

ラプラス確率的フロンティアモデルの ベイズ推定

Bayesian Analysis of Laplace Stochastic Frontier Models

古 澄 英 男*

本論文では、ラプラス確率フロンティアモデルについて考えることにする。ベイズ推定を行うため、ラプラス分布の混合表現を利用して、ラプラス確率的フロンティアモデルに対するギブス・サンプリング法の開発を行う。また、数値実験と実際のデータによる分析を通じて、本論文で提案するギブス・サンプリング法がうまく機能するかどうか検証する。

This paper considers Laplace stochastic frontier models from a Bayesian point of view. For posterior inference, we develop Gibbs sampling methods using the mixture representations of the Laplace distribution. The proposed methods are illustrated by both simulated and real data.

Hideo Kozumi

JEL : C11, C15, D24

キーワード：確率フロンティアモデル、ギブス・サンプリング、混合表現、ラプラス分布

Keywords : Gibbs sampling, Laplace distribution, mixture representation,
stochastic frontier model

1 はじめに

Aigner *et al.* (1977) と Meeusen and van den Broeck (1977) によって提案された確率的フロンティアモデル (stochastic frontier model) は、経済活動の効率性を分析する計量モデルとして様々な分野で用いられている (確率的フロンティアモデルについては、Kumbhakar and Lovell (2000), Parmeter

* 本研究は、JSPS 科研費 25245035 の助成を受けたものである。

and Kumbhakar (2014) に詳しい)。

確率的フロンティアモデルの特徴として、通常の誤差項だけでなく、非効率性を表す非負の誤差項を導入している点を挙げるができる(第 2 節を参照)。また、モデルの推定では、2 つの誤差項に対して特定の確率分布を仮定し、最尤法を利用することが多い。このとき、非負の誤差項には切断正規分布 (Stevenson (1980)), 指数分布 (Meeusen and van den Broeck (1977)), ガンマ分布 (Beckers and Hammond (1987), Greene (1990)), ワイブル分布 (Tsionas (2007)) など様々な確率分布が用いられるが、通常の誤差項に対しては正規分布を仮定することが一般的である。

最近になり、Nguyen (2010) と Horrace and Parmeter (2015) は、通常の誤差項に対してラプラス分布を仮定した確率的フロンティアモデル(以下、ラプラス確率的フロンティアモデル)を提案した。従来の確率的フロンティアモデルが通常の回帰モデルの拡張となっているのに対し、彼らのモデルはメディアン回帰を確率的フロンティアモデルに拡張したものとみなすことができる。したがって、ラプラス確率的フロンティアモデルでは、外れ値などがある場合でも安定した推定結果を得られることが期待される。

ラプラス確率的フロンティアモデルを最尤法によって推定する場合、準ニュートン法などの数値最適化アルゴリズムを利用することは困難である。これは、ラプラス確率的フロンティアモデルの尤度関数に絶対値が含まれるためである。この問題に対処するため、Nguyen (2010) は直接尤度関数を最大化するのではなく、Dempster *et al.* (1977) によって提案された EM アルゴリズム (Expectation-Maximization algorithm) により最尤推定値を求めている。一方、Horrace and Parmeter (2015) では、差分進化 (differential evolution) とよばれる大局的最適化法 (Storn and Price (1997)) によって尤度関数を最大化することを提案している。しかしながら、Nguyen (2010) の方法では繰り返しごとに関数の最大化を数値的に行う必要があり、かなりの計算時間を要する。また、Horrace and Parmeter (2015) の方法もアルゴリズムの調整が必要であり、実行が容易ではないという問題がある。

本論文の目的は、最尤推定法に代わるラプラス確率的フロンティアモデルの

推定方法を開発することである。具体的には、ラプラス確率的フロンティアモデルに対するベイズ推定を考え、マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method: MCMC 法) による推定方法を提案する。そのために本論文では、ラプラス分布に関する混合表現を利用し、ラプラス確率的フロンティアモデルの再定式化を行う。ラプラス分布に対する混合表現はいくつかあるが、本論文では、正規分布と指数分布を用いた混合表現と一様分布とガンマ分布による混合表現の2つを考え、いずれにおいてもラプラス確率的フロンティアモデルがギブス・サンプリング (Gibbs sampling) によって推定できることを示す。さらに、それぞれの混合表現から導かれる2つのギブス・サンプリングについて、効率性の観点から比較・検討も行う。

本論文の構成は以下の通りである。第2節では、ラプラス確率的フロンティアモデルについて説明し、ラプラス分布の混合表現を用いてモデルの再定式化を行う。第3節では、ラプラス確率的フロンティアモデルに対するギブス・サンプリング法を導出する。本論文で提案するギブス・サンプリング法がうまく機能するかどうか調べるため、第4節において数値実験と実際のデータを使った分析を行う。最後に、第5節において結論を述べることにする。

2 ラプラス確率的フロンティアモデル

2.1 モデルと事前分布

Aigner *et al.* (1977) と Meeusen and van den Broeck (1977) によって提案された確率的フロンティアモデルは、

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + v_i - u_i, \quad (i = 1, \dots, n) \quad (1)$$

と表すことができる。ここで、 y_i は被説明変数 (通常は対数値)、 $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$ は $k \times 1$ の説明変数ベクトル、 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ は $k \times 1$ の回帰係数ベクトル、 v_i は誤差項を表す。

(1) 式から明らかなように、確率的フロンティアモデルと通常の回帰モデルとの違いは u_i の存在である。この u_i は、経済主体の非効率性を表す変数であ

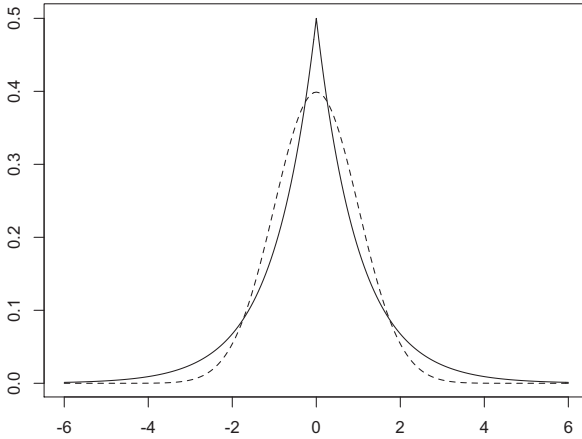


図 1: ラプラス分布 $La(0, 1)$ (実線) と正規分布 $N(0, 1)$ (破線) の確率密度関数

り、非負の確率変数であることが仮定される。また、確率的フロンティアモデルでは、 u_i を片側誤差項 (one-sided error) あるいは非負の誤差項とよび、通常の誤差項 v_i を両側誤差項 (two-sided error) とよぶ。

Nguyen (2010) と Horrace and Parmeter (2015) らが提案したラプラス確率的フロンティアモデルでは、両側誤差項 v_i に対してラプラス分布を仮定する。すなわち、

$$v_i \sim La(0, \sigma_v)$$

であるとする。ここで、ラプラス分布 $La(\mu, \sigma)$ の確率密度関数は、

$$f(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right) \quad (-\infty < x < \infty)$$

で与えられ、その平均は μ 、分散は $2\sigma^2$ である。図 1 には、ラプラス分布 $La(0, 1)$ の確率密度関数が示されており、ラプラス分布が微分不可能な点を持ち、正規分布よりも厚い裾を持っていることが分かる。

次に、非負の誤差項 u_i については、Nguyen (2010) にしたがって、

$$u_i \sim Exp(\sigma_u^{-1}) \tag{2}$$

を仮定することにする。ここで、 $Exp(\gamma^{-1})$ は平均が γ である指数分布を表す。このとき、パラメータを $\theta = (\beta', \sigma_v, \sigma_u)'$ とおけば、ラプラス確率フロンティアモデルの尤度関数は、

$$f(\mathbf{y}|\theta) = \prod_{i=1}^n \int_0^{\infty} \frac{1}{2\sigma_v} \exp\left\{-\frac{|y_i - \mathbf{x}'_i\beta + u_i|}{\sigma_v}\right\} \pi(u_i|\sigma_u) du_i \quad (3)$$

と表すことができる。ただし、 $\pi(u_i|\sigma_u)$ は (2) 式から得られる u_i の確率分布で、

$$\pi(u_i|\sigma_u) = \frac{1}{\sigma_u} \exp\left(-\frac{u_i}{\sigma_u}\right)$$

である。

(3) 式は絶対値と積分を含んでいるため、直接尤度関数の最大化を行いパラメータの最尤推定値を求めることは困難である。そこで最尤推定法に代わり、ラプラス確率的フロンティアモデルのベイズ推定を考えることにする。ベイズ推定では、パラメータ $\beta, \sigma_v, \sigma_u$ に対する事前分布が必要である。本論文では、 $\beta, \sigma_v, \sigma_u$ は互いに独立であるとし、

$$\beta \sim N(\beta_0, \mathbf{B}_0), \quad \sigma_v \sim IG(n_0, s_0), \quad \sigma_u \sim IG(m_0, d_0)$$

を仮定する。ここで、 $\beta_0, \mathbf{B}_0, n_0, s_0, m_0, d_0$ は事前分布のハイパー・パラメータを表す。また、 $IG(\alpha, \beta)$ は逆ガンマ分布を表し、その確率密度関数は

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp\left(-\frac{\beta}{x}\right)$$

で与えられる ($\Gamma(\alpha)$ はガンマ関数である)。

2.2 混合表現

混合表現 I

ラプラス分布は様々な混合表現が可能である (例えば、Kotz *et al.* (2001) を参照)。いま、

$$\epsilon_i \sim N(0, 1), \quad z_i \sim Exp(1/2)$$

とし、 ϵ_i と z_i は互いに独立であるとする。このとき、 $v_i \sim La(0, \sigma_v)$ は、

$$v_i = \sigma_v \sqrt{z_i} \epsilon_i \quad (4)$$

と表すことができる (Park and Casella (2008)). 以下では, ラプラス分布に
 対するこの混合表現を混合表現 I とよぶことにする.

(4) 式において z_i を所与とすれば,

$$v_i \mid z_i \sim N(0, \sigma_v^2 z_i)$$

であることが分かる. よって, $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{z} = (z_1, \dots, z_n)'$, $\mathbf{u} =$
 $(u_1, \dots, u_n)'$ とおけば, \mathbf{z} と \mathbf{u} を所与とした尤度関数は絶対値を使うことなく,

$$f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z}, \mathbf{u}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_v^2 z_i}} \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i)^2}{2\sigma_v^2 z_i} \right\} \quad (5)$$

と表すことができる.

混合表現 II

先の正規分布と指数分布を用いた混合表現 I とは異なる混合表現を考えるため, ϵ_i は区間 $(-1, 1)$ における一様分布 $U(-1, 1)$, z_i はガンマ分布 $Ga(2, 1)$
 にしたがうとする:

$$\epsilon_i \sim U(-1, 1), \quad z_i \sim Ga(2, 1)$$

ここで, ガンマ分布 $Ga(\alpha, \beta)$ の確率密度関数は

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

である. このとき, ラプラス分布 $La(0, \sigma_v)$ にしたがう v_i は, ϵ_i と z_i を用いて

$$v_i = \sigma_v z_i \epsilon_i$$

と表すことができ (Mallick and Yi (2014)), これをラプラス分布の混合表現
 II とよぶことにする.

新たに導入した z_i を所与とすれば,

$$v_i \mid z_i \sim U(-\sigma_v z_i, \sigma_v z_i)$$

となるので, 混合表現 I, II とともに, スケールパラメータに関する混合表現と
 なっていることが分かる. さらに, \mathbf{z} と \mathbf{u} を所与とした尤度関数の別の表現と
 して,

$$f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z}, \mathbf{u}) = \prod_{i=1}^n \frac{1}{2\sigma_v z_i} I(|y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i| < \sigma_v z_i) \quad (6)$$

が得られる。ここで、 $I(\cdot)$ は指示関数を表す。

3 ギブス・サンプリング

(5) 式と (6) 式のいずれの尤度関数を用いても、解析的にパラメータの事後分布を求めることができない。これは、(5) 式と (6) 式から \mathbf{z} と \mathbf{u} を積分によって消去することができないためである。そこで本節では、MCMC 法による推定を考えることにする。

3.1 ギブス・サンプリング I

はじめに、混合表現 I から得られる (5) 式の尤度関数を考え、 $\boldsymbol{\beta}$, σ_v , σ_u , \mathbf{z} , \mathbf{u} の完全条件付き分布をそれぞれ導出することにする。

σ_v の完全条件付き分布については、(3) 式より \mathbf{z} を積分により消去することができ、

$$\sigma_v | \mathbf{y}, \boldsymbol{\beta}, \mathbf{u} \sim IG\left(n_0 + n, s_0 + \sum_{i=1}^n |y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i|\right)$$

となる。また、 σ_u の完全条件付き分布については、 $u_i \sim \text{Exp}(\sigma_u^{-1})$ であることと σ_u の事前分布から、

$$\pi(\sigma_u | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \mathbf{z}, \mathbf{u}) \propto \prod_{i=1}^n \sigma_u^{-1} \exp\left(-\frac{u_i}{\sigma_u}\right) \times \sigma_u^{-(m_0+1)} \exp\left(-\frac{d_0}{\sigma_u}\right)$$

と表すことができる。したがって、

$$\sigma_u | \mathbf{y}, \mathbf{u} \sim IG\left(m_0 + n, d_0 + \sum_{i=1}^n u_i\right)$$

を得る。

通常の線形回帰モデルにおける結果を用いれば、 $\boldsymbol{\beta}$ の完全条件付き分布は

$$\boldsymbol{\beta} | \mathbf{y}, \sigma_v, \mathbf{z}, \mathbf{u} \sim N(\hat{\boldsymbol{\beta}}, \hat{\mathbf{B}})$$

となる。ここで、

$$\hat{\mathbf{B}} = \left(\sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}'_i}{\sigma_v^2 z_i} + \mathbf{B}_0^{-1} \right)^{-1}, \quad \hat{\boldsymbol{\beta}} = \hat{\mathbf{B}} \left\{ \sum_{i=1}^n \frac{\mathbf{x}_i (y_i + u_i)}{\sigma_v^2 z_i} + \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 \right\}$$

である。

片側誤差項 u_i ($i = 1, \dots, n$) の完全条件付き分布は、

$$\pi(u_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{z}) \propto \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i)^2}{2\sigma_v^2 z_i} \right\} \times \exp \left(-\frac{u_i}{\sigma_u} \right)$$

と書くことができる。ここで、

$$\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i)^2}{2\sigma_v^2 z_i} + \frac{u_i}{\sigma_u} = \frac{\left\{ u_i - \left(\mathbf{x}'_i \boldsymbol{\beta} - y_i - \frac{\sigma_v^2 z_i}{\sigma_u} \right) \right\}^2 + (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2}{2\sigma_v^2 z_i}$$

と書き直すことができることから、

$$u_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{z} \sim N \left(\mathbf{x}'_i \boldsymbol{\beta} - y_i - \frac{\sigma_v^2 z_i}{\sigma_u}, \sigma_v^2 z_i \right) I(u_i > 0)$$

を得る。これは、 u_i の完全条件付き分布が、区間 $(0, \infty)$ において切断された正規分布であることを表している。

最後に z_i ($i = 1, \dots, n$) の完全条件付き分布を導出する。(5) 式と $z_i \sim Exp(1/2)$ であることから、 z_i の完全条件付き分布は、

$$\pi(z_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u}) \propto \left(\frac{1}{z_i} \right)^{1/2} \exp \left\{ -\frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i)^2}{2\sigma_v^2 z_i} \right\} \times \exp \left(-\frac{z_i}{2} \right)$$

と表すことができる。いま、

$$\delta_i^2 = \frac{(y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i)^2}{\sigma_v^2}$$

とおけば、

$$\pi(z_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u}) \propto \left(\frac{1}{z_i} \right)^{1/2} \exp \left\{ -\frac{1}{2} \left(\frac{\delta_i^2}{z_i} + z_i \right) \right\}$$

となる。ここで、 $\gamma_i = 1/z_i$ とすれば、

$$\begin{aligned} \pi(\gamma_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u}) &\propto \gamma_i^{1/2} \exp \left\{ -\frac{1}{2} \left(\delta_i^2 \gamma_i + \frac{1}{\gamma_i} \right) \right\} \times \gamma_i^{-2} \\ &\propto \gamma_i^{-3/2} \exp \left\{ -\frac{(\gamma_i - 1/\delta_i)^2}{2\gamma_i/\delta_i} \right\} \end{aligned}$$

とまとめることができる。これは、 γ_i が逆ガウス分布 $IGauss(1/\delta_i, 1)$ にしたがることを示しているので、

$$\frac{1}{z_i} | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \mathbf{u} \sim IGauss \left(\frac{1}{\delta_i}, 1 \right)$$

を得る。ここで、逆ガウス分布 $IGauss(\mu, \lambda)$ の確率密度関数は、

$$f(x) = \left(\frac{\lambda}{2\pi x^3} \right)^{1/2} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}$$

で与えられる。

3.2 ギブス・サンプリング II

次に、混合表現 II から得られる (6) 式の尤度関数にもとづいて、 β , σ_v , σ_u , \mathbf{z} , \mathbf{u} の完全条件付き分布を導出することにする。

容易に確認できるように、 σ_v と σ_u の完全条件付き分布は混合表現 I の場合と同じである。すなわち、

$$\begin{aligned}\sigma_v | \mathbf{y}, \beta, \mathbf{u} &\sim IG\left(n_0 + n, s_0 + \sum_{i=1}^n |y_i - \mathbf{x}'_i \beta + u_i|\right) \\ \sigma_u | \mathbf{y}, \mathbf{u} &\sim IG\left(m_0 + n, d_0 + \sum_{i=1}^n u_i\right)\end{aligned}$$

である。

次に、 β_j ($j = 1, \dots, k$) の完全条件付き分布は、(6) 式より

$$\pi(\beta_j | \mathbf{y}, \sigma_v, \mathbf{z}, \mathbf{u}) \propto \prod_{i=1}^n I(|y_i - \mathbf{x}'_i \beta + u_i| < \sigma_v z_i) \times \pi(\beta_j | \beta_{-j})$$

と書くことができる。ここで、 $\pi(\beta_j | \beta_{-j})$ は $\beta_{-j} = (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k)'$ を所与とした β_j の事前分布を表す。いま、 $\lambda_{ij} = \sum_{r \neq j} x_{ir} \beta_r$ とし、 $x_{ij} \neq 0$ に対して

$$\begin{aligned}b_{ij}^L &= \min \left\{ \frac{y_i - \lambda_{ij} + u_i - \sigma_v z_i}{x_{ij}}, \frac{y_i - \lambda_{ij} + u_i + \sigma_v z_i}{x_{ij}} \right\} \\ b_{ij}^U &= \max \left\{ \frac{y_i - \lambda_{ij} + u_i - \sigma_v z_i}{x_{ij}}, \frac{y_i - \lambda_{ij} + u_i + \sigma_v z_i}{x_{ij}} \right\}\end{aligned}$$

とおくことにする。このとき、

$$\beta_j | \mathbf{y}, \beta_{-j}, \sigma_v, \mathbf{z}, \mathbf{u} \sim \pi(\beta_j | \beta_{-j}) I\left(\max_{i: x_{ij} \neq 0} b_{ij}^L < \beta_j < \min_{i: x_{ij} \neq 0} b_{ij}^U\right) \quad (7)$$

となる。これは、 β_j の完全条件付き分布が、事前分布を区間 $(\max_{i: x_{ij} \neq 0} b_{ij}^L, \min_{i: x_{ij} \neq 0} b_{ij}^U)$ において切断した分布となることを示している。

同様に、 u_i ($i = 1, \dots, n$) の完全条件付き分布は

$$\pi(u_i | \mathbf{y}, \beta, \sigma_v^2, \sigma_u, \mathbf{z}) \propto I(|y_i - \mathbf{x}'_i \beta + u_i| < \sigma_v z_i) \times \exp\left(-\frac{u_i}{\sigma_u}\right)$$

と表すことができ、区間 $(\mathbf{x}'_i \beta - y_i - \sigma_v z_i, \mathbf{x}'_i \beta - y_i + \sigma_v z_i)$ で切断された指数分布 $Exp(\sigma_u^{-1})$ 、すなわち

$$u_i | \mathbf{y}, \beta, \sigma_v^2, \sigma_u, \mathbf{z} \sim Exp(\sigma_u^{-1}) I(|y_i - \mathbf{x}'_i \beta + u_i| < \sigma_v z_i)$$

であることが分かる。さらに, z_i ($i = 1, \dots, n$) の完全条件付き分布も容易に導くことができ,

$$\begin{aligned} \pi(z_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u}) &\propto I(|y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i| < \sigma_v z_i) \times \frac{1}{z_i} \times z_i^{2-1} \exp(-z_i) \\ &= \exp(-z_i) I(|y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i| < \sigma_v z_i) \end{aligned}$$

となる。これより, z_i の完全条件付き分布は, 区間 $(\mathbf{x}'_i \boldsymbol{\beta} - y_i - \sigma_v z_i, \mathbf{x}'_i \boldsymbol{\beta} - y_i + \sigma_v z_i)$ で切断された指数分布

$$z_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u} \sim \text{Exp}(1) I(|y_i - \mathbf{x}'_i \boldsymbol{\beta} + u_i| < \sigma_v z_i)$$

である。

混合表現 I, II のどちらを用いても, すべての完全条件付き分布から容易にサンプリングすることができる。したがって, MCMC 法の 1 つであるギブス・サンプリング (Gelfand and Smith (1990)) によって, ラプラス確率的フロンティアモデルをベイズ推定することができる。本論文では, 混合表現 I にもとづくギブス・サンプリングをギブス・サンプリング I, 混合表現 II によるギブス・サンプリングをギブス・サンプリング II とよぶことにし, これらは以下のようにまとめることができる。

ギブス・サンプリング I, II

1. σ_v を $\pi(\sigma_v | \mathbf{y}, \boldsymbol{\beta}, \mathbf{u})$ からサンプリングする。
2. σ_u を $\pi(\sigma_u | \mathbf{y}, \mathbf{u})$ からサンプリングする。
3. z_i ($i = 1, \dots, n$) を $\pi(z_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{u})$ からサンプリングする。
4. ギブス・サンプリング I :
 $\boldsymbol{\beta}$ を $\pi(\boldsymbol{\beta} | \mathbf{y}, \sigma_v, \mathbf{z}, \mathbf{u})$ からサンプリングする。
 ギブス・サンプリング II :
 β_j ($j = 1, \dots, k$) を $\pi(\beta_j | \mathbf{y}, \boldsymbol{\beta}_{-j}, \sigma_v, \mathbf{z}, \mathbf{u})$ からサンプリングする。
5. u_i ($i = 1, \dots, n$) を $\pi(u_i | \mathbf{y}, \boldsymbol{\beta}, \sigma_v, \sigma_u, \mathbf{z})$ からサンプリングする。

ギブス・サンプリング I と II の違いは、 β のサンプリング方法に見て取れるであろう。すなわち、ギブス・サンプリング I では β を同時にサンプリングするのに対して、ギブス・サンプリング II では各 β_j を個別にサンプリングしている。次節で示すように、この違いはギブス・サンプリングの効率性に影響を与える。また、上述のアルゴリズムにおいて、 σ_v と σ_u の完全条件付き分布が \mathbf{z} に依存していないことは重要である。本論文で導出した 2 つのギブス・サンプリングでは、最初の 3 つのステップを実行することによって、 σ_v 、 σ_u 、 \mathbf{z} を同時にサンプリングしていることになる (σ_v と σ_u のサンプリングの順番は問わないが、 \mathbf{z} はこれらの直後にサンプリングする必要がある)。

4 数値例

4.1 シミュレーション

はじめにシミュレーション実験を行い、前節で導出したギブス・サンプリングがうまく機能するかどうかを検証する。そのために、以下のラプラス確率的フロンティアモデルからデータを発生させた：

$$y_i = \beta_1 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + v_i - u_i, \quad (i = 1, \dots, n)$$

$$x_{ij} \sim N(0, 1), \quad v_i \sim La(0, \sigma_v), \quad u_i \sim Exp(\sigma_u^{-1})$$

ここで、パラメータ値については、 $\beta_j = 1$ ($j = 1, \dots, k$)、 $\sigma_v = \sigma_u = 0.1$ とおき、データと説明変数の数については、 $n = 100, 300$ 、 $k = 2, 5, 10$ の場合を考えた。さらに、 β と σ_v の事前分布については、そのハイパー・パラメータを

$$\beta = \mathbf{0}, \quad \mathbf{B}_0 = 100\mathbf{I}, \quad n_0 = 1, \quad s_0 = 0.1$$

とし、 σ_u の事前分布については、van den Broeck *et al.* (1994) にしたがって、

$$\sigma_u \sim IG(1, -\log r_0), \quad r_0 = 0.875$$

とした。このような設定のもと、繰り返し数が 30,000 回のギブス・サンプリングを走らせ、モデルの推定を行った。なお、パラメータの推定値などを計算

する際には、最初の 10,000 個のサンプルは稼働検査期間 (burn-in period) として捨て、残りの 20,000 個のサンプルを用いている。

表 1 と表 2 には、各パラメータの推定結果が示されている。これらの表から、いずれのギブス・サンプリング法を用いても、事後平均は真の値に近い結果となっていることが分かる。また、標準偏差についても、2つのギブス・サンプリング法でほぼ同じ値となっている。サンプリングの効率性を比較するため、各表には非効率性因子 (*IF*) の値も示されている。非効率性因子の結果を比べると、ギブス・サンプリング I の方がギブス・サンプリング II よりもかなり効率的であることが分かる。さらに、定数項の効率性が他の回帰係数と比べて悪いことも分かる。サンプリングの効率性に関するこれらの結果は、図 2 に示されたパラメータの標本自己相関プロットからも確認することができる。ギブス・サンプリング II について見ると、説明変数の数が増えるにしたがい非効率となる傾向が見て取れる。さらに、データ数が増加するにつれ非効率となることも分かる。

このように、ギブス・サンプリング II の効率性がギブス・サンプリング I よりも悪くなるのは、回帰係数 β_j を個別にサンプリングしなければならず、他の回帰係数と相関を持つためであると考えられる。また、(7) 式から分かるように、 β_j の完全条件付き分布は切断分布となっている。そのため、説明変数やデータの数が増えると、パラメータがより狭い範囲内でしか移動できないことになり、このこともギブス・サンプリング II の効率性が悪くなる一因であると考えられる。

4.2 実データ

次に、Christensen and Greene (1976) で分析された米国電力会社のデータを用いて、前節で説明したギブス・サンプリングがうまく機能するかどうか検証する。ここで使用するデータは、これまでにも Greene (1990), van den Broeck *et al.* (1994), Koop *et al.* (1995) などにおいて分析されている。

先行研究にしたがい、確率的フロンティアモデルとして

$$-\log \frac{C}{P_F} = -\beta_1 - \beta_2 \log Q - \beta_3 (\log Q)^2 - \beta_4 \log \frac{P_L}{P_F} - \beta_5 \log \frac{P_K}{P_F} + v - u$$

表 1: シミュレーションデータの推定結果 ($n = 100$)

k	ギブス・サンプリング I				ギブス・サンプリング II		
		事後平均	事後標準偏差	IF	事後平均	事後標準偏差	IF
2	β_1	1.017	0.025	33.815	1.022	0.022	460.344
	β_2	0.998	0.021	14.587	1.001	0.020	90.988
	σ_u	0.112	0.020	29.436	0.110	0.020	142.585
	σ_v	0.127	0.034	39.526	0.132	0.032	276.572
5	β_1	1.004	0.034	54.468	1.007	0.029	598.534
	β_2	0.992	0.018	8.662	0.992	0.017	72.706
	β_3	0.998	0.019	11.152	0.997	0.020	64.818
	β_4	0.995	0.019	4.448	0.995	0.018	120.064
	β_5	0.991	0.017	12.524	0.992	0.016	143.955
	σ_u	0.112	0.020	29.197	0.111	0.019	192.582
	σ_v	0.106	0.038	55.852	0.109	0.033	404.343
	σ_u	0.106	0.038	55.852	0.109	0.033	404.343
10	β_1	1.006	0.030	35.773	0.999	0.026	781.836
	β_2	0.999	0.016	10.489	1.000	0.015	268.690
	β_3	0.971	0.018	6.927	0.974	0.018	87.592
	β_4	1.004	0.016	8.197	1.004	0.014	93.547
	β_5	0.997	0.018	7.730	1.002	0.018	402.423
	β_6	0.994	0.016	3.531	0.994	0.015	97.016
	β_7	1.016	0.017	1.909	1.013	0.017	258.013
	β_8	1.004	0.015	6.008	1.004	0.015	371.971
	β_9	0.978	0.015	4.249	0.975	0.016	289.292
	β_{10}	0.997	0.015	7.654	0.994	0.015	144.891
	σ_u	0.091	0.019	25.468	0.093	0.019	336.216
	σ_v	0.117	0.034	37.858	0.110	0.031	510.145

を考える。ここで、 C は企業の全費用、 Q は生産量、 P_L 、 P_K 、 P_F はそれぞれ労働、資本、燃料の一単位あたりの価格を表す（データの詳細については、Greene (1990) を参照）。また、パラメータの事前分布については、前項と同じ事前分布を用い、30,000 回のギブス・サンプリングを実行した（最初の 10,000 個のサンプルを稼働検査期間として捨て、残りの 20,000 個を推定などに用いた）。

表 3 には、パラメータの推定結果が示されている。また表には、企業の非効率性に関して、

$$VF = \frac{\text{Var}(u)}{\text{Var}(v) + \text{Var}(u)}$$

表 2: シミュレーションデータの推定結果 ($n = 300$)

k	ギブス・サンプリング I			ギブス・サンプリング II			
		事後平均	事後標準偏差	IF	事後平均	事後標準偏差	IF
2	β_1	0.997	0.016	53.807	0.993	0.014	994.453
	β_2	1.009	0.010	7.172	1.007	0.010	528.858
	σ_u	0.098	0.010	25.599	0.100	0.011	381.788
	σ_v	0.114	0.018	47.234	0.110	0.018	668.099
5	β_1	0.994	0.023	93.118	0.993	0.013	862.109
	β_2	0.992	0.009	5.817	0.994	0.009	97.477
	β_3	0.992	0.010	7.122	0.991	0.010	275.259
	β_4	0.997	0.009	16.985	0.998	0.009	191.796
	β_5	0.983	0.009	11.190	0.985	0.008	279.970
	σ_u	0.112	0.012	46.111	0.113	0.010	225.018
	σ_u	0.084	0.026	105.986	0.084	0.017	570.016
10	β_1	1.006	0.015	40.114	1.005	0.009	563.153
	β_2	1.012	0.008	11.482	1.011	0.008	612.759
	β_3	0.996	0.009	10.952	0.994	0.009	753.699
	β_4	1.002	0.008	10.129	1.001	0.008	689.078
	β_5	1.007	0.009	8.900	1.010	0.008	571.756
	β_6	1.004	0.008	11.688	1.005	0.008	567.865
	β_7	0.999	0.009	11.725	0.996	0.009	568.825
	β_8	1.013	0.008	10.651	1.012	0.008	387.041
	β_9	1.004	0.009	11.556	1.004	0.008	222.331
	β_{10}	1.004	0.008	8.796	1.003	0.007	481.374
	σ_u	0.079	0.010	31.041	0.079	0.008	130.004
	σ_u	0.112	0.017	40.131	0.112	0.012	262.795

と技術的効率性の予測値 TE_f の結果も示されている。ここで、 u_f を非負の誤差項の予測値とすれば、 $TE_f = \exp(-u_f)$ で与えられ、 u_f の計算はその予測分布

$$\pi(u_f|\mathbf{y}) = \int \pi(u_f|\sigma_u)\pi(\sigma_u|\mathbf{y})d\sigma_u$$

にもとづいて行っている。

シミュレーション実験と同じように、どちらのギブス・サンプリング法を用いても、パラメータの事後平均はほぼ同じ値となっている。また、ギブス・サンプリング II の非効率性因子の値が非常に大きく、また図 3 に示された標本自己相関プロットから、ここでもギブス・サンプリング I の方がギブス・サン

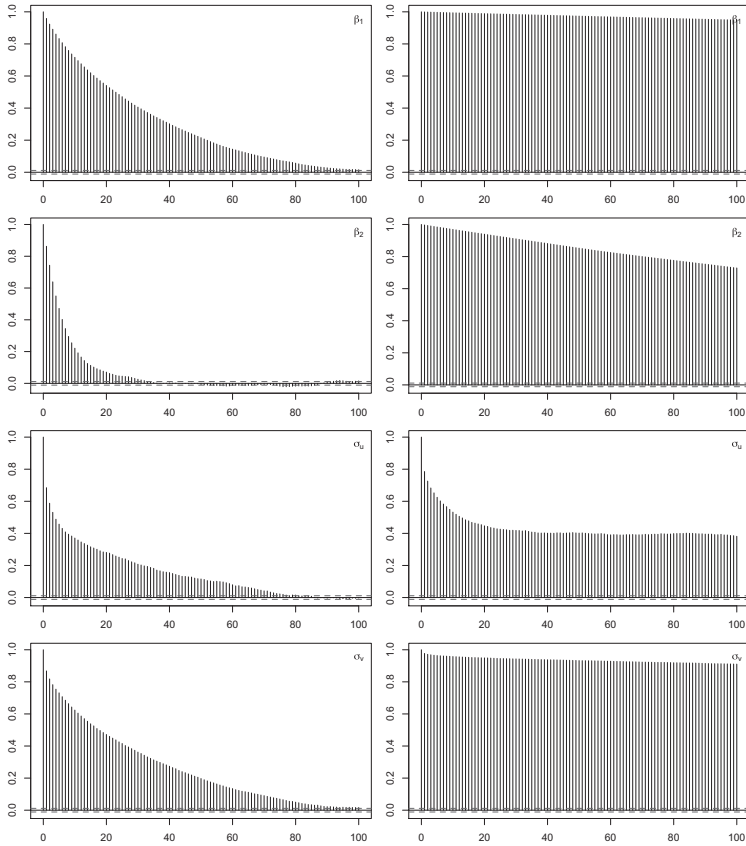


図 2: 標本自己相関プロット (シミュレーションデータ): ギブス・サンプリング I の結果 (左) とギブス・サンプリング II の結果 (右)

プリング II よりも効率的であることが分かる。一方、事後標準偏差や 95%信用区間については、2つのギブス・サンプリング法でいくつか違いが見られる。例えば、ギブス・サンプリング II から計算される生産量 ($\log Q$) の事後標準偏差は、ギブス・サンプリング I のそれよりも小さい値となっている。また、資本価格 ($\log P_k/P_F$) については、ギブス・サンプリング I では 95%信用区間がゼロを含むが、ギブス・サンプリング II ではゼロを含んでいない。この

表 3: 電力会社データの推定結果

ギブス・サンプリング I				
	事後平均	事後標準偏差	95%信用区間	<i>IF</i>
定数項	-7.774	0.366	[-8.462, -7.030]	6.339
$\log Q$	0.445	0.041	[0.359, 0.522]	13.938
$(\log Q)^2$	0.028	0.003	[0.023, 0.034]	13.978
$\log \frac{P_L}{P_F}$	0.291	0.072	[0.150, 0.431]	8.514
$\log \frac{P_K}{P_F}$	0.041	0.060	[-0.077, 0.160]	12.250
σ_v	0.084	0.015	[0.057, 0.115]	30.402
σ_u	0.092	0.029	[0.032, 0.146]	43.782
<i>VF</i>	0.382	0.194	[0.044, 0.739]	42.334
<i>TE_f</i>	0.915	0.084	[0.690, 0.998]	5.272
ギブス・サンプリング II				
	事後平均	事後標準偏差	95%信用区間	<i>IF</i>
定数項	-7.497	0.049	[-7.591, -7.414]	1366.08
$\log Q$	0.416	0.010	[0.397, 0.431]	1872.32
$(\log Q)^2$	0.030	0.001	[0.029, 0.031]	1386.70
$\log \frac{P_L}{P_F}$	0.261	0.014	[0.236, 0.281]	1902.61
$\log \frac{P_K}{P_F}$	0.070	0.032	[0.019, 0.143]	1033.51
σ_v	0.089	0.014	[0.061, 0.116]	348.17
σ_u	0.082	0.026	[0.035, 0.139]	588.12
<i>VF</i>	0.311	0.171	[0.050, 0.697]	592.09
<i>TE_f</i>	0.924	0.077	[0.715, 0.998]	57.38

ような違いは、非効率性因子の結果と合わせて考えると、ギブス・サンプリング II ではパラメータ空間全体からサンプリングできておらず、繰り返し数が十分でない可能性を示唆している。

5 結論

本論文では、Nguyen (2010) と Horrace and Parmeter (2015) によって提案されたラプラス確率的フロンティアモデルを考え、ラプラス分布に対する混合表現を用いてモデルを再定式化し、MCMC 法の 1 つであるギブス・サンプリングによって簡単に推定できることを示した。本論文では、2 つのギブス・サンプリング法を提示したが、シミュレーション実験と実際のデータを使った分析を通して、混合表現 I から導かれるギブス・サンプリングの方が、混合表

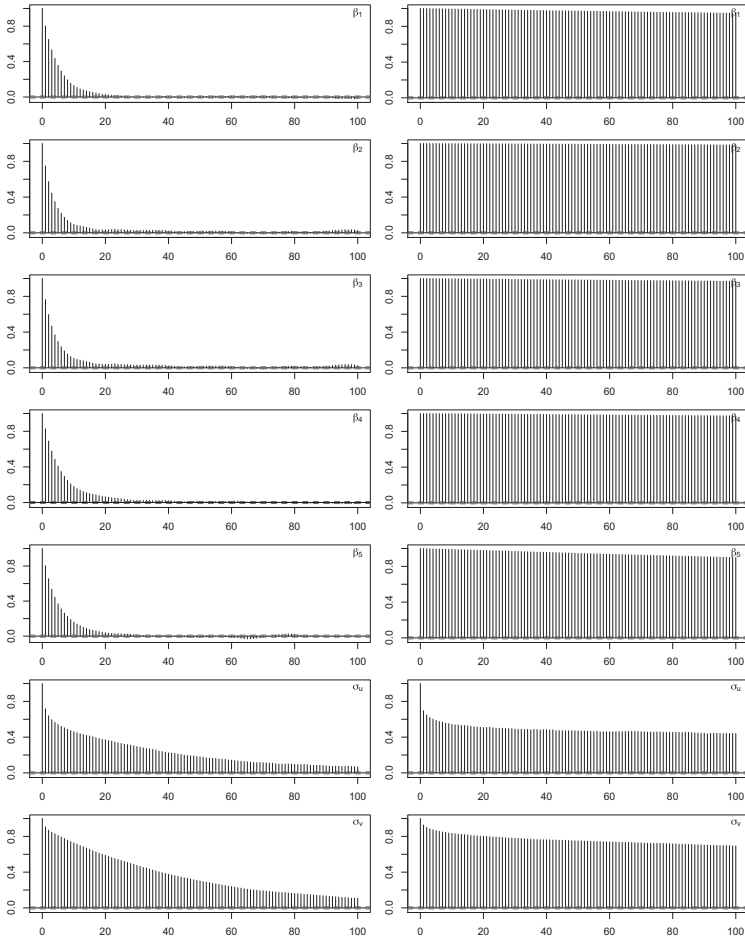


図 3: 標本自己相関プロット (電力会社データ): ギブス・サンプリング I の結果 (左) とギブス・サンプリング II の結果 (右)

現 II から得られるギブス・サンプリングよりも効率的であることが分かった。これは、混合表現 II を用いた場合、回帰係数を個別にサンプリングしなければならず、他の回帰係数と相関を持つためであり、また回帰係数が切断分布か

らサンプリングされているためである。

サンプリングの効率性から見た場合、一様分布とガンマ分布にもとづく混合表現 II は正規分布と指数分布による混合表現 I よりも劣っている。しかしながら、混合表現 II にもいくつか利点がある。例えば、パネルデータの確率的フロンティアモデルでは、不均一分散を考えることが多い。混合表現 II を用いれば、分散構造に含まれるパラメータをギブス・サンプリングによってサンプリングすることが可能である (Damien *et al.* (2000))。また、(7) 式の β_j の完全条件付き分布から分かるように、 β の事前分布からサンプリングできれば、回帰係数を完全条件付き分布からサンプリングすることができる。これにより、例えば縮小型推定で用いられている事前分布を用いても、回帰係数を容易にサンプリングすることができる。したがって、混合表現 II を用いたときのサンプリングの効率性を改善するアルゴリズムを開発することが必要であり、この点については今後の課題としたい。

参考文献

- [1] Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977). "Formulation and estimation of stochastic frontier production functions," *Journal of Econometrics* **6**, 21–37.
- [2] Beckers, D. and Hammond, C. (1987). "A tractable likelihood function for the normal-gamma stochastic frontier model," *Economics Letters* **24**, 33–38.
- [3] Christensen, L. R. and Greene, W. H. (1976). "Economies of scale in U.S. electric power generation," *Journal of Political Economy* **84**, 655–676.
- [4] Damien, P., Qin, Z. S., and Walker, S. G. (2000). "Uniform scale mixture models with applications to variance regression," University of Michigan Working Paper, No. 00-014.
- [5] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society* **B39**, 1–38.

- [6] Gelfand, A. and Smith, A. F. M. (1990). “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association* **85**, 398–409.
- [7] Greene, W. H. (1990). “A gamma distributed stochastic frontier model,” *Journal of Econometrics* **46**, 141–164.
- [8] Horrace, W. C. and Parmeter, C. F. (2015). “A Laplace stochastic frontier model,” *Econometric Reviews*, *forthcoming*.
- [9] Koop, G., Steel, M. F. J., and Osiewalski, J. (1995). “Posterior analysis of stochastic frontier models using Gibbs sampling,” *Computational Statistics* **10**, 353–373.
- [10] Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston.
- [11] Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, New York.
- [12] Mallick, H. and Yi, N. (2014). “A new Bayesian lasso,” *Statistics and Its Interface* **7**, 571–582.
- [13] Meeusen, W. and van den Broeck, J. (1977). “Efficiency estimation from Cobb-Douglas production functions with composed error,” *International Economic Review* **18**, 435–444.
- [14] Nguyen, N. B. (2010). *Estimation of technical efficiency in stochastic frontier analysis*, PhD thesis, Bowling Green State University.
- [15] Park, T. and Casella, G. (2008). “The Bayesian lasso,” *Journal of the American Statistical Association* **103**, 681–686.
- [16] Parmeter, C. F. and Kumbhakar, S. C. (2014). “Efficiency analysis: A primer on recent advances,” *Foundations and Trends in Econometrics* **7**, 191–385.
- [17] Stevenson, R. (1980). “Likelihood functions for generalized stochastic frontier estimation,” *Journal of Econometrics* **13**, 58–66.
- [18] Storn, R. and Price, K. (1997). “Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces,” *Journal of Global Optimization* **11**, 341–359.

- [19] Tsionas, E. G. (2007). “Efficiency measurement with the Weibull stochastic frontier,” *Oxford Bulletin of Economics and Statistics* **69**, 693–706.
- [20] van den Broeck, J., Koop, G., Osiewalski, J., and Steel, M. F. J. (1994). “Stochastic frontier models: A Bayesian perspective,” *Journal of Econometrics* **61**, 273–303.