

再帰的に切断された多変量正規分布を用いた 線形制約下の多変量正規分布からの乱数発生法

宮 脇 幸 治*

A random number generator
of the multivariate normal distributions subject to linear constraints
using the recursively truncated multivariate normal distribution

Koji Miyawaki*

本論文では、線形制約下の多変量正規分布からの乱数発生法について、その代表的な2種類の方法を紹介すると共に新しい発生方法を提案する。提案される方法は再帰的に切断された多変量正規分布を提案分布とする Metropolis-Hastings アルゴリズムである。その特徴は以下の2点を挙げることができる。(1) 同時分布からの乱数発生法である。(2) Metropolis within Gibbs として他のマルコフ連鎖モンテカルロ法に埋め込むことができる。これらの特徴を2種の数値実験により確かめると共に、その応用例として、離散連続選択アプローチに基づくブロック料金制下の需要関数の推定を行う。

This paper reviews two representative random number generators of the multivariate normal distributions subject to linear constraints and proposes another generator of this distribution. Proposed method jointly generates a vector of the random numbers by the Metropolis-Hastings algorithm based on the recursively truncated multivariate normal proposal, and has two following properties: (1) It jointly generates normal variates from its distribution. (2) It can be used as a part of Markov chain Monte Carlo method which is known as Metropolis within Gibbs. Our method is illustrated by two numerical examples and is applied to estimate the demand function under block rate pricing derived by the discrete/continuous choice approach.

Key Words and Phrases: Random number generator, Multivariate normal distribution, Recursively truncated multivariate normal distribution, Markov chain Monte Carlo

1. はじめに

一変量の正規分布からの乱数発生法は、Box and Muller (1958) の方法に始まり、様々な方法が提案されてきた。正規分布は非常に多くの統計的モデルで用いられているため、モデルの模擬実験を行う際に正規乱数を必要することが多い。また近年では、マルコフ連鎖モンテカルロ (MCMC) 法を用いたベイズ推定を行う場合にも正規乱数は頻繁に用いられている。本論文では、MCMC 法を用いたベイズ推定における利用を目的とした正規乱数の発生法、特に線形制約下の多変量正規乱数の発生について分析を行う。

最も代表的な線形制約下の多変量正規分布に従う乱数発生法は Geweke (1991) によるものである。この論文では変数変換を伴ったギブスサンプラー (Gibbs sampler) による方法が提案されている。しかし、目標とする分布によっては得られる標本の自己相関が高くなる場合が

* 東京大学大学院経済学研究科博士課程：〒113-0033 文京区本郷 7-3-1, 日本学術振興会特別研究員

見られ、その状況を改善するために、Rodriguez-Yam, Davis and Scharf (2004) では別の変数変換を伴うギブスサンプラーが提案されている。

これに対して、新しい乱数発生法は同時分布からの乱数発生を行う。具体的には、線形制約式によって切断された多変量正規分布からの乱数発生を直接行うことは難しいため、再帰的に切断された多変量正規分布 (recursively truncated multivariate normal distribution) を提案分布とする Metropolis-Hastings (M-H) アルゴリズムを用いて乱数発生を行う。再帰的に切断された多変量正規分布は多変量分布の一つで、一変量の切断正規分布の積で表されるため乱数発生が容易であるという特徴を持つ。

本論文の構成は以下の通りである。まず2節では線形制約下の多変量正規分布からの乱数発生法、従来の方法及び本論文で提案する方法を説明する。続く3節ではこれらの乱数発生法を数値例を用いて比較する。また、新しい乱数発生法の応用例として離散連続選択アプローチに基づくブロック料金制下の需要関数の推定を行うため、そのモデルについて4節で説明する。このモデルを用いた数値実験を5節で行った後、日本の電力需要データを用いた実証結果を6節で示す。最後に7節で本論文のまとめ及び今後の課題について触れる。

2. 線形制約下の多変量正規分布に従う乱数発生

線形制約 $\mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}$ によって切断された n 変量正規分布に従う確率変数ベクトル $\mathbf{x}$ の乱数を得ることを考える。つまり、

$$\mathbf{x} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}. \quad (2.1)$$

ここで $\boldsymbol{\mu}$, $\mathbf{a}$ と $\mathbf{b}$ は既知の n 次元ベクトル, $\boldsymbol{\Sigma}$ は既知の $n \times n$ 正定値行列, $\mathbf{D}$ は既知の $m \times n$ 行列である。また多変量正規分布の下付き数字はその次元を表す。ただし一変量正規分布の場合はその添字を省略する。このような線形制約下の多変量正規分布からの乱数発生は、MCMC法を用いたベイズ推定において頻繁に必要とされるものであり、その一例は4節のブロック料金制下の需要関数の推定にも見ることができる。以下では、従来用いられてきた乱数発生法のうち代表的なものを2種類紹介すると共に本論文で新しく提案するブロックサンプラーと呼ばれる方法について説明する。

2.1 従来の乱数発生法

従来の乱数発生法ではMCMC法の標準的手法の一つであるギブスサンプラーが応用されてきた。その説明のため、 $n-1$ 次元ベクトル $\mathbf{z}_{-i}$ を n 次元ベクトル $\mathbf{z}$ からその第 i 要素 z_i を除いたものと定義しておく。

Geweke (1991) では、式 (2.1) において $\mathbf{D}$ の次元を $m=n$ とし、 $\text{rank}(\mathbf{D})=n$ と仮定した場合の乱数発生法が提案されている。ただし後述の方法を用いれば、 $\text{rank}(\mathbf{D}) < n$ であっても適用可能である。その乱数発生法は次の3ステップで構成される。

Geweke (1991) の乱数発生法

Step 1. $\mathbf{T} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}'$, $\tilde{\mathbf{a}} = \mathbf{a} - \mathbf{D}\boldsymbol{\mu}$, $\tilde{\mathbf{b}} = \mathbf{b} - \mathbf{D}\boldsymbol{\mu}$ を求める。

Step 2. ギブスサンプラーを用いて $\mathbf{z}$ を以下の切断多変量正規分布より発生させる。

$$\mathbf{z} \sim N_n(\mathbf{0}, \mathbf{T}), \quad \tilde{\mathbf{a}} \leq \mathbf{z} \leq \tilde{\mathbf{b}}. \quad (2.2)$$

ギブスサンプラーは次の3ステップよりなる。

(a). $\mathbf{z}$ に初期値を与える。

(b). 各 $i=1, \dots, n$ に対して、 z_i を次の切断正規分布より発生させる。

$$z_i | \mathbf{z}_{-i} \sim N(\mathbf{t}'_{-i} \mathbf{T}_{-i,-i}^{-1} \mathbf{z}_{-i}, t_{ii} - \mathbf{t}'_{-i} \mathbf{T}_{-i,-i}^{-1} \mathbf{t}_{-i}), \quad \tilde{a} \leq z_i \leq \tilde{b}_i. \quad (2.3)$$

(c). Step 2-(b)に戻る.

Step 3. $\mathbf{x} = \boldsymbol{\mu} + \mathbf{D}^{-1} \mathbf{z}$ と変換する.

ただし $n-1$ 次元ベクトル $\mathbf{t}_{-i}$ は, $\mathbf{T}$ の第 i 列を $\mathbf{t}_i$ とすると, $\mathbf{t}_i$ から第 i 要素を除いたもの, $n-1 \times n-1$ 行列 $\mathbf{T}_{-i,-i}$ は $\mathbf{T}$ から第 i 行, 第 i 列を除いたもの, 及びスカラー t_{ii} は $\mathbf{T}$ の第 (i, i) 要素である.

一方, Rodriguez-Yam et al. (2004) では式 (2.1) に従う別の乱数発生法が提案されている. 具体的な乱数発生法は以下の 3 ステップである.

Rodriguez-Yam et al. (2004) の乱数発生法

Step 1. $\tilde{\mathbf{D}} = \mathbf{D} \mathbf{A}^{-1}$, $\tilde{\boldsymbol{\mu}} = \mathbf{A} \boldsymbol{\mu}$ を求める. ここで $\mathbf{A}$ は $\mathbf{A} \boldsymbol{\Sigma} \mathbf{A}' = \mathbf{I}$ を満たすような $n \times n$ 正則行列とする. 通常は $\boldsymbol{\Sigma}^{-1}$ のコレスキー分解を転置したものを用いる.

Step 2. ギブスサンプラーを用いて $\mathbf{z}$ を以下の切断多変量正規分布より発生させる.

$$\mathbf{z} \sim N_n(\tilde{\boldsymbol{\mu}}, \mathbf{I}), \quad \mathbf{a} \leq \tilde{\mathbf{D}} \mathbf{z} \leq \mathbf{b}. \quad (2.4)$$

ギブスサンプラーは次の 3 ステップよりなる.

(a). $\mathbf{z}$ に初期値を与える.

(b). 各 $i=1, \dots, n$ に対して, z_i を次の切断正規分布より発生させる.

$$z_i | \mathbf{z}_{-i} \sim N(\tilde{\mu}_i, 1), \quad \mathbf{a} - \tilde{\mathbf{D}}_{-i} \mathbf{z}_{-i} \leq \tilde{\mathbf{d}}_i z_i \leq \mathbf{b} - \tilde{\mathbf{D}}_{-i} \mathbf{z}_{-i}. \quad (2.5)$$

(c). Step 2-(b)に戻る.

Step 3. $\mathbf{x} = \mathbf{A}^{-1} \mathbf{z}$ と変換する.

ここで $\tilde{\mathbf{d}}_i$ は $\tilde{\mathbf{D}}$ の第 i 列, $\tilde{\mathbf{D}}_{-i}$ は $\tilde{\mathbf{D}}$ から第 i 列 $\tilde{\mathbf{d}}_i$ を削除した行列を表す.

上の 2 方法は, いずれもギブスサンプラーを用いているために, 条件付き分布からの乱数発生となっている. そのため, 目標とする分布によっては得られる標本の自己相関が高い場合が観察される. Rodriguez-Yam et al. (2004) の方法では, その状況を改善するために別の変数変換を行っているが, 場合によっては, 得られた標本の自己相関が高くなり, 乱数発生がやや非効率的になる場合がある. これに対して, 本論文で提案される方法は同時分布からの乱数発生であり, 目標とする分布によらず効率的な乱数発生が期待される. 次節では, 新しく提案される乱数発生法について説明する.

2.2 ブロックサンプラー

本節では線形制約下の多変量正規分布からの新しい乱数発生法について説明する. 従来の方法とは異なり, 同時分布からの乱数発生を行うためブロックサンプラーと呼ぶことにする. ブロックサンプラーでは, 再帰的に切断された多変量正規分布により目的の切断多変量正規分布を近似する提案分布を構成し, M-H アルゴリズムと棄却サンプリングによって乱数発生を行う. この方法の特徴は次の 2 点を挙げることができる.

1. 同時分布からの乱数発生法である.

2. Metropolis within Gibbs として他の MCMC 法に埋め込むことができる.

以下, 単純な状況から式 (2.1) の状況までの 3 種類の状況において, ブロックサンプラーのアルゴリズムをそれぞれ構成する.

2.2.1 $\mathbf{D}$ が下三角行列かつ $\boldsymbol{\Sigma}$ が単位行列の場合

以下の切断多変量正規分布に従う確率変数ベクトル $\mathbf{z}$ の乱数を発生させることを考える.

$$\mathbf{z} \sim N_n(\mathbf{0}, \mathbf{I}), \quad \tilde{\mathbf{a}} \leq \mathbf{L}\mathbf{z} \leq \tilde{\mathbf{b}}. \quad (2.6)$$

ここで $\tilde{\mathbf{a}}$ 及び $\tilde{\mathbf{b}}$ は n 次元の既知ベクトルであり, $\mathbf{L}$ は $n \times n$ の下三角行列かつその対角成分は正であるとする. つまり, $\mathbf{L}$ の第 (i, j) 要素を l_{ij} とすると, $l_{ij}=0 (i>j)$ かつ $l_{ij}>0$.

式 (2.6) に従う乱数は, M-H アルゴリズムを用いて以下の 4 ステップで発生させることができる.

アルゴリズム BS-1: ブロックサンプラー, $\mathbf{D}$ が下三角行列かつ Σ が単位行列の場合

Step 1. 初期値 $\mathbf{z}^{(0)}$ を $\tilde{\mathbf{a}} \leq \mathbf{L}\mathbf{z}^{(0)} \leq \tilde{\mathbf{b}}$ を満たすように与え, $m=0$ とする.

Step 2. $\mathbf{z}^{(m+1)}$ の候補 $\mathbf{z}^*$ を発生させる.

(a). z_1^* を以下の切断正規分布から発生させる.

$$z_1^* \sim N(0, 1), \quad \frac{\tilde{a}_1}{l_{11}} \leq z_1^* \leq \frac{\tilde{b}_1}{l_{11}}. \quad (2.7)$$

(b). $z_i^* (i=2, \dots, n)$ を以下の条件付き切断正規分布から発生させる.

$$z_i^* | z_1^*, \dots, z_{i-1}^* \sim N(0, 1), \quad \frac{\tilde{a}_i - \sum_{j<i} l_{ij} z_j^*}{l_{ii}} \leq z_i^* \leq \frac{\tilde{b}_i - \sum_{j<i} l_{ij} z_j^*}{l_{ii}}. \quad (2.8)$$

Step 3. Step 2 で得られた候補 $\mathbf{z}^*$ を確率 $\alpha(\mathbf{z}^{(m)}, \mathbf{z}^*)$ で $\mathbf{z}^{(m+1)}$ として受容する. 棄却した場合には $\mathbf{z}^{(m+1)} = \mathbf{z}^{(m)}$ とする.

Step 4. $m \rightarrow m+1$ とし, Step 2 に戻る.

ここで確率 $\alpha(\mathbf{z}^{(m)}, \mathbf{z}^*)$ は, 標準正規分布の累積分布関数 $\Phi(\cdot)$ を用いて,

$$\alpha(\mathbf{z}^{(m)}, \mathbf{z}^*) = \min \left\{ 1, \frac{w(\mathbf{z}^*)}{w(\mathbf{z}^{(m)})} \right\}, \quad (2.9)$$

$$w(\mathbf{z}) = \left\{ \Phi\left(\frac{\tilde{b}_1}{l_{11}}\right) - \Phi\left(\frac{\tilde{a}_1}{l_{11}}\right) \right\} \left[\prod_{i=2}^n \left\{ \Phi\left(\frac{\tilde{b}_i - \sum_{j<i} l_{ij} z_j}{l_{ii}}\right) - \Phi\left(\frac{\tilde{a}_i - \sum_{j<i} l_{ij} z_j}{l_{ii}}\right) \right\} \right], \quad (2.10)$$

と表される.

上の M-H アルゴリズムで用いられている提案分布は, $n \times n$ 単位行列 $\mathbf{I}_n$ を用いて,

$$f(\mathbf{z}) = w(\mathbf{z})^{-1} \phi(\mathbf{z}, \mathbf{I}_n) I(\mathbf{a} \leq \mathbf{L}\mathbf{z} \leq \mathbf{b}),$$

$$\text{ただし } \phi(\mathbf{z}, \mathbf{\Omega}) = \{\det(2\pi\mathbf{\Omega})\}^{-1/2} \exp\left(-\frac{1}{2} \mathbf{z}' \mathbf{\Omega} \mathbf{z}\right), \quad (2.11)$$

となっており, 再帰的に切断された多変量正規分布と呼ばれる. この再帰的に切断された多変量正規分布は GHK simulator で主に用いられる分布である. GHK simulator とは切断多変量正規分布に従う確率変数の関数の期待値を評価するための重点サンプリングを用いた手法を指す (GHK simulator 及び再帰的に切断された多変量正規分布については Hajivassiliou and Ruud, 1994 を参照).

2.2.2 $\mathbf{D}$ が正則行列の場合

次の切断多変量正規分布に従う確率変数ベクトル $\mathbf{x}$ の乱数を得ることを考える.

$$\mathbf{x} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}. \quad (2.12)$$

ここで $\boldsymbol{\mu}$, $\mathbf{a}$ と $\mathbf{b}$ は既知の n 次元ベクトル, $\boldsymbol{\Sigma}$ は既知の $n \times n$ 正定値行列, $\mathbf{D}$ は既知の $n \times n$ の正則行列である. これは式 (2.1) において $m=n$ とした状況であり, Geweke (1991) の方

法が対象とした状況でもある。

この様な状況での乱数発生はアルゴリズム BS-1 に変数変換のステップを加えればよく、以下の3ステップで構成される。

アルゴリズム BS-2：ブロックサンプラー， D が正則行列の場合

Step 1. 下三角行列 L 及びベクトル $\tilde{a} = a - D\mu$, $\tilde{b} = b - D\mu$ を求める。ただし下三角行列 L は $LL' = D\Sigma D'$ を満たす下三角行列とする。通常，下三角行列 L は $D\Sigma D'$ をコレスキー分解したものを用いる。

Step 2. アルゴリズム BS-1 を用いて z を以下の切断多変量正規分布から発生させる。

$$z \sim N_n(0, I), \quad \tilde{a} \leq Lz \leq \tilde{b}. \quad (2.13)$$

Step 3. $x = \mu + D^{-1}Lz$ と変換する。

このブロックサンプラーで発生させた x が式 (2.12) に従うことは、次の同値変形で示される。

$$\begin{aligned} z \sim N_n(0, I), \quad \tilde{a} \leq Lz \leq \tilde{b} &\iff D\mu + Lz \sim N_n(D\mu, D\Sigma D'), \quad a \leq D\mu + Lz \leq b \\ &\iff \mu + D^{-1}Lz \sim N_n(\mu, \Sigma), \quad a \leq D(\mu + D^{-1}Lz) \leq b \\ &\iff x \sim N_n(\mu, \Sigma), \quad a \leq Dx \leq b. \end{aligned} \quad (2.14)$$

アルゴリズム BS-2 を他のサンプラーに埋め込む場合，Step 2 で用いるアルゴリズム BS-1 は初期化 (Step 1) 及びループ (Step 4) の為のステップを除いたもの (Step 2 と 3) でよい。このような方法は Metropolis within Gibbs と呼ばれる。

2.2.3 D が正則行列ではない場合

本小節では式 (2.1) に従う乱数を得ることを考える。乱数発生法の手順としては、まず切断された領域を x の次元 (n) と同数の線形制約式の組で近似する。そうすればアルゴリズム BS-2 を適用できる。次に、アルゴリズム BS-2 を用いて乱数を発生させた後、その乱数を棄却サンプリングを利用して受容または棄却する。以上の手順を踏むことで、式 (2.1) に従う乱数を発生させることができる。

まず切断領域を近似するアルゴリズムを以下に示す。

アルゴリズム BS-A：切断領域の近似

Step 1. $\text{rank}(D) = n$ かつ $m = n$ であれば、 $(\tilde{a}, \tilde{D}, \tilde{b}) = (a, D, b)$ として終了。

Step 2. $\text{rank}(D) = n$ かつ $m > n$ であれば、 $(a^*, D^*, b^*) = (a, D, b)$ として、Step 4 へ。

Step 3. $\text{rank}(D) < n$ であれば、

$$(a^*, D^*, b^*) = \left[\begin{pmatrix} a \\ -\infty \mathbf{1}_n \end{pmatrix}, \begin{pmatrix} D \\ I_n \end{pmatrix}, \begin{pmatrix} b \\ +\infty \mathbf{1}_n \end{pmatrix} \right], \quad (2.15)$$

とする。

Step 4. $k=1$ として、 (a^*, D^*, b^*) から k 番目の制約式から始まる連続した n 本の制約式の組を選ぶ。

$$(\tilde{a}, \tilde{D}, \tilde{b}) = \left[\begin{pmatrix} a_k^* \\ \vdots \\ a_{k+n-1}^* \end{pmatrix}, \begin{pmatrix} d_k^{*'} \\ \vdots \\ d_{k+n-1}^{*'} \end{pmatrix}, \begin{pmatrix} b_k^* \\ \vdots \\ b_{k+n-1}^* \end{pmatrix} \right]. \quad (2.16)$$

Step 5. もし $\det(\tilde{D}) = 0$ であれば $k \rightarrow k+1$ とし Step 4 に戻る。そうでなければ $(\tilde{a}, \tilde{D}, \tilde{b})$ を返して終了。

ここで $-\infty \mathbf{1}_n = (-\infty, \dots, -\infty)'$, $+\infty \mathbf{1}_n = (+\infty, \dots, +\infty)'$, $\mathbf{d}_i'$ は $\mathbf{D}^*$ の第 i 行ベクトルを表す.

もし $\text{rank}(\mathbf{D}) = n$ であれば, 得られた制約式の組 $(\tilde{\mathbf{a}}, \tilde{\mathbf{D}}, \tilde{\mathbf{b}})$ が構成する切断領域 $\{\mathbf{x} | \tilde{\mathbf{a}} \leq \tilde{\mathbf{D}}\mathbf{x} \leq \tilde{\mathbf{b}}\}$ は元の切断領域 $\{\mathbf{x} | \mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}\}$ と同一もしくはその一部である. 一方 $\text{rank}(\mathbf{D}) < n$ の場合, このアルゴリズムは元の制約式の組 $(\mathbf{a}, \mathbf{D}, \mathbf{b})$ に $(-\infty \mathbf{1}_n, \mathbf{I}_n, +\infty \mathbf{1}_n)$ を連結したものから連続した行ベクトルを $\mathbf{x}$ の次元に応じて取り出すものとなっている. そのため, 得られた切断領域 $\{\mathbf{x} | \tilde{\mathbf{a}} \leq \tilde{\mathbf{D}}\mathbf{x} \leq \tilde{\mathbf{b}}\}$ は元の切断領域 $\{\mathbf{x} | \mathbf{a} \leq \mathbf{D}\mathbf{x} \leq \mathbf{b}\}$ を含んでいる.

また以下の命題より, Step 2 は必ず有限回で終了することが保証されている.

命題 1. $m \times n$ 行列 $\mathbf{D} (m \geq n)$ の階数が n である時またその時に限り $\mathbf{D}$ から連続した n 本の一次独立な行ベクトルを取り出すことができる.

命題の証明は付論 A.1 を参照されたい.

切断領域を近似することができれば, 式 (2.1) に従う乱数発生ブロックサンプラーは以下の 3 ステップで記述される.

アルゴリズム BS-3: ブロックサンプラー, $\mathbf{D}$ が正則行列ではない場合

Step 1. アルゴリズム BS-A を用いて近似された切断領域 $\{\mathbf{x} | \tilde{\mathbf{a}} \leq \tilde{\mathbf{D}}\mathbf{x} \leq \tilde{\mathbf{b}}\}$ を得る.

Step 2. アルゴリズム BS-2 によって候補 $\mathbf{x}^*$ を,

$$\mathbf{x}^* \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \tilde{\mathbf{a}} \leq \tilde{\mathbf{D}}\mathbf{x} \leq \tilde{\mathbf{b}}, \quad (2.17)$$

から発生させる.

Step 3. もし $\mathbf{x}^*$ が $\mathbf{a} \leq \mathbf{D}\mathbf{x}^* \leq \mathbf{b}$ を満たせば, $\mathbf{x} = \mathbf{x}^*$ として終了. そうでなければ棄却して Step 2 に戻る.

アルゴリズム BS-A による切断領域の近似が悪い場合, 棄却サンプリングのステップ (Step 3) における受容率が低くなり, 本アルゴリズムの効率が低下する. そのような場合, アルゴリズム BS-A の Step 1 で設定される k の初期値を変更する等, 異なる制約式の組による近似を試す必要がある¹⁾.

3. 数値実験 I

本節では Geweke (1991) の方法, Rodriguez-Yam et al. (2004) の方法及びブロックサンプラーを数値実験により比較する. 数値実験では, 次の 2 変量切断正規分布に従う乱数 $\mathbf{x} = (x_1, x_2)'$ を発生させることを考える.

$$\mathbf{x} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_x), \quad a \mathbf{1}_2 \leq \mathbf{D}_x \mathbf{x} \leq b \mathbf{1}_2, \quad \text{ただし } \boldsymbol{\Sigma}_x = \begin{bmatrix} 10 & \sigma_{12} \\ \sigma_{12} & 0.1 \end{bmatrix}, \quad \mathbf{D}_x = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (3.1)$$

式 (3.1) において, σ_{12} , a 及び b は定数であり, これらの値は表 1 に見られる通りにそれぞれ設定されている. $\boldsymbol{\Sigma}_x$ の構造から σ_{12} は相関係数に等しいことを注意しておく. 表 1 には, 後

表 1 数値実験 I の設定

ケース	(σ_{12}, a, b)	x_1, x_2 の周辺分布の平均及び標準偏差
1	$(-.7, -10, 10)$	$\left\{ (E(x_1), \sqrt{V(x_1)}) = (0, 3.14), (E(x_2), \sqrt{V(x_2)}) = (0, .32) \right\}$
2	$(0, -10, 10)$	
3	$(-.7, -1, 1)$	$\left\{ (E(x_1), \sqrt{V(x_1)}) = (0, .57), (E(x_2), \sqrt{V(x_2)}) = (0, .31) \right\}$
4	$(0, -1, 1)$	

1) 切断領域の近似精度を改善する簡易的な方法として, 以下の方法が考えられる. BS-3 の Step 3 において棄却数の上限を設定し, 上限に達した場合, その Step 1 に戻り BS-A の Step 2-(a) の k を変更する.

の分析に用いるため、解析的に得られる x_1 及び x_2 の周辺分布の平均及び標準偏差もそれぞれ記述してある。また、制約式の構造を表す $\mathbf{D}_x$ が正則であるため、前節で紹介した3種類のブロックサンプラーのうち $\mathbf{D}$ が正則行列の場合のアルゴリズム（アルゴリズム BS-2）を用いればよいことが分かる。この例は Geweke (1991) 及び Rodriguez-Yam et al. (2004) でも分析されている。

これら3種類の乱数発生法を用いて、式 (3.1) に従う乱数を 2×10^5 個の稼動検査期間の後 10^4 個発生させ、得られた全ての MCMC 標本を用いて比較を行った。結果は表 2 にまとめられている。表 2 の各列は、それぞれ事後標本平均、事後標本標準誤差、95%信用区間、非効率性因子及び収束判定のための仮説検定の p -値を表している。

非効率性因子 (inefficiency factor) とは得られた MCMC 標本の効率性を評価する指標であって $1 + 2\sum_{j=1}^{\infty} \rho(j)$ として計算される。ここで $\rho(j)$ はラグ j の標本自己相関である。非効率性因子は得られた MCMC 標本の分散と仮想的な独立標本の分散の比として解釈できるため、独立標本が得られたと仮定した場合と比較して MCMC 標本の効率性を評価することができる（より詳しい議論は Chib, 2001 を参照）。一方 CD (convergence diagnostics) とは Geweke (1992)

表 2 乱数発生法の比較

乱数発生法*	変数	平均	標準偏差	95%信用区間	非効率性因子	CD*
ケース 1. $(\sigma_{12}, a, b) = (-.7, -10, 10)$						
G	x_1	.30	3.02	[-5.63 6.18]	96	.827
	x_2	-.019	.31	[-.62 .58]	43	.790
RY	x_1	.003	3.12	[-6.05 6.10]	1	.264
	x_2	-.002	.32	[-.62 .61]	1	.435
BS	x_1	.016	3.12	[-6.18 6.04]	1	.840
	x_2	-.001	.32	[-.60 .62]	1	.498
ケース 2. $(\sigma_{12}, a, b) = (0, -10, 10)$						
G	x_1	.22	2.99	[-5.63 6.09]	54	.889
	x_2	.003	.32	[-.62 .62]	1	.660
RY	x_1	.028	3.13	[-6.14 6.08]	1	.317
	x_2	-.002	.32	[-.62 .62]	1	.420
BS	x_1	.018	3.13	[-6.23 6.11]	1	.694
	x_2	.000	.31	[-.62 .63]	1	.031
ケース 3. $(\sigma_{12}, a, b) = (-.7, -1, 1)$						
G	x_1	.009	.48	[-.83 .83]	3	.973
	x_2	.002	.20	[-.40 .40]	1	.966
RY	x_1	.004	.49	[-.85 .85]	1	.177
	x_2	.001	.20	[-.39 .39]	14	.738
BS	x_1	.009	.48	[-.84 .84]	1	.850
	x_2	.001	.20	[-.40 .40]	1	.099
ケース 4. $(\sigma_{12}, a, b) = (0, -1, 1)$						
G	x_1	.007	.46	[-.81 .82]	2	.915
	x_2	.003	.26	[-.52 .52]	1	.979
RY	x_1	.004	.47	[-.83 .82]	1	.313
	x_2	-.001	.26	[-.52 .53]	1	.369
BS	x_1	.004	.47	[-.83 .82]	1	.499
	x_2	.000	.26	[-.52 .51]	1	.110

* 乱数発生法の表記は、G, RY, BS がそれぞれ Geweke (1991) の方法、Rodriguez-Yam et al. (2004) の方法、ブロックサンプラーを表す。

** CD は収束判定のための仮説検定の p -値を意味する。

により提案された収束判定のための仮説検定の p -値であり、最初の n_1 個と最後の n_2 個の標本を取り出し、両者の平均が一致しているかを検定するものである。今回は Geweke (1992) で推奨されている、標本の最初の 10%と最後の 50%をそれぞれ n_1 , n_2 として計算を行った。

結果は以下の通り。ケース 1 と 2 では、非効率性因子及び Geweke (1991) の方法による x_1 の事後標本平均を除き、全ての乱数発生法で同様の結果が得られた。一方ケース 3 と 4 では、非効率性因子を除き、全ての乱数発生法で同様な結果が得られた。特に Rodriguez-Yam et al. (2004) の方法及びブロックサンプラーの事後標本平均及び事後標本標準偏差は、全てのケースにおいて表 1 に見られる真の値にほぼ一致している。これらのことから、Rodriguez-Yam et al. (2004) の方法及びブロックサンプラーはうまく機能していると結論できる。

一方、非効率性因子の観点からはブロックサンプラーが最も優れた性能を示している。ブロックサンプラーの結果をより詳しく分析するために、図 1 に、ケース 3 における等高線図、標本自己相関関数、サンプルパス及び事後周辺分布を描いている。この図よりマルコフ連鎖の自己相関の減衰は十分速く、状態空間をくまなく動いていることが見て取れる。その他のケースについてもブロックサンプラーは同様の結果を得ることができた。

次に従来の方法の結果について考察する。まず Geweke (1991) の方法は、先述の通りケース 3 及び 4 についてはうまく機能した。これはケース 3 及び 4 の非効率性因子がケース 1 及び 2 のそれに比べて大きく減少していることから確認できる。このことから、Geweke (1991) の方法は切断領域が狭い場合に比較的有效であるといえる。

一方 Rodriguez-Yam et al. (2004) の方法であるが、ケース 3 を除いて良い性能を示している。ケース 3 の場合を詳しく分析した結果、ギブスサンプラーを用いたことによる影響であることが分かった。図 2 には x へ変数変換を行う前の z の等高線図をケース 3 とケース 4 の場合について描いている。図 2 (a) から分かる通り、ケース 3 の等高線図は北東方向に傾いて切断されている。このような場合、条件付き分布で乱数発生を繰り返すギブスサンプラーではマルコフ連鎖が高い自己相関を持ちやすく、その結果本実験のケース 3 ではやや高い非効率性因子が現れたと考えられる。

4. 応用例：離散連続選択アプローチに基づくブロック料金制下の需要関数

本節以下、ブロックサンプラーを現実の推定問題に応用する。本論文で扱う応用例は離散連続選択アプローチに基づくブロック料金制下の需要関数のベイズ推定である。

ブロック料金制 (block rate pricing) とは、消費量がいくつかのブロックに分けられており、それぞれのブロックに対して異なる単価がつけられている料金体系を指す。同一の財に対して複数の価格をつけるような料金体系のことを非線形料金体系と呼び、ブロック料金制は非線形料金体系の一つである。ブロック料金制のような非線形料金体系は主に規制産業で観察される。例えば、電力、ガスや水道といった財や累進的な構造を持つ税 (所得税など) で採用されている。

単一の価格ではなく複数の価格を持つ財の需要関数を推定する場合において、単純に需要量を消費を行っている価格と所得で回帰した需要関数では、価格の選択行動が無視されているため、推定されるパラメータにバイアスが生じる。ブロック料金制の場合にはブロックと価格を同一視できるため、ブロックの選択を明示的に考慮して推定する必要がある。この問題は Burtless and Hausman (1978) により効用最大化問題として初めて定式化され推定が行われた。彼らのアプローチは離散連続選択アプローチ (discrete/continuous choice approach) と呼ばれ、非線形料金体系下の需要関数の分析における標準的アプローチとなっている。より一般的な分析は、例えば Hanemann (1984), Hausman (1985) や Moffitt (1986) にまとめられている。

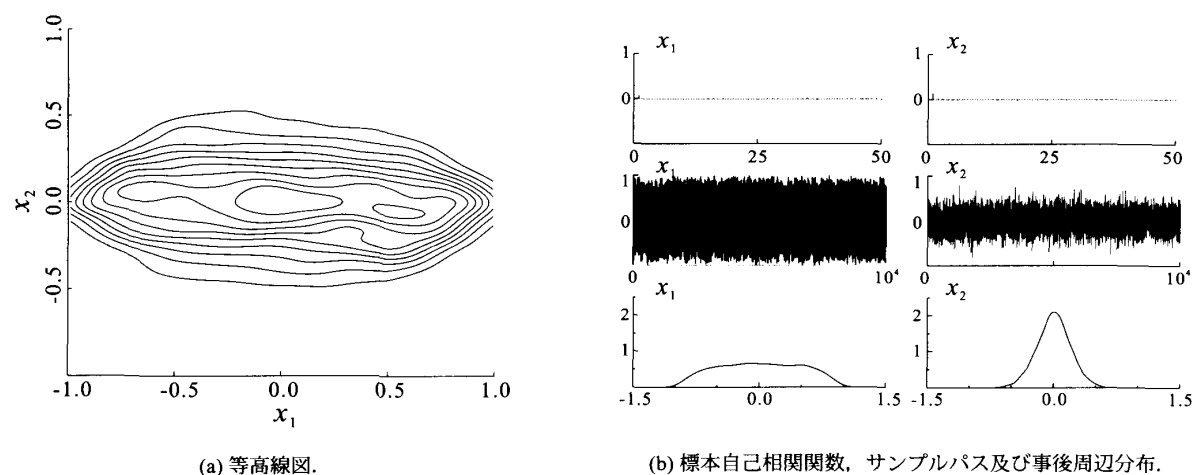
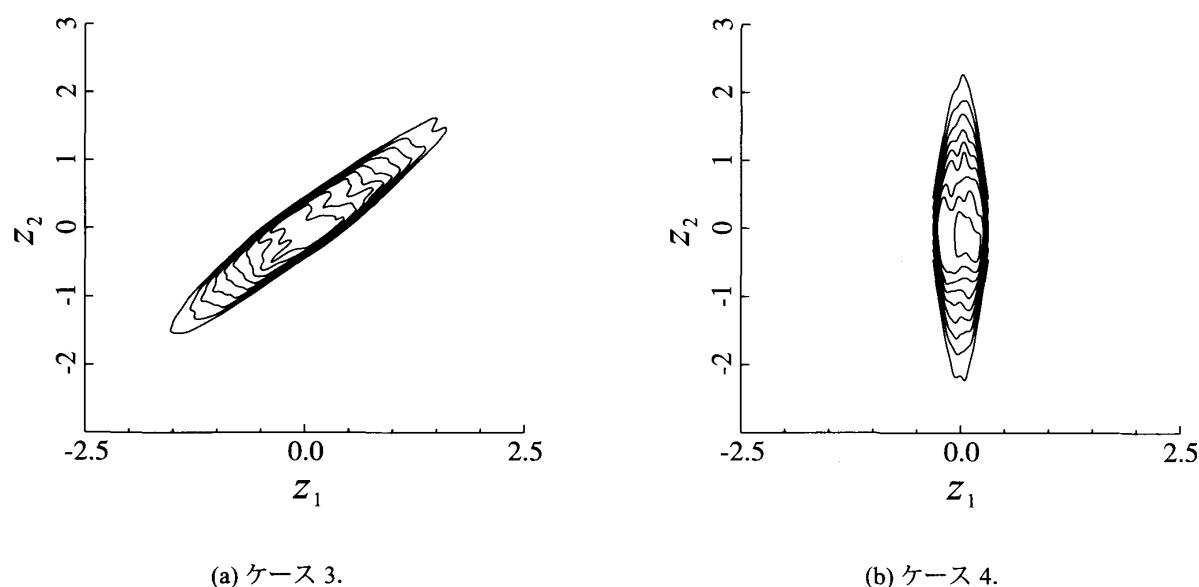


図1 ブロックサンプラーの結果, ケース 3.

図2 z の等高線図, Rodriguez-Yam et al. (2004) の方法.

近年, 離散連続選択アプローチを用いた実証分析はいくつかなされており, 例えば自動車保有 (de Jong, 1990), 労働供給 (Averett and Hotchkiss, 1997), 水道 (Miyawaki, Omori and Hibiki 2006; Olmstead, Hanemann and Stavins, 2007) やガス (Miyawaki and Omori, 2007) などがある. 本論文で実証分析がなされる電力に関しては Reiss and White (2005) や Miyawaki et al. (2006) がある.

以下では, まず離散連続選択アプローチに基づくブロック料金制下の需要関数の導出を行う. 続いて対応する統計的モデルの説明及び事後分布の導出を行う.

4.1 需要関数の導出

まず消費者の効用最大化問題を定義する. 消費者は, 総所得 I の下で, K ブロックのブロック料金制で供給されている財の消費量 $Y > 0$ とそれ以外の財の消費量 $Y_a > 0$ の組を決定する問題に直面している. ブロック料金制は以下の 3 種類の変数から構成される.

1. 第 k ブロックの価格 $P_k (k=1, \dots, K)$.
2. 第 k ブロックの使用量の上限, 閾値 $\bar{Y}_k (k=1, \dots, K)$. ただし $\bar{Y} \equiv 0$, $\bar{Y}_k \equiv \infty$ と定義する.
3. 固定費用 FC .

ブロック料金制を考えているので、多くの場合価格 P_k は消費量に関して単調に変化し、(1) $P_k < P_{k+1}$ の場合を逓増型ブロック料金制 (increasing block rate pricing), (2) $P_k > P_{k+1}$ の場合を逓減型ブロック料金制 (decreasing block rate pricing), と呼ぶ. 日本では水道や電力が逓増型, ガスが逓減型の料金体系で供給されている. 一方, 合成財 Y_a の価格は 1 に規準化されているとする.

この時, 適切に定義された選好の下で, 消費者は効用 $U(Y, Y_a)$ を最大にするような財の消費量の組 (Y, Y_a) を選ぶ. より正確には, 以下の消費者の効用最大化問題を解く.

$$\begin{aligned} \max_{Y, Y_a} U(Y, Y_a) \quad \text{subject to } Y_a + c(Y) &\leq I, \\ c(Y) &= FC + P_k(Y - \bar{Y}_{k-1}) + \sum_{j=1}^{k-1} P_j(\bar{Y}_j - \bar{Y}_{j-1}), \quad \text{if } \bar{Y}_{k-1} < Y \leq \bar{Y}_k. \end{aligned} \quad (4.1)$$

図3は3ブロックの場合の逓増型と逓減型の予算線をそれぞれ描いている. 図より, いずれのブロック料金制においても予算線が区分的線形になっていることが分かる.

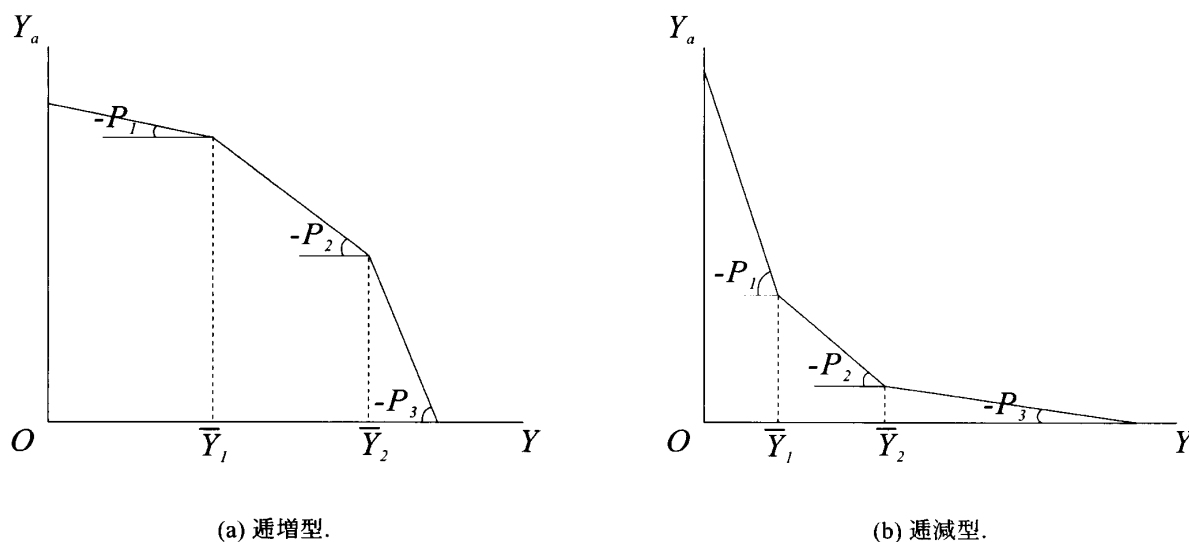


図3 3ブロック料金制下の予算制約集合.

このような非線形予算制約下の最大化問題を解くための一般的な方法として, 離散連続選択アプローチがある. 離散連続選択アプローチによる需要関数の導出の前に, 図を用いた直感的な説明を行う. 図によって効用最大化問題を解く場合, 無差別曲面 (2財モデルの場合, 無差別曲線) と呼ばれる「効用を一定水準に固定した時の財の組の軌跡」を用いる. 最も単純な2財 (Y_1, Y_2) モデルの場合, 効用最大化問題は,

$$\max_{Y_1, Y_2} U(Y_1, Y_2) \quad \text{subject to } P_1 Y_1 + P_2 Y_2 \leq I, \quad (4.2)$$

であり, 予算線と無差別曲線は図4(a)及び図4(b)のように書くことができる. ただし (P_1, P_2) はそれぞれ1財, 2財の価格を表す. 適当な選好を仮定すれば, 無差別曲線は北東方向に移動するにつれて効用の水準が上がる (図4(b)参照) ので, 予算制約を満たしつつ効用を最大にする財の組 (需要量) は図4(c)に見られるように, 無差別曲線が予算線と接する部分になることが分かる. 図4(c)の場合, 最適な需要量は点Eであり, その時の間接効用 (最大化された効用の水準) は V^* となる.

ブロック料金制の場合にも同様に無差別曲線を書くことができ, 図5のようになる. この図より, 逓増型の場合, 選好によって2種類の最適需要が存在することが分かる. つまり無差

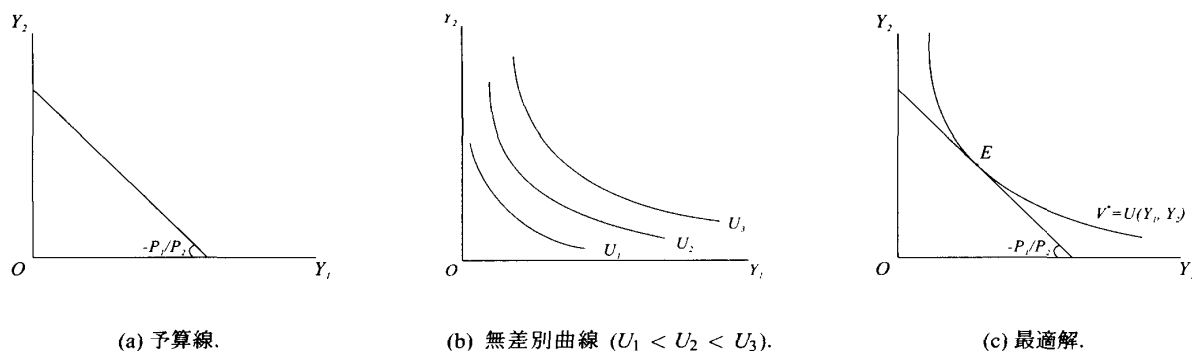


図4 予算線, 無差別曲線及び最適解.

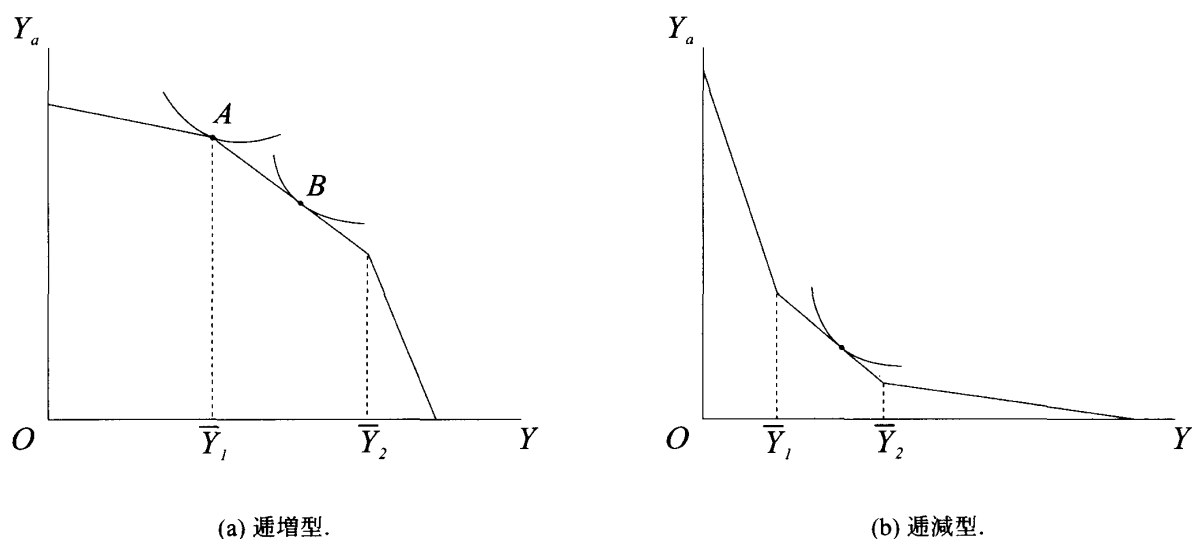


図5 ブロック料金制下の効用最大化問題.

別曲線が点Aのように閾値で接する場合と点Bのようにブロックの内側で予算線に接する場合がある。一方逓減型の場合、最適需要は予算線に接するものしか存在しない。

実際、離散連続選択アプローチにより最適需要を求めれば、図と同様な結論を得ることができる。離散連続選択アプローチの特徴は、区分的線形な予算制約下の最大化問題を線形制約下の最大化問題の集合に置き換えることにある。具体的には以下の2段階の最大化問題を解くことで解を得る。

離散連続選択アプローチ

Step 1. 各 $k=1, \dots, K$ に対して以下の最大化問題の解及び最大値を求める。

$$V_k^* = \sup_{Y, Y_a} U(Y, Y_a) \quad \text{subject to } (Y, Y_a) \in \{(Y, Y_a) \mid \bar{Y}_{k-1} < Y \leq \bar{Y}_k, Y_a + c(Y) \leq I\}. \quad (4.3)$$

Step 2. $V^* = \max\{V_1^*, \dots, V_K^*\}$ を求め、得られた V^* に対応する Step 1 の解を元の最大化問題 (4.1) の解とする。

第一段階が連続量の最大化問題であり第二段階が離散的な選択であるため離散連続選択アプローチと呼ばれる。

明示的な需要関数の導出には、離散連続選択アプローチの Step 1 で用いた効用最大化問題 (4.3) の予算制約をやや緩める一方で、元の効用最大化問題 (4.1) の予算制約集合の性質を

用いる。まず Step 1 の効用最大化問題 (4.3) の予算制約集合を含む新しい効用最大化問題を考える。新しい効用最大化問題は条件付き効用最大化問題 (conditional utility maximization problem) と呼ばれ、各ブロック ($k=1, \dots, K$) に対して定義される。第 k 番目のブロックに対する条件付き効用最大化問題は、

$$\max_{Y, Y_a} U(Y, Y_a) \quad \text{subject to } Y_a + P_k Y \leq Q_k, \quad (4.4)$$

と定式化される。ここで Q_k ($k=1, \dots, K$) は帰属所得 (virtual income) と呼ばれる変数であり、 $Q_k \equiv I - FC - \sum_{j=1}^{k-1} (P_j - P_{j+1}) \bar{Y}_j$ である。以下ではこの条件付き最大化問題の解及び最大値をそれぞれ条件付き需要 (conditional demand) 及び条件付き間接効用 (conditional indirect utility) と呼び、 Y_k 及び V_k で表す。

条件付き効用最大化問題を定義すれば、元の効用最大化問題 (4.1) の予算制約集合の性質を用いることで需要関数を明示的に導出することができる。逓増型と逓減型で異なる性質を用いるため、逓増型の場合の需要関数のみを導出する。逓減型の場合は補論 A.2 を参照されたい。逓増型ブロック料金制下の需要関数は、元の効用最大化問題 (4.1) の予算制約集合の凸性を用いて、

$$Y = \begin{cases} Y_k, & \text{if } \bar{Y}_{k-1} < Y_k < \bar{Y}_k \text{ and } k=1, \dots, K, \\ \bar{Y}_k, & \text{if } \bar{Y}_{k+1} < \bar{Y}_k < Y_k \text{ and } k=1, \dots, K-1, \end{cases} \quad (4.5)$$

となる。図 5 (a) で見た通り、逓増型ブロック料金制下の需要関数は、選好によって閾値 $\bar{Y}_k$ と条件付き需要 Y_k のいずれかを選ぶ形になっている。

4.2 クロスセクションモデル

本節以下、添字 i ($i=1, \dots, n$) により第 i 番目の消費者を表す。統計的モデルは消費者ごとに料金体系が異なる状況を想定して構成されているため、消費者特有の変数 (需要量や所得等) に加えて料金体系の変数にも消費者を表す添字をつける。また、逓増型のブロック料金制下の需要関数のモデルに焦点を絞る。逓減型の場合は補論 A.2 を参照されたい。

まず条件付き需要関数が対数線形モデルであると仮定する。つまり第 i 番目の消費者の条件付き需要関数は、

$$\ln Y_{ik} = \beta_1 \ln P_{ik} + \beta_2 \ln Q_{ik} \iff y_{ik} = \beta_1 p_{ik} + \beta_2 q_{ik} \equiv \mathbf{x}'_{ik} \boldsymbol{\beta}, \quad (4.6)$$

と仮定する。上式の右側は、モデルの簡略化のため対応する変数の対数値を小文字で表したものである。小文字の変数は以下の記述にも用いるため全て定義しておく。 ($y_i, y_{ik}, \bar{y}_{ik}, p_{ik}, q_{ik}$) $\equiv$ ($\ln Y_i, \ln Y_{ik}, \ln \bar{Y}_{ik}, \ln P_{ik}, \ln Q_{ik}$)。また $\mathbf{x}_{ik} = (p_{ik}, q_{ik})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ である。対数線形モデルは需要関数の分析にしばしば用いられるモデルであり、価格及び帰属所得の係数はそれぞれ需要の価格弾力性及び帰属所得弾力性を表す。

続いてデータ発生過程を特定化する。離散連続選択アプローチに基づく需要関数のモデルには、次の 2 つの確率項を仮定することが通常である (例えば Moffitt, 1986 参照)。一つは消費者の選好の異質性 (heterogeneity of preferences) と呼ばれる潜在変数 w_i^* 、もう一つは測定誤差 u_i である。

消費者の選好の異質性は条件付き需要に対して加法的に加わる確率項であり、それは線形構造の項と正規分布に従う誤差項とに分けられる。より正確には、

$$w_i^* = \mathbf{z}_i' \boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N(0, \sigma_w^2), \quad (4.7)$$

と仮定される。ここで $\mathbf{z}_i$ 及び $\boldsymbol{\delta}$ は d 次元ベクトルであり、それぞれ選好の異質性の説明変数

及びその係数ベクトルを表す。選好の異質性は、条件付き需要を価格と帰属所得に加えて、消費者に特有な変数によって説明するための潜在変数である。その結果、逓増型のブロック料金制下の需要関数は、

$$y_i^* = \begin{cases} y_{ik} + w_i^*, & \text{if } \bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{ik} \text{ and } k=1, \dots, K_i, \\ \bar{y}_{ik}, & \text{if } y_{i,k+1} + w_i^* \leq \bar{y}_{ik} \leq y_{ik} + w_i^* \text{ and } k=1, \dots, K_i-1, \end{cases} \quad (4.8)$$

となる。以下 y_i^* を潜在需要 (unobserved demand) と呼ぶことにする。

選好の異質性は最適な潜在需要の状態を決める潜在変数であるとも見ることもできる。潜在需要の状態とは、逓増型のブロック料金制下の需要関数の場合、どのブロックの条件付き需要もしくは閾値を選んでいるかの2種類を指す。より正確には、逓増型の場合、 $s_i^* = 1, \dots, 2K_i - 1$ をとる離散変数であって、

$$s_i^* = \begin{cases} 2k-1, & \text{if } y_i^* < y_{ik} + w_i^* \text{ and } k=1, \dots, K_i, \\ 2k, & \text{if } y_i^* = \bar{y}_{ik} \text{ and } k=1, \dots, K_i-1, \end{cases} \quad (4.9)$$

である。もし状態 s_i^* が奇数であれば、第 i 番目の消費者にとっては条件付き需要に選好の異質性を加えた需要量が最適であり、偶数であれば、閾値を潜在需要として選択している。

式 (4.9) を変形すれば、状態に応じて選好の異質性の取りうる値の範囲 $(-\infty, \infty)$ を排反な区間に分割することができる。つまり、

$$\begin{aligned} s_i^* = 2k-1 &\iff \bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{ik} \iff w_i^* \in R_{i,2k-1} = (\bar{y}_{i,k-1} - y_{ik}, \bar{y}_{ik} - y_{ik}), \\ s_i^* = 2k &\iff y_{i,k+1} + w_i^* \leq \bar{y}_{ik} \leq y_{ik} + w_i^* \iff w_i^* \in R_{i,2k} = (\bar{y}_{ik} - y_{ik}, \bar{y}_{ik} - y_{i,k+1}), \end{aligned} \quad (4.10)$$

と変形できる。その結果、最適な潜在需要の状態は潜在変数である選好の異質性によって決定されていると解釈できる。

ここで状態 s_i^* は潜在変数であり、観測できないことに注意する。つまり、潜在需要の状態は選好の異質性が決まればその値も決まるが、 w_i^* が潜在変数であるため、 s_i^* も潜在変数である。本モデルにおける状態 s_i^* のように、潜在変数でモデルを拡大して MCMC 推定を行う手法はデータ拡大法と呼ばれている。潜在変数によってモデルを拡大する理由は潜在変数以外のパラメータの条件付き事後分布からの乱数発生を容易にするためである。Albert and Chib (1993) を始めとし、離散選択が含まれているモデルはデータ拡大法で推定されることが多い。

上の条件 (4.10) から、 β のパラメータ領域に関する新しい条件が導かれる。つまり選好の異質性の区間 R_{ik} が排反に分けられるために、以下の条件が必要となる。

$$\beta \in \bigcap_{i=1}^n SR_i, \quad SR_i = \bigcap_{j=1}^{K_i-1} \{\beta | \mathbf{x}'_{i,j+1} \beta \leq \mathbf{x}'_{ij} \beta\}. \quad (4.11)$$

この条件は分離条件と呼ばれ、Miyawaki et al. (2006) で初めて導入された。分離条件は統計的モデルが多項トービットモデルであり、その閾値にパラメータが含まれている限り常に必要な条件である。しかしその数は選択枝数 (ブロック料金制のモデルの場合ブロック数) と観測値数に比例して増える。そのため MCMC 法以外の推定方法を用いて全ての制約を評価して推定することは困難であり、分離条件の存在は MCMC 法を用いて推定する理由の一つとなっている。

もう一つの確率項、測定誤差は潜在需要 y_i^* に加わるものであり、

$$y_i = y_i^* + u_i, \quad u_i \sim \text{i.i.d. } N(0, \sigma_w^2), \quad (4.12)$$

とする.

以上の準備の下, 逓増型のブロック料金制下の需要関数のクロスセクションモデルは以下の式で表すことができる.

$$\begin{aligned} y_{ik} &= \mathbf{x}'_{i,k} \boldsymbol{\beta}, \quad k=1, \dots, K_i, \\ w_i^* &= \mathbf{z}'_i \boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N(0, \sigma_v^2), \\ s_i^* &= \begin{cases} 2k-1, & \text{if } w_i^* \in R_{i,2k-1} \text{ and } k=1, \dots, K_i, \\ 2k, & \text{if } w_i^* \in R_{i,2k} \text{ and } k=1, \dots, K_i-1, \end{cases} \\ y_i^* &= \begin{cases} y_{ik} + w_i^*, & \text{if } s_i^* = 2k-1 \text{ and } k=1, \dots, K_i, \\ \bar{y}_{ik} & \text{if } s_i^* = 2k \text{ and } k=1, \dots, K_i-1, \end{cases} \\ y_i &= y_i^* + u_i, \quad u_i \sim \text{i.i.d. } N(0, \sigma_u^2). \end{aligned} \quad (4.13)$$

この統計的モデルは第V型の多項トービットモデルに分類される.

4.3 事前分布, 事後分布及びギブスサンプラー

前節で得られたモデルを推定するために, 本論文ではMCMC法を用いた推定手法を提案する. まず, パラメータの事前分布として正規-ガンマ分布を仮定する. つまり, $\boldsymbol{\beta}$ 及び $\boldsymbol{\delta}$ はそれぞれ σ_u^2 及び σ_v^2 で条件付けられた多変量正規分布, σ_u^2 及び σ_v^2 は逆ガンマ分布であるとする. より正確には,

$$\begin{aligned} \boldsymbol{\beta} | \sigma_u^2 &\sim N_2(\boldsymbol{\mu}_{\beta,0}, \sigma_u^2 \boldsymbol{\Sigma}_{\beta,0}), \quad \boldsymbol{\delta} | \sigma_v^2 \sim N_d(\boldsymbol{\mu}_{\delta,0}, \sigma_v^2 \boldsymbol{\Sigma}_{\delta,0}), \\ \sigma_u^2 &\sim IG\left(\frac{n_{u,0}}{2}, \frac{S_{u,0}}{2}\right), \quad \sigma_v^2 \sim IG\left(\frac{n_{v,0}}{2}, \frac{S_{v,0}}{2}\right), \end{aligned} \quad (4.14)$$

と仮定する. ここで, $\boldsymbol{\mu}_{\beta,0}$ は2次元の既知ベクトル, $\boldsymbol{\Sigma}_{\beta,0} = \text{diag}(\sigma_{\beta_1,0}^2, \sigma_{\beta_2,0}^2)$ は 2×2 既知正定値行列, $\boldsymbol{\mu}_{\delta,0}$ は d 次元の既知ベクトル, $\boldsymbol{\Sigma}_{\delta,0}$ は $d \times d$ 既知正定値行列, $n_{u,0}$, $S_{u,0}$, $n_{v,0}$ 及び $S_{v,0}$ は既知の正定数を表す. これらの事前分布をまとめて $\pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2)$ と表記する.

続いて尤度関数を導出する. 第 i 番目の消費者の需要 y_i の尤度を潜在変数 s_i^* と w_i^* で補った尤度関数 $f(y_i, s_i^*, w_i^* | \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2)$ は, 測定誤差をモデル化した条件付き密度 $f(y_i | s_i^*, w_i^*, \boldsymbol{\beta}, \sigma_u^2)$ と潜在需要の状態と選好の異質性の条件付き密度 $f(s_i^*, w_i^* | \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v^2)$ の積で表わされ,

$$\begin{aligned} f(y_i, s_i^*, w_i^* | \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2) &= f(y_i | s_i^*, w_i^*, \boldsymbol{\beta}, \sigma_u^2) f(s_i^*, w_i^* | \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_v^2) \\ &\propto (\sigma_u \sigma_v)^{-1} \exp \left[-\frac{1}{2} \{ \sigma_u^{-2} (y_i - y_i^*)^2 + \sigma_v^{-2} (w_i^* - \mathbf{z}'_i \boldsymbol{\delta})^2 \} \right] I(w_i^* \in R_{i,s_i^*}) I(\boldsymbol{\beta} \in SR_i), \end{aligned} \quad (4.15)$$

と書ける. ここで $I(A)$ は定義関数であり, A が真の場合1, そうでない場合は0を取る. 各消費者の尤度関数を導出できれば, 事後分布を導出することができる. 事後分布は,

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2, \mathbf{s}^*, \mathbf{w}^* | \mathbf{y}) &\propto \pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2) \\ &\times (\sigma_u \sigma_v)^{-n} \exp \left[-\frac{1}{2} \{ \sigma_u^{-2} (\mathbf{y} - \mathbf{y}^*)' (\mathbf{y} - \mathbf{y}^*) + \sigma_v^{-2} (\mathbf{w}^* - \mathbf{Z} \boldsymbol{\delta})' (\mathbf{w}^* - \mathbf{Z} \boldsymbol{\delta}) \} \right] \\ &\times \prod_{i=1}^n \{ I(w_i^* \in R_{i,s_i^*}) I(\boldsymbol{\beta} \in SR_i) \}, \end{aligned} \quad (4.16)$$

と書ける. ここで, $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_n^*)'$, $\mathbf{s}^* = (s_1^*, s_2^*, \dots, s_n^*)'$, $\mathbf{w}^* = (w_1^*, w_2^*, \dots, w_n^*)'$, $\mathbf{Z} = (z_1^*, z_2^*, \dots, z_n^*)'$ とした.

事後分布 (4.16) は非常に多くの線形制約式で切断された複雑な分布であるため解析的に分

析を行うことは困難である。そのため、本論文ではギブスサンプラーを用いてこの事後分布からの MCMC 標本を得、得られた標本を用いて分析を行う。ギブスサンプラーのアルゴリズムは以下の 5 ステップで記述される。

アルゴリズム MCMC：クロスセクションモデルの MCMC アルゴリズム

Step 1. パラメータ及び潜在変数 $(\beta, \delta, s^*, w^*, \sigma_u^2, \sigma_v^2)$ に初期値を与える。

Step 2. (σ_u^2, β) を s^*, w^* を所与として発生させる。

(a). σ_u^2 を s^*, w^* を所与として逆ガンマ分布より発生させる。

(b). β を s^*, w^*, σ_u^2 を所与として切断多変量正規分布より発生させる。

Step 3. (σ_v^2, δ) を w^* を所与として発生させる。

(a). σ_v^2 を w^* を所与として逆ガンマ分布より発生させる。

(b). δ を w^*, σ_v^2 を所与として多変量正規分布より発生させる。

Step 4. (s_i^*, w_i^*) を $\beta, \delta, \sigma_u^2; \sigma_v^2 (i=1, \dots, n)$ を所与として発生させる。

(a). s_i^* を $\beta, \delta, \sigma_u^2; \sigma_v^2$ を所与として多項分布より発生させる。

(b). w_i^* を $\beta, \delta, s_i^*, \sigma_u^2; \sigma_v^2$ を所与として切断正規分布より発生させる。

Step 5. Step 2 に戻る。

Step 2-(b) を除いた、それぞれのステップで用いられる条件付き事後分布は通常の分布（正規分布，逆ガンマ分布，多項分布）となる。よって， β の標本を除き，パラメータ及び潜在変数の標本を得るのは容易である。具体的な条件付き事後分布は補論 A.3 にまとめている。

Miyawaki et al. (2006) では，Step 2-(b) で必要となる線形制約下の多変量正規分布から直接乱数を発生することを避け， β_1 と β_2 の条件付き事後分布をそれぞれ解析的に導出し，それらを用いたアルゴリズムを提案している (Miyawaki et al, 2006, Algorithm 1.1)。しかし β_1 と β_2 の MCMC 標本の自己相関が非常に高く，それらの事後分布への収束が遅いことが分かったため，彼らの論文では一般化ギブスステップ (generalized Gibbs step; GGS) を用いてアルゴリズムの改善を行った (Miyawaki et al, 2006, Algorithm 1.2)。しかし，ブロックサンプラーを用いれば，Step 2-(b) で必要となる切断多変量正規分布からの標本を直接得ることができる。

実際ブロックサンプラーが適用できることを確認するために， β の条件付き事後分布が式 (2.1) の形になっていることを確認しておく。回帰係数 β の条件付き事後分布は，線形制約 $\beta \in R_\beta$ で切断された多変量正規分布 $N_2(\mu_{\beta,1}, \sigma_u^2 \Sigma_{\beta,1})$ である。ただし $R_\beta, \mu_{\beta,1}$ 及び $\Sigma_{\beta,1}$ は補論 A.3 の式 (A.18) で定義される。通増型のブロック料金制の場合に限って第 i 消費者に関する切断領域を整理すれば，

$$[a_i, D_i, b_i] = \begin{cases} \left[\begin{pmatrix} \bar{y}_{i,k-1} - w_i^* \\ -\infty \\ \vdots \\ -\infty \end{pmatrix}, \begin{bmatrix} x'_{ik} \\ (x_{i2} - x_{i1})' \\ \vdots \\ (x_{iK_i} - x_{iK_i-1})' \end{bmatrix}, \begin{pmatrix} \bar{y}_{i,k} - w_i^* \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right], & \text{if } s_i^* = 2k-1, \\ \left[\begin{pmatrix} \bar{y}_{i,k} - w_i^* \\ -\infty \\ -\infty \\ \vdots \\ -\infty \end{pmatrix}, \begin{bmatrix} x'_{ik} \\ x'_{ik+1} \\ (x_{i2} - x_{i1})' \\ \vdots \\ (x_{iK_i} - x_{iK_i-1})' \end{bmatrix}, \begin{pmatrix} \infty \\ \bar{y}_{i,k} - w_i^* \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right], & \text{if } s_i^* = 2k, \end{cases} \quad (4.17)$$

となるので，

$$\beta \in R_\beta \Leftrightarrow \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \leq \begin{bmatrix} D_1 \\ \vdots \\ D_n \end{bmatrix} \beta \leq \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad (4.18)$$

と書ける．よって式 (2.1) の形に書けることが確かめられた．逓減型の場合も同様に式 (2.1) の形に書くことができる．

最後に，上のアルゴリズムの Step 4 では潜在変数 (s_i^*, w_i^*) を同時分布から発生させている．前節でも述べたように，潜在変数は他のパラメータの標本を容易に得るためにモデルを拡大する変数であり，上のギブスサンプラー中でその標本を発生する必要がある．しかし，状態 s_i^* は選好の異質性 w_i^* が与えられればその値が決まってしまうため，「 s_i^* を所与とした w_i^* の発生」と「 w_i^* を所与とした s_i^* の発生」を繰り返し行うギブスサンプラーを用いることはできない．そのため同時分布からの標本発生が必要となる．つまり，いずれかの変数を積分消去した周辺分布から標本を発生させた後，それを所与として残りの変数を発生させる必要がある．本アルゴリズムでは， w_i^* を積分消去した s_i^* の周辺分布の方が s_i^* を積分消去した w_i^* のそれより容易に標本を発生できることから Step 4 の手順で標本を得ることにした．

5. 数値実験 II

本節では，3 節に引き続き，ブロックサンプラーの有効性を数値実験によって確認する．次節で日本の家庭用電力需要関数の推定を行うため，電力の料金体系である逓増型ブロック料金制の場合のモデル (式 4.13) の下で，実験データを発生させる．そして，アルゴリズム MCMC にブロックサンプラー (アルゴリズム BS-3) を用いたもので推定を行い，ブロックサンプラーの評価を行う．同様な数値実験は Miyawaki et al. (2006) でもなされており，比較のため彼らの論文と同設定，同事前分布で実験を行う．

実験の設定は表 3 にある値もしくは分布からの乱数とした．また事前分布は，

$$\begin{aligned} \beta | \sigma_u^2 &\sim N_2(\mathbf{0}, 100\sigma_u^2 \mathbf{I}), \quad \delta | \sigma_u^2 \sim N_2(\mathbf{0}, 100\sigma_v^2 \mathbf{I}), \\ \sigma_u^2 &\sim IG(0.01, 0.01), \quad \sigma_v^2 \sim IG(0.01, 0.01), \end{aligned} \quad (5.1)$$

とした．

表 3 数値実験 II の設定

観測値数	100
パラメータの真の値	$(\beta_1, \beta_2, \delta_0, \delta_1, \sigma_u, \sigma_v) = (-0.6, 0.3, 0.1, 0.1, 0.3, 0.1)$
所得	$I_i \sim N(3, 0.3^2) ^*$
価格	$P_{i1} \sim N(2, 0.4^2) ^*, \Delta_i \sim N(0.7, 0.2^2) ^*, P_{i2} = P_{i1} + \Delta_i$
閾値	$\bar{Y}_{i1} = 2$
固定費用	$FC = 0$
選好の異質性の変数	$z_{i1} = 1, z_{i2} \sim N(2.5, 1)$

* $|N(\mu, \sigma^2)|$ は平均 μ ，分散 σ^2 の正規分布の絶対値を取ったもの．

以上の設定の下，アルゴリズム MCMC にブロックサンプラーを用いたものに従って MCMC 標本を発生させた．最初の 4×10^4 個を稼動検査期間として棄てた後， 10^5 個の標本を発生させ，その全ての標本を用いた推定結果を図 6 及び表 4 にまとめた．図 6 は各パラメータの周辺事後分布を描いたもの，表 4 は事後分布を用いた推論を行うための統計量をまとめたものである．Miyawaki et al. (2006) との比較のために，非効率性因子の列には彼らの論文の Table 1 の下段に報告されているものを角括弧内に併記してある．やや σ_u の CD が低いものの，1%の有意

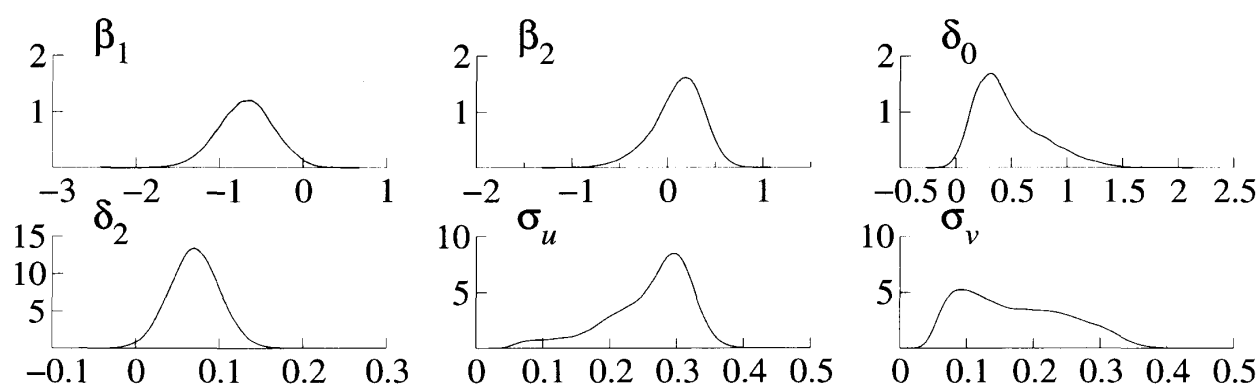


図6 周辺事後分布（数値実験Ⅱ）.

表4 数値実験Ⅱの推定結果

パラメータ	真の値	平均	標準偏差	95%信用区間	非効率性因子*	CD**
β_1	-.6	-.68	.33	[-1.34 - .057]	118 [345]	.737
β_2	.3	.11	.27	[- .50 .56]	433 [731]	.759
δ_0 (定数項)	.1	.47	.30	[-.050 1.19]	1090 [1034]	.926
δ_1	.1	.071	.030	[.012 .13]	11 [25]	.881
σ_u (測定誤差)	.3	.26	.064	[.092 .35]	167 [152]	.038
σ_v (選好の異質性)	.1	.17	.079	[.057 .33]	159 [144]	.038

* 角括弧内の値は、Miyawaki et al. (2006) の Table 1 の下段で報告されている GGS の場合の非効率性因子。

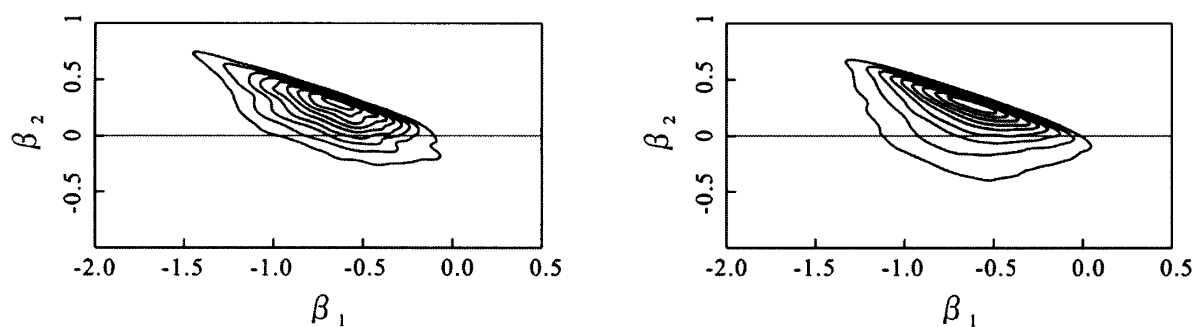
** CD は収束判定のための仮説検定の p -値を意味する。

水準で見ればいずれのパラメータの標本もその周辺事後分布に収束していると見ることができる。実際、いずれのパラメータの標本自己相関関数も十分早く減衰しており、そのサンプルパスも状態空間をくまなく移動していたため、得られた標本は目的の事後分布からの標本とみなしてよい。

続いて、その他の得られた結果に関して考察を行う。まず、各パラメータの周辺事後分布がそれぞれ真の値をモード付近に含んでいることが分かる。また、ブロック料金制下の需要関数のモデルには2種類の分散がパラメータとして存在するが、それぞれの周辺事後分布は、真の値を含んで分かれて推定されている。以上より、アルゴリズム MCMC にブロックサンプラーを用いたものはうまく機能しているといえる。Miyawaki et al. (2006) の Table 1 の下段の結果と比較しても事後標本平均及び事後標本標準誤差はほぼ同じであった。ただし β_1 と β_2 の 95%信用区間がやや広がっていた。これはこの後に示す (β_1, β_2) の等高線プロットからも見て取れる（図7参照）。

一方非効率性因子は彼らの結果と比べて改善を示しており、特に β_1 の非効率性因子が大きく減少していることが分かった。また β_2 についても、 β_1 ほどではないが、非効率性因子の減少が確認された。これらは同時分布から β の乱数を発生していることによる。一方、その他のパラメータ $(\delta, \sigma_u, \sigma_v)$ の非効率性因子は、GGS を用いた場合とほぼ同程度の大きさとなった。選好の異質性の説明変数のうち定数項の係数 δ_0 の非効率性因子は最大であり、更なる改善が期待される。

次に、ブロックサンプラーの受容率について述べる。本実験において、その棄却サンプリングのステップ（アルゴリズム BS-3 の Step 3）の受容率は約 27%であった。この比率は切断領域の近似の程度に強く影響を受けていると考えられる。実際アルゴリズム BS-A において、選



(a) Miyawaki et al. (2006) の GGS を用いた方法.

(b) ブロックサンプラー.

図7 (β_1, β_2) の等高線図 (数値実験 II).

ばれる制約式の組を変えればこの受容率は変動する. より良く切断領域を近似する制約式の組を探索することは, より効率的に乱数発生を行うために重要である. しかし, 本実験では切断に用いられる線形制約式の数, 選好の異質性に起因するものが潜在需要の状態 s_i^* に応じて 100~200 本あり, 分離条件に起因するものが 100 本あるため, 良い近似領域を構成する線形制約式を探索することは計算負荷の観点から難しい. 計算負荷が低く良い近似領域の構成は今後の課題であろう.

最後に (β_1, β_2) の MCMC 標本を用いて等高線図を描いた. 図7は Miyawaki et al. (2006) の方法と本論文の方法で得られた等高線図である. いずれの方法も似通った等高線図となっており, 特に北東方向が急斜面になっていることが共通している. ただし本論文の方法で発生させた標本の等高線プロットがやや南西方向に膨らんでおり, その結果 β_1 と β_2 の 95% 信用区間がやや広がったことが分かる. 3 節でも触れたが, 上図のように長細く北東もしくは北西のいずれかの方向に傾いているような領域からの標本を得る場合, 条件付き分布からの標本発生を繰り返すギブスサンプラーでは得られた標本の自己相関が高くなってしまふ. それに対して, ブロックサンプラーのように同時分布からの標本発生を行うことができれば, このような状況であってもギブスサンプラーより短いマルコフ連鎖で十分であり, 同時分布から発生するパラメータに関しては一般化ギブスステップより効率的な推定を行うことができる. 本実験におけるブロックサンプラーの効率性の高さの理由は, この等高線図からも見て取ることができる.

6. 日本の家庭用電力需要関数の推定

6.1 日本の電気事業及びデータの説明

まず日本の電気事業について簡単に説明する. 電気事業は発電部門と送電部門に分かれるが, 従来, 共に 10 社の一般電気事業者による地域独占供給が行われてきた. 独占供給を行う理由は規模の経済性である. しかし, 1995 年, 1999 年及び 2003 年の 3 度に渡る電気事業制度改革により地域独占が崩れ一部自由化が導入された. 具体的には高圧需要家 (電圧 6000 V 以上かつ契約電力 50 kW 以上の需要家) に対する電力供給が自由化された. 2006 年時点での事業者の概要は表 5 にまとめられている.

表 5 日本の電気事業者（2006 年）

事業者名	事業者数	発電設備と送電網の保有状況	対象需要家
一般電気事業者	10	発送電設備共に保有	全ての需要家
卸電気事業者等	50	発電設備のみ保有，一般電気事業者への卸売	} 高圧需要家
高圧需要家特定電気事業者	6	発送電設備共に保有	
特定規模電気事業者	23	発電設備のみ保有，一般電気事業者の送電網で託送	

一方，本論文で分析対象とする家庭用電力は用途別では電灯と呼ばれる需要であり，先の自由化の対象外となっている．そのため家庭用電力に関しては，従来同様地域独占であり規制料金での供給が行われている．その料金体系は通増型ブロック料金制が採用されている．通増型ブロック料金制の採用理由としては，エネルギー問題の観点からの理由があげられる．つまり，より多い需要量に対してより高い料金を設定することで電気を節約するインセンティブを持たせる目的があると考えられている．

続いて電力需要関数の推定を行う 1999 年のデータについて説明する．これらのデータは Miyawaki et al. (2006) で電力需要関数の推定に用いられたものと同じである．用いるデータは都道府県ごとに集計されたデータであり，被説明変数及び説明変数は以下の通り．被説明変数は総務省「家計調査年報」で報告されている一家計が一月に平均的に使用する電気料金を，各都道府県において家庭用電力の供給を行っている電力会社の料金表により量（ m^3 ）に変換したものをを用いる．料金表は電力会社ごとに異なっており，また同一電力会社内においても複数の料金表が提供されている．本論文の推定には，その中で最も標準的とされる従量電灯 B という料金区分のうち契約容量 30 アンペア（A）の料金表を用いる．先述の通り，この料金体系は通増型ブロック料金制であり，そのブロック数は各電力会社により異なり 3-4 ブロックであった．

説明変数は表 6 にまとめられているものをを用いる．価格としては従量電灯 B，契約容量 30A の料金表を使用し，帰属所得としては「家計調査年報」の消費支出と従量電灯 B，30A の料金表を用いて計算したものを使用する．選好の異質性の説明変数には「家計調査年報」より平均世帯人数及び気象庁発表の平均気温を用いる．

表 6 電力需要関数の推定に用いられる説明変数

変数名	対応する係数	変数の説明（単位）
価格	β_1	従量電灯 B，30A の価格（ 10^3 円/kWh）
帰属所得	β_2	消費支出と従量電灯 B，30A の料金表を用いて計算（ 10^3 円）
w_i^* の説明変数	δ_1	平均世帯人数（人）
	δ_2	平均気温（ $^{\circ}\text{C}$ ）

6.2 日本の家庭用電力需要関数の推定

数値実験 II 同様，式 (4.13) のモデルを用いて推定を行う．事前分布は，

$$\begin{aligned} \beta|\sigma_u^2 &\sim N_2(\mathbf{0}, 100\sigma_u^2 I), \quad \delta|\sigma_v^2 \sim N_2(\mathbf{0}, 100\sigma_v^2 I), \\ \sigma_u^2 &\sim IG(0.01, 0.01), \quad \sigma_v^2 \sim IG(0.01, 0.01), \end{aligned} \quad (6.1)$$

とした．事後分布からの MCMC 標本をアルゴリズム MCMC にブロックサンプラー（アルゴリズム BS-3）を用いたものから発生させ推定を行った．発生させた標本のうち最初の 4×10^5 個を稼動検査期間として棄てた後， 10^6 個の標本を得，得られた 10^6 標本を 100 個置きに拾い 10^4 標本に減らした結果が図 8 及び表 7 にまとめられている．

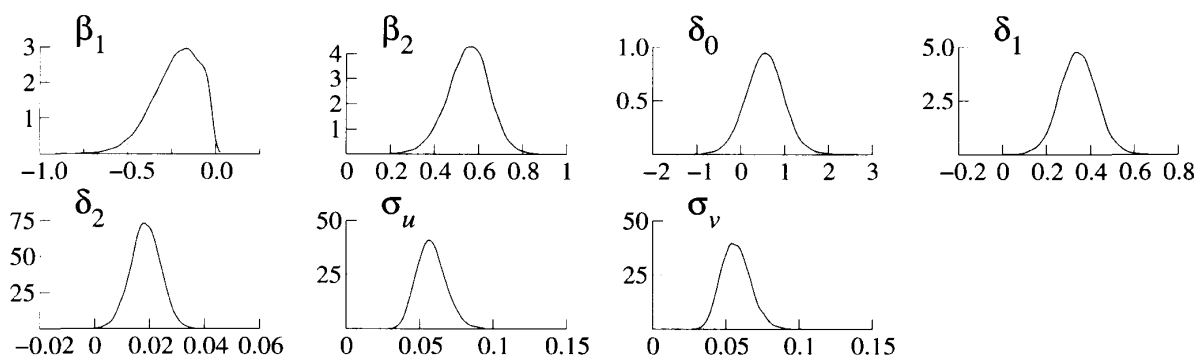


図8 周辺事後分布（電力）.

表7 日本の家庭用電力需要関数（1999年）の推定結果

パラメータ	平均	標準偏差	95%信用区間	非効率性因子*	CD**
β_1 (価格)	-.22	.13	[-.51 - .031]	8.19 [70.50]	.446
β_2 (帰属所得)	.55	.094	[.36 .73]	18.99 [69.68]	.976
δ_0 (定数項)	.55	.43	[-.29 1.41]	32.44 [64.55]	.720
δ_1 (世帯人数)	.34	.083	[.18 .51]	10.79 [17.99]	.997
δ_2 (気温)	.019	.005	[.008 .029]	1.15 [4.56]	.446
σ_u (測定誤差)	.058	.010	[.041 .079]	1.98 [3.12]	.261
σ_v (選好の異質性)	.057	.010	[.040 .078]	2.39 [2.98]	.888

* 角括弧内の値は、Miyawaki et al. (2006) の Table 11 で報告されている非効率性因子。

** CD は収束判定のための仮説検定の p -値を意味する。

推定結果は、非効率性因子を除いて Miyawaki et al. (2006) の Table 11 と同様な結果になっている。つまり、電力需要の価格弾力性は $\Pr(\beta_1 < 0 | y) > .95$ より 95%以上の事後確率で負と推定され、所得弾力性は $\Pr(\beta_2 > 0 | y) > .95$ より同様に正と推定された。事後標本平均の絶対値でみれば、価格に比べて所得の方が電力需要に与える影響がやや大きいことが分かる。また、選好の異質性の変数のうち、平均世帯人数及び平均気温のいずれも電力需要に正の効果を持ち、その一単位あたりの効果は平均世帯人数の方が大きいことが分かった。分散のパラメータ (σ_u , σ_v) は、数値実験 II の結果 (表 4) とは異なり、それらの周辺事後分布がほぼ同一となった。分散のパラメータは消費者のブロック選択行動を通じて推定される。しかし、Miyawaki et al. (2006) でも触れられているが、電力は最も料金の高いブロックで消費を行う場合がほとんどである。そのため、離散連続選択モデルはブロック選択を考慮した部分の影響を失い、2つの誤差項を含む回帰モデルに退化したと推測される。実際、潜在変数 s_i^* の MCMC 標本は全ての i に対して $s_i^* = 2K_i - 1$ となっていた。その結果、分散のパラメータを識別できなくなり、これらの周辺事後分布がほぼ同一になったと考えられる。

続いて非効率性因子に関して分析する。回帰係数 β に加えて選好の異質性の変数の係数 δ についても Miyawaki et al. (2006) の Table 11 と比べて改善が見られた。回帰係数 β に関しては、数値実験 II 同様、ブロックサンプラーの効果によると考えられる。一方選好の異質性のパラメータに関しては、これは数値実験では観察できなかった改善であり、計算機上の乱数の性質による影響が大きいのではないかと考えられる。また、依然として非効率性因子は δ_0 のものが最大であり、今後の改善が期待される。

7. おわりに

本論文では、線形制約下の多変量正規分布からの乱数発生法について、その代表的な 2 種類

の方法を紹介し、ブロックサンプラーと呼ばれる新しい発生方法を提案した。また、ブロックサンプラーの有効性を2種の数値実験によって確かめると共に、その応用例として、離散連続選択アプローチに基づくブロック料金制下の需要関数の推定を行った。提案されたブロックサンプラーは、MCMC法を用いたベイズ推定の実証研究において、従来の方法と比べて効率的であるため、様々な状況への応用が期待される。

今後の課題は、ブロックサンプラーにおける切断領域の近似の精度の改善である。ブロックサンプラーは任意の数の線形制約式に対応するために切断領域を近似するステップを含む（アルゴリズムBS-A）。この近似の精度が低い場合、ブロックサンプラーによる乱数発生効率性が低くなることが観察されている。最も簡単な対応方法としては、様々な線形制約式の組を試すようにすることである。しかし線形制約式の順序によっては、この方法が機能しない場合も考えられる。ブロックサンプラーの効率性を高めるために、切断領域の近似の精度の向上は重要である。

謝 辞

まず、2人の査読者から有益な提案及び指摘を得ることができ、本論文を大幅に改善することができました。また、古澄英男教授（神戸大学）、福重元嗣教授（大阪大学）、寒河江雅彦准教授（岐阜大学）及び編集長である大森裕浩准教授（東京大学）から有益なコメントを頂きました。ここに深く感謝致します。また、本研究は日本学術振興会特別研究員制度より助成を受けています。最後に、全ての計算結果はOx 4.1aを用いて得られたものです（Doornik, 2002）。

A. 補 論

A.1 命題1の証明

($\Leftarrow$) 階数の定義と $\text{rank}(\mathbf{D}) < n$ より明らか。

($\Rightarrow$) 命題の対偶を示すことで証明する。対偶は「 $\mathbf{D}$ から取り出すことのできる全ての連続した n 本の行ベクトルが一次従属であれば、 $\text{rank}(\mathbf{D}) < n$ 」である。ここで $\mathbf{D}$ の第 i 行ベクトルの転置を $\mathbf{d}_i (m \times 1)$ とする。つまり $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_m)'$ 。

まず、最初の n 本の行ベクトルが一次従属であること及び2番目から $n+1$ 番目までの行ベクトルが一次従属であることから、以下の関係式が成立する。

$$\mathbf{d}_1 = [\mathbf{d}_2, \dots, \mathbf{d}_n] \mathbf{c}_1, \quad \mathbf{d}_{n+1} = [\mathbf{d}_2, \dots, \mathbf{d}_n] \mathbf{c}_{n+1}. \quad (\text{A.1})$$

ただし $\mathbf{c}_i (i=1, \dots, m)$ は少なくとも一つの要素が0でない $n-1$ 次元ベクトルを表す。よって、 $\mathbf{D}$ の最初の $n+1$ 行を取り出した行列 $(\mathbf{d}_1, \dots, \mathbf{d}_{n+1})'$ の階数は $n-1$ 以下となる。

続いて、 $\mathbf{D}$ の最初の $n+2$ 行を取り出した行列 $(\mathbf{d}_1, \dots, \mathbf{d}_{n+2})'$ の階数を考える。まず最初の n 本の行ベクトル及び2番目から $n+1$ 番目までの行ベクトルがそれぞれ一次従属であるため、

$$\begin{aligned} \mathbf{d}_1 &= [\mathbf{d}_2, \dots, \mathbf{d}_n] \mathbf{c}_1, \quad \mathbf{d}_2 = [\mathbf{d}_3, \dots, \mathbf{d}_{n+1}] \mathbf{c}_2 \\ \Rightarrow \mathbf{d}_1 &= [(\mathbf{d}_3, \dots, \mathbf{d}_{n+1}) \mathbf{c}_2, \mathbf{d}_3, \dots, \mathbf{d}_n] \mathbf{c}_1 \equiv [\mathbf{d}_3, \dots, \mathbf{d}_{n+1}] \tilde{\mathbf{c}}_1, \end{aligned} \quad (\text{A.2})$$

となっている。よって、

$$\mathbf{d}_1 = [\mathbf{d}_3, \dots, \mathbf{d}_{n+1}] \tilde{\mathbf{c}}_1, \quad \mathbf{d}_2 = [\mathbf{d}_3, \dots, \mathbf{d}_{n+1}] \mathbf{c}_2, \quad \mathbf{d}_{n+2} = [\mathbf{d}_3, \dots, \mathbf{d}_{n+1}] \mathbf{c}_{n+2}, \quad (\text{A.3})$$

となるため、行列 $(\mathbf{d}_1, \dots, \mathbf{d}_{n+2})'$ の階数は $n-1$ 以下であることが分かる。

同様に考えれば、行列 $(\mathbf{d}_1, \dots, \mathbf{d}_{n+1})' (i=3, \dots, m-n)$ の階数はいずれも $n-1$ 以下であることを示せる。よって $i=m-n$ の場合から $\text{rank}(\mathbf{D}) \leq n-1 < n$ 。

A.2 逓減型ブロック料金制の場合

本節は逓減型のブロック料金制下の需要関数及びクロスセクションモデルを簡単に説明する。事後分布に関しては、以下の変数の定義を用いれば逓増型の場合と同じ式 (4.16) で書けるため、省略する。詳しい議論は Miyawaki and Omori (2007) を参照されたい。

まず需要関数は、

$$Y = Y_k, \text{ if } V_k = \max_j V_j \text{ and } k = 1, \dots, K, \quad (\text{A.4})$$

となるが、消費者理論の双対性を用いればさらに変形することができ、

$$Y = Y_k, \text{ if } \begin{cases} \bar{Y}_{k-1} < Y_k < \bar{Y}_k, \\ Y_k > H_j, & \text{for } j < k, \\ Y_k < H_j, & \text{for } j > k, \end{cases} \quad (\text{A.5})$$

となる。ここで $H_j = (Q_j - Q_k) / (P_j - P_k)$ である。4 節の図 5 (b) で見た通り、消費者は最適需要として条件付き需要（無差別曲線がブロックの内側で予算線に接するもの）のみを選んでゐる。

続いてクロスセクションモデルを導出する。条件付き需要に対数線形モデルを仮定し、選好の異質性を加えれば、

$$y_i^* = y_{ik} + w_i^*, \text{ if } \begin{cases} \bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{i,k}, \\ y_{ik} + w_i^* > h_{ij}, & \text{for } j < k, \\ y_{ik} + w_i^* < h_{ij}, & \text{for } j > k, \end{cases} \quad (\text{A.6})$$

となる。小文字の変数は対数を取ったものを表している。潜在需要の状態 s_i^* は 1 から K_i までの自然数をとる離散変数で定義され、 $k = 1, \dots, K_i$ に対して、

$$s_i^* = k, \text{ if } \begin{cases} \bar{y}_{i,k-1} < y_{ik} + w_i^* < \bar{y}_{i,k}, \\ y_{ik} + w_i^* > h_{ij}, & \text{for } j < k, \\ y_{ik} + w_i^* < h_{ij}, & \text{for } j > k, \end{cases} \quad (\text{A.7})$$

である。選好の異質性に関して整理すれば、

$$\begin{aligned} s_i^* = k &\iff \begin{cases} \bar{y}_{i,k-1} < y_i^* < \bar{y}_{i,k} \\ y_i^* > h_{ij}, & \text{for } j < k \\ y_i^* < h_{ij}, & \text{for } j > k \end{cases} \\ &\iff w_i^* \in R_{ik} = \left\{ \max \left(\bar{y}_{i,k-1}, \max_{j < k} h_{ij} \right) - y_{ik}, \min \left(\bar{y}_{i,k}, \min_{j > k} h_{ij} \right) - y_{ik} \right\}, \end{aligned} \quad (\text{A.8})$$

と変形できる。逓減型の場合も分離条件が必要となり、

$$\beta \in \bigcap_{i=1}^n SR_i, \quad SR_i = \bigcap_{j=1}^{K_i-1} \{ \beta | x'_{i,j+1} \beta \leq x'_{i,j} \beta + \bar{y}_{ij}^+ \}, \quad (\text{A.9})$$

となる。ただし変数 $\bar{y}_{ij}^+$ は、 $\bar{y}_{ij}^+ = \max(\bar{y}_{ij}, \max_{l < j+1} h_{il}) - \min(\bar{y}_{ij}, \min_{l > j} h_{il})$ ($i = 1, \dots, n, j = 1, \dots, K_{i-1}$) と定義される。

潜在需要に測定誤差を加えれば、クロスセクションモデルは、

$$\begin{aligned}
y_{ik} &= \mathbf{x}'_{i,k} \boldsymbol{\beta}, \quad k=1, \dots, K_i, \\
w_i^* &= \mathbf{z}'_i \boldsymbol{\delta} + v_i, \quad v_i \sim \text{i.i.d. } N(0, \sigma_v^2), \\
s_i^* &= k, \quad \text{if } w_i^* \in R_{ik} \text{ and } k=1, \dots, K_i, \\
y_i &= y_i^* + u_i = y_{is_i^*} + w_i^* + u_i, \quad u_i \sim \text{i.i.d. } N(0, \sigma_u^2),
\end{aligned} \tag{A.10}$$

となる。この統計的モデルは第 II 型の多項トービットモデルに分類される。

A.3 条件付き事後分布

本節はアルゴリズム MCMC で用いられる具体的な条件付き事後分布を各ステップに従って説明する。逓増型、逓減型のブロック料金制に関わらず同一の条件付き事後分布を導出するために、料金制によって異なる集合 $(\mathcal{A}, R_{\beta,i})$ 及び多項分布のパラメータ $(m_{is}, \theta_{is}, \tau_s^2)$ を先に定義しておく²⁾。

逓増型の場合：

$$\mathcal{A} = \{i | s_i^* = 2k-1\}, \tag{A.11}$$

$$R_{\beta,i} = \begin{cases} \{\boldsymbol{\beta} | \mathbf{x}'_{ik} \boldsymbol{\beta} < \bar{y}_{il} - w_i^*, \mathbf{x}'_{ik} \boldsymbol{\beta} > \bar{y}_{i,k-1} - w_i^*\}, & \text{if } s_i^* = 2k-1, \\ \{\boldsymbol{\beta} | \mathbf{x}'_{i,k+1} \boldsymbol{\beta} \leq \bar{y}_{ik} - w_i^*, \mathbf{x}'_{ik} \boldsymbol{\beta} \geq \bar{y}_{ik} - w_i^*\}, & \text{if } s_i^* = 2k, \end{cases} \tag{A.12}$$

$$(m_{is}, \theta_{is}, \tau_s^2) = \begin{cases} \left(\frac{\sigma_u^{-2} \sigma_v^{-2} (y_i - y_{ik} - \mathbf{z}'_i \boldsymbol{\delta})^2}{\sigma_u^{-2} + \sigma_v^{-2}}, \frac{\sigma_u^{-2} (y_i - y_{ik}) + \sigma_v^{-2} \mathbf{z}'_i \boldsymbol{\delta}}{\sigma_u^{-2} + \sigma_v^{-2}}, \{\sigma_u^{-2} + \sigma_v^{-2}\}^{-1} \right), & \text{if } s = 2k-1, \\ (\sigma_u^{-2} (y_i - \bar{y}_{ik})^2, \mathbf{z}'_i \boldsymbol{\delta}, \sigma_v^2), & \text{if } s = 2k. \end{cases} \tag{A.13}$$

逓減型の場合：

$$\mathcal{A} = \{i | i=1, \dots, n\}, \tag{A.14}$$

$$R_{\beta,i} = \{\boldsymbol{\beta} | \bar{y}_{ik-1} - w_i^* < \mathbf{x}'_{ik} \boldsymbol{\beta} < \bar{y}_{il} - w_i^*, \mathbf{x}'_{ik} \boldsymbol{\beta} > h_{ij} - w_i^* (j < k), \mathbf{x}'_{ik} \boldsymbol{\beta} < h_{ij} - w_i^* (j > k)\}, \tag{A.15}$$

$$(m_{is}, \theta_{is}, \tau_s^2) = \left(\frac{\sigma_u^{-2} \sigma_v^{-2} (y_i - y_{is} - \mathbf{z}'_i \boldsymbol{\delta})^2}{\sigma_u^{-2} + \sigma_v^{-2}}, \frac{\sigma_u^{-2} (y_i - y_{is}) + \sigma_v^{-2} \mathbf{z}'_i \boldsymbol{\delta}}{\sigma_u^{-2} + \sigma_v^{-2}}, \{\sigma_u^{-2} + \sigma_v^{-2}\}^{-1} \right), \tag{A.16}$$

A.3.1 Step 2. $(\sigma_u^2, \boldsymbol{\beta})$ を s^*, w^* を所与として発生させる。

$\boldsymbol{\beta}$ を積分消去すれば、 σ_u^2 の条件付き事後分布は逆ガンマ分布 $IG\left(\frac{n_{u,1}}{2}, \frac{S_{u,1}}{2}\right)$ となる。ただし、 $n_{u,1} = n_{u,0} + n$,

$$S_{u,1} = S_{u,0} + \boldsymbol{\mu}'_{\beta,0} \boldsymbol{\Sigma}_{\beta,0}^{-1} \boldsymbol{\mu}_{\beta,0} + \sum_{i \in \mathcal{A}} (y_i - w_i^*)^2 + \sum_{i \notin \mathcal{A}} (y_i - \bar{y}_{ik})^2 - \boldsymbol{\mu}'_{\beta,1} \boldsymbol{\Sigma}_{\beta,1}^{-1} \boldsymbol{\mu}_{\beta,1}. \tag{A.17}$$

この逆ガンマ分布で得られた σ_u^2 の標本を所与とすれば、 $\boldsymbol{\beta}$ の条件付き事後分布は線形制約 $\boldsymbol{\beta} \in R_\beta$ で切断された多変量正規分布 $N_2(\boldsymbol{\mu}_{\beta,1}, \boldsymbol{\Sigma}_u^2 \boldsymbol{\Sigma}_{\beta,1})$ となる。ただし、 R_β , $\boldsymbol{\mu}_{\beta,1}$ 及び $\boldsymbol{\Sigma}_{\beta,1}$ は以下の通り。

$$R_\beta = \bigcap_{i=1}^n \{R_{\beta,i} \cap SR_i\}, \quad \boldsymbol{\mu}_{\beta,1} = \boldsymbol{\Sigma}_{\beta,1} \left(\boldsymbol{\Sigma}_{\beta,0}^{-1} \boldsymbol{\mu}_{\beta,0} + \sum_{i \in \mathcal{A}} (y_i - w_i^*) \mathbf{x}_{ik} \right), \quad \boldsymbol{\Sigma}_{\beta,1}^{-1} = \boldsymbol{\Sigma}_{\beta,0}^{-1} + \sum_{i \in \mathcal{A}} \mathbf{x}_{ik} \mathbf{x}'_{ik}. \tag{A.18}$$

²⁾ 逓減型の場合、 τ_s^2 は s に依存しないが、記法の統一のために τ_s^2 としている。

A.3.2 Step 3. (σ_v^2, δ) を w^* を所与として発生させる.

δ を積分消去すれば, σ_v^2 の条件付き事後分布は逆ガンマ分布 $IG\left(\frac{n_{v,1}}{2}, \frac{S_{v,1}}{2}\right)$ となる. ただし, $n_{v,1} = n_{v,0} + n$,

$$S_{v,1} = S_{v,0} + \mu'_{\delta,0} \Sigma_{\delta,0}^{-1} \mu_{\delta,0} + w'^* w^* - \mu'_{\delta,1} \Sigma_{\delta,1}^{-1} \mu_{\delta,1}. \quad (\text{A.19})$$

この逆ガンマ分布で得られた σ_v^2 の標本を所与とすれば, δ の条件付き事後分布は多変量正規分布 $N_d(\mu_{\delta,1}, \sigma_v^2 \Sigma_{\delta,1})$ となる. ここで,

$$\mu_{\delta,1} = \Sigma_{\delta,1} (\Sigma_{\delta,0}^{-1} \mu_{\delta,0} + z' w^*), \quad \Sigma_{\delta,1}^{-1} = \Sigma_{\delta,0}^{-1} + Z' Z. \quad (\text{A.20})$$

A.3.3 Step 4. (s_i^*, w_i^*) を $\beta, \delta, \sigma_u^2, \sigma_v^2 (i=1, \dots, n)$ を所与として発生させる.

選好の異質性 w_i^* を積分消去すれば, 潜在需要 y_i^* の状態を表す s_i^* の条件付き事後分布は多項分布に従い,

$$Pr(s_i^* = s | \beta, \delta, \sigma_u^2, \sigma_v^2) \propto \tau_s [\Phi\{\tau_s^{-1}(RU_{is} - \theta_{is})\} - \Phi\{\tau_s^{-1}(RL_{is} - \theta_{is})\}] \exp\left(-\frac{m_{is}}{2}\right). \quad (\text{A.21})$$

ただし (RL_{ik}, RU_{ik}) は選好の異質性の区間 R_{ik} の下限, 上限をそれぞれ表す.

この多項分布で得られた $s_i^* = s$ を所与とすれば, w_i^* の条件付き事後分布は領域 R_{is} で切断された正規分布 $N(\theta_{is}, \tau_s^2)$ となる.

参 考 文 献

- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, **88**, 669-679.
- Averett, S. L. and Hotchkiss, J. L. (1997). Female labor supply with a discontinuous, nonconvex budget constraint: Incorporation of a part-time/full-time wage differential, *The Review of Economics and Statistics*, **79**, 461-470.
- Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, *Annals of Mathematical Statistics*, **29**, 610-611.
- Burtless, G. and Hausman, J. A. (1978). The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment, *Journal of Political Economy*, **86**, 1103-1130.
- Chib, S. (2001). Markov chain Monte Carlo methods: Computation and inference, In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 57, 3569-3649, North-Holland; Amsterdam.
- de Jong, G. C. (1990). An indirect utility model of car ownership and private car use, *European Economic Review*, **34**, 971-985.
- Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox* (3rd ed.), Timberlake Consultants Press and Oxford; London.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities, In E. M. Keramidas (Ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 571-578, Fairfax, VA: Interface Foundation of North America.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, 169-193, Oxford University Press; Oxford.
- Hajivassiliou, V. A. and Ruud, P. A. (1994). Classical estimation methods for ldv models using simulation, In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 40, 2383-2441, North Holland; Oxford.
- Hanemann, W. M. (1984). Discrete/continuous models of consumer demand, *Econometrica*, **52**, 541-562.
- Hausman, J. A. (1985). The econometrics of nonlinear budget sets, *Econometrica*, **53**, 1255-1282.
- Miyawaki, K. and Omori, Y. (2007). Duality-based analysis of residential gas demand under decreasing block rate pricing, University of Tokyo CIRJE Discussion Paper Series CIRJE-F-506.

- Miyawaki, K., Omori, Y., and Hibiki, A. (2006). Bayesian estimation of demand functions under block rate pricing, University of Tokyo CIRJE Discussion Paper Series CIRJE-F-424.
- Moffitt, R. (1986). The econometrics of piecewise-linear budget constraint, *Journal of Business and Economic Statistics*, **4**, 317-328.
- Olmstead, S. M., Hanemann, W. M., and Stavins, R. N. (2007). Water demand under alternative price structures, *Journal of Environmental Economics and Management*, **54**, 181-198.
- Reiss, P. C. and White, M. W. (2005). Household electricity demand, revisited, *Review of Economic Studies*, **72**, 853-883.
- Rodriguez-Yam, G., Davis, R. A., and Scharf, L. L. (2004). Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression, Technical report, Columbia University, Available at: <http://www.stat.columbia.edu/~rdavis/papers/CLR.pdf>.