

Title	マルコフ連鎖モンテカルロ法によるベイズ分析：回帰モデルへの応用
Sub Title	Bayesian analysis via Markov chain Monte Carlo : applications to regression models
Author	中妻, 照雄(Nakatsuma, Teruo)
Publisher	慶應義塾経済学会
Publication year	2002
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.94, No.4 (2002. 1) ,p.745(181)- 771(207)
JaLC DOI	10.14991/001.20020101-0181
Abstract	
Notes	論説
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-20020101-0181

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

マルコフ連鎖モンテカルロ法によるベイズ分析

—— 回帰モデルへの応用 ——

中 妻 照 雄

1 はじめに

ベイズ統計学（Bayesian statistics）は推定すべき確率モデルの未知のパラメータを確率変数であるとみなし、データが観測された下での条件付分布（事後分布）を使ってパラメータの推定やモデルに関する仮説の検証などを行う統計学である。ベイズ統計学は古典的統計学と違って、研究者が持つ確率モデルやパラメータに関する事前情報を事前分布の形でモデルの推定、仮説の検証、モデルによる予測などに活用できる特徴を持っている。例えば将来の日銀の金融政策に関する専門家の意見を加味しつつ将来の金利予測を行うといったことが容易にできる。

さらにベイズ統計学の枠組みでは、推定されたモデルを使って意思決定を行う際に、確率モデルが想定する不確実性に加えてモデルの推定に伴う不確実性も考慮しつつ意思決定を行うことができる。モデルのパラメータの値はあくまでも推定されたものであって、真の値ではない。さらに想定しているモデル自体が現実正しいのかもさえない。したがって現実の問題解決に当たっては、これらのモデルに関する不確実性も考慮しつつモデルを予測等に利用するのが妥当であろう。例えば投資家はポートフォリアを組む際に資産価値を決定するモデルが想定する資産保有のリスクを考慮して投資計画を立てる。ベイズ統計学の枠組みでは、この資産価値を決定するモデル自体に関する不確実性も投資に伴うリスクの一種として投資計画に反映させることを整合性の取れた自然な形で行うことができる。

このようにベイズ統計学は統計分析の手法として魅力的なものであるが、経済の実証分析の分野では、古典的統計学によるものに比べて実際に応用されることはそれほど多くはない。一般的な統計学や計量経済学の教科書が古典的統計学に偏りすぎた内容になっており、ベイズ統計学があまり広く知られていないという理由もあろう。しかし、現実には過去のある時期において実証分析を行うとする研究者にとってベイズ統計学の手法はあまり使い勝手が良いものではなかったという事実が、その普及の大きな妨げになっていたことは否定できない。ベイズ統計学ではパラメータの推定

等が解析的に求まるモデルは限られており、多くのモデルでは推定を行う際に複雑な数値計算をかなりの時間をかけて行う必要がある。このため現実の実証分析に使うには時間と手間がかかり過ぎるという問題がベイズ統計学にはあった。この問題を解決したのがマルコフ連鎖モンテカルロ (Markov chain Monte Carlo, MCMC) 法である。MCMC 法はマルコフ連鎖サンプリング (Markov chain sampling) という任意の確率分布から乱数を発生させる方法を使って事後分布からパラメータの乱数を発生し、それを使ってモンテカルロ積分を行う手法の総称である。マルコフ連鎖サンプリングによるモンテカルロ積分法なので「マルコフ連鎖モンテカルロ法」の名前がついた。MCMC 法の特徴は複雑なモデルに対しても適応可能であり、その実行もパソコンレベルのコンピュータがあれば手軽に行うことができることである。このためベイズ統計学の研究者の間で1990年代以降爆発的に普及し、今ではベイズ統計学の計算手法のデファクト・スタンダードとでも言うべき存在となった。

本稿では、経済の実証分析で広く使われる回帰モデルの例にして MCMC 法によるベイズ分析を説明する。特に、最尤法などでは推定の難しい多値選択モデルや複雑な制限従属変数モデルを重点的に説明したい。2 節ではベイズ統計学における分析の基本的な枠組みを説明する。3 節ではモンテカルロ積分法とマルコフ連鎖サンプリングの代表格であるギブズ・サンプラー、メトロポリス・ヘイスティングズ・アルゴリズム、データ拡大法を解説する。4 節と 5 節では多値選択モデルと制限従属変数モデルへの MCMC 法の応用を概観する。最後に、6 節では興味のある読者がさらに MCMC 法について学ぶための文献を幾つか紹介する。

2 ベイズ分析の基礎

2.1 事前分布、尤度、事後分布

ベイズ統計学では確率モデルの未知のパラメータを確率変数として扱う。ここでパラメータ・ベクトルを θ とし、パラメータ空間を Θ とする。分析に使うデータ $\mathbf{y}=(y_1, \dots, y_n)$ が与えられたとすると、パラメータの条件付分布 $p(\theta|\mathbf{y})$ はベイズの定理を適用すると、

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int_{\Theta} p(\mathbf{y}|\theta)p(\theta)d\theta} = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \quad (1)$$

となる。 $p(\mathbf{y})$ は未知のパラメータ θ に依存しないで、通常は

(1) θ が連続な確率変数である場合には $p(\theta|\mathbf{y})$ は確率密度関数となるので、それを「分布」と呼ぶのは不正確である。しかし、本稿では文脈から判断できる場合には「密度」と「分布」を明確に区別せず「分布」で統一する。

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

と表記される。ここで $p(\mathbf{y}|\boldsymbol{\theta})$ は尤度 (likelihood) と呼ばれ、 $p(\boldsymbol{\theta})$ は事前分布 (prior distribution) と呼ばれる。パラメータの条件付分布 $p(\boldsymbol{\theta}|\mathbf{y})$ は事後分布 (posterior distribution) と呼ばれ、ベイズ統計学では $p(\boldsymbol{\theta}|\mathbf{y})$ に基づいてパラメータに関する統計分析を進めることになる。

ここでベイズ統計学における事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ の意味を説明しよう。尤度 $p(\mathbf{y}|\boldsymbol{\theta})$ はパラメータ $\boldsymbol{\theta}$ の値が与えられた下で観測されたデータ $\mathbf{y}$ が実現する条件付確率密度である。この尤度は観測されたデータが持っているパラメータに関する情報を表したものであると解釈される。一方、事前分布 $p(\boldsymbol{\theta})$ はパラメータに関する事前情報を表していると解釈される。事前情報とはデータが観測される前に研究者が持っているパラメータに関する情報であり、通常は確率分布⁽²⁾の形で表現される。つまり、事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ はベイズの定理によってパラメータに関する事前情報 (事前分布) をデータからの情報 (尤度) で更新して得られたものであると解釈される。

事前分布の例を考えてみよう。ある研究者が、あるパラメータが 10 ± 3 ぐらいの値をとるという事前情報を持っているとする。このとき事前分布として区間 $(7, 13)$ 上の一様分布を使うことができよう。この場合の事前分布は、 $(7, 13)$ 内の全ての値が同じ密度を持つことから、区間内の全ての値がパラメータの値として同じように相応しいという情報を暗に含んでいることに注意しよう。無論、これが唯一の事前分布というわけではない。正規分布では平均から標準偏差の3倍以上乖離する確率がゼロに近いことから、 $\mathcal{N}(10, 1)$ という事前分布を考えることもできる。この場合には10に近い値が分布の裾の方の値と比べてパラメータの値として相応しいという事前情報を暗に含むことになる。

先の例に限らず一般に如何なる事前分布が望ましいかを万人が納得する形では決めることは難しい⁽³⁾。そのため事前分布の選択は研究者の主観的な判断に任されることになる。しかし、このことはベイズ統計学が使えない分析手法であることを意味しない。(1) の関係式を繰り返し適用すると、後で更に新しいデータが観測されたときに事後分布を次の分析の事前分布として使うことができる。そして、この「事前分布→事後分布→次の分析の事前分布」の更新作業を新しいデータが観測される度に次々と繰り返していくことができる。よってベイズ統計学のアプローチを新しく観測されたデータを基いて事後分布を次々と更新していく過程として解釈できるのである。この観点からす

(2) 実は $p(\boldsymbol{\theta})$ が必ずしも確率密度である (積分して1になる) 必要はない。確率密度でない $p(\boldsymbol{\theta})$ は変則的事前分布 (improper prior) と呼ばれる。

(3) 事前分布の選択に「客観性」を持たせる一つの考え方として、できるだけパラメータに関する事前情報が乏しい状態から分析を始めることで事前情報による影響を極力減らし、「データをして語らしめる」ことを目指そうとするアプローチがある。このような事前分布については Box and Tiao [7] や Robert [20] などで詳しく解説されている。

ると、ベイズ統計学は科学的分析の手順として妥当なものと思われる。むしろ研究者の持つ事前情報を分析に積極的に活用できる点でベイズ統計学は有用な統計分析の枠組みの一つであるといえよう。

事後分布の例として回帰モデルにおける事後分布を導出しよう。古典的正規回帰モデル

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (2)$$

を考える。ここで標本のサイズは n で説明変数の数は k ($n > k$) とする。 $N_n(\cdot)$ は n 変量正規分布を意味し、 $\mathbf{I}_n$ は n 次元の単位行列である。このとき尤度は

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}) &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\ &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\{\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}\right], \end{aligned} \quad (3)$$

$$\hat{\boldsymbol{\beta}} \equiv (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}, \quad (4)$$

$$s^2 \equiv \frac{1}{n-k} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad (5)$$

として与えられる。ここで $\nu \equiv n - k$ である。パラメータ $(\boldsymbol{\beta}, \sigma^2)$ の事前密度を

$$p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}, \quad (6)$$

にすると、事後分布は

$$p(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2}\{\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}\right], \quad (7)$$

として与えられる。

パラメータが m 個あるときには $p(\boldsymbol{\theta}|\mathbf{y})$ は m 個の確率変数の同時確率密度になる。実際の実証分析において全てのパラメータが分析対象になるわけではない。例えば、回帰分析では攪乱項の分散 σ^2 は分析対象でない場合が多い。このとき分析対象外のパラメータの影響を除いて分析対象となるパラメータのみの事後分布を構築することができる。ここでパラメータ $\boldsymbol{\theta}$ が $\boldsymbol{\theta} = [\boldsymbol{\theta}_1; \boldsymbol{\theta}_2]$ と関心のあるパラメータ $\boldsymbol{\theta}_1$ と関心のないパラメータ $\boldsymbol{\theta}_2$ に分けられるとする。 $\boldsymbol{\theta}_2$ のパラメータ空間を Θ_2 として $p(\boldsymbol{\theta}|\mathbf{y}) = p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})$ を $\boldsymbol{\theta}_2$ について積分すると、 $\boldsymbol{\theta}_1$ の事後分布は

$$p(\boldsymbol{\theta}_1|\mathbf{y}) = \int_{\Theta_2} p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}) d\boldsymbol{\theta}_2, \quad (8)$$

として与えられる。特に $\boldsymbol{\theta}_1$ が単一のパラメータである場合には、 $p(\boldsymbol{\theta}_1|\mathbf{y})$ は $\boldsymbol{\theta}_1$ の周辺事後分布 (marginal posterior distribution) と呼ばれる。古典的正規回帰モデルの例では、回帰係数 $\boldsymbol{\beta}$ の事後

分布は自由度 ν の k 変量 t 分布

$$p(\boldsymbol{\beta}|\mathbf{y}) \propto \left[1 + \frac{1}{\nu s^2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right]^{-(\nu+k)/2}, \quad (9)$$

になることが知られている (Zellner [23] などを参照)。これを $t_k[\hat{\boldsymbol{\beta}}, s^2(\mathbf{X}'\mathbf{X})^{-1}, \nu]$ と表記する。このように対象外のパラメータを積分して事後分布を分析対象となるパラメータに集約することは、ベイズ分析ではよく行われる作業である。

2.2 事後分布による分析

次に、事後分布を用いたパラメータに関するベイズ分析の解説をしよう。ベイズ分析におけるパラメータの点推定にはパラメータの事後分布の中心の特性値を使う。この中心の特性値としてよく使われるのが事後分布の平均 $E(\boldsymbol{\theta}|\mathbf{y})$ である。例えば、 $\boldsymbol{\theta}$ の第 j 番目のパラメータを θ_j ($j=1, \dots, m$) とし、その周辺事後分布を $p(\theta_j|\mathbf{y})$ とすると、 θ_j の周辺事後分布の平均は

$$E(\theta_j|\mathbf{y}) = \int_{\theta_j} \theta_j p(\theta_j|\mathbf{y}) d\theta_j, \quad (10)$$

として与えられる。これを θ_j の点推定として使う。パラメータの周辺事後分布の平均は事後平均と呼ばれる。事後平均以外にも事後分布の中央値やモードを点推定として使うこともある。⁽⁴⁾ 古典的正規回帰モデルの場合、回帰係数 β の事後分布は $t_k[\hat{\boldsymbol{\beta}}, s^2(\mathbf{X}'\mathbf{X})^{-1}, \nu]$ であるから、平均、中央値、モードは全て $\hat{\boldsymbol{\beta}}$ である。つまり、最小二乗推定値が回帰係数の点推定となっている。もちろん古典的統計学における最小二乗推定量とは解釈が異なるのは言うまでもない。

一方、ベイズ分析における区間推定には事後分布のパーセント点 (例えば2.5%点と97.5%点) を使う。これを信用区間 (credible interval) という。古典的正規回帰モデルの例では、個々の回帰係数 β_j は自由度 ν の t 分布に従うので、 $100 \times (1-\alpha) \%$ 信用区間は古典的統計学における $100 \times (1-\alpha) \%$ 信頼区間と数値上は同値になる。つまり、 $\hat{\beta}_j - t_{\alpha/2} \text{SE}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2} \text{SE}(\hat{\beta}_j)$ である。ここで $t_{\alpha/2}$ は $T \sim t(\nu)$ に対して $\Pr\{T > t_{\alpha/2}\} = \alpha/2$ となる値である。 $\text{SE}(\hat{\beta}_j)$ は古典的統計学において最小二乗推定量 $\hat{\beta}_j$ の標準誤差と呼ばれるものである。

また、事後分布を使ってパラメータがある範囲の値をとる確率を評価することもできる。例えば、パラメータ θ_j が正の値をとる確率は、

(4) もし事前密度が $p(\boldsymbol{\theta}) \propto \text{constant}$ であるならば、 $p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})$ となる。このとき事後分布のモードは尤度を最大にするパラメータの値に一致する。よって、古典的統計学における最尤推定量は、ベイズ統計学において事後分布のモードをパラメータの点推定に使うことと数値上は同値になる。しかし、両者の解釈は全く違うことに注意しよう。

$$\Pr\{\theta_j > 0 | \mathbf{y}\} = \int_0^\infty p(\theta_j | \mathbf{y}) d\theta_j, \quad (11)$$

である。 $\Pr\{\theta_j > 0 | \mathbf{y}\}$ のように事後分布で評価した事象の確率を事後確率という。もし事後確率 $\Pr\{\theta_j > 0 | \mathbf{y}\}$ が 1 に近い値であれば、 θ_j が正の値をとるといふ仮説はデータから支持されるといえるだろう。このように事後分布を用いてパラメータに関する仮説の検証もできるのである。

さらに、事後分布を使って予測を行うこともできる。新しい観測値 $\tilde{y}$ とパラメータの $\boldsymbol{\theta}$ の同時事後分布 $p(\tilde{y}, \boldsymbol{\theta} | \mathbf{y})$ は

$$p(\tilde{y}, \boldsymbol{\theta} | \mathbf{y}) = p(\tilde{y} | \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} | \mathbf{y}), \quad (12)$$

となる。 $p(\tilde{y} | \boldsymbol{\theta}, \mathbf{y})$ は $\mathbf{y}$ と $\boldsymbol{\theta}$ が与えられた下での $\tilde{y}$ の条件付分布であり、データを生成した確率モデルそのものをさす。このとき $\tilde{y}$ の予測分布 (predictive distribution) は

$$p(\tilde{y} | \mathbf{y}) = \int_{\boldsymbol{\theta}} p(\tilde{y}, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} p(\tilde{y} | \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad (13)$$

として与えられる。予測分布は特に時系列モデルで予測を行う際に有用である。例えば $\{y_t\}_{t=1}^T$ というデータがあったとすると、 y_{T+1} の予測分布は

$$p(y_{T+1} | y_1, \dots, y_T) = \int_{\boldsymbol{\theta}} p(y_{T+1} | \boldsymbol{\theta}, y_1, \dots, y_T) p(\boldsymbol{\theta} | y_1, \dots, y_T) d\boldsymbol{\theta},$$

として与えられる。

ベイズ分析についてはまだまだ説明すべきことがあるが、多くの教科書 (Berger [5], Bernardo and Smith [6], Box and Tiao [7], Robert [20], Zellner [23] など) で詳しく解説されるので、ここでは省略する。

3 マルコフ連鎖モンテカルロ法

3.1 モンテカルロ法

前節で紹介したように、ベイズ分析では周辺事後分布の導出や事後平均の評価などで事後分布 $p(\boldsymbol{\theta} | \mathbf{y})$ の積分を評価することが多い。幾つかの比較的簡単な確率モデルでは積分公式によって値が求まることもあるが、そのような例は限られている。よって、一般には数値積分を用いることになる。

数値積分には様々な方法があるが、ここではモンテカルロ (Monte Carlo) 法による積分を考える。ベイズ分析で考える積分の多くは

$$E[f(\boldsymbol{\theta})] = \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad (14)$$

の形をしていることに注意しよう。例えば、 θ_j の事後平均であれば $f(\boldsymbol{\theta}) = \theta_j$ であり、 $\Pr\{a < \theta_j \leq b\}$ を求めるのであれば $f(\boldsymbol{\theta}) = \mathbf{1}_{(a,b)}(\theta_j)$ である。ここで $\mathbf{1}_A(x)$ は $x \in A$ が真であるときに 1、そうでなければ 0 をとる関数である。 $E[f(\boldsymbol{\theta})]$ をモンテカルロ法で求めるには、事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ から $\boldsymbol{\theta}$ の乱数 $\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(R)}\}$ を発生させて、

$$\hat{E}[f(\boldsymbol{\theta})] = \frac{1}{R} \sum_{r=1}^R f(\boldsymbol{\theta}^{(r)}), \quad (15)$$

を計算する。大数の法則により、 $R \rightarrow \infty$ で

$$\hat{E}[f(\boldsymbol{\theta})] \xrightarrow{a.s.} E[f(\boldsymbol{\theta})], \quad (16)$$

が示されるので、 R を十分大きくとれば $\hat{E}[f(\boldsymbol{\theta})]$ で $E[f(\boldsymbol{\theta})]$ を近似できることがわかる。

事後分布からの $\boldsymbol{\theta}$ の乱数を容易に発生させることができる場合には、モンテカルロ法は手軽な数値積分法として便利である。特に $\boldsymbol{\theta}$ の次元が高い場合には、現時点ではモンテカルロ法が唯一の実行可能な数値積分法といっても差し支えない。しかし、実際に使われる多くの確率モデルでは事後分布から $\boldsymbol{\theta}$ をうまく発生させることができない。このような場合に適用される乱数の発生法が以下に説明するギブズ・サンプラー (Gibbs sampler) とメトロポリス・ヘイスティングズ (Metropolis-Hastings, MH) アルゴリズムに代表されるマルコフ連鎖サンプリング法である。マルコフ連鎖モンテカルロ (MCMC) 法という名称は、マルコフ連鎖サンプリング法によるモンテカルロ法という意味である。

3.2 ギブズ・サンプラー

まず、ギブズ・サンプラーの解説から始めよう。ギブズ・サンプラーは Geman and Geman [11] により画像処理のための統計的手法として開発された。もっともギブズ・サンプラーのアイデアは Geman and Geman [11] の前から様々な分野で考えられてきた。(大森 [1] などを参照。)

ギブズ・サンプラーによって事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ から $\boldsymbol{\theta}$ の乱数を発生させる方法を説明しよう。パラメータ・ベクトル $\boldsymbol{\theta} = [\theta_1, \dots, \theta_m]'$ から θ_j を除いたものを $\boldsymbol{\theta}_{-j} = [\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m]'$ と表記しよう。ここで事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ から $\boldsymbol{\theta}$ の乱数を同時に発生させることはできないが、全ての j ($j = 1, \dots, m$) に対して条件付事後分布 $p(\theta_j|\boldsymbol{\theta}_{-j}, \mathbf{y})$ から乱数を容易に発生できると仮定する。このときギブズ・サンプラーを使うと $p(\boldsymbol{\theta}|\mathbf{y})$ から乱数を発生させることが可能となる。ギブズ・サンプラーは以下の手順で実行される。

1. 初期値 $(\theta_1^{(0)}, \dots, \theta_m^{(0)})$ を決める。
2. $\theta_1^{(r+1)}$ を $p(\theta_1|\theta_2^{(r)}, \dots, \theta_m^{(r)}, \mathbf{y})$ から発生させる。

3. $\theta_2^{(r+1)}$ を $p(\theta_2|\theta_1^{(r+1)}, \theta_3^{(r)}, \dots, \theta_m^{(r)}, \mathbf{y})$ から発生させる。

⋮

4. $\theta_m^{(r+1)}$ を $p(\theta_m|\theta_1^{(r+1)}, \dots, \theta_{m-1}^{(r+1)}, \mathbf{y})$ から発生させる。

5. 以上を繰り返す。

ギブズ・サンプラーを十分な回数繰り返せば、 $\boldsymbol{\theta}^{(r)} = [\theta_1^{(r)}, \dots, \theta_m^{(r)}]'$ の分布が $p(\boldsymbol{\theta}|\mathbf{y})$ に収束することが証明されている。しかし、ギブズ・サンプラーで発生させた乱数が事後分布に収束するまでには、ある程度の回数ギブズ・サンプラーを繰り返す必要がある。そのため最初の数百回から数千回のサンプリングで出た乱数は捨てて（これを“burn-in”という）、その後のサンプリングによる乱数を実際にモンテカルロ積分に利用する乱数 $\{\boldsymbol{\theta}^{(r)}\}_{r=1}^R$ とする方法がよく使われる。ギブズ・サンプラーの収束に関する解説は Robert and Casella [21] などを参照。

一般にギブズ・サンプラーの中で発生させる順序はどんなものでもかまわない。また、発生させる順序を途中でランダムに変更することもできる。さらに、別に確率変数を一つずつ発生させなくてもよい。例えば、 $\boldsymbol{\theta} = [\boldsymbol{\theta}_1; \boldsymbol{\theta}_2]$ の事後分布 $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})$ で $\boldsymbol{\theta}_1$ と $\boldsymbol{\theta}_2$ をそれぞれ $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{y})$ と $p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{y})$ から容易に発生できるのであれば、

1. $\boldsymbol{\theta}_1^{(r+1)}$ を $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(r)}, \mathbf{y})$ から発生させる。

2. $\boldsymbol{\theta}_2^{(r+1)}$ を $p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(r+1)}, \mathbf{y})$ から発生させる。

というギブズ・サンプラーを実行することも可能である。これで発生させた $\boldsymbol{\theta}^{(r)} = [\boldsymbol{\theta}_1^{(r)}; \boldsymbol{\theta}_2^{(r)}]$ の分布は事後分布に収束する。このように幾つかのパラメータををまとめて発生させることをブロック化 (blocking) という。うまくブロック化をすればギブズ・サンプラーで乱数を効率的に発生させることが知られている。

$\boldsymbol{\beta}$ の σ^2 が与えられた下での条件付事後分布は

$$p(\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X}) \propto \sigma^{-k} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right], \quad (17)$$

となる。これは k 変量正規分布 $\mathcal{N}_k[\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}' \mathbf{X})^{-1}]$ である。一方、 σ^2 の $\boldsymbol{\beta}$ が与えられた下での条件付事後分布は

$$p(\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(n/2+1)} \exp\left(-\frac{n\hat{\sigma}^2}{2\sigma^2}\right), \quad (18)$$

$$\hat{\sigma}^2 \equiv \frac{1}{n}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (19)$$

である。これは逆ガンマ分布 $IG(n/2, n\hat{\sigma}^2/2)$ になっている。

このとき $(\boldsymbol{\beta}, \sigma^2)$ を事後分布から発生させるギブズ・サンプラーは以下のように与えられる。

1. $\boldsymbol{\beta}^{(r+1)}$ を $\mathcal{N}_k[\hat{\boldsymbol{\beta}}, \sigma^{2(r)}2(\mathbf{X}'\mathbf{X})^{-1}]$ から発生させる。
2. $\sigma^{2(r+1)}$ を $IG(n/2, n\hat{\sigma}^{2(r+1)}/2)$ から発生させる。

ここで $\hat{\sigma}^{2(r)} = n^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(r)})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(r)})$ である。このギブズ・サンプラーでは回帰係数 $\boldsymbol{\beta} = [\beta_1, \dots, \beta_k]'$ は一つのブロックとして扱われている。わざわざ $\beta_1, \dots, \beta_k$ を一つ一つ正規分布から発生させることは可能であるが、計算上非効率的であるしギブズ・サンプラーの収束も遅くなる傾向があるので望ましい方法ではない。

古典的正規回帰モデルでは事後分布の平均などの公式がわかっているのでギブズ・サンプラーを使うメリットはない。しかし、SUR (seemingly unrelated regression) モデルのベイズ分析ではギブズ・サンプラーが威力を発揮する。SUR モデルは m 本の方程式からなる体系である。

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{bmatrix}. \quad (20)$$

これを

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

と表記する。SUR モデルでは攪乱項 $\boldsymbol{\epsilon}$ の平均と共分散行列は

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Sigma} \otimes \mathbf{I}_n,$$

と仮定される。したがって、尤度は

$$p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right], \quad (21)$$

である。

$$\mathbf{A} = \begin{bmatrix} (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1) & \cdots & (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)'(\mathbf{y}_m - \mathbf{X}_m\boldsymbol{\beta}_m) \\ \vdots & \ddots & \vdots \\ (\mathbf{y}_m - \mathbf{X}_m\boldsymbol{\beta}_m)'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1) & \cdots & (\mathbf{y}_m - \mathbf{X}_m\boldsymbol{\beta}_m)'(\mathbf{y}_m - \mathbf{X}_m\boldsymbol{\beta}_m) \end{bmatrix}, \quad (22)$$

と定義すると、尤度は

$$p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{A})\right], \quad (23)$$

と書き直される。事前分布を

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(m+1)/2}, \quad (24)$$

とすると、事後分布は

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{X}) \propto |\boldsymbol{\Sigma}|^{-(n+m+1)/2} \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right], \quad (25)$$

$$\propto |\boldsymbol{\Sigma}|^{-(n+m+1)/2} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{A}) \right], \quad (26)$$

となる。SUR モデルで $\boldsymbol{\beta}$ の周辺事後分布を解析的に求めることはできない。代わりに、ギブズ・サンプラーによって $\boldsymbol{\beta}$ の事後平均などを評価する。Zellner [23] によると、 $\boldsymbol{\beta}$ と $\boldsymbol{\Sigma}$ の条件付事後分布は、

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}_{km}(\hat{\boldsymbol{\beta}}, \boldsymbol{\Omega}), \quad (27)$$

$$\boldsymbol{\Sigma} | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim IW(\mathbf{A}, n, m), \quad (28)$$

として与えられる。ここで

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{X}]^{-1} \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{y}, \quad \boldsymbol{\Omega} = [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{X}]^{-1}, \quad (29)$$

であり、 $IW(\cdot)$ は逆ウィシャート分布である。よって (27) と (28) から $\boldsymbol{\beta}^{(r)}$ と $\boldsymbol{\Sigma}^{(r)}$ を順次発生させていくことで SUR モデルのギブズ・サンプラーを実行できる。

ギブズ・サンプラーによって発生させたパラメータの乱数を使って事後分布の特性値などを推定する際には、ラオブラックウェル化推定量 (Rao-Blackwellized estimator) が便利である。ここでパラメータ θ_j の条件付事後分布の平均 $E(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y})$ の公式があると仮定する。このとき θ_j の事後平均 $E(\theta_j | \mathbf{y})$ のラオブラックウェル化推定量は

$$\hat{E}(\theta_j | \mathbf{y}) = \frac{1}{R} \sum_{r=1}^R E(\theta_j | \boldsymbol{\theta}_{-j}^{(r)}, \mathbf{y}), \quad (30)$$

となる。ラオブラックウェル化推定量は事後分布の特定値だけでなく周辺事後密度関数の推定にも利用できる。周辺事後密度関数 $p(\theta_j | \mathbf{y})$ のラオブラックウェル化推定量は

$$\hat{p}(\theta_j | \mathbf{y}) = \frac{1}{R} \sum_{r=1}^R p(\theta_j | \boldsymbol{\theta}_{-j}^{(r)}, \mathbf{y}), \quad (31)$$

として与えられる。ただし、これを使うには $p(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y})$ の基準化定数の公式が判っている必要がある。そうでない場合には基準化定数を別途推定しなければならない。

3.3 メトロポリス－ヘイスティングズ・アルゴリズム

次に、メトロポリス－ヘイスティングズ (MH) アルゴリズムを説明しよう。MH アルゴリズムは、Metropolis et al. [18] によって提案され、Hastings [13] によって一般化されたマルコフ連鎖サンプリング法の代表格である。先に紹介したギブズ・サンプラーも実は MH アルゴリズムの一種であることが知られている。

ここで事後分布 $p(\boldsymbol{\theta}|\mathbf{y})$ から乱数を発生させることを考える。この事後分布から直接 $\boldsymbol{\theta}$ を発生させることはできないが、代わりに別の分布 $q(\boldsymbol{\xi}|\boldsymbol{\theta})$ からは乱数を容易に発生させることができると仮定する。このとき MH アルゴリズムを使って事後分布から乱数を発生させることができる。MH アルゴリズムは以下の手順で実行される。

1. 初期値 $\boldsymbol{\theta}^{(0)}$ を決める。
2. $\boldsymbol{\theta}^{(r+1)}$ の候補 $\boldsymbol{\xi}$ を $g(\boldsymbol{\xi}|\boldsymbol{\theta}^{(r)})$ から発生させる。
3. $\boldsymbol{\theta}^{(r+1)}$ を以下のようにして決める。

$$\boldsymbol{\theta}^{(r+1)} = \begin{cases} \boldsymbol{\xi} & \text{with probability } \lambda(\boldsymbol{\theta}^{(r)}, \boldsymbol{\xi}), \\ \boldsymbol{\theta}^{(r)} & \text{with probability } 1 - \lambda(\boldsymbol{\theta}^{(r)}, \boldsymbol{\xi}). \end{cases} \quad (32)$$

ここで

$$\lambda(\boldsymbol{\theta}^{(r)}, \boldsymbol{\xi}) = \min \left\{ \frac{p(\boldsymbol{\xi}|\mathbf{y})}{p(\boldsymbol{\theta}^{(r)}|\mathbf{y})} \frac{q(\boldsymbol{\theta}^{(r)}|\boldsymbol{\xi})}{q(\boldsymbol{\xi}|\boldsymbol{\theta}^{(r)})}, 1 \right\}. \quad (33)$$

4. 2-3 を繰り返す。

この手順を十分な回数繰り返せば、 $\boldsymbol{\theta}^{(r)}$ の分布は $p(\boldsymbol{\theta}|\mathbf{y})$ に収束することが証明されている。

MH アルゴリズムの利点の一つは、それをギブズ・サンプラーと組み合わせて使うことができることである。ギブズ・サンプラーを事後分布に適用するには、すべてのパラメータの条件付事後分布から容易に乱数を発生できることが必要である。しかし、いつも条件付事後分布から乱数を発生できるとは限らない。この問題はギブズ・サンプラーの中で MH アルゴリズムを使うことで解決できる。条件付事後分布 $p(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m, \mathbf{y})$ から乱数の発生が容易でないが、別の分布 $q(\xi_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m, \mathbf{y})$ からは容易に乱数を発生できると仮定する。このとき以下の手順で $\theta_1, \dots, \theta_m$ を事後分布から発生させることができる実行される。

1. 初期値 $(\theta_1^{(0)}, \dots, \theta_m^{(0)})$ を決める。
2. $\theta_j^{(r+1)} (j=1, \dots, m)$ の候補 ξ_j を $q(\xi_j|\theta_1^{(r+1)}, \dots, \theta_{j-1}^{(r+1)}, \theta_{j+1}^{(r)}, \dots, \theta_m^{(r)}, \mathbf{y})$ から発生させる。
3. $\theta_j^{(r+1)}$ を以下のようにして決める。

$$\theta_j^{(r+1)} = \begin{cases} \xi_j & \text{with probability } \lambda(\theta_j^{(r)}, \xi_j), \\ \theta_j^{(r)} & \text{with probability } 1 - \lambda(\theta_j^{(r)}, \xi_j). \end{cases} \quad (34)$$

ここで

$$\lambda(\theta_j^{(r)}, \xi_j) = \min \left\{ \frac{w(\xi_j)}{w(\theta_j^{(r)})}, 1 \right\}, \quad (35)$$

$$w(x) \equiv \frac{p(x | \theta_1^{(r+1)}, \dots, \theta_{j-1}^{(r+1)}, \theta_{j+1}^{(r)}, \dots, \theta_m^{(r)}, \mathbf{y})}{q(x | \theta_1^{(r+1)}, \dots, \theta_{j-1}^{(r+1)}, \theta_{j+1}^{(r)}, \dots, \theta_m^{(r)}, \mathbf{y})}, \quad (35)$$

である。

以上を十分な回数繰り返せば、ギブズ・サンプラーと同様に $(\theta_1^{(r)}, \dots, \theta_m^{(r)})$ の分布が $p(\theta_1, \dots, \theta_m | \mathbf{y})$ に収束することが証明されている。

例として攪乱項が AR(1) に従う回帰モデル

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \epsilon_t, \quad \epsilon_t \sim \begin{cases} \mathcal{N}(\rho \epsilon_{t-1}, \sigma^2) & (t=2, \dots, T), \\ \mathcal{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right) & (t=1). \end{cases} \quad (37)$$

のマルコフ連鎖サンプリング法を考える。事前分布を

$$p(\boldsymbol{\beta}, \sigma^2, \rho) \propto \sigma^{-2} \mathbf{1}_{(-1,1)}(\rho), \quad (38)$$

とおく。ここで $\mathbf{1}_{(-1,1)}(\rho)$ とするのは AR(1) 過程に定常性を仮定していることを意味する。事後分布は

$$p(\boldsymbol{\beta}, \sigma^2, \rho | \mathbf{y}, \mathbf{X}) \propto (\sigma^2)^{-(T/2+1)} \exp \left[-\frac{1}{2\sigma^2} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})' (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \right] \mathbf{1}_{(-1,1)}(\rho), \quad (39)$$

となる。ここで $\tilde{\mathbf{y}} \equiv [\tilde{y}_1, \dots, \tilde{y}_T]'$, $\tilde{\mathbf{X}} \equiv [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_T]'$ であり、 $\tilde{y}_t$ と $\tilde{\mathbf{x}}_t$ の各要素は以下の変数変換で求められたものである。

$$\tilde{z}_t = \begin{cases} z_t - \rho z_{t-1} & (t=2, \dots, T), \\ \sqrt{1-\rho^2} z_t & (t=1). \end{cases} \quad (40)$$

(39) は ρ が与えられた下では $\tilde{\mathbf{y}}$ と $\tilde{\mathbf{X}}$ の古典的正規回帰モデルになっているので、 $\boldsymbol{\beta}$ と σ^2 の条件付事後分布はそれぞれ

$$\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{y}, \mathbf{X} \sim \mathcal{N}_k[\tilde{\boldsymbol{\beta}}, \sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}], \quad (41)$$

$$\sigma^2|\boldsymbol{\beta}, \rho, \mathbf{y}, \mathbf{X} \sim IG\left(\frac{T}{2}, \frac{T\tilde{\sigma}^2}{2}\right), \quad (42)$$

となる。ここで $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$ および $\tilde{\sigma}^2 \equiv T^{-1}(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})$ である。

ρ の条件付事後分布から ρ を直接発生させるのは難しいので、MH アルゴリズムをギブズ・サンプラーに組み合わせて使うことにする。 $e_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$ と定義し $t=1$ の項を事後分布から無視すると、

$$\begin{aligned} q(\rho|\boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X}) &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{t=2}^T(e_t - \rho e_{t-1})^2\right] \mathbf{1}_{(-1,1)}(\rho) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{t=2}^T e_{t-1}^2\right)(\rho - \hat{\rho})^2\right] \mathbf{1}_{(-1,1)}(\rho), \end{aligned} \quad (43)$$

となる。ここで $\hat{\rho} = \sum_{t=2}^T e_{t-1}e_t / \sum_{t=2}^T e_{t-1}^2$ である。この分布 $q(\rho|\boldsymbol{\beta}, \sigma^2, \mathbf{y}, \mathbf{X})$ は切断正規分布 $N[\hat{\rho}, \sigma^2 / \sum_{t=2}^T e_{t-1}^2] \mathbf{1}_{(-1,1)}(\rho)$ であり、これから簡単に ρ を発生させることができる。したがって MH アルゴリズムを適用できる。この場合、MH アルゴリズムの採択確率 $\lambda(\rho, \xi)$ は無視した $t=1$ の項の比を使って

$$\lambda(\rho, \xi) = \min\left\{\sqrt{\frac{1-\xi^2}{1-\rho^2}} \exp\left[\frac{(y_1 - \mathbf{x}_1'\boldsymbol{\beta})^2}{2\sigma^2}(\xi^2 - \rho^2)\right], 1\right\}, \quad (44)$$

となる。よってマルコフ連鎖サンプリングは以下のようになる。

1. 初期値 $(\boldsymbol{\beta}^{(0)}, \sigma^{2(0)}, \rho^{(0)})$ を決める。
2. $\boldsymbol{\beta}^{(r+1)}$ を (41) から発生させる。
3. $\sigma^{2(r+1)}$ を (42) から発生させる。
4. $\rho^{(r+1)}$ の候補 ξ を (43) から発生させ、確率 $\lambda(\rho^{(r)}, \xi)$ で ξ と $\rho^{(r)}$ のどちらを $\rho^{(r+1)}$ にするかを決定する。

3.4 潜在変数とデータ拡大法

今まではデータは全て観測されたものとして分析を行ってきた。しかし、実際の応用では一部のデータが欠損してしまう場合も多い。ここでデータ $\mathbf{y}$ が観測されたデータ $\mathbf{y}_1$ と欠損値 $\mathbf{y}_2$ に分けられるとする。すると、観測されたデータ $\mathbf{y}_1$ が与えられた下でのパラメータ $\boldsymbol{\theta}$ の事後分布は

$$p(\boldsymbol{\theta}|\mathbf{y}_1) = \int p(\boldsymbol{\theta}, \mathbf{y}_2|\mathbf{y}_1) d\mathbf{y}_2, \quad (45)$$

となる。もし欠損したデータ y_2 の数が多くなると積分の評価が難しくなる。これに対処する方法が Tanner and Wong [22] によるデータ拡大 (data augmentation) 法である。

条件付事後分布 $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2)$ と $p(\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1)$ から $\boldsymbol{\theta}$ と $\mathbf{y}_2$ を容易に発生できるとすると、データ拡大法のアルゴリズムは以下のように与えられる。

1. 初期値 $\boldsymbol{\theta}^{(0)}$ と $\mathbf{y}^{(0)}$ を決める。
2. $\boldsymbol{\theta}^{(r+1)}$ を $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2^{(r)})$ から発生させる。
3. $\mathbf{y}_2^{(r+1)}$ を $p(\mathbf{y}_2|\boldsymbol{\theta}^{(r+1)}, \mathbf{y}_1)$ から発生させる。
4. 2-3 を繰り返す。

これによって得られる $(\boldsymbol{\theta}^{(r)}, \mathbf{y}_2^{(r)})$ の同時分布は $p(\boldsymbol{\theta}, \mathbf{y}_2|\mathbf{y}_1)$ に収束する。したがって $p(\boldsymbol{\theta}|\mathbf{y}_1)$ に関する積分は $\{\boldsymbol{\theta}^{(r)}\}_{r=1}^R$ を使ったモンテカルロ法によって評価される。定義より、データ拡大法はギブズ・サンプラーの一種であるのは自明であろう。よって、データ拡大法と MH アルゴリズムを組み合わせることも可能である。

例としてデータ拡大法のトービット・モデル (Tobit model) への応用を説明しよう。回帰モデル

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (46)$$

において、 $y_i > 0$ である y_i のみが観測される場合がある。このような回帰モデルをトービット・モデルという。トービット・モデルは潜在変数 (latent variable) モデルの一つと解釈することができる。潜在変数 y_i^* が古典的回帰モデル

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (47)$$

によって生成されると仮定する。しかし、実際には $y_i^* > 0$ のときのみ y_i^* が観測され、 $y_i^* \leq 0$ のときには y_i^* は観測されない (検閲される) とする。このような潜在変数の回帰モデルとしてトービット・モデルを解釈することができる。この検閲された従属変数をデータ拡大法によって補完することでトービット・モデルを MCMC 法で分析できるようになる。

まず回帰モデル (47) を観測された y_i^* の部分と検閲された y_i^* の部分に分ける。

$$\begin{aligned} \mathbf{y}_1^* &= \mathbf{X}_1 \boldsymbol{\beta} + \boldsymbol{\epsilon}_1 && \text{(観測された従属変数の回帰式),} \\ \mathbf{y}_2^* &= \mathbf{X}_2 \boldsymbol{\beta} + \boldsymbol{\epsilon}_2 && \text{(検閲された従属変数の回帰式).} \end{aligned}$$

$\mathbf{y}_1^*$ は観測された潜在変数の n_1 次元ベクトルで、 $\mathbf{y}_2^*$ は検閲された潜在変数の $(n - n_1)$ 次元ベクトルであると仮定する。そして、 $\mathbf{X}_j$ と $\boldsymbol{\epsilon}_j (j=1, 2)$ はそれぞれに対応した説明変数と攪乱項である。 $\mathbf{y}_1^*$ の実現値 (実際に観測された値) を $\mathbf{y}_1$ とする。このとき尤度は、

$$p(\mathbf{y}_1, \mathbf{y}_2^* | \beta, \sigma^2, \mathbf{X}) \propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y}_1 - \mathbf{X}_1 \beta)' (\mathbf{y}_1 - \mathbf{X}_1 \beta) - \frac{1}{2\sigma^2} (\mathbf{y}_2^* - \mathbf{X}_2 \beta)' (\mathbf{y}_2^* - \mathbf{X}_2 \beta) \right] \quad (48)$$

$$\propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X} \beta)' (\mathbf{y} - \mathbf{X} \beta) \right], \quad (49)$$

となる。ここで $\mathbf{y} = [\mathbf{y}_1; \mathbf{y}_2^*]$, $\mathbf{X} = [\mathbf{X}_1; \mathbf{X}_2]$ である。事前分布は $p(\beta, \sigma^2) \propto \sigma^{-2}$ とすると、 β と σ^2 の検閲された $\mathbf{y}_2^*$ が与えられた下での条件付事後分布は

$$\beta | \sigma^2, \mathbf{y}_1, \mathbf{y}_2^*, \mathbf{X} \sim \mathcal{N}_k[\hat{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}], \quad (50)$$

$$\sigma^2 | \beta, \mathbf{y}_1, \mathbf{y}_2^*, \mathbf{X} \sim IG\left(\frac{n}{2}, \frac{n\hat{\sigma}^2}{2}\right), \quad (51)$$

として与えられる。ここで $\hat{\beta}$ と $\hat{\sigma}^2$ は古典的正規回帰モデルのギブズ・サンプラーで定義された値である。結局、 $\mathbf{y}$ の検閲された部分 $\mathbf{y}_2^*$ がデータ拡大法によって得られた値に置き換えられている点を除けば、古典的正規回帰モデルの場合と全く同じであることがわかる。検閲された y_i^* の (β , σ^2) が与えられた下での条件付事後分布は

$$y_i^* | \beta, \sigma^2, \mathbf{y}_1, \mathbf{X} \sim \mathcal{N}^-(\mathbf{x}_i' \beta, \sigma^2), \quad (i = n_1 + 1, \dots, n) \quad (52)$$

となる。ここで $\mathcal{N}^-(\cdot)$ は負の切断正規分布である。したがって (50), (51), (52) から順次 $\beta^{(r)}$, $\sigma^{2(r)}$, $y_i^{*(r)}$ を発生させていくことでトービット・モデルのギブズ・サンプラーが実行される。

トービット・モデルでは一部の従属変数が検閲されていた。これに対して全ての従属変数に関して符号や順序など限られた情報以外わからないようなモデルも存在する。そのようなモデルの一例がプロビット・モデル (probit model) である。プロビット・モデルは 0 か 1 のみをとる従属変数 y_i を説明変数 $\mathbf{x}_i$ と係数 β の線形結合 $\mathbf{x}_i' \beta$ によって説明するモデルである。プロビットモデルでは y_i が 0 および 1 になる確率は

$$\Pr\{y_i = 1\} = \Phi(\mathbf{x}_i' \beta), \quad \Pr\{y_i = 0\} = 1 - \Phi(\mathbf{x}_i' \beta), \quad (i = 1, \dots, n) \quad (53)$$

として与えられる。ここで $\Phi(\cdot)$ は標準正規分布の累積分布関数である。

プロビットモデルは潜在変数 y_i^* の回帰モデルとみなすことができることを示そう。潜在変数 y_i^* は潜在変数回帰モデル

$$y_i^* = \mathbf{x}_i' \beta + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1), \quad (54)$$

によって決定され、観測される y_i との間に以下のような関係があるとする。

$$\begin{cases} y_i=1 & \text{if } y_i^* > 0, \\ y_i=0 & \text{otherwise.} \end{cases} \quad (55)$$

すると,

$$\begin{cases} y_i=1 & \text{if } -\epsilon_i < \mathbf{x}_i' \boldsymbol{\beta}, \\ y_i=0 & \text{otherwise.} \end{cases}$$

と書き表せる。 $\Pr\{-\epsilon_i < \mathbf{x}_i' \boldsymbol{\beta}\} = \Phi(\mathbf{x}_i' \boldsymbol{\beta})$ なので、これはプロビット・モデルそのものであることがわかる。

もし潜在変数 $\mathbf{y}^* = [y_1^*, \dots, y_n^*]'$ をパラメータと同じく未知の確率変数であるとする、事前分布 $p(\boldsymbol{\beta}) \propto \text{constant}$ として $(\boldsymbol{\beta}, \mathbf{y}^*)$ の事後分布は、

$$p(\boldsymbol{\beta}, \mathbf{y}^* | \mathbf{y}, \mathbf{X}) \propto \exp\left[-\frac{1}{2}(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})\right] \times \prod_{i=1}^n \{y_i \mathbf{1}_{(0,\infty)}(y_i^*) + (1-y_i) \mathbf{1}_{(-\infty,0]}(y_i^*)\}, \quad (56)$$

となる。(56) の右辺の後半は y_i が 1 か 0 によって y_i^* の分布が正か負に切断されることを意味している。したがって、 $\mathbf{y}^*$ が与えられた下での $\boldsymbol{\beta}$ の条件付事後分布は

$$\boldsymbol{\beta} | \mathbf{y}^*, \mathbf{y}, \mathbf{X} \sim \mathcal{N}_k[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^*, (\mathbf{X}'\mathbf{X})^{-1}], \quad (57)$$

である。一方、 $\boldsymbol{\beta}$ が与えられた下での y_i^* ($i=1, \dots, n$) の条件付事後分布は

$$y_i^* | \boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{N}^+(\mathbf{x}_i' \boldsymbol{\beta}, 1) & \text{if } y_i = 1, \\ \mathcal{N}^-(\mathbf{x}_i' \boldsymbol{\beta}, 1) & \text{if } y_i = 0, \end{cases} \quad (58)$$

である。ここで $\mathcal{N}^+(\cdot)$ および $\mathcal{N}^-(\cdot)$ はそれぞれ正と負の切断正規分布を意味する。よって (57), (58) から順次 $\boldsymbol{\beta}^{(r)}$, $y_i^{*(r)}$ を発生させることでプロビット・モデルのギブズ・サンプラーが実行される。

本節で例として挙げたトービット・モデルやプロビット・モデルは簡単に最尤法で推定できるため、MCMC 法によるベイズ分析を無理に使う必要は感じられないかもしれない。しかし、次節以降で紹介する多値選択モデルや制限従属変数モデルは最尤法で推定するのが難しく MCMC 法が威力を発揮するモデルとなっている。

4 多値選択モデルのベイズ分析

4.1 順序型プロビットモデル

観測されたデータ y_i が $0, 1, 2, \dots$ という 3 つ以上の値をとる場合に, y_i を x_i で説明しようとするモデルが多値選択モデルである。このような多値選択モデルには選択肢の順序に意味があるモデルと意味がないモデルがある。前者の例として順序型プロビット・モデル (ordered probit model), 後者の例として多項プロビット・モデル (multinomial probit model) を考える。

まず順序型プロビット・モデルを説明する。順序型プロビット・モデルは選択肢の順序に意味がある問題 (例えば企業の格付けなど) に適用される。順序型プロビット・モデルではプロビット・モデルで使ったものと同じ潜在変数回帰モデル (54) で決定される潜在変数 y_i^* と観測されたデータ y_i の間に

$$y_i = j \quad \text{if} \quad \gamma_{j-1} < y_i^* \leq \gamma_j \quad (j=1, \dots, J+1), \quad (59)$$

という関係があると仮定する。ここで γ_j は $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_J < \gamma_{J+1} = \infty$ という閾値であり, 未知のパラメータとして推定の対象になる。閾値の識別のため γ_1 は 0 とおき, x_i には定数項を含めない。ここで $\boldsymbol{\gamma} = [\gamma_2, \dots, \gamma_J]'$ とする。

事前分布をプロビット・モデルのときと同じもの $p(\boldsymbol{\beta}) \propto \text{constant}$ を使うと, $(\boldsymbol{\beta}, \mathbf{y}^*)$ の事後分布は,

$$p(\boldsymbol{\beta}, \mathbf{y}^* | \boldsymbol{\gamma}, \mathbf{y}, \mathbf{X}) \propto \exp \left[-\frac{1}{2} (\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}) \right] \times \prod_{i=1}^n \sum_{j=1}^J \mathbf{1}(y_i = j) \mathbf{1}_{(\gamma_{j-1}, \gamma_j]}(y_i^*), \quad (60)$$

となる。順序型プロビットモデルの事後分布 $p(\boldsymbol{\beta}, \mathbf{y}^* | \boldsymbol{\gamma}, \mathbf{y}, \mathbf{X})$ からのサンプリングにはギブズ・サンプラーを使う。 $\boldsymbol{\beta}$ と y_i^* の条件付事後分布は

$$\boldsymbol{\beta} | \boldsymbol{\gamma}, \mathbf{y}^*, \mathbf{y}, \mathbf{X} \sim \mathcal{N}_k[(\mathbf{X}'\mathbf{X})\mathbf{X}'\mathbf{y}^*, (\mathbf{X}'\mathbf{X})^{-1}], \quad (61)$$

$$y_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(x_i'\boldsymbol{\beta}, 1) \mathbf{1}_{(\gamma_{y_i-1}, \gamma_{y_i}]}(y_i^*), \quad (62)$$

である。閾値 γ_j ($j=2, \dots, J$) は, 残りの閾値 $\boldsymbol{\gamma}_{-j} = [\gamma_2, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_J]'$ が与えられた下で以下のような条件付事後分布を持つ。

$$\gamma_j | \boldsymbol{\gamma}_{-j}, \boldsymbol{\beta}, \mathbf{y}^*, \mathbf{y}, \mathbf{X} \sim \mathcal{U}(a_j, b_j). \quad (63)$$

この一様分布の下限と上限は

$$a_j = \max \left\{ \max_{1 \leq i \leq n} (y_i^* : y_i = j), \gamma_{j-1} \right\}, \quad b_j = \min \left\{ \min_{1 \leq i \leq n} (y_i^* : y_i = j+1), \gamma_{j+1} \right\},$$

で与えられる。よって (61), (62), (63) から順次 $\beta^{(r)}$, $y_i^{*(r)}$, $\gamma_j^{(r)}$ を発生させることで順序型プロビット・モデルのギブズ・サンプラーが実行される。以上は Albert and Chib [3] の方法であるが、これ以外にも順序型プロビット・モデルに対して様々なマルコフ連鎖サンプリング法が提案されている。詳しくは Chen et al. [8] を参照。

4.2 多項プロビット・モデル

次に多項プロビット・モデルを説明する。多項プロビット・モデルもまた3つ以上の選択肢に対するプロビット・モデルの一つである。順序型プロビット・モデルとの違いは選択肢の順序に意味がある場合だけに応用が限定されないことである。例えば、消費者の商品やサービスの選択をモデル化するのに多項プロビット・モデルが応用される。 $J+1$ 個の選択肢からなる多項プロビット・モデルは以下のようにして与えられる。まず $J+1$ 個の潜在変数 y_{ij}^* の回帰式

$$\begin{aligned} y_{i0}^* &= \mathbf{z}_{i0}'\beta + u_{i0}, \\ y_{i1}^* &= \mathbf{z}_{i1}'\beta + u_{i1}, \\ &\vdots \\ y_{iJ}^* &= \mathbf{z}_{iJ}'\beta + u_{iJ}, \end{aligned} \tag{64}$$

を考える。そして、選択肢 j ($j=0, 1, \dots, J$) は

$$y_{ij}^* = \max_{0 \leq h \leq J} y_{ih}^*, \tag{65}$$

であるときのみ選択されると仮定する。このとき観測された y_i ($y_i=0, 1, \dots, J$) と潜在変数 ($y_{i0}^*, \dots, y_{iJ}^*$) の間には

$$y_i = \begin{cases} j & \text{if } y_{ij}^* = \max_{0 \leq h \leq J} y_{ih}^*, \\ 0 & \text{otherwise} \end{cases} \tag{66}$$

という関係が成り立つ。多項プロビット・モデルを消費者行動への応用で考えると、 y_{ij}^* は選択肢 j を選ぶことによって消費者 i が得る効用として解釈され、実際に消費者 i が選んだ選択肢 y_i は消費者 i にとって最も高い効用が得られる選択肢であったと解釈することになる。

実際には ($y_{i0}^*, \dots, y_{iJ}^*$) の間の大小関係しか識別されないので、 y_{i0}^* を基準にして $w_{ij} = y_{ij}^* - y_{i0}^*$ ($j=1, \dots, J$) という新しい潜在変数を定義すると、($w_{i1}, \dots, w_{iJ}$) と y_i の間には

$$y_i = \begin{cases} j & \text{if } w_{ij} = \max_{1 \leq h \leq J} w_{ih}, \\ 0 & \text{if } \max_{1 \leq h \leq J} w_{ih} < 0, \end{cases} \quad (67)$$

という関係が成り立つ。これより、 $\epsilon_{ij} = u_{ij} - u_{i0}$ および $x_{ij} = z_{ij} - z_{i0}$ と定義すると、

$$\begin{aligned} w_{i1} &= \mathbf{x}'_{i1} \boldsymbol{\beta} + \epsilon_{i1}, \\ &\vdots \\ w_{iJ} &= \mathbf{x}'_{iJ} \boldsymbol{\beta} + \epsilon_{iJ}, \end{aligned} \quad (68)$$

という潜在変数回帰モデルを考えることができる。ここで $\mathbf{w}_i = [w_{i1}, \dots, w_{iJ}]'$, $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{iJ}]'$, $\boldsymbol{\epsilon}_i = [\epsilon_{i1}, \dots, \epsilon_{iJ}]'$ とし、攪乱項 $\boldsymbol{\epsilon}_i$ の分布を $\boldsymbol{\epsilon}_i \sim \mathcal{N}_J(\mathbf{0}, \boldsymbol{\Sigma})$ と仮定する。

多項プロビット・モデルでは、プロビット・モデルと同様にスケールが識別されないので、 $\boldsymbol{\Sigma}$ に対して何らかの制約が必要になる。よく使われる制約は $\boldsymbol{\Sigma}$ の (1, 1) 要素を 1 にすることである。つまり、

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{\Phi} + \sigma_{11}^{-1} \boldsymbol{\gamma} \boldsymbol{\gamma}' \end{bmatrix}, \quad (69)$$

において、 $\sigma_{11} = 1$ とすることである。この制約をギブズ・サンプラーに反映させるやり方には 2 通りある。

1. 識別制約を入れないギブズ・サンプラーを適用した後、変数変換を施す。
2. 識別制約を入れたギブズ・サンプラーを適用する。

前者のアプローチをとっている研究に McCulloch and Rossi [16] があり、後者のアプローチをとっている研究に McCulloch, Polson, and Rossi [17] や Nobile [19] がある。ここでは McCulloch and Rossi [16] の方法を説明する。

モデルのパラメータ $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ に対する事前分布を $\boldsymbol{\beta} \sim \mathcal{N}_k(\bar{\boldsymbol{\beta}}, \mathbf{A}^{-1})$, $\boldsymbol{\Sigma} \sim IW(\mathbf{S}, \nu, J)$ とする。ここで $\mathbf{w} = [\mathbf{w}'_1, \dots, \mathbf{w}'_n]'$ および $\mathbf{X} = [\mathbf{X}'_1, \dots, \mathbf{X}'_n]'$ とする。 $\mathbf{w}$ が与えられた下での $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ の条件付事後分布は

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{w}, \mathbf{y}, \mathbf{X}) &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{w}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{w}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \\ &\quad \times \exp \left[-\frac{1}{2} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \mathbf{A} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) \right] \times |\boldsymbol{\Sigma}|^{-(\nu+J+1)/2} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) \right] \\ &\propto |\boldsymbol{\Sigma}|^{-(n+\nu+J+1)/2} \exp \left[-\frac{1}{2} \{ (\mathbf{w} - \mathbf{X} \boldsymbol{\beta})' (\mathbf{I}_n \otimes \boldsymbol{\Sigma}^{-1}) (\mathbf{w} - \mathbf{X} \boldsymbol{\beta}) + (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' \mathbf{A} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) + \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) \} \right], \end{aligned} \quad (70)$$

となる。したがって β と Σ の条件付事後分布は

$$\beta|\Sigma, \mathbf{w}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}_k[\Sigma_\beta\{\mathbf{X}'(\mathbf{I}_n \otimes \Sigma^{-1})\mathbf{w} + \mathbf{A}\bar{\beta}\}, \Sigma_\beta], \quad (71)$$

$$\Sigma|\beta, \mathbf{w}, \mathbf{y}, \mathbf{X} \sim IW(\mathbf{E}\mathbf{E}' + \mathbf{S}, n + \nu, J), \quad (72)$$

である。ここで $\Sigma_\beta = \{\mathbf{X}'(\mathbf{I}_n \otimes \Sigma^{-1})\mathbf{X} + \mathbf{A}\}^{-1}$ および $\mathbf{E} = [\mathbf{w}_1 - \mathbf{X}_1\beta, \dots, \mathbf{w}_n - \mathbf{X}_n\beta]$ としている。こうして発生させた β と Σ に対し $(\beta, \Sigma) \rightarrow (\beta/\sqrt{\sigma_{11}}, \Sigma/\sigma_{11})$ という変数変換を施すと $(\beta/\sqrt{\sigma_{11}}, \Sigma/\sigma_{11})$ は識別される。

次に $\mathbf{w}_i$ の発生方法を考える。 $\mathbf{w}_i$ の同時条件付事後分布を考えると、

$$\mathbf{w}_i|\beta, \Sigma, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{N}_J(\mathbf{X}_i\beta, \Sigma)\mathbf{1}(w_{ij} > w_{i1}, \dots, w_{ij} > w_{ij}) & \text{if } y_i = j, \\ \mathcal{N}_J(\mathbf{X}_i\beta, \Sigma)\mathbf{1}(w_{i1} < 0, \dots, w_{ij} < 0) & \text{if } y_i = 0, \end{cases} \quad (73)$$

となる。しかし、 J が大きくなると J 変量切断正規分布から $\mathbf{w}_i$ を発生させることは難しくなる。かわりに w_{ij} を一つ一つ条件付事後分布から発生させることを考える。そうすると w_{ij} の条件付事後分布は 1 変量切断正規分布になる。つまり、

$$w_{ij}|\mathbf{w}_{i,-j}, \beta, \Sigma, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{N}(\mu_{ij}, \tau_{ij}^2)\mathbf{1}(w_{ij} > \max(\mathbf{w}_{i,-j})) & \text{if } y_i = j, \\ \mathcal{N}(\mu_{ij}, \tau_{ij}^2)\mathbf{1}(w_{ij} < \max(\mathbf{w}_{i,-j}, 0)) & \text{if } y_i \neq j, \end{cases} \quad (74)$$

ここで $\mathbf{w}_{i,-j}$ は w_{ij} をのぞいた $\mathbf{w}_i$ の要素であり、 μ_{ij} と τ_{ij}^2 はそれぞれ $\mathcal{N}_J(\mathbf{X}_i\beta, \Sigma)$ における $\mathbf{w}_{i,-j}$ が与えられた下での w_{ij} の条件付期待値と条件付分散である。

$$\mathbf{w}_i = \begin{bmatrix} \mathbf{w}_{i(-j)} \\ w_{ij} \end{bmatrix}, \quad \mathbf{X}_i\beta = \begin{bmatrix} \mathbf{X}_{i(-j)}\beta \\ \mathbf{x}'_{ij}\beta \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{(-j)(-j)} & \boldsymbol{\sigma}_{(-j)j} \\ \boldsymbol{\sigma}_{j(-j)} & \sigma_{jj} \end{bmatrix},$$

と並べ替えると、 μ_{ij} と τ_{ij}^2 は

$$\mu_{ij} = \mathbf{x}'_{ij}\beta + \boldsymbol{\sigma}_{j(-j)}\Sigma_{(-j)(-j)}^{-1}(\mathbf{w}_{i(-j)} - \mathbf{X}_{i(-j)}\beta), \quad (75)$$

$$\tau_{ij}^2 = \sigma_{jj} - \boldsymbol{\sigma}_{(-j)j}\Sigma_{(-j)(-j)}^{-1}\boldsymbol{\sigma}_{j(-j)}, \quad (76)$$

として与えられる。まとめると、条件付事後分布 (71), (72), (74) より、 $\beta^{(r)}$, $\Sigma^{(r)}$, $w_{ij}^{(r)}$ を順次発生させることで多項プロビット・モデルのギブズ・サンプラーを実行することができる。

4.3 多変量プロビット・モデル

多変量プロビット・モデル (multivariate probit model) は、通常のプロビット・モデルが一つの設問に対する Yes/No の回答を説明するモデルであるのに対し、複数の設問に対する Yes/No の回答を同時に説明するモデルである。各設問に対する J 本の同時方程式体系

$$\begin{bmatrix} y_{i1}^* \\ \vdots \\ y_{ij}^* \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_{i1} & & \\ & \ddots & \\ & & \mathbf{x}'_{ij} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{ij} \end{bmatrix}, \quad \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{ij} \end{bmatrix} \sim \mathcal{N}_J(\mathbf{0}, \mathbf{R}), \quad (77)$$

を考える。ここで $\mathbf{R}$ は相関係数行列である。多変量プロビット・モデルでは設問 j に対する回答 y_{ij} は、通常のプロビット・モデルと同じく

$$y_{ij} = \begin{cases} 1, & \text{if } y_{ij}^* > 0, \\ 0, & \text{if } y_{ij}^* \leq 0, \end{cases}$$

によって決定される。 J 個の設問がある多変量プロビット・モデルを古典的統計学の枠組みで最尤推定するためには、尤度関数を評価するために J 変量正規分布の積分を計算しなければならない。そのため J が大きくなると最尤推定値の計算は、モンテカルロ法以外では実行不可能になってしまう。そこで MCMC 法を使った分析法が有用となる。

ここで、 $\mathbf{y}_j^* = [y_{1j}^*, \dots, y_{nj}^*]'$, $\mathbf{X}_j = [\mathbf{x}_{1j}, \dots, \mathbf{x}_{nj}]'$, $\boldsymbol{\epsilon}_j = [\epsilon_{1j}, \dots, \epsilon_{nj}]'$ ($j=1, \dots, J$) とおく。すると潜在変数 y_{ij}^* が与えられた下では多変量プロビット・モデルは SUR モデル

$$\underbrace{\begin{bmatrix} \mathbf{y}_1^* \\ \vdots \\ \mathbf{y}_J^* \end{bmatrix}}_{\mathbf{y}^*} = \underbrace{\begin{bmatrix} \mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_J \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_J \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_J \end{bmatrix}}_{\boldsymbol{\epsilon}}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_{nJ}(\mathbf{0}, \mathbf{R} \otimes \mathbf{I}_n), \quad (78)$$

に帰着されてしまう。2 節の SUR モデルの結果より、 $\boldsymbol{\beta}$ の条件付事後分布は

$$\boldsymbol{\beta} | \mathbf{R}, \mathbf{y}^*, \mathbf{y}, \mathbf{X} \sim N_k[\{\mathbf{X}'(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{y}^*, \{\mathbf{X}'(\mathbf{R}^{-1} \otimes \mathbf{I}_n)\mathbf{X}\}^{-1}], \quad (79)$$

と求まる。

$\mathbf{R}$ は相関係数行列であるため逆ウィシャート分布から出すことはできない。代わりに MH アルゴリズムを使うことになる。相関係数行列のサンプリングについては Chib and Greenberg [9]などを参照。

y_{ij}^* の発生方法は多項プロビット・モデルにおける w_{ij} と基本的に同じである。 $\mathbf{y}_i^* = [y_{i1}^*, \dots, y_{ij}^*]'$ とし、 $\mathbf{y}_{i,-j}^*$ を y_{ij}^* を $\mathbf{y}_i^*$ から除いたものとする。すると y_{ij}^* の $\mathbf{y}_{i,-j}^*$ が与えられた下での条件付事後分布は 1 変量切斷正規分布になる。つまり、

$$y_{ij}^* | \mathbf{y}_{i,-j}^*, \boldsymbol{\beta}, \mathbf{R}, \mathbf{y}, \mathbf{X} \sim \begin{cases} \mathcal{N}^+(\mu_{ij}, \tau_{ij}^2) & \text{if } y_{ij} = 1, \\ \mathcal{N}^-(\mu_{ij}, \tau_{ij}^2) & \text{if } y_{ij} = 0. \end{cases} \quad (80)$$

μ_{ij} と τ_{ij}^2 は, (75) と (76) の要領で求められた y_{ij}^* の $y_{i,-j}^*$ が与えられた下での条件付期待値と条件付分散である。よって, 条件付事後分布 (79), (80) から $\beta^{(r)}$ と $y_{ij}^{*(r)}$ を発生させ, $\mathbf{R}^{(r)}$ を MH アルゴリズムを使って発生させることで多変量プロビット・モデルのマルコフ連鎖サンプリングを行うことができる。

5 制限従属変数モデルのベイズ分析

5.1 付随的切断

トービット・モデルに代表される従属変数が何らかの理由で検閲されるモデルを制限従属変数モデル (limited dependent variable model) という。従属変数の検閲はトービット・モデルのように特定の値を境に起きるとは限らない。本節では, 付随的切断 (incidental truncation) と不均衡モデル (disequilibrium model) を例にして MCMC 法による制限従属変数モデルのベイズ分析を説明する。

まず付随的切断の話から始めよう。2つの回帰モデル

$$\begin{cases} z_i^* = \mathbf{w}_i' \boldsymbol{\gamma} + u_i, \\ y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \end{cases} \quad \begin{bmatrix} u_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right), \quad (81)$$

を考える。 $\boldsymbol{\beta}$ は k 次元ベクトルで, $\boldsymbol{\gamma}$ は h 次元ベクトルであるとする。このとき付随的切断とは y_i^* は $z_i^* > 0$ のときのみ観測されることをさす。ここで y_i^* が観測されたかどうかを表すダミー変数 z_i

$$z_i = \begin{cases} 1, & \text{if } z_i^* > 0, \\ 0, & \text{if } z_i^* \leq 0, \end{cases}$$

を導入する。 z_i^* は観測されない潜在変数であるが, z_i は全て観測されていると仮定する。そして, $z_i = 1$ のとき y_i^* は観測され, $z_i = 0$ のとき y_i^* は観測されていないことになる。観測された y_i^* のベクトルを y_1 とし, 観測されなかった y_i^* のベクトルを y_2^* とする。

経済の実証分析で付随的切断の代表的な例は, 労働供給と留保賃金の問題である。標準的な新古典派労働供給理論では, 労働者は留保賃金を超える賃金がオファーされたときに正の労働供給を行うとされる。したがって, (81) では, y_i^* が労働供給関数に対応し, z_i^* はオファーされた賃金が留保賃金を上回っている ($z_i^* > 0$) かどうかを決定する式となっている。

古典的統計学の枠組みで付随的切断モデルを推定する方法として有名なのが Heckman [14] による2段階推定である。しかし, Heckman の2段階推定量は一致性はあるが効率的ではないことが知られている。そこで MCMC 法による付随的切断モデルの推定を考える。

事前分布を $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho) \propto \sigma^{-2} \mathbf{1}_{(-1,1)}(\rho)$ とすると, 事後分布は

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho, \mathbf{y}_i^*, \mathbf{z}^* | \mathbf{y}_i, \mathbf{z}, \mathbf{X}, \mathbf{W}) &\propto (\sigma^2)^{-(n/2+1)} (1-\rho^2)^{-n/2} \\
&\times \exp \left[-\sum_{i=1}^n \frac{(y_i^* - \mathbf{x}_i' \boldsymbol{\beta})^2 - 2\rho\sigma(y_i^* - \mathbf{x}_i' \boldsymbol{\beta})(z_i^* - \mathbf{w}_i' \boldsymbol{\gamma}) + \sigma^2(z_i^* - \mathbf{w}_i' \boldsymbol{\gamma})^2}{2\sigma^2(1-\rho^2)} \right] \\
&\times \prod_{i=1}^n \{z_i \mathbf{1}_{(0,\infty)}(z_i^*) + (1-z_i) \mathbf{1}_{(-\infty,0]}(z_i^*)\},
\end{aligned} \tag{82}$$

である。ここで $\mathbf{X}$ が各行が $\mathbf{x}_i'$ になっている $(n \times k)$ 行列であり、 $\mathbf{W}$ は各行が $\mathbf{w}_i'$ になっている $(n \times h)$ 行列である。

まず潜在変数 z_i^* と観測されなかった y_i^* の発生方法を考える。 z_i^* と y_i^* の条件付事後分布は

$$\begin{aligned}
p(z_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho, y_i^*, z_i, \mathbf{x}_i, \mathbf{w}_i) \\
\propto \exp \left[-\frac{\{z_i^* - \mathbf{w}_i' \boldsymbol{\gamma} - (\rho/\sigma)(y_i^* - \mathbf{x}_i' \boldsymbol{\beta})\}^2}{2(1-\rho^2)} \right] \times \{z_i \mathbf{1}_{(0,\infty)}(z_i^*) + (1-z_i) \mathbf{1}_{(-\infty,0]}(z_i^*)\},
\end{aligned} \tag{83}$$

$$\begin{aligned}
p(y_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho, z_i^*, z_i, \mathbf{x}_i, \mathbf{w}_i) \\
\propto \exp \left[-\frac{\{y_i^* - \mathbf{x}_i' \boldsymbol{\beta} - \rho\sigma(z_i^* - \mathbf{w}_i' \boldsymbol{\gamma})\}^2}{2\sigma^2(1-\rho^2)} \right],
\end{aligned} \tag{84}$$

である。したがって

$$z_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho, \mathbf{y}^*, \mathbf{z}, \mathbf{X}, \mathbf{W} \sim \begin{cases} \mathcal{N}^+[\mathbf{w}_i' \boldsymbol{\gamma} + (\rho/\sigma)(y_i^* - \mathbf{x}_i' \boldsymbol{\beta}), 1-\rho^2] & \text{if } z_i=1, \\ \mathcal{N}^-[\mathbf{w}_i' \boldsymbol{\gamma} + (\rho/\sigma)(y_i^* - \mathbf{x}_i' \boldsymbol{\beta}), 1-\rho^2] & \text{if } z_i=0, \end{cases} \tag{85}$$

$$y_i^* | \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \rho, \mathbf{y}^*, \mathbf{z}, \mathbf{X}, \mathbf{W} \sim \mathcal{N}[\mathbf{x}_i' \boldsymbol{\beta} + \rho\sigma(z_i^* - \mathbf{w}_i' \boldsymbol{\gamma}), \sigma^2(1-\rho^2)], \tag{86}$$

となる。トービット・モデルとは異なり、欠損値 y_i^* は切断されず代わりに潜在変数 z_i^* が切断されていることに注意しよう。

次に回帰係数 $\boldsymbol{\gamma}$ と $\boldsymbol{\beta}$ の発生方法を考える。最初に $\boldsymbol{\gamma}$ の条件付事後分布を考える。 z_i^* の条件付事後分布の形より、

$$z_i^* - \frac{\rho}{\sigma}(y_i^* - \mathbf{x}_i' \boldsymbol{\beta}) = \mathbf{w}_i' \boldsymbol{\gamma} + v_i, \quad v_i \sim \mathcal{N}(0, 1-\rho^2), \tag{87}$$

という回帰モデルを考えればよいことがわかる。よって

$$\boldsymbol{\gamma} | \boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}^*, \mathbf{z}^*, \mathbf{z}, \mathbf{X}, \mathbf{W} \sim \mathcal{N}_h[(\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \bar{\mathbf{z}}, (1-\rho^2)(\mathbf{W}' \mathbf{W})^{-1}], \tag{88}$$

である。ここで $\bar{\mathbf{z}}$ は $z_i^* - \frac{\rho}{\sigma}(y_i^* - \mathbf{x}_i' \boldsymbol{\beta})$ を第 i 要素に持つ n 次元ベクトルである。同様に $\boldsymbol{\beta}$ を発生させるには

$$y_i - \rho\sigma(z_i^* - \mathbf{w}_i' \boldsymbol{\gamma}) = \mathbf{x}_i' \boldsymbol{\beta} + v_i, \quad v_i \sim \mathcal{N}[0, \sigma^2(1-\rho^2)], \tag{89}$$

という回帰モデルを考える。よって

$$\beta|\gamma, \sigma^2, \rho, \mathbf{y}^*, \mathbf{z}^*, \mathbf{z}, \mathbf{X}, \mathbf{W} \sim \mathcal{N}_k[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}}, \sigma^2(1-\rho^2)(\mathbf{X}'\mathbf{X})^{-1}], \quad (90)$$

である。ここで $\bar{\mathbf{y}}$ は $y_i - \rho\sigma(z_i^* - \mathbf{w}_i'\gamma)$ を第 i 要素に持つ n 次元ベクトルである。

最後に、 σ^2 と ρ については標準的な分布にならないので MH アルゴリズムを使うことになる。まとめると、(90) と (88) から $\beta^{(r)}$ と $\gamma^{(r)}$ を発生させ、(85) から $z_i^{*(r)}$ 、(86) から観測されなかった $y_i^{*(r)}$ を発生させ、 $\sigma^{2(r)}$ と $\rho^{(r)}$ は MH アルゴリズムを使って発生させることを順次繰り返すことで、付随的切断モデルのマルコフ連鎖サンプリングを行う。

5.2 不均衡モデル

完全競争市場においては需給と供給が均衡するように市場価格が決まる。しかし、何らかの理由で価格変化が硬直的である場合、瞬時には需給が均衡しないと考えるのが現実的である。このような状況をモデル化することを考える。

$$\begin{cases} d_t = \mathbf{x}_{1t}'\beta_1 + \epsilon_{1t}, \\ s_t = \mathbf{x}_{2t}'\beta_2 + \epsilon_{2t}, \\ q_t = \min\{d_t, s_t\}, \end{cases} \quad \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim \mathcal{N}_2(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \quad (91)$$

ここで $\mathbf{d} = [d_1, \dots, d_T]'$, $\mathbf{s} = [s_1, \dots, s_T]'$, $\mathbf{q} = [q_1, \dots, q_T]'$, $\mathbf{X}_j = [\mathbf{x}_{j1}, \dots, \mathbf{x}_{jT}]'$ ($j=1, 2$) とする。この不均衡モデルの特徴は、

1. 超過需給が存在する ($d_t > s_t$) ときには、供給 s_t が q_t として観測されるが、需要 d_t は観測されない。
2. 超過供給が存在する ($d_t \leq s_t$) ときには、需要 d_t が q_t として観測されるが、供給 s_t は観測されない。
3. どの時点で需要 d_t と供給 s_t のどちらが観測されているか不明である。

である。特に最後の点は

$$z_t = \begin{cases} 1, & \text{if } s_t = q_t, \\ 0, & \text{if } d_t = q_t, \end{cases} \quad (92)$$

というダミー変数が欠損していることに等しい。この不均衡モデルは最尤推定が難しいことで知られている。Maddala [15] を参照。しかし、MCMC 法を使えば比較的簡単に推定が可能である。不均衡モデルのマルコフ連鎖サンプリングは、 z_t をデータ拡大法によって補完し、さらに需給不均衡のため観測されなかった需要 d_t と供給 s_t もサンプリングによって補完する方法をとる。

事前分布を $p(\beta_1, \beta_2, \Sigma) \propto \text{constant}$ とすると、事後分布は

$$\begin{aligned} p(\beta_1, \beta_2, \Sigma, \mathbf{d}, \mathbf{s} | \mathbf{z}, \mathbf{q}, \mathbf{X}_1, \mathbf{X}_2) \\ \propto \prod_{t=1}^T \{ z_t f_{1|2}(d_t - \mathbf{x}'_{1t} \beta_1 | q_t - \mathbf{x}'_{2t} \beta_2) f_2(q_t - \mathbf{x}'_{2t} \beta_2) \mathbf{1}_{(q_t, \infty)}(d_t) \\ + (1 - z_t) f_{2|1}(s_t - \mathbf{x}'_{2t} \beta_2 | q_t - \mathbf{x}'_{1t} \beta_1) f_1(q_t - \mathbf{x}'_{1t} \beta_1) \mathbf{1}_{(q_t, \infty)}(s_t) \}, \end{aligned} \quad (93)$$

となる。ここで

$$f_{j|h}(\epsilon_j | \epsilon_h) \propto (\sigma_j^2 - \sigma_{jh}^2 / \sigma_h^2)^{-1/2} \exp \left[-\frac{\{\epsilon_j - (\sigma_{jh} / \sigma_h^2) \epsilon_h\}^2}{2(\sigma_j^2 - \sigma_{jh}^2 / \sigma_h^2)} \right], \quad (94)$$

$$f_h(\epsilon_h) \propto (\sigma_h^2)^{-1/2} \exp \left(-\frac{\epsilon_h^2}{2\sigma_h^2} \right), \quad (95)$$

($j, h=1, 2$) である。さらに

$$\begin{aligned} \zeta_t^{ED} &\equiv \Pr\{d_t > q_t | s_t = q_t\} \\ &= \Pr\{\epsilon_{1t} > q_t - \mathbf{x}'_{1t} \beta_1 | \epsilon_{2t} = q_t - \mathbf{x}'_{2t} \beta_2\} \\ &= \Phi \left[-\frac{q_t - \mathbf{x}'_{1t} \beta_1 - (\sigma_{12} / \sigma_2^2)(q_t - \mathbf{x}'_{2t} \beta_2)}{\sqrt{\sigma_1^2 - \sigma_{12}^2 / \sigma_2^2}} \right], \end{aligned} \quad (96)$$

$$\begin{aligned} \zeta_t^{ES} &\equiv \Pr\{s_t \geq q_t | d_t = q_t\} \\ &= \Pr\{\epsilon_{2t} \geq q_t - \mathbf{x}'_{2t} \beta_2 | \epsilon_{1t} = q_t - \mathbf{x}'_{1t} \beta_1\} \\ &= \Phi \left[-\frac{q_t - \mathbf{x}'_{2t} \beta_2 - (\sigma_{12} / \sigma_1^2)(q_t - \mathbf{x}'_{1t} \beta_1)}{\sqrt{\sigma_2^2 - \sigma_{12}^2 / \sigma_1^2}} \right], \end{aligned} \quad (97)$$

と定義すると、ベイズの定理より超過需要が存在する ($z_t=1$ となる) 条件付事後確率 π_t^{ED} は

$$\pi_t^{ED} = \frac{\zeta_t^{ED} f_2(q_t - \mathbf{x}'_{2t} \beta_2)}{\zeta_t^{ED} f_2(q_t - \mathbf{x}'_{2t} \beta_2) + \zeta_t^{ES} f_1(q_t - \mathbf{x}'_{1t} \beta_1)}, \quad (98)$$

となる。よって各観測値を確率 π_t^{ED} と $\pi_t^{ES} \equiv 1 - \pi_t^{ED}$ で超過需要と超過供給のケースにランダムに分類すればよい。分類ができれば観測されなかった需要 d_t と供給 s_t が分かるので、それを

$$d_t \sim \mathcal{N} \left[\mathbf{x}'_{1t} \beta_1 + \frac{\sigma_{12}}{\sigma_2^2} (q_t - \mathbf{x}'_{2t} \beta_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \right] \mathbf{1}_{(q_t, \infty)}(d_t) \quad \text{if } z_t = 1, \quad (99)$$

$$s_t \sim \mathcal{N} \left[\mathbf{x}'_{2t} \beta_2 + \frac{\sigma_{12}}{\sigma_1^2} (q_t - \mathbf{x}'_{1t} \beta_1), \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \right] \mathbf{1}_{(q_t, \infty)}(s_t) \quad \text{if } z_t = 0, \quad (100)$$

という切断正規分布から発生させる。 (d_t, s_t) がすべて与えられた下では、不均衡モデルは単なる

2 方程式の SUR モデルに帰着されるため、 β_1 , β_2 および攪乱項の分散共分散行列 Σ は SUR モデルのギブズ・サンプラーによって発生することができる。以上を繰り返すことで、不均衡モデルにおけるマルコフ連鎖サンプリングを行うことができる。

6 おわりに

本稿ではマルコフ連鎖モンテカルロ (MCMC) 法によるベイズ分析を回帰モデルへの応用を例にして説明した。今回紹介したモデルは多値選択モデルや制限従属変数モデルなど経済の実証分析では重要なものばかりである。しかし、これ以外にも経済の実証分析で広く使われるモデルに MCMC 法を使ったベイズ分析を応用できるものも数多く存在する。例えば、時系列分析の応用については Bauwens, Lubrano, and Richard [4] や渡部 [2], 生存分析への応用は Ibrahim, Chen, and Sinha [12] に詳しく説明されている。

本稿ではベイズ統計学についての詳しい説明は省略した。ベイズ統計学について詳しく知りたい読者は、Box and Tiao [7] や Zellner [23] といった古典的教科書に加えて、Beger [5], Bernardo and Smith [6], Robert [20] などの解説書を参照されたい。また近年の MCMC 法の急速な普及を反映して、MCMC 法に特化したベイズ分析の本 (例えば Congdon [10]) も出始めている。MCMC 法の解説書としては、本文でも触れた Robert and Casella [21] や Chen et al. [8] などがある。日本語で書かれた MCMC 法に関するサーベイとしては大森 [1] を参照されたい。

(経済学部専任講師)

参 考 文 献

- [1] 大森裕浩「マルコフ連鎖モンテカルロ法の最近の展開」『日本統計学会誌』, 31, 2001, 305-344。
- [2] 渡部敏明『ボラティリティ変動モデル』朝倉書店, 2000。
- [3] Albert, J. H., and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 1991, 669-679.
- [4] Bauwens, L., M. Lubrano, and J. Richard, *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, 1999.
- [5] Berger, J. O., *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, 1985.
- [6] Bernardo, J. M., and A. F. M. Smith, *Bayesian Theory*, Wiley, 1994.
- [7] Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Wiley, 1973.
- [8] Chen, M.-H., Q.-M. Shao, and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, 2000.
- [9] Chib, S., and E. Greenberg, "Analysis of multivariate probit models," *Biometrika*, 85, 1998, 347-361.
- [10] Congdon, P., *Bayesian Statistical Modelling*, Wiley, 2001.

- [11] Geman, S., and D. Grmen, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 1984, 721-741.
- [12] Ibrahim, J. G., M.-H. Chen, and D. Sinha, *Bayesian Survival Analysis*, Springer-Verlag, 2001.
- [13] Hastings, W. K., "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 1970, 97-109.
- [14] Heckman, J. J., "Sample selection bias as a specification error," *Econometrica*, 47, 1979, 153-161.
- [15] Maddala, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, 1983.
- [16] McCulloch, R. E., and P. E. Rossi, "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64, 1994, 207-240.
- [17] McCulloch, R. E., N. G. Polson, and P. E. Rossi, "A Bayesian analysis of the multinomial probit model with fully identified parameters," *Journal of Econometrics*, 99, 2000, 173-193.
- [18] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, 21, 1953, 1087-1092.
- [19] Nobile, A., "Comments: Bayesian multinomial probit models with a normalization constraint," *Journal of Econometrics*, 99, 2000, 335-345.
- [20] Robert, C. P., *The Bayesian Choice*, 2nd ed., Springer-Verlag, 2001.
- [21] Robert, C. P., and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, 1999.
- [22] Tanner, M. A., and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 1987, 528-540.
- [23] Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, Wiley, 1971.