ベイジアン因子分析モデルとその評価

九州大学大学院数理学府　廣瀬　慧
九州大学大学院数理学府　川野　秀一
九州大学大学院数理学研究院　小西　貞則
東京外国語大学 外国語学部　市川　雅教

1　はじめに

因子分析モデルは，観測データから直接観測できない潜在因子を見出すモデルであり，心理学をはじめとして諸科学の様々な分野で応用されている．

モデルの推定法として最尤法が用いられるが，最尤法で推定すると，独自因子の分散の推定値が負となる不適解問題が生じることがある．この問題に対処するための一つの方法として，ベイズアプローチに基づくモデルの推定法が提案されている．これは，パラメータに事前分布を設定することによって制約条件を課し，不適解を防ぐ手法である．事前分布は，Martin and McDonald (1975)，Akaike (1987) などにより提案されている．

因子分析モデルにおける実際上重要な問題の一つとして，適切な因子数の選択がある．適切な因子数を選択するための方法として，AIC や BIC の適用が考えられる．しかし，ベイズアプローチによって推定されたモデルに対して，これらの評価基準を直接適用することはできない．また，ベイズ推定による推定量は事前分布を特徴付けるハイパーパラメータに依存するため，適切なハイパーパラメータの選択も重要な問題となる．したがって，ベイズ推定に伴うハイパーパラメータの値を適切に選択し，不適解問題に対処するとともに因子数の選択を行うことが本質的となる．

本報告では，不適解を防ぐための事前分布を導出し，一般化ベイズ型モデル評価基準 GBIC (Konishi *et al.*, 2004) を導出して適切な因子数及びハイパーパラメータの値を同時に選択する方法を提案する．また，提案したモデリング手法を，実データへの適用を通して検証する．

2　ベイジアン因子分析モデル

p 次元観測変数ベクトルを $\boldsymbol{X} = (X_1, \cdots, X_p)'$，$k$ 次元潜在変数ベクトルを $\boldsymbol{F} = (F_1, \cdots, F_k)'$ とする．因子分析モデルでは，観測変数と潜在変数の間に次のような線形性を仮定する．

$$\boldsymbol{X} - \boldsymbol{\mu} = L\boldsymbol{F} + \boldsymbol{\varepsilon}. \tag{1}$$

ただし，$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$ は平均ベクトル，$\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_p)'$ は独自因子ベクトル，$L = (\ell_{ij})$ は $p \times k$ 因子負荷行列である．以下，一般性を失うことなく $\boldsymbol{\mu} = \boldsymbol{0}$ の場合を考える．

潜在変数ベクトル $\boldsymbol{F}$ と独自因子ベクトル $\boldsymbol{\varepsilon}$ に対して，次の多変量正規分布を仮定する．

$$\boldsymbol{F} \sim N_k(\boldsymbol{0}, I), \quad \boldsymbol{\varepsilon} \sim N_p(\boldsymbol{0}, \Psi) \tag{2}$$

ただし，$\Psi = \mathrm{diag}(\psi_1^2, \cdots, \psi_p^2)$ とする．さらに，$\boldsymbol{F}, \boldsymbol{\varepsilon}$ は互いに独立とする．これらの仮定のもとで，$\boldsymbol{X}$ は平均ベクトル $\boldsymbol{0}$，分散共分散行列 $LL' + \Psi$ の p 次元正規分布

$$\boldsymbol{X} \sim N_p(\boldsymbol{0}, LL' + \Psi) \tag{3}$$

に従う．

N 個の p 次元データ $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ が観測されたとき，モデルのパラメータ L, Ψ を最尤法によって推定する．(3) 式より，対数尤度関数 $l(L, \Psi)$ は

$$l(L, \Psi) = -\frac{N}{2}\left\{p \log(2\pi) + \log|\Sigma| + \mathrm{tr}(\Sigma^{-1}S)\right\} \tag{4}$$

で与えられる．ただし，$\Sigma = LL' + \Psi$ とし，S は標本分散共分散行列である．対数尤度関数 $l(L, \Psi)$ の最大化によって最尤推定値 $\hat{L}, \hat{\Psi}$ を得る．

因子分析モデルのパラメータを最尤法で推定すると，独自分散の推定値が負となる不適解問題が生じることがある．そこで，不適解を防ぐためにベイズアプローチに基づくモデルの推定を検討する．ベイズアプローチでは，パラメータに事前分布を仮定し，事後分布のモードまたは事後平均で推定するが，本報告では事後分布のモードで推定を行う．事後分布のモードは，対数尤度に罰則を加えた罰則付き対数尤度関数

$$l_\rho(L, \Psi) = l(L, \Psi) - \rho \frac{N}{2} H(L, \Psi) \tag{5}$$

の最大化により得られる推定値と一致する．ただし，$H(L, \Psi)$ は罰則項で，独自分散の推定値が 0 に近づくのを防ぐ役割を果たす．また，ρ は事前分布のハイパーパラメータである．

ベイズ推定では事前分布の設定が重要な問題となる．(5) 式より，事前分布を設定することは，罰則項 $H(L, \Psi)$ を定めることと同等であるので，罰則項の設定が本質的となる．Akaike (1987) は，独自分散が 0 に近づくと，$\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$ の大きな固有値が過度に大きくなることを示し，それを防ぐための罰則項を導入した．本報告では，Akaike (1987) に基づき，以下の罰則項を用いることを提案する．

$$H(L, \Psi) = \mathrm{tr}(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}) \tag{6}$$

このとき事前分布の密度関数は

$$f(\psi_i^{-2}) = \frac{N \rho s_{ii}}{2} \exp\left\{ -\frac{N \rho s_{ii}}{2} \psi_i^{-2} \right\} \tag{7}$$

で与えられる．ただし，s_{ii} は標本分散共分散行列 S の第 (i, i) 成分である．提案した (**??**) 式の事前分布は，独自分散 ψ_i^{-2} に指数分布を仮定している．

3　モデル評価基準

ベイジアン因子分析モデルでは，適切なハイパーパラメータ ρ と因子数 k の選択が重要な問題となる．しかし，多くの論文ではこれらの値を主観的に与えており，観測データに基づいて客観的に選択するための方法が必要と考える．本報告では，ハイパーパラメータと因子数を同時に決定するために一般化ベイズ型モデル評価基準 GBIC (Konishi *et al.*, 2004) を導出する．この評価基準を最小にするようなハイパーパラメータと因子数を適切なモデルとして選択する．

参考文献

[1] Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

[2] Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*. **91**(1), 27-43.

[3] Martin, J.K. and McDonald, R.P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, **40**, 505-517.