

ベイズ的アプローチによる因子平均の群間差・線形対比の有意性検定

星 野 崇 宏¹ 橋 本 貴 充² 繁 柊 算 男³

教育心理学研究において、いくつかの群の差異の原因を考える際にはしばしば多母集団の因子分析を行い、因子平均の差異で観測変量の差異を説明することが試みられる。しかし、既存の統計的手法による因子平均の群間差の検定では、群ごとに因子平均が異なるモデルと同じであるモデルの比較をするという間接的な χ^2 検定を用いる。この検定は漸近理論を用いているため、データが大量に存在することが前提となるが、教育心理学研究においてはそれが望めない場合が多い。本研究ではベイズ的アプローチによって、平均構造付きの多母集団因子分析モデルと、因子平均の群間差、線形対比の事後分布を数値的に導出する手法を提案した。導出された事後分布によって、点仮説が信頼区間の中に存在すれば保持し、そうでなければ棄却するという有意性検定が可能である。この方法により片側検定も可能になる。また χ^2 検定では不可能な第一種の誤りの統制が可能であり、ヘイウッドケースが存在しないという利点がある。既存の χ^2 検定と今回提案されたベイズ有意性検定を比較するため、シミュレーションによる検定力の比較を行った。実データへの応用例として、WAIS-Rの素点の横断データを用いて加齢による変化を解析し、既存の手法では扱えない複合仮説の評価を行った。

キーワード：構造方程式モデル、ベイズ推測、有意性検定、多母集団、群間差

問題と目的

因子平均の群間差の検定

教育心理学研究においてはしばしば、教授法や年齢差などの外的基準によって分類される複数群の群間差が研究の対象となる。研究の関心が観測変量における群間差に向けられる場合、 t 検定や分散分析などといった既存の統計的仮説検定の手法が利用できる。しかし教育心理学研究においては観測変量の背後に仮定される心理学的構成概念についての仮説の検証が研究の目的になることが多い。この場合、心理学的構成概念に対応する潜在変量を分析することが必要である。そこで、複数の群において変数から同一の因子が測定されると仮定されるとき、平均構造付きの多母集団の因子分析(Sörbom, 1974)が行われ、群間の因子平均の差が関心の対象になることがしばしばある。例えばMarsh & Grayson (1990)では、公立高校とカトリック系高校の生徒に対して、初年度、3年次でのテスト得点、人種、社会経済的地位を測定した。そして測定変数から数学、言語能力の変化の差について平均構造付きの確認的因子分析を用いて解析した。

このような研究目的のための解析手法としてしばしば

ば因子得点の平均値の t 検定や分散分析が行われる場合がある。しかし、因子得点は因子分析のモデルを用いて推定されたものであり、推定の誤差(標準誤差)を含む。従って因子得点の t 検定や分散分析をすることはデータそのものでなくデータからの推定値を用いて統計的仮説検定をするということになり、妥当な推論を引き出すことが難しくなる。これが因子平均の群間差の直接的な検定が要請される所以である。しかし既存の推定法や検定法は、提案された推定量や検定統計量が、すでに定義されているいくつかの優れた性質(不変性、不偏性、許容性、最強力性など)をどれくらい満たすかという探索的な基準によって発展を遂げてきた経緯がある。標本理論による検定法の構成においては検定の一樣最強力性や不偏性などが要請される。最尤理論による尤度比検定においても尤度比の分布が求められなくてはならない。しかし因子平均の検定のような複雑な状況においてはどちらも実際上不可能である。従って、これまでは群間に因子平均の差が存在するかどうかを検証する際、群間で因子平均が共通であるモデル(帰無仮説)と群ごとに因子平均が異なるモデル(対立仮説)の比較という観点から χ^2 検定が行われてきた。この検定は、帰無仮説が正しいときに、尤度比の-2倍が漸近的に、つまり標本数が大きいときに χ^2 分布に従うという性質を用いて行われる検定である(Serfling, 1980)。

¹ 東京大学 総合文化研究科 bayesian@jasmine.ocn.ne.jp
² 東京大学 総合文化研究科 hashimoto@bayes.c.u-tokyo.ac.jp
³ 東京大学 総合文化研究科 kshige@bayes.c.u-tokyo.ac.jp

多母集団因子分析へのベイズ解析の適用

近年、数値的手法の発達とともに、このような複雑な状況下でも一貫した統計学的手法を与え、かつ研究者の関心のある仮説を直接評価できるベイズ理論がさまざまな統計モデルに適用されてきている。ベイズ的手法では、統計的手法は全て事後分布によって行われる (Berger, 1985; Gelman, Carlin, Stern & Rubin, 1995)。あるパラメータについての推測を行うには、そのパラメータの周辺事後分布を求める必要がある。しかし、因子分析のような複雑なモデルにおいては、その解析的導出が困難である。そのため、因子分析のベイズ解析に対しては近年理論的発達の著しい数値的解法であるマルコフチェーン・モンテカルロ法 (MCMC 法) が導入されている。MCMC 法による数値積分法のなかでも特にその容易さや出力結果の柔軟な利用が可能なギブスサンプリング法が頻繁に利用されている。Arminger & Muthén (1998) や Scheines, Hoijsink & Boomsma (1999) ではギブスサンプリングと他のサンプリング法を組み合わせる数値的に因子分析モデルの各母数の周辺事後分布を導出している。Shigemasu & Nakamura (1993) の研究では、因子得点をギブスサンプリングの過程で生成することにより、ギブスサンプリングに必要な条件付き事後分布がすべて多変量正規分布や逆ウィッシュアート分布になることを利用し、純粋なギブスサンプリングだけによる数値解析を可能にした。繁樹・大森・星野 (1999) の研究では純粋なギブスサンプリングにおける因子負荷の条件付き分布の正規性を利用し、回転の不定性を除去するための母数の一部固定が可能になり、共分散構造分析への応用がなされた。また、Shi & Lee (1998) では観測変数に多値変数を含む因子分析への適用が行われた。本研究では Shigemasu & Nakamura (1993) の方法を拡張し、ギブスサンプリングを用いた平均構造付きの多母集団の因子分析モデルのベイズ解析を提案する。ギブスサンプリングの利点として、パラメータの関数の周辺事後分布が容易に導出できるという点がある。また本研究ではこの性質を用い、応用研究において要請の強い「因子平均の群間差の直接的検定」が「因子平均の群間差」の事後分布を用いたベイズ有意性検定 (Lindley 1965, Kim 1991, 繁樹 1985) によって可能であることを示す。そしてこの方法と既存の χ^2 検定の検定力の比較をシミュレーションによって実行する。ベイズ的アプローチによる多母集団の確認的因子分析の利点は大きくすると以下の 4 点にまとめることができる。

1 漸近論を使用していないため、有意性検定におけ

る第一種の誤りの制御が理論上完全になされる。また、被験者数が少なくても実行できる。

- 2 ハイウッドケースが発生しない (ベイズ解析では分散要素の母数の推定値を必ず 0 より大にできる)
- 3 因子平均群間差・線形対比の複合仮説の直接評価が可能である。
- 4 因子平均群間差・線形対比の片側検定が可能である (χ^2 検定では不可能)。

方 法

モデル

多母集団の平均構造付き因子分析モデルについて考える。 J 個の集団に対して、 j 番目の集団における観測変数ベクトルを x_j とおくと、因子得点 f_j が平均 v_j 、分散共分散行列 Φ_j をもつ多変量正規分布に従うと仮定する。

そして誤差 e_j が平均 0、分散共分散行列 Ψ_j の多変量正規分布であるとする。但し、 Ψ_j は対角行列である。また、観測変数 x_j と因子 f_j が以下のような関係にあるとする。

$$x_j = \Lambda_j f_j + e_j \quad (1)$$

但し、 x_j 、 e_j は p 次元、 f_j は q 次元のベクトルである。すると、 x_j の分布は平均 $\Lambda_j v_j$ 、分散共分散行列 $\Lambda_j \Phi_j \Lambda_j' + \Psi_j$ の多変量正規分布になる。上記のモデル設定が多母集団の因子分析である。さて、因子平均の群間比較が行われる条件としては、「因子の不変性」 (Meredith, 1993) が要請される。具体的には因子空間が群間で共通である必要があるため、

$$\Lambda_1 = \Lambda_2 = \cdots = \Lambda_K = \Lambda$$

という制約が必要である。

無情報事前分布

ベイズ的アプローチは統計モデルの母数 θ の分布を直接的に求め、それにもとづいて推論する統計手法である。具体的には尤度 $l(X|\theta)$ と事前分布 $p(\theta)$ から、事後分布 $p(\theta|X)$ を求める。そして θ の事後分布から直接的に推論を行う。従って、ベイズ的アプローチを行うためには事前分布の設定が必要である。事前分布の設定には主観性が伴うという側面があり、ベイズ的アプローチの利用を敬遠させる最大の要因となっている。しかし著者らは、無情報事前分布やデータに対してほとんど影響を及ぼさない漠然とした事前分布の導入によってその設定の恣意性を除去できると考える。

本研究においては、全ての母数に対して事前分布として無情報事前分布を設定する。無情報事前分布とは、推測しようとする母数について未知の状況を表現する、

情報を有さない事前分布である (Jeffreys 1961, Berger 1985)。このモデルにおける Λ と各 v_j の無情報事前分布は定数である。また、 Φ_j 、 Ψ_j の無情報事前分布は

$$\Phi_j \sim |\Phi_j|^{-\frac{q(q+1)}{2}} \quad (2)$$

$$\Psi_j \sim |\Psi_j|^{-\frac{p(p+1)}{2}} \quad (3)$$

である。

ギブスサンプリングによる周辺事後分布の導出

本研究の解析の対象であるモデルにおけるギブスサンプリングは、以下のようなステップで行う。

- [1] 各母数の初期値を用意する。(ここで繰り返しカウン
トは $l=0$)

l 回目の繰り返しにおける、母数の値を $\{Z^l, (F_j^l, \Phi_j^l, \Psi_j^l, v_j^l) (j=1 \sim J), \Lambda^l\}$ とする。ここでは F_j は、 j 群の標本の因子得点すべてを集めた行列とする。

- [2] 各被験者に対して j 群の i 番目の被験者の因子得点 $f_{i(j)}$ を、

$$f_{i(j)}^{l+1} \sim p(f_{i(j)} | x_{i(j)}, \Phi_j^l, \Psi_j^l, v_j^l, \Lambda^l) \quad (4)$$

から生成する。ここで $p(\alpha|\beta)$ は β を条件付きとする α の分布(確率密度関数)を表わす。また、 $x_{i(j)}$ は j 群の個体 i の観測データベクトルである。

- [3] 因子負荷 Λ を生成する。

$$\Lambda^{l+1} \sim p(\Lambda | X, \Psi_j^l, F_j^l (j=1 \sim J)) \quad (5)$$

ここで、 X は全ての群の被験者の観測値を集めることによってできる行列である。

- [4] 各群ごとに母数を生成する。

$$\Phi_j^{l+1} \sim p(\Phi_j | X, F_j^l, v_j^l) \quad (6)$$

ここで誤差分散行列 Ψ_j は対角行列であるので、対角要素ごとに母数生成をし、それを行列の要素に代入し、非対角要素は 0 に固定すればよい。よって、ここで各 j 群に対して誤差分散行列の第 k 対角要素 ψ_{jk} を、

$$\psi_{jk}^{l+1} \sim p(\psi_{jk} | X, F_j^l, \Lambda^l) \quad (7)$$

から生成する。

$$v_j^{l+1} \sim p(v_j | X, F_j^l, \Phi_j^l) \quad (8)$$

- [5] [2] から [4] までを収束基準が満たされ、かつ推測に十分な数のサンプルが得られるまで繰り返す (各繰り返しごとに $l=l+1$ として繰り返しカウントを増やす。)(ギブスサンプリングに関しては、付録 1 に説明を述べる。)

上記のギブスサンプリングを実行する際に必要な各パラメータの条件付き事後分布は以下になる。因子負荷 Λ については、因子負荷行列の転置行列の行ベクトルを繋ぎ合わせるによって構成されるベク

トルを λ (つまり $\lambda = \text{vec}(\Lambda')$) とすると、 λ の条件付き事後分布は全ての群のデータと因子得点 F_j 、誤差分散 Ψ_j を所与とする多変量正規分布となり、その分散共分散行列は

$$(\sum_j [\Psi_j^{-1} \otimes \sum_i^{n_j} f_i f_i'])^{-1} \quad (9)$$

であり、またこれを V_λ とおくと、平均は

$$V_\lambda \sum_j \text{vec}([\sum_i^{n_j} x_i f_i'] [\sum_i^{n_j} f_i f_i']^{-1})^t \quad (10)$$

である。但し、ここで n_j は第 j 群の被験者数である。また j 群の誤差分散行列の第 k 対角要素 ψ_{jk} については、 j 群のデータと因子得点 F_j 、因子負荷 Λ を所与とする逆 χ^2 分布となり、その自由度は $n_j + p$ 、尺度パラメータは、行列

$$\sum_i^{n_j} (x_i - \Lambda f_i)(x_i - \Lambda f_i)^t \quad (11)$$

の kk 要素となる。 j 群の因子共分散行列 Φ_j の条件付き事後分布は、 j 群のデータと因子得点 F_j 、因子平均 v_j を所与とする逆ウィッシュャート分布となり、その自由度は $n_j + q$ 、尺度パラメータ行列は

$$\sum_i^{n_j} (f_i - v_j)(f_i - v_j)^t \quad (12)$$

になる。また、 j 群の因子平均 v_j の条件付き事後分布は、 j 群のデータと因子得点 F_j 、因子共分散行列 Φ_j を所与とした多変量正規分布となり、その平均は

$$\Phi_j^{-1} \bar{f}_j \quad (13)$$

(但し、 $\bar{f}_j$ は j 群の被験者の因子得点の平均ベクトル) であり、分散共分散行列は

$$\frac{1}{n_j} \Phi_j \quad (14)$$

である。また、 j 群の個体 i の因子得点ベクトル $f_{i(j)}$ の条件付き事後分布は、 j 群の個体 i の観測データ $x_{i(j)}$ と j 群の因子共分散 Φ_j 、誤差分散 Ψ_j 、因子平均 v_j 、そして因子負荷 Λ を所与とする多変量正規分布となり、その分散共分散行列は

$$(\Lambda_j' \Psi_j^{-1} \Lambda_j + \Phi_j^{-1})^{-1} \quad (15)$$

またこれを V_{fi} とおくと、平均は

$$V_{fi} (\Lambda' \Psi_j^{-1} x_{i(j)} + \Phi_j^{-1} v_j) \quad (16)$$

となる。(上記の条件付き事後分布の導出の概略は付録 2 を参照。)

以上のようにすべての条件付き分布が多変量正規分布や逆ウィッシュャート分布になるという性質は、因子得点を一種の母数として考えることにより可能になる。因子負荷や因子平均の条件付き分布が多変量正規分布になるという性質によって初めて、本研究における様々な手法が可能となる。さてここで、ギブスサン

リングによって得られた $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}$ が, L 回の繰り返し以降収束したとすると, $L+1$ 回から T 回までの $T-L$ 個の値は θ の周辺事後分布 $p(\theta|x)$ からの無作為標本と見なすことができる。よって, ギブスサンプリングによって得られたサンプルの経験分布をパラメータの周辺事後分布と見なすことができる。また, ある関数 $f(\cdot)$ に対して, $\theta^{(i)}$ の関数値 $\frac{1}{(T-L)} \sum_{L+1}^T f(\theta^{(i)})$ は $f(\theta)$ の期待値に概収束する (Tanner, 1996)。これらの性質から, ギブスサンプリングによるベイズ的アプローチでは, 関心のある母数の周辺事後分布を導出することができるために, 既存の標本理論や最尤法においては不可能であった, または漸近論に依存してしか解決されえなかった様々な推測問題の解決が可能になる。またここで, 確認的因子分析のための因子負荷等の一部母数固定については, 繁樹, 大森, 星野 (1999) の方法を用いる。概略を示すと, 因子負荷をベクトルに直したものである λ に対して, 固定される要素をベクトルの下部に集めるような置換をする行列 R を掛ける。つまり,

$$R\lambda = \begin{pmatrix} \lambda_1^* \\ \lambda_2^* \end{pmatrix} \quad (17)$$

また, これに対応する平均を

$$\begin{pmatrix} \mu_{\lambda_1}^* \\ \mu_{\lambda_2}^* \end{pmatrix} \quad (18)$$

また分散共分散行列を

$$\begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} \quad (19)$$

とする。ここで $\lambda_2^* = c$ という制約を置く。 λ の条件付き事後分布の平均が μ_i , 分散共分散行列が Σ_i であるとする, $\lambda^* | \lambda_2^* = c$ の条件付き分布は平均が

$$\mu_{\lambda_1}^* - \Sigma_{12}^* \Sigma_{22}^{*-1} (c - \mu_{\lambda_2}^*) \quad (20)$$

そして分散共分散行列が

$$\Sigma_{11}^* - \Sigma_{12}^* \Sigma_{22}^{*-1} \Sigma_{21}^* \quad (21)$$

の多変量正規分布になる。これによって, $\lambda^* | \lambda_2^* = c$ の条件付き事後分布から λ^* を生成し, 因子負荷の当該要素に代入することで, 確認的因子分析が可能になる。また, これを拡張することにより, 因子負荷の要素間の等号制約などの線形制約も可能である (Hoshino, in press)。

ギブスサンプリングによる因子平均の群間差の周辺事後分布の導出

因子平均の群間差の有意性検定は, ギブスサンプリングによって求まる因子平均の群間差の周辺事後分布を直接利用すればいい。具体的には, j 群の因子平均を v_j とし, k 群の因子平均を v_k とすると, これらの条件付き事後分布は互いに独立であるので, それらの平

均を μ_{vj} , μ_{vk} , 分散共分散行列を Σ_{vj} , Σ_{vk} とすると,

$$v_j - v_k \sim N(\mu_{vj} - \mu_{vk}, \Sigma_{vj} + \Sigma_{vk}) \quad (22)$$

となる。ギブスサンプリングの各繰り返しごとにこの分布から $v_j - v_k$ を発生させることによって因子平均の群間差の周辺事後分布を数値的に導出できる。

また v_j の r 番目の要素 v_{jr} (j 群の r 番目の因子の平均) の条件付き分布は, v_j の条件付き分布が多変量正規分布であることから, μ_{vj} の r 要素を平均, Σ_{vj} の第 r 対角要素を分散とする正規分布である。このことから, 各繰り返しにおいて $v_{jr} - v_{kr}$ を発生させれば, 数値的に周辺事後分布の95%信頼区間を求めることができ, 母数についての点仮説が信頼区間外であれば棄却するという形式でベイズ有意性検定を実行できる。ここで $(1-\alpha)\%$ 信頼区間の導出は, 最短な信頼区間を採用する方法と, 両側とも $\alpha/2$ になるような方法の2つがあるが, 本研究では母数の単調変換に対して不変である後者の手法を用いる。このベイズ有意性検定は, 伝統的な有意性検定に対応するベイズ的アプローチの手法であり, Bernardo & Smith (1994, p.412) のモデル棄却の方法として考えることもできる。

同時にすべての因子についての群間差の検定を行いたい場合は, 多次元の同時信頼区間を求め, 点仮説がその区間に入るかどうかで有意性検定を行う。しかし実際は多次元の同時信頼区間の導出は困難である。そこで, 因子数が m であるとき, ボンフェロニの不等式を用いて各因子平均の群間差の検定の有意確率を m で割ることにより有意水準全体のコントロールが出来る。また, ギブスサンプリングによって生成された標本は周辺事後分布からの無作為標本であると見なせるので, その等高線を描くことにより同時信頼区間の近似も可能である。

因子平均の線形対比の周辺事後分布の導出

同様に, 計画的線形対比についての有意性検定も同様の方法によって構成できる。これは, 対比についての仮説が $Cv=0$ の形で表わされるとき, 新たに母数として ξ を $\xi=Cv$ と置き, ξ についての周辺事後分布から推測を行えばよい (渡部, 1999)。但し, $v^t = (v_1^t, v_2^t, \dots, v_k^t)$ である。例えば群が3群あるとき第1因子の平均についての対比仮説

$$H_0: v_{21} = \frac{1}{2}(v_{11} + v_{31}) \quad (23)$$

の有意性検定を行いたいときは,

$$\xi = v_{21} - \frac{1}{2}(v_{11} + v_{31}) = \begin{pmatrix} -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = Cv \quad (24)$$

とする。ここで、各繰り返しごとに ν に行列 C を掛けたものを ξ の周辺事後分布からの無作為標本と見なせる。また、 ν の条件付き分布は、各 ν_j が他の母数を条件付きとすると互いに独立なので、

$$\nu \sim N\left(\begin{pmatrix} \mu_{\nu 1} \\ \mu_{\nu 2} \\ \vdots \\ \mu_{\nu K} \end{pmatrix}, \begin{pmatrix} \Sigma_{\nu 1} & 0 & \cdot & 0 \\ 0 & \Sigma_{\nu 2} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \Sigma_{\nu K} \end{pmatrix}\right). \quad (25)$$

である。

教育心理学研究においては一般に第2群の因子平均が第1、第3群の平均より大きい小さいかなどが関心の対象になることは多い。このような場合、 ξ の周辺事後分布の導出をすれば、有意性検定だけでなく ξ が0より大きい確率などの複合仮説が直接的に評定できるということがこの方法の利点である。線形対比の点仮説は χ^2 検定で評価できるが、複合仮説対複合仮説の評価は本研究で提案されたベイズ法による直接的な確率評価でしか行えない。

シミュレーションデータを用いた χ^2 検定との検定力の比較

ここでは、本研究において提案された有意性検定と最尤法による推定値を用いた漸近的な χ^2 検定とを比較するために、因子平均の真の群間差ごとにシミュレーションデータを100回ずつ生成し、おのおのに対して両方の検定(有意水準5%)を実行した。ここで、最尤法による推定で不適解の生じたデータも分析の対象とした。これは、データが不適解になるかどうかということが群間差の関数になる可能性があり、この場合不適解をデータから除去するとシミュレーションの結果にバイアスを与えてしまう可能性があるためである。因子平均の真の群間差と、差が0であるという仮説が棄却された割合をプロットすることにより、数値的に検定力曲線が導出できる。これによって両方の検定を比較する。標本数の影響を知るために、2種類の標本数で実行した。ここでは比較のために群の数は2群で、観測変量は6変量、因子は3因子であり、因子負荷は群間で共通とした。因子負荷は以下のように設定した。

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0^* & 0^* \\ \lambda_{21} & \lambda_{22} & 0^* \\ 0^* & \lambda_{32} & 0^* \\ 0^* & \lambda_{42} & 0^* \\ 0^* & \lambda_{52} & \lambda_{53} \\ 0^* & \lambda_{62} & \lambda_{63} \end{pmatrix} \quad (26)$$

但し、 0^* は、母数を0に固定したことを意味する。また、発生されたシミュレーションデータの各母数の真値は付録3に掲載した。ギブスサンプリングにおい

ては生成された数値列が分布として収束するかが重要な問題となる。収束判定方法は多数提案されているが、ここでは Gelman & Rubin (1992) の方法を用いた。この方法は同じデータを用いた複数の数値列を用い、その列内分散と列間分散を用いて収束判定を行う方法である。本研究では初期値、乱数を変えて10回ギブスサンプリングを実行し、この方法を適用することにより、本研究におけるギブスサンプリングの数値列の収束を確認した⁴。

ここで数値的に検定力曲線を構成するため、因子平均の第1要素の差 $\nu_{21} - \nu_{11}$ を0, 0.25, 0.5, 0.75, 1, 1.5, 2の7段階で変化させ、それぞれ100データセットを生成した。また標本数は、2群それぞれ $N=80$ (計160)、 $N=40$ (計80)の2つを用意した。従って全データセット数は1400である。各データに対してギブスサンプリングを、500回の収束のための繰り返しの後さらに3000回行い、それを周辺事後分布からの無作為標本と見なして95%信頼区間を計算し、ベイズ有意性検定を行った。シミュレーションデータによって得られた検定力曲線はFIGURE 1に掲載した。

結果

検定力という観点からすると一見、今回提案された方法と χ^2 検定では大きな差がないといえる。実際、フィッシャーの直接法によって検定を行ったが、全ての群間差において差が無いという仮説が棄却されなかった。FIGURE 1 から、 $N=160$ (各群80) の場合に χ^2 検定の検定力はベイズ有意性検定に対して一様に高くはないことがわかる。ここで注意すべきことは、第一種の誤りの確率の統制がベイズ有意性検定ではなされているのに対して、 χ^2 検定は漸近的な検定であるために、中程度以下の大きさの標本数のデータではその保証がなされないということである。実際、ベイズ有意性検定と χ^2 検定それぞれの、100データセット中での棄却回数を表にしたTABLE 1からも、差の真値が0で

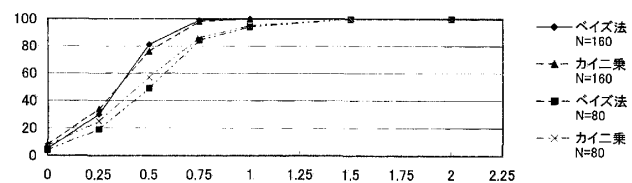


FIGURE 1 ベイズ有意性検定とカイ二乗検定の検定力曲線

⁴ この方法を含め、様々な収束判定法が統計ソフト S-plus 上で動くモジュールである BOA (Bayesian Output Analysis) によって実行可能である (入手は <http://www.pmeu.uiowa.edu/BOA> を参照のこと)。

あるときは、 $N=80, 160$ ともに χ^2 検定においては第一種の誤りが有意水準より大きくなっている。また、FIGURE 1 において $N=80$ ではベイズ有意性検定が検定力において劣っているように見えるのは、 χ^2 検定では第一種の誤りの統制がなされていないからである。特に、 $N=80$ (各群 40) の場合、 χ^2 検定の結果は漸近論からも保証されない。またここで、事後分布のモードが 0 より大きいか小さいかによって棄却域を片側に設定するならば、ベイズ有意性検定の方が検定力が一樣に高いという結果が得られた (TABLE 2)。この方法は標本理論からは、データを得た後に棄却域を定めるという点で批判を免れない (但し p 値を得てから有意水準を定めるのも同様であると言えよう) が、ベイズ的な事後推論からは棄却域を一貫した論理で構成できる手法であるといえる。

WAIS-R 素点データによる知能の年齢差の解析

知能の測定に関しては様々な議論がなされているが、標準化がなされ、かつ教育場面や臨床場面で具体的に利用されている信頼性の高い検査の一つとして日本語版 WAIS-R (品川・小林・藤田・前川, 1991) が挙げられる。日本語版 WAIS-R には動作性知能検査と言語性知能検査という 2 つの下位検査があり、それぞれ別々の心理学的構成概念を測定するとされる。一般的に利用される際には運動性知能因子と言語性知能因子はそれぞれの下位項目の合計得点によって測定されると仮定され、利用されている。しかし潜在変量に関心がある場合、因子分析を用いて解析する方がより統計学的に

TABLE 1 ベイズ有意性検定とカイ二乗検定の比較

標本数	N=160		N=80	
	ベイズ	カイ二乗	ベイズ	カイ二乗
差の大きさ				
0	5	8	4	7
0.25	30	34	19	25
0.5	81	76	49	57
0.75	99	98	84	86
1	100	100	94	95
1.5	100	100	100	100
2	100	100	100	100

TABLE 2 モードの位置による片側のベイズ有意性検定とカイ二乗検定の比較

標本数	N=160		N=80	
	ベイズ	カイ二乗	ベイズ	カイ二乗
差の大きさ				
0	7	8	8	7
0.25	38	34	34	25
0.5	88	76	76	57
0.75	99	98	98	86
1	100	100	100	95
1.5	100	100	100	100
2	100	100	100	100

妥当である。ここでは、新たに提案された方法の解析例として、WAIS-R の素点の横断的データから、言語性知能因子より動作性知能因子の方が加齢による変化が大きく、因子平均の差の検定は有意になりやすいという仮説を検討する。

ここで、モデルとして平均構造付きの 2 因子 (言語性知能因子と動作性知能因子) の確認的因子分析を考える。ここで群間で因子負荷は共通とする。(ここで本データにおいて、「因子負荷が群間で共通している」という仮説に対する χ^2 検定では $p=0.18$, またモデルの適合度指標である GFI は 0.927 であった)。またここでは WAIS-R の下位項目分類通りに、「知識」「数唱」「単語」「算数」「理解」「類似」の 6 下位尺度が言語性知能因子に対して因子負荷を持ち、「絵画完成」「絵画配列」「積木模様」「組合せ」「符号」の 5 下位尺度が動作性知能因子に対して因子負荷を持つと仮定する。今回は WAIS-R の標準化の際に得られた 50 歳代、60 歳代の素点データ 263 人分を解析する。具体的には 50 歳代 (109 人)、60 歳代前半 (59 人) と 60 歳代後半 (95 人) の 3 群に分け、群間の差や線形対比を考える。ここで 50 歳代、60 歳代前半、60 歳代後半の因子平均をそれぞれ $\nu_1=(\nu_{11}, \nu_{12})^t, \nu_2=(\nu_{21}, \nu_{22})^t, \nu_3=(\nu_{31}, \nu_{32})^t$ と置く。このとき、50 歳代と 60 歳代前半の言語性知能因子の群間差 $\nu_{21}-\nu_{11}$ のギブスサンプリング (繰り返し数 10000 回) による周辺事後分布は FIGURE 2 のようになる (縦線は 0 の位置を表す)。また、縦軸は 10,000 回中の頻度を表す。同様に動作性知能因子の群間差の周辺事後分布は FIGURE 3 のようになる。言語性知能因子の平均の差については Pr (群間差が 0 より大きい) = 0.4267 であり、片側で棄却域を定めても、5% 有意水準で棄却されない。しかし、動作性知能因子の平均の差については Pr (群間差が 0 より大きい) = 0.0457 であり、片側で棄却域を定めると、5% 有意水準で棄却されることがわかる (但し χ^2 検定では棄却されない)。このことから、言語性知能因子の群間差は有意でないが、動作性知能因子の群間差は有意に 50 歳代の方が大きいといえる。

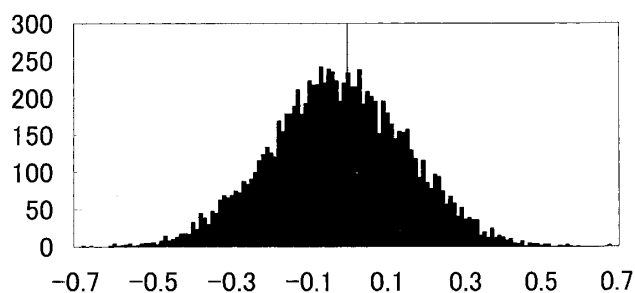


FIGURE 2 言語性知能の群間差の事後分布

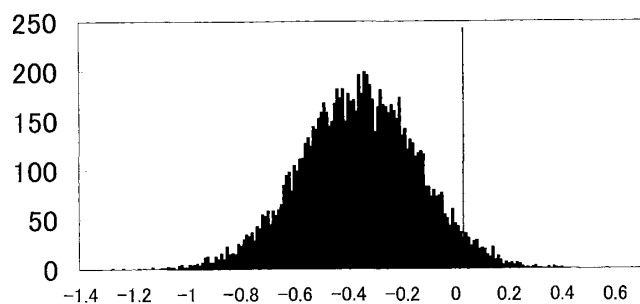


FIGURE 3 動作性知能の群間差の事後分布

ベイズ的アプローチの長所は、研究者の問題意識にあわせた柔軟な解析を提供できるという点にある。例えば、動作性知能因子について、50歳代と60歳代後半の平均よりも60歳代前半の平均の方が高いかどうかという線形対比についての複合仮説の推論も容易に行える。これは、「動作性の知能は50歳代から60歳代で低下するという知見は既に得られた。では、コンスタントに低下するのか ($\nu_{22} = (\nu_{12} + \nu_{32})/2$)、50歳代から60歳代前半で大きく低下するのか ($\nu_{22} < (\nu_{12} + \nu_{32})/2$)、それとも60歳代前半から後半にかけての低下が大きいのか ($\nu_{22} > (\nu_{12} + \nu_{32})/2$) ?」 (FIGURE 4) といった研究者の関心に直接答えることのできる方法である。統計学的には、仮説

$$H_0: \xi = \nu_{22} - \frac{1}{2}(\nu_{12} + \nu_{32}) > 0 \quad (27)$$

が、ギブスサンプリングによる ξ の周辺事後分布 (FIGURE 5) から直接的に評価できる。ここで実質科学的な意味を考えると、帰無仮説 $\xi > 0$ と仮説 $\xi \leq 0$ は同等であり、一方が他方に優先する有意性検定にはなじまない。このような状況では、仮説 $\xi \neq 0$ の検定とは異なり、仮説 $\xi > 0$ の事後確率を直接的に導出することが有意義である。ベイズ的アプローチでは ξ の事後分布からこの確率が容易にわかり、結果として、 $Pr(\xi > 0) =$

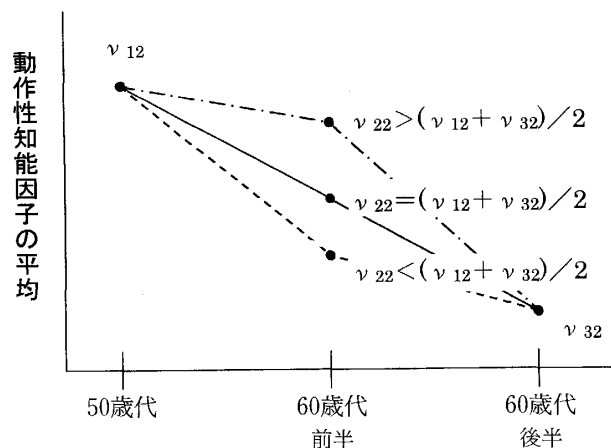


FIGURE 4 線形対比による動作性知能の変化の仮説

0.7405となる。従って、動作性知能因子では60歳代前半から後半にかけての落ち込みの方が50歳代から60歳代前半にかけての落ち込みより大きいことがわかる (既存の方法でも $\xi = 0$ の両側検定なら可能である)。また、ギブスサンプリングの収束に関しては様々な指標 (レビューとしては Cowles & Carlin, 1996) を用いて解析し、収束の確認を行った。例として ξ の移動平均 (FIGURE 6) を示す。

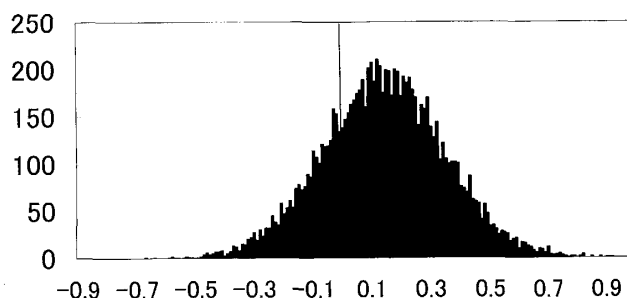
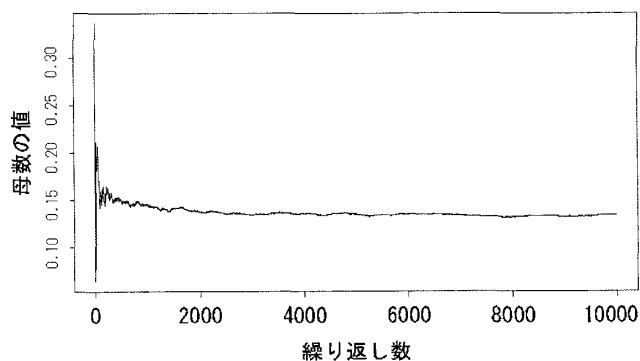


FIGURE 5 線形対比の事後分布

FIGURE 6 対比 ξ の移動平均

結 語

本研究ではギブスサンプリングを用いたベイズ的アプローチによる多母集団の因子分析が提案された。また因子平均の群間差の事後分布を用いたベイズ有意性検定がシミュレーションによって χ^2 検定と同様の (またはそれ以上の) 検定力を持ち、かつ χ^2 検定ではなされない第一種の誤りの統制を比較的小標本でもできることがわかった。またヘイウッドケースが生じず、線形対比の複合仮説に対する直接的評価を可能にするなどの利点から、この手法は有用であると考えられる。また、WAIS-R のデータを用いた解析例では教育心理学研究においてしばしば関心の対象とされる線形対比の複合仮説がこの手法によって自然な形で評価されることが示された。ベイズ的アプローチの問題点として事

前分布の設定が挙げられるが、このような複雑な問題に対しても無情報な事前分布の設定が可能であることが示された。本方法を実際のデータ解析に用いる際に注意すべき点を、以下に列挙する。

1 因子の不変性

因子平均に関する推論が必要である場合の解析方法を行うためには、因子の不変性 (Meredith, 1993) が成立するかどうかの確認が必要である。確認方法としては χ^2 検定やベイズ法によるモデルの事後確率の導出 (Raftery, 1996) がある。

2 収束判定の問題

ギブスサンプリングによる数列の収束のために、繰り返しの初めの数百回は捨て、その後の数千回を推測にあてるとというのが一般的な方針であるが、収束がなされているかどうかはすでに提案されている様々な収束判定基準方法 (例えば丹後 2000 参照) で判定する必要がある。

3 モデルの識別性

モデルの識別性がないと、情報のある事前分布を導入しなければ、一般的にギブスサンプリングは収束しない。

4 初期値の問題

初期値としてデータから考えてあまりにも極端な値を設定すると、収束しない可能性がある。

5 本研究での線形対比の検定は非計画的非直交対比を扱えない。対比 $\xi = Cv$ の C を事前に決定している必要がある。

しかし、これらは収束判定を除き最尤法や最小二乗法による解析時にも必要な注意事項であり、収束判定を除けば既存の解析方法と比較して特別な注意事項や前提条件は必要ではない。

今後への課題としては実際の研究で本方法を容易に利用できるようにすることが挙げられるが、標本数に応じて計算量が増加するという問題があるため、アルゴリズムのさらなる改善が必要である。

引用文献

- Arminger, G., & Muthén, B.O. 1998 A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis Hastings algorithm. *Psychometrika*, **63**, 271—300.
- Berger, J.O. 1985 *Statistical decision theory and Bayesian analysis*. New York : Springer-Verlag.
- Bernardo, J.M., & Smith, A.F.M. 1994 *Bayesian Theory*. New York : Wiley.
- Cowles, M.K., & Carlin, B.P. 1996 Markov chain Monte Carlo convergence diagnostics : A Comparative review. *Journal of the American Statistical Association*, **91** 883—904.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. 1995 *Bayesian data analysis*. London : Chapman & Hall.
- Gelman, A., & Rubin, D.B. 1992 Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457—511.
- Hoshino, T. (in press) Bayesian inference for finite mixtures in confirmatory factor analysis. *Behaviormetrika*.
- Jeffreys, H. 1961 *Theory of probability (3rd ed)*. London ; Oxford University Press.
- Kim, D. 1991 A Bayesian significance test of the stationarity of regression parameter. *Biometrika*, **78**, 667—675.
- Lindley, D.V. 1965 *Introduction to probability and statistics : Inference*. New York : Cambridge University Press.
- Marsh, H.W., & Grayson, D. 1990 Public/Catholic differences in the high school and beyond data : A multigroup structural equation modeling approach to testing mean differences. *Journal of Educational Statistics*, **15**, 199—235.
- Meredith, W. 1993 Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, **58**, 525—543.
- Raftery, A.E. 1996 Hypothesis testing and model selection. In W.R.Gilks, S.Richardson, & D. Spiegelhalter (Eds.), Markov chain Monte Carlo in Practice. London : Chapman & Hall. Pp. 441—464.
- Scheines, R., Hoijtink, H., & Boomsma, A. 1999 Bayesian estimation and testing of structural equation models. *Psychometrika*, **64**, 37—52.
- Serfling, R.J. 1980 *Approximation theorems of mathematical statistics*. John Wiley & Sons.
- Shi, J-Q., & Lee, S-Y. 1998 Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, **51**, 233—252.

Shigemasu, K., & Nakamura, T. 1993 A Bayesian numerical estimation procedure in factor analysis model. *E.S.T. Research Report*, 93—6, Tokyo Institute of Technology.

繁樹算男 1985 ベイズ統計入門 東京大学出版会

繁樹算男 1995 意思決定の認知統計学 朝倉書店

繁樹算男・大森拓哉・星野崇宏 1999 Bayesian inference for structural equation modeling. 日本行動計量学会第 27 回大会発表抄録集, 75—78.

品川不二郎・小林重雄・藤田和弘・前川久男 1991 WAIS-R 成人知能検査 日本文化科学社

Sörbom, D. 1974 A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229—239.

丹後俊郎 2000 統計モデル入門 朝倉書店

Tanner, M.A. 1996 *Tools for statistical inference* (3rd ed.). Heidelberg and New York :

Springer-Verlag.

渡部 洋 1999 ベイズ統計学入門 福村出版

謝 辞

WAIS-R の素点データ利用に際しては、武蔵野女子大学の大六一志助教授、そして筑波大学の前川久男教授には大変お世話になりました。大学入試センターの前川眞一教授には SAS/IML プログラムと最尤法による多母集団因子分析法について御教示いただきました。また、日本文化科学社には WAIS-R 標準化の際のデータ利用を承諾して頂きました。この場をお借りして、お礼を申し上げます。

付 記

本方法のプログラムは SAS/IML で作成しました。利用希望の方は、星野までご連絡下さい。

なお、付録 1～3 は次頁に掲載してあります。

(2000.5.11 受稿, 10.16 受理)

A Bayesian Significance Test for the Difference and Linear Combination of Factor Means Between Groups

TAKAHIRO HOSHINO, TAKAMITSU HASHIMOTO, AND KAZUO SHIGEMASU (DEPARTMENT OF COGNITIVE AND BEHAVIORAL SCIENCE, UNIVERSITY OF TOKYO) JAPANESE JOURNAL OF EDUCATIONAL PSYCHOLOGY, 2001, 49, 31—40

The present article proposes Bayesian inference for multiple group factor analysis, via a Gibbs sampling algorithm. When this method is used, a significance test for the difference of factor means between groups and a test for linear contrast can be done to determine whether a point hypothesis is in the Bayesian confidence interval of the posterior distribution. Bayesian inference is not used in research in educational psychology because of its arbitrariness in the selection of the prior distribution. However, all the methods described in the present article can be done with a noninformative prior distribution, thus excluding subjectivity. As a result, researchers will be able to use Bayesian inference easily and objectively. In order to compare the proposed method with a χ^2 asymptotic likelihood test in terms of their respective power, 1400 simulation data sets for $N=160$ and $N=80$ were generated. The proposed method has some advantages over the χ^2 test in that control of Type-1 errors is complete with small-sized samples, and also in that the Haywood case problem does not occur. The proposed method was used to analyze the relationship of age and intelligence as measured by the WAIS-R, and a composite hypothesis was evaluated.

Key Words : structural equation modeling, Bayesian inference, significance test, multiple groups, difference between groups

付録1 ギブスサンプリングについて

ギブスサンプリングは、積分することが難しい分布を、その分布に従う乱数を発生させて数値的に積分する方法の一つである。以下にその手順を示す。

まず、母数ベクトル θ を、 $\theta^t = (\theta_1^t, \theta_2^t, \dots, \theta_m^t)$ と m 個に分割する。各 θ_j はスカラーでもベクトルでもよい。

次に、 $\theta_1, \theta_2, \dots, \theta_m$ の初期値 $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_m^{(0)}$ を用意し、以下のステップを繰り返す。

1. θ_1 以外の母数を $\theta_j = \theta_j^{(i)}$ に固定した条件付事後分布 $p(\theta_1 | \theta_2^{(i)}, \theta_3^{(i)}, \dots, \theta_m^{(i)}, x)$ より $\theta_1^{(i+1)}$ を発生させる。但し x は観測値ベクトルである。
2. $p(\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_m^{(i)}, x)$ より $\theta_2^{(i+1)}$ を発生させる。
3. $\theta_3^{(i+1)}, \dots, \theta_m^{(i+1)}$ を順次発生させた後、 $p(\theta_m | \theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_{m-1}^{(i+1)}, x)$ より $\theta_m^{(i+1)}$ を発生させる。
4. 1. に戻る。

このサイクルを繰り返すことによって発生される $\theta_1, \theta_2, \dots, \theta_m$ が、 $p(\theta | x)$ からの無作為標本と考えることができる。

ギブスサンプリングでは初期値の影響で、始め $p(\theta | x)$ から外れた値をとることがあるが、何度も繰り返すうちに $p(\theta | x)$ からの無作為標本と考えてよい値に収束する。 L 回の繰り返し以降収束したならば、 θ_j の周辺密度関数は

$$p(\theta_j | x) = \frac{1}{K} \sum_{i=L+1}^{L+K} p(\theta_j | \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i)}, \dots, \theta_m^{(i)}, x)$$

の式で推定される。

また、発生させた標本の経験分布は、その母数の周辺事後分布とみなすことができる。

さらに詳しくは繁樹 (1995), Tanner (1996) を参照。

付録2 各パラメータの条件付き事後分布を導出する方法の概略

ベイズの定理より、群間で共通なパラメータ Λ の条件付き事後分布は、

$$\begin{aligned} & p(\Lambda | \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J, X, F) \\ & \propto p(\Lambda | \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J) p(X, F | \Lambda, \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J) \\ & = p(\Lambda | \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J) \prod_{j=1}^J p(X_j, F_j | \Lambda, \nu_j, \Phi_j, \Psi_j) \\ & = p(\Lambda | \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J) \prod_{j=1}^J \prod_{i=1}^{n_j} p(x_{i(j)} | f_{i(j)}, \Lambda, \Psi_j) p(f_{i(j)} | \nu_j, \Phi_j) \\ & p(\Lambda | \nu_1, \dots, \nu_J, \Phi_1, \dots, \Phi_J, \Psi_1, \dots, \Psi_J) \propto \text{定数}, \\ & x_{i(j)} | f_{i(j)}, \Lambda, \Psi_j \sim N(\Lambda f_{i(j)}, \Psi_j), \\ & f_{i(j)} | \nu_j, \Phi_j \sim N(\nu_j, \Phi_j) \text{ より, } \Lambda \text{ の条件付き事後分布, すなわち } \lambda = \text{vec}(\Lambda^t) \text{ の条件付き事後分布は, 多変量正規分布の性質より,} \end{aligned}$$

$$N(V_\lambda \sum_{j=1}^J \text{vec}[\sum_{i=1}^{n_j} x_{i(j)} f_{i(j)}^t] [\sum_{i=1}^{n_j} x_{i(j)} f_{i(j)}^t]^{-1}, V_\lambda)$$

但し、

$$V_\lambda = (\sum_{j=1}^J [\Psi_j^{-1} \otimes \sum_{i=1}^{n_j} f_{i(j)} f_{i(j)}^t])^{-1}$$

となる。

また、パラメータ Φ_j の条件付き事後分布は、

$$\begin{aligned} & p(\Phi_j | \Lambda, \nu_j, \Psi_j, X_j, F_j) \\ & \propto p(\Phi_j | \Lambda, \nu_j, \Psi_j) p(X_j | F_j, \Lambda, \Psi_j) p(F_j | \nu_j, \Phi_j) \\ & \propto p(\Phi_j | \Lambda, \nu_j, \Psi_j) \prod_{i=1}^{n_j} p(f_{i(j)} | \nu_j, \Phi_j) \end{aligned}$$

ここで、 $p(\Phi_j | \Lambda, \nu_j, \Psi_j) \propto |\Phi_j|^{-\frac{q(q+1)}{2}}$ より、 Φ_j の条件付き事後分布は、密度関数が、

$$|\sum_{i=1}^{n_j} (f_i - \nu_j)(f_i - \nu_j)^t|^{-\frac{n_j+q}{2}} |\Phi_j|^{-\frac{n_j+2q+1}{2}} \exp\{-\frac{1}{2} \Phi_j^{-1} \sum_{i=1}^{n_j} (f_i - \nu_j)(f_i - \nu_j)^t\}$$

に比例する逆ウィッシュャート分布になる。

パラメータ $\nu_j, \Psi_j, f_{i(j)}$ の条件付き事後分布も、ベイズの定理と多変量正規分布の性質から同様に導出することができる。

付録3 シミュレーション研究における母数の真値

$$\Lambda = \begin{bmatrix} 2.5 & 0 & 0 \\ 1.5 & 1.5 & 0 \\ 0 & 2.5 & 0 \\ 0 & 2.5 & 0 \\ 0 & 1.5 & 1.5 \\ 0 & 1.5 & 1.5 \end{bmatrix} \quad \Phi_1 = I \quad \Phi_2 = \begin{bmatrix} 1.5 & 0.5 & -0.5 \\ 0.5 & 1.5 & 0.5 \\ -0.5 & 0.5 & 1.5 \end{bmatrix}$$

$$\Psi_1 = \text{diag}\{0.5, 1, 0.5, 0.5, 1, 1\} \quad \Psi_2 = \text{diag}\{1, 0.5, 0.5, 1, 0.5, 0.5\}$$

$$v_1 = [0, 0, 0] \quad v_2 = [a, 0, -1]$$

*ここで、シミュレーション研究では a を変化させることにより群間差を変化させている。