpha} p_{\alpha j}^{(n+1)} - \min_{\beta} p_{\beta j}^{(n+1)} = \max_{\alpha} (p_{\alpha j}^{(n+1)} - \min_{\beta} p_{\beta j}^{(n+1)}) \\ &= \max_{\alpha} [\max_{\beta} (p_{\alpha j}^{(n+1)} - p_{\beta j}^{(n+1)})] = \max_{\alpha, \beta} (p_{\alpha j}^{(n+1)} - p_{\beta j}^{(n+1)}) = \max_{\alpha, \beta} (\sum_i p_{\alpha i} p_{ij}^{(n)} - \sum_i p_{\beta i} p_{ij}^{(n)}) \\ &= \max_{\alpha, \beta} \sum_i (p_{\alpha i} - p_{\beta i}) p_{ij}^{(n)} \quad \dots (*) \end{aligned}$$

だから、上式の右辺を、 $p_{\alpha i} - p_{\beta i} \geq 0$ である部分 Σ_i^+ と < 0 である部分 Σ_i^- に分け、状態の数を I とすると

$$\sum_i^+ (p_{\alpha i} - p_{\beta i}) = \sum_i^+ p_{\alpha i} - \sum_i^+ p_{\beta i} = 1 - \sum_i^- p_{\alpha i} - \sum_i^+ p_{\beta i} = 1 - \sum_i^- \varepsilon - \sum_i^+ \varepsilon = 1 - \sum_i \varepsilon = 1 - I\varepsilon$$

また明らかに $\sum_i^+ (p_{\alpha i} - p_{\beta i}) + \sum_i^- (p_{\alpha i} - p_{\beta i}) = 1 - 1 = 0$ すなわち

$\sum_i^+ (p_{\alpha i} - p_{\beta i}) = -\sum_i^- (p_{\alpha i} - p_{\beta i})$ だから任意の α, β に対して、

$$\begin{aligned} \sum_i (p_{\alpha i} - p_{\beta i}) p_{ij}^{(n)} &= \sum_i^+ (p_{\alpha i} - p_{\beta i}) p_{ij}^{(n)} + \sum_i^- (p_{\alpha i} - p_{\beta i}) p_{ij}^{(n)} \\ &\leq M(j, n) \sum_i^+ (p_{\alpha i} - p_{\beta i}) + m(j, n) \sum_i^- (p_{\alpha i} - p_{\beta i}) \\ &= M(j, n) \sum_i^+ (p_{\alpha i} - p_{\beta i}) - m(j, n) \sum_i^+ (p_{\alpha i} - p_{\beta i}) \\ &= \{M(j, n) - m(j, n)\} \sum_i^+ (p_{\alpha i} - p_{\beta i}) \leq \{M(j, n) - m(j, n)\} (1 - I\varepsilon) \end{aligned}$$

したがって

$$M(j, n+1) - m(j, n+1) = \max_{\alpha, \beta} \sum_i (p_{\alpha i} - p_{\beta i}) p_{ij}^{(n)} \leq \{M(j, n) - m(j, n)\} (1 - I\varepsilon) \leq \dots \leq (1 - I\varepsilon)^{n+1}$$

ここで明らかに $\varepsilon = \min_{i,j} p_{ij} \leq 1/I$ すなわち $0 < I\varepsilon \leq 1$ だから $0 \leq 1 - I\varepsilon < 1$ である。したがって上式の右辺は 0 に収束する： $M(j, N) - m(j, N) \xrightarrow{N \rightarrow \infty} 0$ 。すなわち $p_{ij}^{(n)}$ は j に依存する或る π_j に収束する。

なお定理の後半部分の証明は以下の通り。

$$\begin{aligned} \pi^{(0)} P^N &= (\pi_1^{(0)} \pi_2^{(0)}) \begin{bmatrix} p_{11}^{(N)} & p_{12}^{(N)} \\ p_{21}^{(N)} & p_{22}^{(N)} \end{bmatrix} = (\pi_1^{(0)} p_{11}^{(N)} + \pi_2^{(0)} p_{21}^{(N)}, \pi_1^{(0)} p_{12}^{(N)} + \pi_2^{(0)} p_{22}^{(N)}) \\ &\xrightarrow{N \rightarrow \infty} (\pi_1^{(0)} \pi_1 + \pi_2^{(0)} \pi_1, \pi_1^{(0)} \pi_2 + \pi_2^{(0)} \pi_2) = (\pi_1, \pi_2) = \pi \end{aligned}$$

(証明終)

定理 4 P -不変な分布 $\pi(\theta)$ に対して次の①～⑥は同値である。

- ① [不変条件, 空間平均] $\pi = \pi P$
- ② [詳細釣り合条件] $\pi_1 p_{12} = \pi_2 p_{21}$
- ③ 任意の初期分布 $\pi^{(0)}$ に対して $\pi^{(N)} \stackrel{def}{=} \pi^{(0)} P^N \xrightarrow{N \rightarrow \infty} \pi$
- ④ 任意の i に対して $p_{ij}^{(N)} \xrightarrow{N \rightarrow \infty} \pi_j$
- ⑤ [時間平均] $w(N, j) = \# \{n \leq N; \theta^{(n)} = j\} / N \xrightarrow{N \rightarrow \infty} \pi_j$
- ⑥ 各期の実現値 $\theta^{(n)} = 0, 1$ に対して $1/N \cdot \sum_{n=1}^N \theta^{(n)} \xrightarrow{N \rightarrow \infty} E[\pi]$

(証明)

[①⇔②]

$$\begin{aligned} \pi &= P\pi \\ \Leftrightarrow (\pi_1 \ \pi_2) &= (\pi_1 p_{11} + \pi_2 p_{21} \ \pi_1 p_{12} + \pi_2 p_{22}) \\ \Leftrightarrow \pi_1(p_{11} + p_{12}) &= \pi_1 = \pi_1 p_{11} + \pi_2 p_{21} \\ \Leftrightarrow \pi_1 p_{12} &= \pi_2 p_{21} \end{aligned}$$

[①⇒③]

P は既約だから任意の初期分布 π_0 に対して或る N が存在して $\pi = \pi_0 P^N$ が成立する。このとき①より $\pi = \pi P = \pi_0 P^{N+1} = \pi_0 P^{N+2} = \dots$ だから, $\pi_0 P^N \xrightarrow{N \rightarrow \infty} \pi$

[③⇒①]

$$\pi^{(N+1)} = \pi P^{N+1} = \pi P^N P \text{ だから, 両辺の } N \text{ に関する極限をとると } \pi = \pi P$$

[③⇒④]

$$\pi^{(0)} = (1 \ 0) \text{ のとき } \pi^{(0)} P^N = (1 \ 0) \begin{pmatrix} p_{11}^{(N)} & p_{12}^{(N)} \\ p_{21}^{(N)} & p_{22}^{(N)} \end{pmatrix} = (p_{11}^{(N)} \ p_{12}^{(N)}) \xrightarrow{N \rightarrow \infty} (\pi_1 \ \pi_2). \text{ また } \pi^{(0)} = (0 \ 1) \text{ のとき}$$

同様に $\pi^{(0)} P^N \xrightarrow{N \rightarrow \infty} (\pi_1 \ \pi_2)$ である。したがって任意の i に対して $p_{ij}^{(N)} \xrightarrow{N \rightarrow \infty} \pi_j$ が成立する。

[④⇒③]

$$\begin{aligned} \pi^{(0)} P^N &= (\pi_1^{(0)} \ \pi_2^{(0)}) \begin{pmatrix} p_{11}^{(N)} & p_{12}^{(N)} \\ p_{21}^{(N)} & p_{22}^{(N)} \end{pmatrix} = (\pi_1^{(0)} p_{11}^{(N)} + \pi_2^{(0)} p_{21}^{(N)} \ \pi_1^{(0)} p_{12}^{(N)} + \pi_2^{(0)} p_{22}^{(N)}) \\ &\xrightarrow{N \rightarrow \infty} (\pi_1^{(0)} \pi_1 + \pi_2^{(0)} \pi_1 \ \pi_1^{(0)} \pi_2 + \pi_2^{(0)} \pi_2) = (\pi_1 \ \pi_2) = \pi \end{aligned}$$

[④⇒⑤]

$$v(n, j) \stackrel{def}{=} \begin{cases} 1, & X_n = j \\ 0, & X_n \neq j \end{cases}$$

よって確率変数 $v(n, j)$ を定義すると, 明らかに $w(N, j) = \sum_{n=1}^N v(n, j)$ が成立する。

また $E[v(n, j) | X_0 = i] = 1 \times \Pr(X_n = j | X_0 = i) + 0 \times \Pr(X_n \neq j | X_0 = i) = \Pr(X_n = j | X_0 = i) = p_{ij}^{(n)}$ だから

MCMC アルゴリズムによるロジットモデルのベイズ推定に関する若干の考察

$$\frac{1}{N} E[w(N, j) | X_0] = \frac{1}{N} \sum_{n=1}^N E[v(n, j) | \pi^{(0)}] = \frac{1}{N} \sum_{n=1}^N \hat{p}_i^{(n)} \xrightarrow{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \pi_j = \pi_j$$

[⑤⇒⑥]

N が十分大きいとき

$$\frac{1}{N} \sum_{n=1}^N \theta^{(n)} \approx \frac{1}{N} \sum_{j \in S} (j E[w(N, j)]) = \sum_{j \in S} \left(j \frac{E[w(N, j)]}{N} \right) \xrightarrow{N \rightarrow \infty} \sum_{j \in S} (j \pi_j) = E[\pi]$$

(証明終)

次にエルゴード的なマルコフ過程の収束に関する定理を紹介する。証明はマルコフ過程に関する本格的な成書 (例えば [森村・高橋, 1979] や [小山, 2011] など) を参照されたい。

定義 4 $\{X^{(n)}\}_n$ を, 状態空間 S が実数全体 \mathfrak{R} である様な確率変数列とする。さらに各確率変数 $X^{(n)}$ の確率密度関数を $\pi^{(n)}(\theta)$ とする。ここで任意の $A \subset \mathfrak{R}$ および $n \geq 0$ に対して次式が成立するとき, $\{X^{(n)}\}_n$ はマルコフ過程 (Markov process) であるという。

$$\Pr (X^{(n)} \in A | X^{(0)}, X^{(1)}, \dots, X^{(n-1)}) = \Pr (X^{(n)} \in A | X^{(n-1)})$$

斉時的なマルコフ過程に対して次式で定義される関数 K を, 推移カーネル (transition kernel) と呼ぶ。

$$K(\theta, \varphi) \stackrel{def}{=} \Pr (X^{(n)} = \varphi | X^{(n-1)} = \theta), \quad \theta, \varphi \in \mathfrak{R}$$

$$K(\theta, A) \stackrel{def}{=} \int_A \Pr (X^{(n)} = \varphi | X^{(n-1)} = \theta) d\varphi, \quad \theta \in \mathfrak{R}, A \subset \mathfrak{R}$$

マルコフ過程が既約かつ非周期的ならば, エルゴード的 (ergodic) であるという。

定理 5 エルゴード的なマルコフ過程に対して, 不変分布が唯一とつ存在する。

定理 6 K -不変な分布 $\pi(\theta)$ に対して, 次の①~⑥は同値である。

- ① [不変条件] $\pi(\theta) = \int_{\mathfrak{R}} \pi(\varphi) K(\varphi, \theta) d\varphi$
- ② [詳細釣り合条件] $\pi(\theta) K(\theta, \varphi) = \pi(\varphi) K(\varphi, \theta)$
- ③ 任意の初期分布 $\pi^{(0)}(\theta)$ に対して $\pi^{(N)}(\theta) \stackrel{def}{=} \int_{\mathfrak{R}} \pi^{(0)}(\varphi) K^N(\varphi, \theta) d\varphi \rightarrow \pi(\theta)$
- ④ $K^N(\varphi, \theta) \xrightarrow{N \rightarrow \infty} \pi(\varphi)$
- ⑤ 任意の $A \subset \mathfrak{R}$ に対して $w(N, A) \stackrel{def}{=} \# \{n \leq N; \theta^{(n)} \in A\}$ とすると
 $E[w(N, A) | X_0] / N \xrightarrow{N \rightarrow \infty} E[\pi]$
- ⑥ 各期の実現値 $\theta^{(n)} \in \mathfrak{R}$ に対して $\sum_{n=1}^N \theta^{(n)} / N \xrightarrow{N \rightarrow \infty} E[\pi]$

従って前節の定理 1 と上の定理 6 から, 酔歩 Metropolis-Hastings アルゴリズムによるベイ

ズ推定の正当性が、容易に導かれる。

系 $\pi(\theta)$ をターゲット分布, $\{\theta^{(n)}\}_n$ を酔歩 Metropolis-Hastings アルゴリズムによるサンプル系列とすると, $\sum_{n=1}^N \theta^{(n)} / N \xrightarrow{N \rightarrow \infty} E[\pi] = E[\Pr(\theta|\delta)] = \hat{\theta}_B$ が成立する。

5. 酔歩 Metropolis-Hastings サンプリングにおけるチューニングについて

前節の系により, 酔歩 Metropolis-Hastings アルゴリズムによるサンプル系列の平均は, ターゲット分布の期待値, 従って事後分布の期待値, すなわちベイズ推定値に収束する。つまり burn-in 期間 (切り捨てる初期のサンプル数) を M とすると $\sum_{n=M+1}^N \theta^{(n)} / (N-M) \approx E[\pi] = E[\Pr(\theta|\delta)]$ が成立する。ここでサンプリングを実行するためには $\theta^{(0)}$ (サンプル系列の初期値), N, M (サンプリング回数および burn-in 期間), σ (酔歩の歩幅を規定する標準偏差) を定める必要がある。これらを定めることはチューニング (tuning) と呼ばれるが, その方法は他の数値解析と同様に極めて恣意的である。MCMC アルゴリズムにおいては, チューニングの目安として, 次のような指標がある。

- ・ $\{\theta^{(n)}\}_n$ の目視 [照井, p.78]
- ・ $\pi(\theta)$ の目視 [小西, p.192]
- ・ 自己相関因数 (ACF) [小西, p.194]
- ・ 平均採用確率 $\sum_{n=1}^N \alpha_n / N$ [小西, p.195]
 $\dim \theta = 1, 2$ ならば 35~40%, $\dim \theta \geq 3$ ならば 25~35% が妥当
- ・ 採用率 $\sum_{n=1}^N \delta_{u_n \leq \alpha_n} / N$
- ・ Gewerke の判定法 [照井, p.78] サンプル系列を二分割し平均値の差の検定を実施

そこで本節では, 特に酔歩の歩幅の標準偏差 σ の定め方について知見を得ることを目的に, R 言語を用いたシミュレーションを行う。ただし, ここではターゲット分布として, ロジットモデルの誤差項として用いられる (平均 0 の) ガンベル分布を採用する。ただし次式において γ はオイラー定数を表す。

$$\pi(\theta) \stackrel{def}{=} e^{-(\theta+\gamma)} \exp(-e^{-(\theta+\gamma)})$$

具体的には $\theta^{(0)} = 6$, $N = 450$, $M = 50$ を固定し, $\sigma = 0.1, 0.5, 1, 2, 3, 4, 5, 10, 100$ と 9 通りに変化させながら, 酔歩 Metropolis-Hastings アルゴリズムによるサンプリングを実行し, $n \leq M = 50$ を切り捨てて $N - M = 400$ 個のサンプルを採用し, その平均・分散・平均採用確率・採用率等を計算した。その結果をまとめたのが, 次の表である。なおサンプリングを実行するための R プログラ

表 2 シミュレーション結果

σ	$\theta^{(n)}$ の 平均	$\theta^{(n)}$ の 分散	平均 採用確率	採用率	
0.1	4.1906	0.1839	0.9605	0.9644	サンプルが 4 の周辺に集中し、ACF も全く減衰しない。N=10000 にしたところ、M>3000 である事が分かった
0.5	-0.2152	1.8265	0.8424	0.8577	M>100, ACF が緩やかに減衰
1	-0.1411	1.0733	0.7189	0.7044	M>50, ACF の減衰率が上がる
2	0.2380	2.1820	0.5350	0.5511	
3	-0.0278	1.5768	0.4113	0.3911	
4	0.1482	2.3450	0.3377	0.3400	
5	-0.0571	1.3146	0.2938	0.2777	
10	-0.2383	1.5260	0.1257	0.1333	
100	0.0634	1.2328	0.0178	0.0177	平均採用確率・採用率がともに低すぎて、シミュレーション結果が毎回異なる

ムは付録 1 に、そしてサンプル系列の時系列プロット等は付録 2 に掲載した。

今回のシミュレーションによって得られた知見は以下の通りである。

- ・平均採用確率と採用率は、常にほぼ等しい。これはある程度、大数の法則によって説明できる。
- ・ σ が小さすぎると $\theta \approx \theta^{(n-1)}$ 従って $\pi(\theta) \approx \pi(\theta^{(n-1)})$ となり、採用確率 $\alpha^{(n)}$ および ACF (自己相関係数) が 1 に近づく。これは、初期値 $\theta^{(0)}$ の近くばかりをサンプリングしている事を意味する。 $\theta^{(n)}$ の平均は真の平均 0 と異なった。この場合は burn-in 期間 M が長くなるので、サンプル数 N を大きくする必要がある
- ・ σ が大きすぎると平均採用確率や採用率が小さくなる。これは $\theta^{(n)} = \theta^{(n-1)}$ であるケースが増え、サンプリングの効率が低下している事を意味する。 $\theta^{(n)}$ の平均は真の平均 0 と異なった。この場合も burn-in 期間 M が長くなる。
- ・今回は $\sigma = 3$ の時が最も良好であった、ちなみに採用率は約 0.4 である。

6. おわりに

本稿では、ロジットモデル等において、選好に関する消費者の異質性を積極的に取り込むために用いられる MCMC アルゴリズム (特にそのひとつである酔歩 Metropolis-Hastings アルゴリズム) によるベイズ推定について、若干の考察を行った。3 節で酔歩 Metropolis-Hastings ア

ルゴリズムを詳しく説明した後に、4 節では生成されたサンプル系列の平均がターゲット分布、従って事後分布の期待値、すなわちベイズ推定値に収束する事を、2 値のマルコフ連鎖に置き換えて詳しく証明した。さらに5 節ではパラメータを推定する際のチューニングについて、とくにパラメータ更新のステップ幅を規定する標準偏差がアルゴリズムのパフォーマンスに与える影響について、シミュレーション結果に基づいて考察した。その結果、標準偏差が小さすぎても大きすぎてもパフォーマンスが低下するので、サンプル系列の目視や平均採用確率等を参考にしながら試行錯誤する事が重要である事が確認された。5 節で述べた通り、MCMC アルゴリズムも他の数値解析と同様、チューニングの方法は極めて恣意的である。従ってサンプル系列の収束性が数学的に証明されたとしても、実行に当たっては、いくつかの指標を参考にしながら、試行錯誤を繰り返す事が重要であると考えられる。

注

- 1) 本稿ではこれ以降、状態空間が2 値、すなわち $S=\{1,2\}$ である場合を考える。
- 2) N 期までに状態 j を訪問する回数を表す。

参 考 文 献

- 伊藤清『確率論』岩波書店, 1991.
- 稲田有香・室町泰徳 (2006)「マーケティング分野におけるベイズ推定モデルに関する基礎的研究」『土木計画学研究・講演集』(CD-ROM).
- 伊庭幸人・種村正美『計算統計Ⅱ』岩波書店, 2005.
- 片平秀貴・杉田善弘 (1994)「マーケティング・サイエンスの最近の動向」『オペレーションズ・リサーチ』Vol.39, No.4.
- 小西貞則・越智義道・大森裕浩『計算統計学の方法』朝倉書店, 2008.
- 小山昭雄『確率論』(新装版 経済数学教室 9) 岩波書店, 2011.
- 佐武一郎『線形代数学』裳華房, 1974.
- 土田尚弘 (2010)「マーケティング・サイエンスにおける離散選択モデルの展望」『経営と制度』No.8.
- 照井伸彦『ベイズモデリングによるマーケティング分析』東京電機大学出版局, 2008.
- 舟尾暢男『The R Tips』(第二版) オーム社, 2009.
- 森村英典・高橋幸雄『マルコフ解析』日科技連出版社, 1979.
- Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Household Heterogeneity," *Journal of Econometrics*, Vol.89.
- Malhotra, Naresh K. (1984), "The Use of Linear Logit Models in Marketing Research," *Journal of Marketing Research*, Vol.21, No.1.
- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes 3rd Edition*, Cambridge University Press, 2007.
- Rossi, Peter E., Greg M. Allenby, and Robert McCulloch, *Bayesian Statistics and Marketing*, Wiley, 2006.

付録1 シミュレーションの R プログラム

```
#####
##          Gumbel 分布の 酔歩 M-H サンプリング          ##
#####

##### 1. ターゲット分布である Gumbel 分布の密度関数 Gumbel( $\theta$ ) の定義, ただし平均=0 とする
gamma <- 0.577208593849      # 変数  $\gamma$  に, Euler 定数の近似値を代入
Gumbel <- function(theta){   # 関数 Gumbel( $\theta$ ) の定義
  exp(-(theta+gamma))*exp(-exp(-(theta+gamma)))
                                # 定義式
}                                # 関数 Gumbel の定義の終わり

##### 2. M-H サンプリングの準備
N <- 450                      # 変数 N に, サンプリング回数を代入
M <- 50                       # 変数 M に, burn-in 期間を代入
theta0 <- 6                   # 変数  $\theta_0$  に, サンプルの初期値を代入
sigma <- 3                    # 変数  $\sigma$  に, 酔歩の歩幅の標準偏差を代入

theta <- numeric(N)          # サイズ N のベクトル変数  $\theta$  を用意 サンプル数列を格納
theta[1] <- theta0           #  $\theta$  の第 1 成分 theta[1] に, 初期値  $\theta(0)$  を代入
```

```

alpha <- numeric(N)          # サイズ N ベクトル変数  $\alpha$  を用意 採用確率  $\alpha[n]$  を格納
accept <- numeric(N)        # サイズ N のベクトル変数 accept を用意 採用ダミーを格納

z <- rnorm(N, mean=0, sd=sigma)
  # サイズ N の正規乱数ベクトル変数  $z \sim N(0, \sigma^2)$  を用意 歩幅  $z[n]$  を決定
u <- runif(N)                # サイズ N の一様乱数ベクトル変数 u を用意...採用・棄却を決定

##### 3. M-H サンプリングの実行
for ( n in 2:N ) {          # 反復計算回数を表す変数 n が 2~N の場合に関して
##### 3-1. 採用確率  $\alpha[n]$  の計算
  theta.temp <- theta[n-1]+z[n] # 変数  $\theta_{temp}$  に、 $\theta[n-1]+z[n]$  を代入
  if (Gumbel(theta[n-1])>0) { # Gumbel( $\theta[n-1]$ ) > 0 の場合は
    alpha[n] <- min( Gumbel(theta.temp) / Gumbel(theta[n-1]), 1 )
                                #  $\alpha[n]$  に、min(Gumbel 関数の比, 1) を代入
  }                               # Gumbel( $\theta[n-1]$ ) > 0 の場合の終わり
  else {                          # そうでない、すなわち Gumbel( $\theta[n-1]$ ) = 0 の場合は
    alpha[n] <- 1                  #  $\alpha[n]$  に、1 を代入
  }                               # そうでない場合の終わり
##### 3-2.  $\theta_{temp}$  の採用・棄却
  if ( u[n] <= alpha[n] ) {       #  $u[n] \leq \alpha[n]$  の場合は
    theta[n] <- theta.temp        # 採用:  $\theta[n]$  に、 $\theta_{temp}$  を代入
    accept[n] <- 1                # 採用ダミーに、1 を代入
  }                               #  $u[n] \leq \alpha[n]$  の場合の終わり
  else {                          # そうでない、すなわち  $u[n] > \alpha[n]$  の場合は
    theta[n] <- theta[n-1]        # 棄却:  $\theta[n]$  に、 $\theta[n-1]$  を代入, 採用ダミーは 0 のまま
  }                               # そうでない場合の終わり
}                               # 反復計算の終わり

##### 4. サンプリング結果のグラフ表示
##### 4-1. 準備
par(mfrow=c(3,1))            # 1 ページにグラフを 3 つ縦に並べる
theta.band <- c(-3,7)        # 分布の定義域の範囲
p.band <- c(0,0.6)           # 分布の値域の範囲
##### 4-2. サンプリング系列  $\theta$  の時系列プロット
plot(theta, type="l", ylim=theta.band, xlab="n") # 縦軸の範囲は theta.band に

```

```

従う
title(                                     # グラフのタイトル (超絶技巧なので、初心者は理解不要)
  substitute(
    list(theta[0], N, M, sigma)
    == group("(", list(theta0temp, Ntemp, Mtemp, sigmatemp), ")"),
    list(theta0temp=theta0, Ntemp=N, Mtemp=M, sigmatemp=sigma)
  )
)
# グラフのタイトルの終わり
##### 4-3. burn-in 期間 (1:M) を除いたサンプリング系列  $\theta$  のコログラム (自己相関, ACF)
acf(theta[-(1:M)], main=" ") # タイトルは空欄
##### 4-4. ノンパラメトリックに推定した密度関数のプロット
plot(Gumbel, xlim=theta.band, ylim=p.band, xlab=" ", ylab=" ")
# 参考のため、Gumbel 関数のグラフをプロット
segments(-gamma, 0, -gamma, Gumbel(-gamma), lty=3)
# 参考のため、Gumbel 関数の最頻値に縦点線を引く
# A(- $\gamma$ , 0) B(- $\gamma$ , Gumbel(- $\gamma$ ))
par(new=T) # 上のグラフに下のグラフを上書きする
plot(density(theta[-(1:M)]), xlim=theta.band, ylim=p.band, type="p", main=" ")
# ノンパラメトリックに推定した密度関数を  $\circ$  でプロット
# rug(theta[-(1:M)])
par(new=F); par(mfrow=c(1,1)) # 既定値に戻す

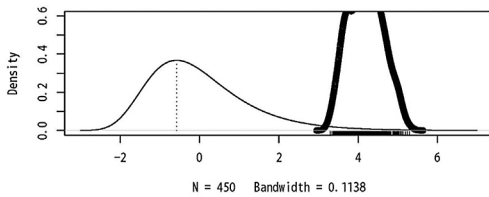
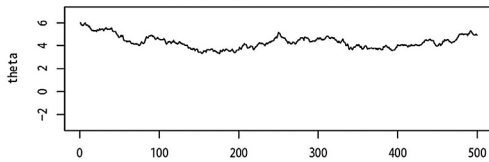
##### 5. ターゲット分布の平均値等の計算
mean(theta[-(1:M)]) # burn-in 期間を除くサンプルの平均値=ターゲット分布の平均値
var(theta[-(1:M)]) # burn-in 期間を除くサンプルの分散=ターゲット分布の分散
mean(alpha[-(1:M)]) # burn-in 期間を除く採用確率の平均値
mean(accept[-(1:M)]) # burn-in 期間を除く採用ダミーの平均値=平均採用率

##### 6. MonteCarlo シミュレーション
Ginv <- function(x) -log(-log(x))-gamma # Gumbel 関数の累積関数の逆関数
Ginv(x) の定義
mean(Ginv(runif(N))) # N 個の一樣乱数 runif(N) を使った MC

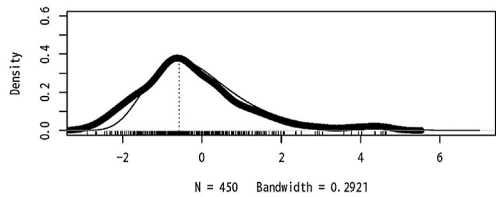
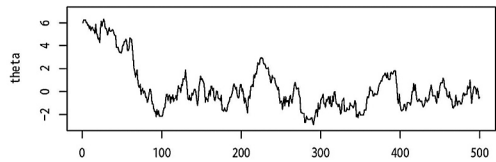
```


付録 2 シミュレーション結果のグラフ

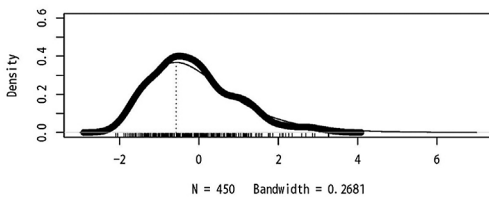
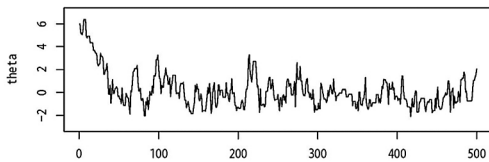
初期値・サンプリング数・burn-in 期間をそれぞれ $\theta^{(0)}=6$, $N=450$, $M=50$ と固定し, 酔歩の歩幅を規定する標準偏差を $\sigma=0.1, 0.5, 1, 2, 3, 4, 5, 10, 100$ と 9 通りに変化させながら, 酔歩 Metropolis-Hastings アルゴリズムによるサンプリングを実行し, $n \leq M=50$ を切り捨てて $N-M=400$ 個のサンプルを採用した結果である。上段はサンプル系列の時系列プロット (横軸が n , 縦軸が θ) である。また下段はターゲット分布の実際の密度関数および, サンプル系列を用いてノンパラメトリックに推定した密度関数である。横軸上の点はサンプル系列である。



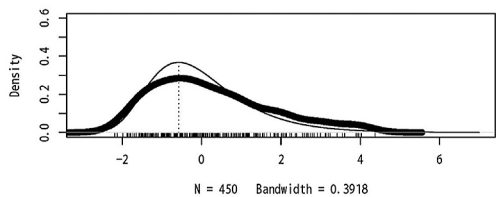
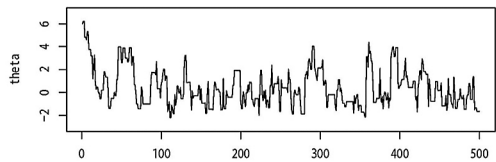
$\sigma=0.1$



$\sigma=0.5$

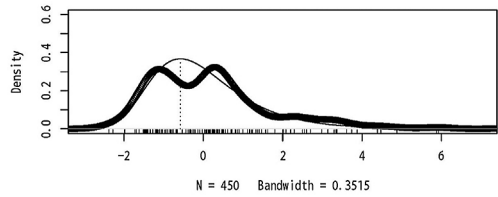
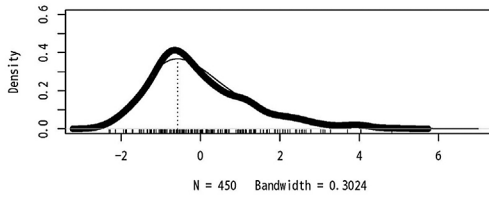
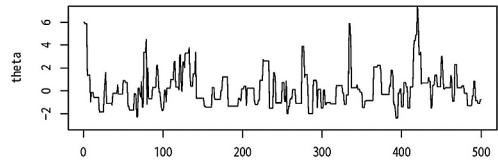
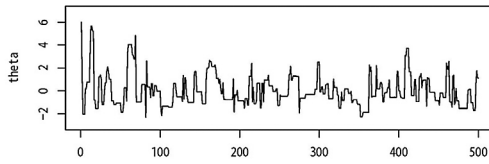


$\sigma=1$



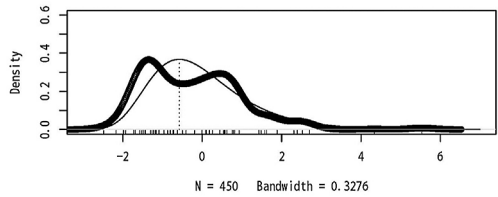
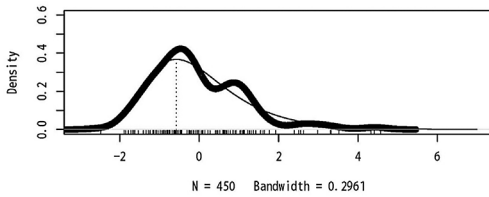
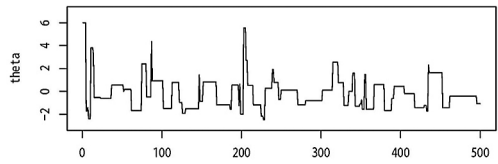
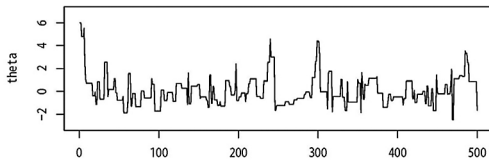
$\sigma=2$

MCMC アルゴリズムによるロジットモデルのベイズ推定に関する若干の考察



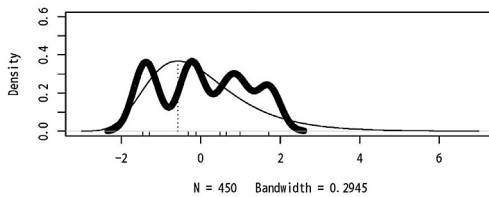
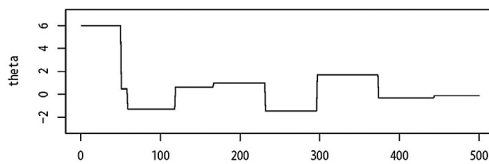
$\sigma = 3$

$\sigma = 4$



$\sigma = 5$

$\sigma = 10$



$\sigma = 100$

—2011年10月21日受領—