

社会システム分析のための統合化プログラム 33 —最尤推定とMCMC—

福井正康

福山平成大学経営学部経営学科

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム *College Analysis* を作成してきた。今回は最尤推定法によるパラメータの自動推定について議論し、共分散構造分析や非線形最小2乗法のパラメータの初期値設定に MCMC を利用する問題を考えた。また Hamiltonian モンテカルロ法を MCMC による乱数発生に追加した。

キーワード：College Analysis、最尤推定、MCMC、Hamiltonian モンテカルロ法

URL：http://www.heisei-u.ac.jp/ba/fukui/

1. はじめに

我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム *College Analysis* を作成してきた^[1]。今回はこれまでに作成した「分布と検定」のプログラムにパラメータの自動推定機能を追加する問題、非線形最小2乗法や共分散構造分析の初期値設定に MCMC (Markov Chain Monte Carlo) を利用する問題、「MCMC 乱数発生」に Hamiltonian モンテカルロ法を加える問題について考察する。

これまで *College Analysis* の「分布と検定」のプログラムでは、検定したい分布のパラメータについて手動で設定してきた^[2]。しかし、初めからはっきりとパラメータが分かることはあまり考えられないので、自動的に調べる機能は必要である。もちろん他の統計ソフトでも類似した機能がある。我々もこのような形にプログラムを改めることを考えたが、あわせて各分布に対する最尤推定をどのようにするか効率的方法を考えてみた。この報告では我々が利用した方法を失敗例なども含め紹介し、利用者がプログラミングを行う際の助けとしたい。

次に、非線形最小2乗法や共分散構造分析の解を求める際のパラメータの初期値設定の問題について考える。これまでは一様乱数によって初期値を与え、収束するものの中から最適な解を選ぶ方法を採用していたが、共分散構造分析のように多くのパラメータを含む場合、乱数による初期値設定には限界がある。そこで、評価関数を確率分布に見立てた MCMC による初期値設定を考えた。即ち、MCMC により発生させた乱数の中で評価関数を最小にするものを初期値として採用する。問題は乱数発生にかかる時間であるが、正確な乱数を発生させる必要はないので、初期段階で捨てる数や発生乱数の数を少なくして対応できる。これが意外に良い結果をもたらし、うまく収束する初期値が得られることが分かった。

因子分析については、これまで主成分法と主因子法だけで時代遅れの感があったので、今回

共分散構造分析を基にした最小2乗法と最尤法を組み込んだ。初期値設定については、当初MCMCを予定していたが、主成分法で与えた値が良い結果を与えることが分かり、計算速度の関係からそちらを採用した。

最後に、MCMCの乱数発生法で、我々はこれまでMetropolis-Hastings法を用いてきたが、Hamiltonianモンテカルロ法も実際に扱ってみた。パラメータが予想外に大きな値の時に早く乱数域に到達でき、採択率も向上するということがあったので、二つの方法はどちらが優れているのか比較してみた。

2. 最尤推定法

2.1 分布と検定での利用

得られたデータ x_λ ($\lambda = 1, \dots, N$) が、特定の分布に従うかどうかを調べる際、分布のパラメータが既知であるとは限らない。多くの場合、与えられたデータを用いて各種分布のパラメータを推定し、その後に検定の問題を考えることになると思われる。そのため、メニュー「分析-基本統計-分布と検定」で表示される図1の分析実行画面に、パラメータを自動的に推定する機能を加えた。分布を選んで「推定」ボタンをクリックすると左のテキストボックスに推定値が表示される。

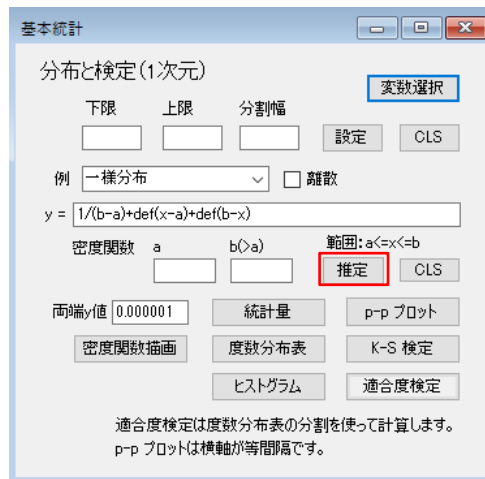


図1 分布と検定実行画面

ここでは少し長くなるが、分布毎に、密度関数、尤度関数、対数尤度関数、スコアベクトル、情報行列を具体的に示し、このプログラムで使っているパラメータの推定法を具体的に与えておく。

正規分布 ($-\infty < x < \infty$)

$$\text{密度関数: } f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x_\lambda - \mu)^2 / 2\sigma^2]$$

$$\text{尤度関数: } L = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{\lambda=1}^N (x_\lambda - \mu)^2\right]$$

$$\text{対数尤度: } \log L = -\frac{1}{2\sigma^2} \sum_{\lambda=1}^N (x_\lambda - \mu)^2 - \frac{N}{2} \log(2\pi\sigma^2)$$

スコアベクトル \mathbf{U} と情報行列 \mathfrak{I}

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \partial \log L / \partial \mu \\ \partial \log L / \partial \sigma^2 \end{pmatrix},$$

$$\mathfrak{I} = - \begin{pmatrix} \partial^2 \log L / \partial \mu^2 & \partial^2 \log L / \partial \mu \partial \sigma^2 \\ \partial^2 \log L / \partial \mu \partial \sigma^2 & \partial^2 \log L / \partial (\sigma^2)^2 \end{pmatrix}$$

$$\partial \log L / \partial \mu = \frac{1}{\sigma^2} \sum_{\lambda=1}^N (x_\lambda - \mu) = 0$$

$$\mu = \frac{1}{N} \sum_{\lambda=1}^N x_\lambda$$

$$\partial \log L / \partial \sigma^2 = \frac{1}{2\sigma^4} \sum_{\lambda=1}^N (x_\lambda - \mu)^2 - \frac{N}{2\sigma^2} = 0$$

$$\sigma^2 = \frac{1}{N} \sum_{\lambda=1}^N (x_\lambda - \mu)^2$$

この場合、最後の2式によってパラメータを解析的に求めることが可能であるが、プログラムでは練習としてニュートン・ラフソン法を用いて計算を試している。

$$\partial^2 \log L / \partial \mu^2 = -\frac{N}{\sigma^2}$$

$$\partial^2 \log L / \partial \mu \partial \sigma^2 = -\frac{1}{\sigma^4} \sum_{\lambda=1}^N (x_\lambda - \mu) \rightarrow 0$$

$$\partial^2 \log L / \partial (\sigma^2)^2 = -\frac{1}{\sigma^6} \sum_{\lambda=1}^N (x_\lambda - \mu)^2 + \frac{N}{2\sigma^4} \rightarrow -\frac{N}{2\sigma^4}$$

初期値は収束の状態を見るため、故意に $\mu_0 = 0$, $\sigma_0^2 = 1$ を用いている。

χ^2 分布 ($0 < x < \infty$)

パラメータが離散的なので、特別な方法を利用する。

$$\text{密度関数: } f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} \exp(-x/2)$$

$$\text{尤度関数: } L = \frac{1}{2^{Nn/2} \Gamma(n/2)^N} \prod_{\lambda=1}^N x_\lambda^{n/2-1} \exp(-x_\lambda/2)$$

$$\text{対数尤度: } \log L = \frac{n-2}{2} \sum_{\lambda=1}^N \log(x_\lambda) - \frac{1}{2} \sum_{\lambda=1}^N x_\lambda - \frac{Nn}{2} \log 2 - N \log \Gamma(n/2)$$

$\chi^2 \square \chi_n^2$ のとき、 $E(\chi^2) = n$ の性質を用いて、

$$n = \lfloor x + 0.5 \rfloor \quad \text{注) } \lfloor x \rfloor \text{ は } x \text{ を越えない最大の整数}$$

これを元に $(1 \leq) n - 5 \leq n_{\max} \leq n + 5$ の範囲で最大の対数尤度を与える自由度 n_{\max} を求め
ている。

F分布 ($0 < x < \infty$)

この分布もパラメータが離散的なので、特別な方法を利用する。

$$\text{密度関数: } f(x) = \frac{1}{B(n_1/2, n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{x^{n_1/2-1}}{(1+xn_1/n_2)^{(n_1+n_2)/2}}$$

$$\text{尤度関数: } L = \frac{1}{B(n_1/2, n_2/2)^N} \left(\frac{n_1}{n_2}\right)^{Nn_1/2} \prod_{\lambda=1}^N \frac{x_\lambda^{n_1/2-1}}{(1+x_\lambda n_1/n_2)^{(n_1+n_2)/2}}$$

$$\begin{aligned} \text{対数尤度: } \log L &= \left(\frac{n_1}{2} - 1\right) \sum_{\lambda=1}^N \log(x_\lambda) - \frac{n_1 + n_2}{2} \sum_{\lambda=1}^N \log(1 + x_\lambda n_1/n_2) \\ &\quad + \frac{Nn_1}{2} \log(n_1/n_2) - N \log B(n_1/2, n_2/2) \end{aligned}$$

$$E[X] = \frac{n_2}{n_2 - 2} \quad (n_2 > 2), \quad V[X] = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)} \quad (n_2 > 4) \quad \text{を利用して、}$$

$$n_2 = \frac{2E[X]}{E[X] - 1}, \quad n_1 = \frac{2n_2^2(n_2 - 2)}{(n_2 - 2)^2(n_2 - 4)V[X] - 2n_2^2}$$

これを元に、ぶれが大きいと思われるので、 $(1 \leq) n_i - 20 \leq n_{i\max} \leq n_i + 20$ の範囲で対数
尤度を最大化する $n_{i\max}$ を求めている。

t分布 ($-\infty < x < \infty$)

この分布もパラメータが離散的なので、特別な方法を利用する。

$$\text{密度関数: } f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

$$\text{尤度関数: } L = \left(\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\right)^N \prod_{\lambda=1}^N \left(1 + \frac{x_\lambda^2}{n}\right)^{-(n+1)/2}$$

$$\text{対数尤度: } \log L = -\frac{n+1}{2} \sum_{\lambda=1}^N \log\left(1 + \frac{x_\lambda^2}{n}\right) + N \log\left(\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\right)$$

平均： $E[X] = 0$

分散： $V[X] = \frac{n}{n-2}$ を利用して、 $n = \frac{2V[X]}{V[X]-1}$

これを元に $(1 \leq) n - 5 \leq n_{\max} \leq n + 5$ の範囲で最大の対数尤度を与える自由度 n_{\max} を求めている。

ガンマ分布 ($0 < x < \infty$)

密度関数： $f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} \exp(-x/b)$

尤度関数： $L = \frac{1}{[b^a \Gamma(a)]^N} \prod_{\lambda=1}^N x_{\lambda}^{a-1} \exp(-x_{\lambda}/b)$

対数尤度： $\log L = (a-1) \sum_{\lambda=1}^N \log x_{\lambda} - \frac{1}{b} \sum_{\lambda=1}^N x_{\lambda} - Na \log b - N \log \Gamma(a)$

$\partial \log L / \partial a = \sum_{\lambda=1}^N \log x_{\lambda} - N \log b - N \Gamma'(a) / \Gamma(a)$

$\partial \log L / \partial b = 1/b^2 \sum_{\lambda=1}^N x_{\lambda} - Na/b$

$\partial^2 \log L / \partial a^2 = -N \left[\Gamma''(a) / \Gamma(a) - \Gamma'(a)^2 / \Gamma(a)^2 \right]$

$\partial^2 \log L / \partial a \partial b = -N/b$

$\partial^2 \log L / \partial b^2 = -2/b^3 \sum_{\lambda=1}^N x_{\lambda} + Na/b^2$

初期値は $a_0 = 0.5$, $b_0 = 0.5$ を用いている。

逆ガンマ分布 ($0 \leq x \leq 1$)

密度関数： $f(x) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-b/x)$

尤度関数： $L = \left[b^a / \Gamma(a) \right]^N \prod_{\lambda=1}^N x_{\lambda}^{-a-1} \exp(-b/x_{\lambda})$

対数尤度： $\log L = -(a+1) \sum_{\lambda=1}^N \log x_{\lambda} - b \sum_{\lambda=1}^N (1/x_{\lambda}) + Na \log b - N \log \Gamma(a)$

$\partial \log L / \partial a = - \sum_{\lambda=1}^N \log x_{\lambda} + N \log b - N \Gamma'(a) / \Gamma(a)$

$$\partial \log L / \partial b = -\sum_{\lambda=1}^N (1/x_{\lambda}) + N a/b$$

$$\partial^2 \log L / \partial a^2 = -N \left[\Gamma''(a)/\Gamma(a) - \Gamma'(a)^2/\Gamma(a)^2 \right]$$

$$\partial^2 \log L / \partial a \partial b = N/b, \quad \partial^2 \log L / \partial b^2 = -N a/b^2$$

初期値は $a_0 = 0.5$, $b_0 = 0.5$ を用いている。

ベータ分布 ($0 \leq x \leq 1$)

$$\text{密度関数: } f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

$$\text{尤度関数: } L = \frac{1}{B(a,b)} \prod_{\lambda=1}^N x_{\lambda}^{a-1}(1-x_{\lambda})^{b-1}$$

$$\text{対数尤度: } \log L = (a-1) \sum_{\lambda=1}^N \log x_{\lambda} + (b-1) \sum_{\lambda=1}^N \log(1-x_{\lambda}) - N \log B(a,b)$$

$$\partial \log L / \partial a = \sum_{\lambda=1}^N \log x_{\lambda} - N \frac{B(a,b)_a}{B(a,b)}$$

$$\partial \log L / \partial b = \sum_{\lambda=1}^N \log(1-x_{\lambda}) - N \frac{B(a,b)_b}{B(a,b)}$$

$$\partial^2 \log L / \partial a^2 = -N \left(\frac{B(a,b)_{aa}}{B(a,b)} - \frac{B(a,b)_a^2}{B} \right)$$

$$\partial^2 \log L / \partial a \partial b = -N \left(\frac{B(a,b)_{ab}}{B(a,b)} - \frac{B(a,b)_a B(a,b)_b}{B(a,b)^2} \right)$$

$$\partial^2 \log L / \partial b^2 = -N \left(\frac{B(a,b)_{bb}}{B(a,b)} - \frac{B(a,b)_b^2}{B} \right)$$

ここで右下の文字は、そのパラメータでの微分を表し、2つある場合は2階微分を表す。

初期値の設定で、平均値が0に近い場合は $a = 1, b = 5$ 、1に近い場合は $a = 5, b = 1$ 、0.5に近い場合は $a = 0.5, b = 0.5$ などを使う。パラメータの値を小さい方から大きい方へ近づけて行くことは問題ないが、大きい方から小さい方へ近づけて行く際にはエラーが出る可能性がある。

ワイブル分布 ($0 < x < \infty$)

最初に通常のパラメータ a, b を使って最尤法を試みた。

密度関数： $f(x) = (a/b)(x/b)^{a-1} \exp[-(x/b)^a]$

$$L = (a/b)^N \prod_{\lambda=1}^N (x_{\lambda}/b)^{a-1} \exp[-(x_{\lambda}/b)^a]$$

尤度関数：

$$= a^N b^{-Na} \prod_{\lambda=1}^N x_{\lambda}^{a-1} \exp[-x_{\lambda}^a b^{-a}]$$

対数尤度： $\log L = (a-1) \sum_{\lambda=1}^N \log x_{\lambda} - b^{-a} \sum_{\lambda=1}^N x_{\lambda}^a + N \log a - Na \log b$

$$\begin{aligned} \partial \log L / \partial a &= \sum_{\lambda=1}^N \log x_{\lambda} + \log b \cdot b^{-a} \sum_{\lambda=1}^N x_{\lambda}^a - b^{-a} \sum_{\lambda=1}^N \log x_{\lambda} \cdot x_{\lambda}^a \\ &\quad + N/a - N \log b \end{aligned}$$

$$\partial \log L / \partial b = ab^{-a-1} \sum_{\lambda=1}^N x_{\lambda}^a - Na/b$$

$$\begin{aligned} \partial^2 \log L / \partial a^2 &= -(\log b)^2 b^{-a} \sum_{\lambda=1}^N x_{\lambda}^a + 2 \log b \cdot b^{-a} \sum_{\lambda=1}^N \log x_{\lambda} \cdot x_{\lambda}^a \\ &\quad - b^{-a} \sum_{\lambda=1}^N (\log x_{\lambda})^2 \cdot x_{\lambda}^a - N/a^2 \end{aligned}$$

$$\partial^2 \log L / \partial a \partial b = (1 - a \log b) b^{-a-1} \sum_{\lambda=1}^N x_{\lambda}^a + ab^{-a-1} \sum_{\lambda=1}^N \log x_{\lambda} \cdot x_{\lambda}^a - N/b$$

$$\partial^2 \log L / \partial b^2 = -a(a+1)b^{-a-2} \sum_{\lambda=1}^N x_{\lambda}^a + Na/b^2$$

この方法は、収束が思うように行かず、エラーとなった。

上記の失敗を踏まえ、生存時間分析で用いた $b^{-a} = e^{\beta}$ というパラメータの変換を利用した推定法を利用した。

密度関数： $f(x) = (a/b)(x/b)^{a-1} \exp[-(x/b)^a]$

$$L = (a/b)^N \prod_{\lambda=1}^N (x_{\lambda}/b)^{a-1} \exp[-(x_{\lambda}/b)^a]$$

尤度関数：

$$= a^N b^{-Na} \prod_{\lambda=1}^N x_{\lambda}^{a-1} \exp[-x_{\lambda}^a b^{-a}] = a^N e^{N\beta} \prod_{\lambda=1}^N x_{\lambda}^{a-1} \exp[-x_{\lambda}^a e^{\beta}]$$

$$\begin{aligned} \log L(a, b) &= \sum_{\lambda=1}^N \left[(\log a + (a-1) \log x_{\lambda} + \beta) - x_{\lambda}^a e^{\beta} \right] \\ &= (a-1) \sum_{\lambda=1}^N \log x_{\lambda} - e^{\beta} \sum_{\lambda=1}^N x_{\lambda}^a + N \log a + N \beta \\ \frac{\partial}{\partial a} \log L &= \sum_{\lambda=1}^N \log x_{\lambda} - e^{\beta} \sum_{\lambda=1}^N \log x_{\lambda} \cdot x_{\lambda}^a + N/a \\ \frac{\partial}{\partial \beta} \log L &= -e^{\beta} \sum_{\lambda=1}^N x_{\lambda}^a + N \\ \frac{\partial^2}{\partial a^2} \log L &= -e^{\beta} \sum_{\lambda=1}^N (\log x_{\lambda})^2 x_{\lambda}^a - N/a^2 \\ \frac{\partial}{\partial a \partial \beta} \log L &= -e^{\beta} \sum_{\lambda=1}^N \log x_{\lambda} x_{\lambda}^a, \quad \frac{\partial^2}{\partial \beta^2} \log L = -e^{\beta} \sum_{\lambda=1}^N x_{\lambda}^a \end{aligned}$$

初期値は $a_0 = 2$, $\beta = 2$ を用いている。

指数分布 ($0 < x < \infty$)

密度関数: $f(t) = a \exp(-ax) \quad (x \geq 0)$

尤度関数: $L = a^N \prod_{i=1}^N \exp(-ax_{\lambda}) = a^N \exp\left(-a \sum_{\lambda=1}^N x_{\lambda}\right)$

対数尤度: $\log L = N \log a - a \sum_{i=1}^N x_{\lambda}$

$\frac{\partial}{\partial a} \log L = \frac{N}{a} - \sum_{\lambda=1}^N x_{\lambda} = 0$ これより、 $a = N / \sum_{\lambda=1}^N x_{\lambda}$

$\frac{\partial^2}{\partial a^2} \log L = -\frac{N}{a^2}$

推定値は解析的に求まるが、練習として最尤法を利用した。初期値は $a = 0.1$ を用いている。

ポアソン分布 ($x = 0, 1, 2, \dots$)

確率関数: $P(x) = e^{-a} a^x / x!$

尤度関数: $L = \prod_{\lambda=1}^N e^{-a} a^{x_{\lambda}} / x_{\lambda}! = e^{-Na} \prod_{\lambda=1}^N a^{x_{\lambda}} / x_{\lambda}!$

対数尤度: $\log L = -Na + \log a \sum_{\lambda=1}^N x_{\lambda} - \sum_{\lambda=1}^N \log x_{\lambda}!$

$$\partial \log L / \partial a = -N + \frac{1}{a} \sum_{\lambda=1}^N x_{\lambda}, \quad \partial^2 \log L / \partial a^2 = -\frac{1}{a^2} \sum_{\lambda=1}^N x_{\lambda}$$

初期値は $a = 0.1$ を用いている。

2項分布 ($x = 0, 1, 2, \dots, n$)

まず以下の関係を使って、度数 n を求める。

$$E[X] = np, \quad V[X] = npq, \quad n = \frac{E[X]^2}{E[X] - V[X]}$$

次に最尤法を使って、確率 p を求める。

$$\text{確率関数: } P(x) = {}_n C_x p^x (1-p)^{n-x}$$

$$\text{尤度関数: } L = \prod_{\lambda=1}^N {}_n C_{x_{\lambda}} p^{x_{\lambda}} (1-p)^{n-x_{\lambda}}$$

$$\text{対数尤度: } \log L = \sum_{\lambda=1}^N \log {}_n C_{x_{\lambda}} + \log p \sum_{\lambda=1}^N x_{\lambda} + \log(1-p) \sum_{\lambda=1}^N (n-x_{\lambda})$$

$$\partial \log L / \partial p = \frac{1}{p} \sum_{\lambda=1}^N x_{\lambda} - \frac{1}{1-p} \sum_{\lambda=1}^N (n-x_{\lambda})$$

$$\partial^2 \log L / \partial p^2 = -\frac{1}{p^2} \sum_{\lambda=1}^N x_{\lambda} - \frac{1}{(1-p)^2} \sum_{\lambda=1}^N (n-x_{\lambda})$$

これも解析的に解を求めることができるが、最尤法の練習とする。

初期値は $p = 0.5$ を用いている。

以上の計算は教科書的であり、分析としてはあまり意味がないが、今後のプログラミングの教材用に残しておく。

2.2 共分散構造分析での利用

共分散構造分析での最尤法については参考文献 [3]や[4] で詳しく述べられているが、ここで簡単に振り返っておく。我々はまず観測値を与える確率変数 \mathbf{x}_{λ} ($\lambda = 1, 2, \dots, N$) がそれぞれ独立に n 変量正規分布に従うと考える。パラメータから作られる共分散行列を $\Sigma(\boldsymbol{\theta})$ とすると、 \mathbf{x}_{λ} の確率密度関数は以下で与えられる。

$$f(\mathbf{x}_{\lambda}, \boldsymbol{\theta}) = (2\pi)^{-p/2} |\Sigma(\boldsymbol{\theta})|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}_{\lambda} - \boldsymbol{\mu})' \Sigma(\boldsymbol{\theta})^{-1} (\mathbf{x}_{\lambda} - \boldsymbol{\mu})\right]$$

N 回の独立な観測に関する確率密度関数は以下で与えられる。

$$f(\mathbf{x}, \boldsymbol{\theta}) = \prod_{\lambda=1}^N f(\mathbf{x}_{\lambda}, \boldsymbol{\theta})$$

最尤法ではこの確率密度関数に実測値 $\hat{\mathbf{x}}_{\lambda}$ を代入した尤度関数 $f(\boldsymbol{\theta})$ を最大化するようにパラ

メータを決定する。実際には計算の簡単化のため、尤度関数を対数変換した対数尤度関数の符号を変えたものを最小化する。符号を変えた対数尤度関数は以下で与えられる。

$$\begin{aligned} -\log f(\boldsymbol{\theta}) &= -\sum_{\lambda=1}^N \log f(\hat{\mathbf{x}}_{\lambda}, \boldsymbol{\theta}) \\ &= \frac{1}{2} \sum_{\lambda=1}^N (\hat{\mathbf{x}}_{\lambda} - \bar{\mathbf{x}})' \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\hat{\mathbf{x}}_{\lambda} - \bar{\mathbf{x}}) - \frac{N}{2} \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}| + const. \\ &= \frac{N}{2} \left(\text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{S}) - \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}| \right) + const. \end{aligned}$$

ここに、

$$\mathbf{S} = \frac{1}{N} \sum_{\lambda=1}^N (\hat{\mathbf{x}}_{\lambda} - \bar{\mathbf{x}})(\hat{\mathbf{x}}_{\lambda} - \bar{\mathbf{x}})'$$

通常最尤法の評価関数としては、上の対数尤度関数に定数を加えた以下の式が用いられることが多い。

$$f_{ML}(\boldsymbol{\theta}) = \text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{S}) - \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{S}| - n$$

これは $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ と \mathbf{S} が一致する場合 0 となる。

これらの評価関数の最小化法には様々な方法が用いられるが、現在我々は、パラメータの初期値設定に MCMC の Metropolis-Hastings 法、パラメータの推定に Newton-Raphson 法を発展させた Levenberg-Marquart 法を用いている。

2.3 因子分析での利用

因子分析における最尤法は、図 1 のようなデータ構造のモデルを、共分散構造分析の最尤法を用いて分析することである。

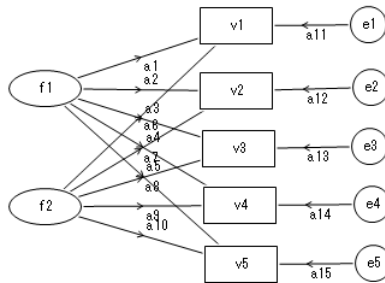


図 2 因子分析の構造モデル

観測変数の数が比較的小さい場合はモデルが識別されない状況が生じる。因子分析の場合は観測変数が多い（即ちパラメータ数も多い）分析も見られることから、パラメータの初期設定に前節で述べたような MCMC を使うのはあまり得策ではない。我々は主成分分析法を用いてパラメータの初期設定を行っているが、現在のところ時間的にも解の収束も良好な結果が得られている。

3. MCMC による乱数発生

共分散構造分析の初期値設定について、College Analysis ではこれまで乱数による設定を用いてきた。そのため、パラメータの数が多い場合には、最適な初期値を得ることが難しく、計算に時間がかかったり、正しい解が得られなかったりしたことがあった。そこで今回我々は共分散構造分析のモデルを、変数を標準化した相関係数モデルに限り、MCMC によって初期値を求める方法を導入した。即ち、MCMC によって密度関数が評価関数に比例する乱数を発生させ、その最小値（密度関数としては最大値）をとるパラメータを初期値とする方法である。

MCMC の酔歩乱数による Metropolis-Hastings 法^[5]（以後 MH 法と略す）では、パラメータが初期値と大きく異なると、その後のステップで、なかなか乱数発生域に到達できないという問題点があった。しかし、標準化モデルではパラメータはほぼ ± 1 の範囲に入り、初期値設定には向いている。

一方、非線形最小 2 乗法では、パラメータの数は少ないが、値がどれほどになるか予想できない。そのため MH 法では不十分なように思われ、乱数発生による初期値設定のメニューも残すようにしている。しかし、MCMC の中には Hamiltonian モンテカルロ法^[6]（以後 HMC 法と略す）という手法があり、このような場合には有効であると言われている。

我々はこのような経緯を背景として、MCMC 乱数発生のプログラムに HMC 法も導入した。これによって 2 つのモンテカルロ法の違いが利用者に理解できるのではないかと考える。HMC 法の特徴として、乱数発生のスピードは遅いが、精度は MH 法に勝るのではないと思われる。問題は遠くの乱数域に早く到達するかどうかであるが、まだ結論は得られていない。問題は 2 つの方法とも、密度関数の大きさの比を利用する点である。乱数域から遠く離れていると、理論上密度関数は小さな値を持っているが、数値計算上は丸め誤差により、その値が 0 になる。計算上この値での割り算が現れることから、計算不能の状態が発生する。我々はこれを回避するための方法も考える必要がある。ここではこれら 2 つの方法をできるだけ平易に説明し、プログラムの実行結果を比較する。

3.1 MH 法の考え方

過去のデータから順次確率的に決まって行く時系列のデータを $x^{(t)}$ とし、このデータの従う分布の密度関数を $f^{(t)}(x)$ とする。このデータの決定過程を確率過程というが、マルコフ連鎖とは、1 期先のデータ $x^{(t+1)}$ が、それまでのデータの履歴によらず、 $x^{(t)}$ の値だけから推移確率（推移核） $p(x^{(t+1)}|x^{(t)})$ によって、決まるような確率過程をいう。ある一般的な条件のマルコフ連鎖に従うデータは、時間と共に、一定の分布 $f(x)$ に推移することが知られている。この性質を利用して乱数を発生させる方法がマルコフ連鎖モンテカルロ法である。

ここでは密度関数 $y = f(x)$ の乱数を発生させる問題を考える。下図のように、まずある点 x_1 を起点として平均 0 の正規分布（特にこれに限らないが非対称な分布の場合は下の式が異なる）の乱数を 1 つ発生し、その点を x_2 とする。

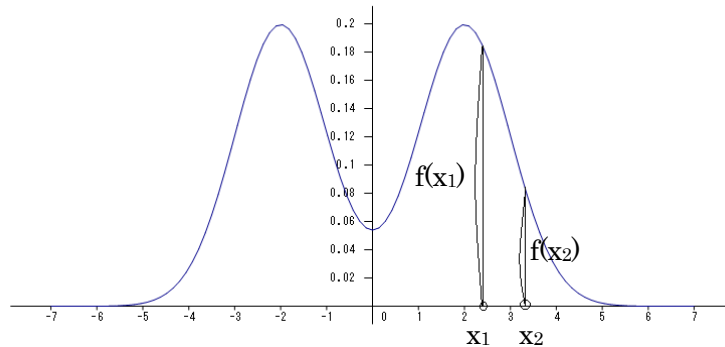


図1 乱数発生

次にこの値と2点での関数の高さから、以下のような量を計算する。

$$\alpha = \begin{cases} \min[f(x_2)/f(x_1), 1] & \text{分母} > 0 \\ 1 & \text{分母} = 0 \end{cases}$$

コンピュータで、 $[0, 1]$ 区間の一様乱数を発生させ、その値がこの α より小さいなら x_2 を採択し、大きいなら、改めて x_1 より再度やり直す。(これは確率 α で x_2 を採択すると言ってもよい。)

これを続けて行くと、最終的に密度関数が $y = f(x)$ で与えられる乱数に収束することになる。なぜなら密度関数の高さが高い x_1 の点から高さの低い $x_2 \pm d/2$ の点に遷移する確率は $f(x_2)/f(x_1)d$ 、高さの低い x_2 の点から高さの高い $x_1 \pm d/2$ の点に遷移する確率は d になり、両者の比 $f(x_2)/f(x_1):1 = f(x_2):f(x_1)$ は、推移する先の点の密度関数の高さの比になる。これがすべての2点で成り立っているため、各点の出現比率は密度関数の高さに比例する。即ち、これは密度関数で与えられる分布の乱数を発生したことになる。この方法は n 次元の場合にも容易に拡張できる。

実際の計算では、正しい乱数にするために、最初のいくつかの点(通常1000点以上)は捨てるのが望ましい。

3.2 HMC法の考え方

MCMCによる乱数発生では、初期値の設定は重要である。MH法の酔歩乱数では、正規分布を使って最尤値に1歩ずつ近づけていくために、初期値が最尤値から離れた位置だと大きな標準偏差が必要である。しかし、最尤値に近いところで良い精度を出そうとすると適当な大きさの標準偏差が必要となる。これらの相反する条件を解決する手法として期待されるのがHMC法である。

HMC法は、変数を q_α ($\alpha = 1, 2, \dots, n$)とした目的の分布と、変数を p_α ($\alpha = 1, 2, \dots, n$)とした独立な標準正規分布を合成した分布の密度関数を、力学のHamiltonian H の関数式 e^{-H} とみなし、Hamiltonian (エネルギー)の保存則を利用して変数 q_α を決めて行く方法で

ある。ここでは初めから n 次元の問題として解説する。

今、発生させたい乱数の密度関数を $f(\mathbf{q})$ とし、それに独立な標準正規分布の密度関数を $g(\mathbf{p}) = 1/(2\pi)^{n/2} \exp(-\sum p_\alpha^2/2)$ とすると、合成関数 $f(\mathbf{q})g(\mathbf{p})$ は以下ようになる。

$$f(\mathbf{q})g(\mathbf{p}) \propto \exp\left[-h(\mathbf{q}) - \sum p_\alpha^2/2\right] \equiv \exp[-H(\mathbf{q}, \mathbf{p})]$$

ここに、 $h(\mathbf{q}) = -\log f(\mathbf{q})$ はポテンシャルエネルギー、 $\sum p_\alpha^2/2$ は質点の運動エネルギーに相当する。但し、質点の質量はすべて 1 としている。この Hamiltonian の元で、運動方程式は以下となる。

$$\frac{dp_\alpha}{dt} = -\frac{\partial H}{\partial q_\alpha}, \quad \frac{dq_\alpha}{dt} = \frac{\partial H}{\partial p_\alpha} = p_\alpha$$

この運動に際して、Hamiltonian は以下のように不変である。

$$\frac{dH}{dt} = \sum_{\alpha=1}^n \left[\frac{dq_\alpha}{dt} \frac{\partial H}{\partial q_\alpha} + \frac{dp_\alpha}{dt} \frac{\partial H}{\partial p_\alpha} \right] = \sum_{\alpha=1}^n \left[-\frac{dq_\alpha}{dt} \frac{dp_\alpha}{dt} + \frac{dp_\alpha}{dt} \frac{dq_\alpha}{dt} \right] = 0$$

Hamiltonian の不変性から、2つの時点 t, t' ($t < t'$) で関数間の関係は以下となる。

$$f(\mathbf{q}')g(\mathbf{p}') = f(\mathbf{q})g(\mathbf{p})$$

ここに、上式では以下のように時間 t', t が略されている。

$$\mathbf{q}' = \mathbf{q}(t'), \mathbf{p}' = \mathbf{p}(t'), \quad \mathbf{q} = \mathbf{q}(t), \mathbf{p} = \mathbf{p}(t)$$

我々は変数 \mathbf{q} を初期値として与え、独立な n 個の正規乱数を発生させ、それを変数 \mathbf{p} の初期値とする。これらを使ってハミルトンの運動方程式を解き、新しい変数 \mathbf{q}', \mathbf{p}' を求める。その際、位置 \mathbf{q} で $\mathbf{p} \pm \mathbf{d}/2$ の乱数を発生させる確率は $g(\mathbf{p})d^n$ であるため、位置 \mathbf{q}' の近傍に到達する確率も $g(\mathbf{p})d^n$ である。またこの過程を逆にたどることを考えると、位置 \mathbf{q}' で $\mathbf{p}' \pm \mathbf{d}/2$ の乱数を発生させ、位置 \mathbf{q} の近傍に到達する確率は $g(\mathbf{p}')d^n$ であるため、位置 \mathbf{q} から位置 \mathbf{q}' に到達する確率とその逆の確率の比は $g(\mathbf{p}) : g(\mathbf{p}')$ となる。ここで、上に述べた関係 $f(\mathbf{q}')g(\mathbf{p}') = f(\mathbf{q})g(\mathbf{p})$ を使うと、この比は $f(\mathbf{q}') : f(\mathbf{q})$ となり、到達する位置の発生させたい密度関数の大きさに比例することになる。これがすべての2つの位置の間で成り立っていることから、 \mathbf{q} の値が得られる確率は $f(\mathbf{q})$ に比例する。これは密度関数 $f(\mathbf{q})$ で乱数が発生したことになる。

この手法はマルコフ連鎖を意識して利用しているわけではないが、関連を考えてみよう。マルコフ連鎖では1つの状態 (\mathbf{p}, \mathbf{q}) から他の状態 $(\mathbf{p}', \mathbf{q}')$ に推移する場合、推移は推移核

$S(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p})$ を用いて以下の形で表される。

$$f(\mathbf{q}')g(\mathbf{p}') = S(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p})f(\mathbf{q})g(\mathbf{p})$$

運動が可逆過程であることから、推移も可逆的となり、

$$f(\mathbf{q})g(\mathbf{p}) = S(\mathbf{q}, \mathbf{p} | \mathbf{q}', \mathbf{p}')f(\mathbf{q}')g(\mathbf{p}')$$

これらの関係より、

$$S(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p}) = S(\mathbf{q}, \mathbf{p} | \mathbf{q}', \mathbf{p}') = 1 \quad (\text{確率 } 1 \text{ でこの推移が起こる})$$

$$S(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p}) f(\mathbf{q}) g(\mathbf{p}) = S(\mathbf{q}, \mathbf{p} | \mathbf{q}', \mathbf{p}') f(\mathbf{q}') g(\mathbf{p}')$$

となり、詳細つり合い条件は自動的に満たされる。

この推移を実際に計算するには、オイラー法を拡張したリープ・フロッグ法を用いる。その際、微分を差分で置き換えるため誤差が生じ、以下のような関係になるとする。

$$f(\mathbf{q}') g(\mathbf{p}') = r f(\mathbf{q}) g(\mathbf{p})$$

これを補正するために MH 法の考え方を利用する。

$$\alpha(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p}) = \begin{cases} \min \left[\frac{f(\mathbf{q}') g(\mathbf{p}')}{f(\mathbf{q}) g(\mathbf{p})} = r, 1 \right] & \text{分母} > 0 \\ 1 & \text{分母} = 0 \end{cases}$$

即ちリープ・フロッグ法で新しく得られた変数 \mathbf{q}' については、上を与えた確率 $\alpha(\mathbf{q}', \mathbf{p}' | \mathbf{q}, \mathbf{p})$ で採択の可否を決める。 r は 1 に近い値のため、採択率はかなり高くなる。これによって、原理的には 1 回の乱数発生で通常の MH 法の移動よりも遠くに移動できる。

最後に、オイラー法とリープ・フロッグ法の計算法を与えておく。

n 次元オイラー法

$$p_\alpha(t+1) = p_\alpha(t) - \varepsilon \partial h / \partial q_\alpha$$

$$q_\alpha(t+1) = q_\alpha(t) + \varepsilon p_\alpha(t)$$

n 次元リープ・フロッグ法

$$p_\alpha(2t+1) = p_\alpha(2t) - (\varepsilon/2) dh/dq_\alpha \Big|_{q(2t)}$$

$$q_\alpha(2t+2) = p_\alpha(2t) + \varepsilon p_\alpha(2t+1)$$

$$p_\alpha(2t+2) = p_\alpha(2t+1) - (\varepsilon/2) dh/dq_\alpha \Big|_{q(2t+2)}$$

さて、実際の計算での HMC 法の使い勝手はどうであろうか。標準正規乱数を発生させるごとにリープ・フロッグ法を用いるため、計算量（特に微分の部分）がかなり多くなる。採択率は上がるが、その分計算量が増えるため、計算時間は MH 法に比べて長くなっている。しかし、乱数の精度から見ると改良されているのではないかと思われる。

3.3 2つの方法の比較

MCMC による乱数生成の新しい実行画面は、メニュー [分析-基本統計-ユーティリティ-MCMC 乱数生成] を選択すると図 2 のように示される。

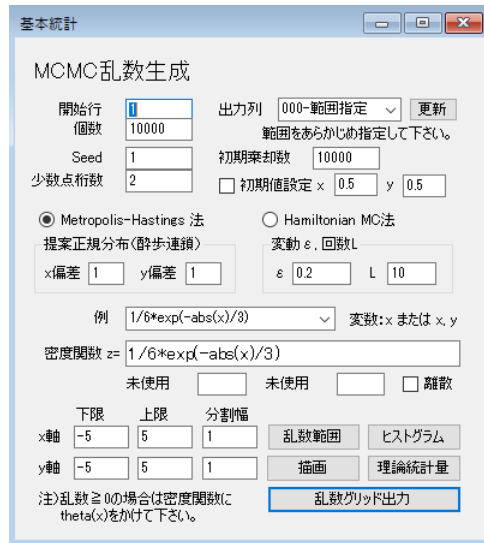


図2 MCMC 乱数生成実行画面

このメニューにある密度関数 $z = \exp(-|x|/3)/6$ について、MH法を使って、棄却数を10000個として、10000個の乱数を発生させたヒストグラムと統計量を図3に示す。

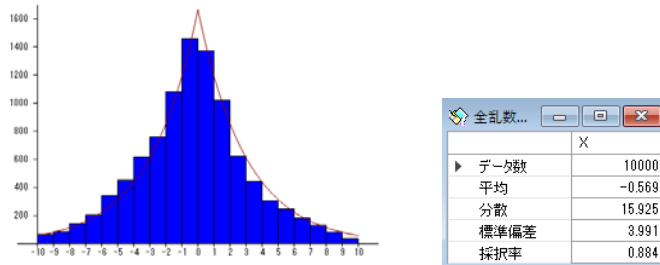


図3 MH法による乱数発生

同様に、HMC法を使って、同じ棄却数で10000個の乱数を発生させたヒストグラムと統計量を図4に示す。但し、変動の微量量 ϵ を0.2、リープ・フロッグの回数を10としている。

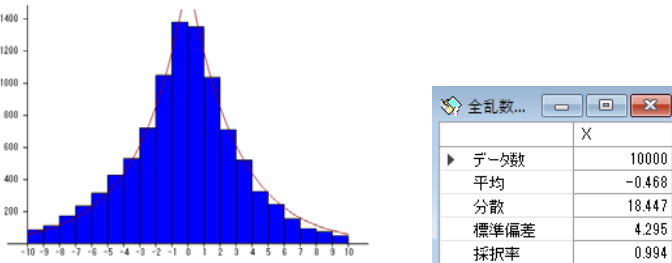


図4 MCH法による乱数発生

どちらの適合性が高いか調べるために、2つの方法で10回ずつ10000個の乱数を発生させ、適合指標としてKS検定のK統計量を比較する。結果を表1に示す。値の小さな方が適合性が良

いと判断できる。

表1 K統計量によるMH法とHMC法の適合度の違い

	MH法	HMC法
1	5.217	3.504
2	6.400	1.971
3	1.712	1.339
4	6.556	2.645
5	2.88	1.13
6	4.626	3.224
7	4.416	0.857
8	2.331	1.533
9	3.353	2.23
10	1.253	1.255

これを比較すると、有意に ($p < 0.05$) 差が見られた。

次に、メニューの中にある密度関数 $z = \frac{1}{2\sqrt{2\pi}} \left[e^{-(x+2)^2/2} + e^{-(x-2)^2/2} \right]$ について、

MH法を使って、棄却数を10000個として、10000個の乱数を発生させたヒストグラムと統計量を図5に示す。

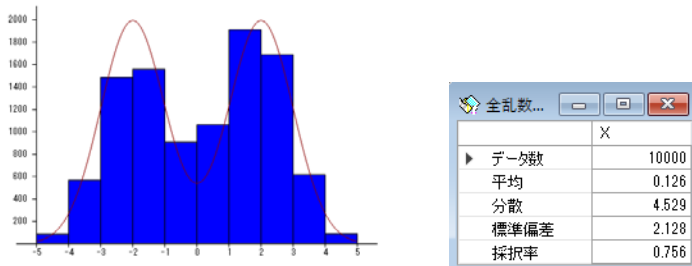


図3 MH法による乱数発生

同様に、HMC法を使って、同じ棄却数で10000個の乱数を発生させたヒストグラムと統計量を図6に示す。但し、変動の微量 ϵ を0.2、リーブフロッグの回数を10としている。

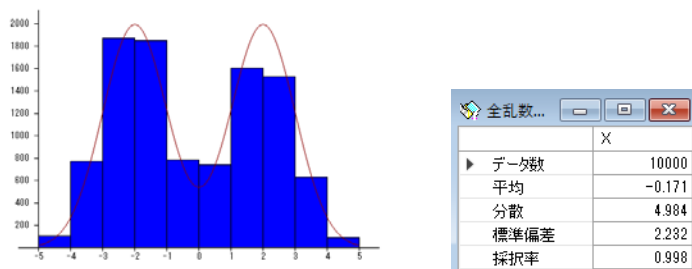


図6 MCH法による乱数発生

どちらの適合性が高いか調べるために、前と同様に、2つの方法で10回ずつ10000個の乱数を発生させ、適合指標としてKS検定のK統計量を比較する。結果を表2に示す。

表 2 K 統計量による MH 法と HMC 法の適合度の違い

	MH 法	HMC 法
1	5.452	4.117
2	5.153	1.196
3	2.175	1.988
4	3.691	1.066
5	4.041	1.904
6	4.295	2.928
7	6.09	2.497
8	1.539	0.756
9	5.769	1.483
10	2.318	1.951

これを比較すると、有意に ($p < 0.01$) 差が見られた。

2つの例で、MH法とHMC法の精度を比較するとHMC法が良い結果になるようである。同様に実行時間を10万個と20万個の発生で比較するとHMC法はMH法に比べて、どちらの場合も約25倍時間がかかっている(手動で測定したため厳密ではない)。これらのことから、正確な乱数を発生させる必要がある場合はHMC法を、最尤法の初期値設定など、精度が低くてもスピードが必要な場合はMH法を利用するとよいことが分かった。

4. おわりに

我々はまず、これまで作成して来た College Analysis 中の「分布と検定」のプログラムにパラメータの自動推定機能を追加する問題を考えた。あるデータがどんな分布に従うか考える際、これまでの経験や傾向から類推するが、それは分布の形までで、細かいパラメータについては不明な場合が多い。そのため、パラメータを推定することは重要である。ここでは、分布を指定して、パラメータを推定する方法を与えているが、基本的には最尤法に基づいている。しかし、分布によっては最尤法の使えない場合やもっと効率の良い方法もある。また最尤法による場合でも、ニュートン・ラフソン法の初期値を与える方法も重要である。我々は分布ごとにプログラムの中で用いた方法を紹介した。もちろん、ここで述べた方法が最良であるとは限らないが、今後同様のプログラムを作成する場合には指針になると考える。

共分散構造分析の解を求める際のパラメータの初期値の設定にはMCMC、特にMH法が有効である。初期値設定には特に正確な乱数が必要なわけではなく、近似的に分布に従えば事足りる。しかし発生のスピードは重要であるため、MH法が有利である。また、標準化モデルでは、パラメータの範囲が限定されており、MH法の弱点も問題にならない。このようなことから我々は共分散構造分析の初期値設定にはMH法を利用している。

因子分析については、新しく最尤法と最小2乗法を組み込んだが、当初初期値設定はMCMCを考えていた。しかし、通常の共分散構造分析とパラメータ数が大きく異なり、MCMCでは対応が難しかった。即ち、変数の数が20以上あるような因子分析を共分散構造分析として扱うと、パラメータ数は「変数の数×(因子数+1)」となるため、極端に多くなる。このことから、我々

は主成分法を初期値設定に利用して計算を行った。これにより、多少大きなモデルも解決可能になった。

分布の中心が MCMC の乱数発生の初期値から遠く離れている場合、MH 法でも HMC 法でも初期値の位置で密度関数 $f(\mathbf{x})$ が、計算機のまるめ誤差によって 0 になることが大きな問題である。酔歩乱数列の場合、これによって乱数は正しい次の位置に移動できなくなる。我々はこの問題に対して、予め初期値を与えるか、徐々にパラメータの値を変化させて $f(\mathbf{x}) \neq 0$ の位置を求める方法を採用している。しかし、後者の方法はパラメータ数が多い場合困難である。これは今後の重要な課題である。

参考文献

- [1] 福井正康, College Analysis リファレンスマニュアル, <http://www.heisei-u.ac.jp/mi/fukui/>
- [2] 福井正康, 孟紅燕, 呉夢, 崔永杰, 社会システム分析のための統合化プログラム 21 – 乱数生成と検定 –, 福山平成大学経営研究, 第 10 号, (2014) 63-78.
- [3] 豊田秀樹, 共分散構造分析 [入門編] – 構造方程式モデリング –, 朝倉書店, 1998.
- [4] 福井正康, 陳文龍, 王嘉琦, 社会システム分析のための統合化プログラム 12 – 共分散構造分析 (中間報告) –, 福山平成大学経営研究, 6 号, (2010) 99-116.
- [5] 豊田秀樹, マルコフ連鎖モンテカルロ法 (統計ライブラリ), 朝倉書店, 2008.
- [6] 豊田秀樹, 基礎からのベイズ統計学: ハミルトニアンモンテカルロ法による実践的入門, 浅倉書店, 2015.

Multi-purpose Program for Social System Analysis 33 - Maximum Likelihood Method and MCMC -

Masayasu FUKUI

*Department of Business Administration, Faculty of Business Administration,
Fukuyama Heisei University*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we discuss the automatic parameter estimation by using maximum likelihood estimation method. And we consider an idea of parameter setting by using MCMC in the covariance structure analysis and the nonlinear least squares method. We also add the Hamiltonian Monte Carlo method to the program of random number generation.

Keywords: College Analysis, maximum likelihood estimation, MCMC, Hamiltonian Monte Carlo method

URL: <http://www.heisei-u.ac.jp/ba/fukui/>

福井 正康