Group Lasso に基づく因子分析モデリング

九州大学大学院数理学府　廣瀬　慧
九州大学大学院数理学研究院　小西　貞則

1　はじめに

因子分析モデルとは，観測される多変量データの相関構造から，観測されない共通因子を見出すモデルであり，心理学，社会科学，生命科学をはじめとする諸科学の様々な分野で応用されている．因子分析モデルは通常最尤法によって推定されるが，観測変数の数や因子数が多くなると，モデルに含まれるパラメータ数が膨大となるため，推定が不安定になりやすい．さらに，因子分析を行う上で不要な観測変数が存在するとき，その変数が原因で推定が不安定になり，時には不適解（独自分散の推定値が負になる現象）が発生し，不適切なモデルを推定することがある（たとえば，狩野，1998）．また，因子分析モデルの目的は，推定されたモデルから観測されない共通因子を解釈することにあるが，因子数が変わると，推定されたモデルの解釈も変わる．

本報告では，これらの問題点に着目し，次の 3 つの項目を同時に実行できる新たな因子分析モデリング手法を提案する．

- 安定した推定量を求める．
- 因子分析を行う上で不要な観測変数が存在した時にその変数を削除する．
- 適切な因子数を選択する．

まず，対数尤度関数に罰則項を加えた罰則付き対数尤度関数の最大化によってモデルを推定することにより，推定の安定化を図る．ここで，どのような罰則項を用いるかが重要となるが，因子分析を行う上で不要な観測変数が存在した時にその変数を削除するため，Group Lasso (Yuan and Lin, 2006) の考え方に基づいて，因子負荷行列の各行ベクトルに L_1 ノルムの罰則項を導入する．また，このモデルを EM アルゴリズム (Rubin and Thayer, 1982) と 2 次近似 (Fan and Li, 2001) を組み合わせたアルゴリズムを用いて推定する．さらに，このモデルに含まれる平滑化パラメータの値および因子数を客観的に選択するために，モデル評価基準 (Konishi and Kitagawa, 2008) を導出する．提案した一連のモデリング手法を実データに適用し，その特徴と有効性を検証する．

2　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関係を仮定する．

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

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 $p \times k$ 因子負荷行列とする．以下，一般性を失うことなく $\boldsymbol{\mu} = \boldsymbol{0}$ の場合を考える．ここで，

潜在変数ベクトル $\boldsymbol{F}$ と独自因子ベクトル $\boldsymbol{\varepsilon}$ は互いに独立とし，それぞれ次の多変量正規分布に従うとする．

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

ただし，$\Psi = \mathrm{diag}(\psi_1^2, \cdots, \psi_p^2)$ とする．この仮定のもとで，観測変数ベクトル $\boldsymbol{X}$ は平均ベクトル $\mathbf{0}$，分散共分散行列 $LL' + \Psi$ の p 次元正規分布

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

に従う．これより，第 i 番目の観測変数 X_i の分散は

$$\mathrm{Var}(X_i) = \sum_{j=1}^{k} \ell_{ij}^2 + \psi_i^2 \tag{4}$$

と表される．ここで，$\sum_{j=1}^{k} \ell_{ij}^2$ は共通性と呼ばれ，第 i 番目の観測変数 X_i の分散の中で，共通因子で表される分散の大きさを表す．(4) 式は，第 i 番目の観測変数 X_i の分散が共通性と独自分散の和で書けることを意味する．

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ことを考える．(3) 式より，対数尤度関数 $l(L, \Psi)$ は

$$l(L, \Psi) = -\frac{N}{2} \left\{ p \log(2\pi) + \log|\Sigma| + \mathrm{tr}(\Sigma^{-1} S) \right\} \tag{5}$$

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

$$S = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n' \tag{6}$$

とする．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．ただし，(5) 式を最大にするパラメータ L, Ψ を解析的に求めることは困難であるので，反復法によって推定する必要がある．本報告では EM アルゴリズム (Rubin and Thayer, 1982) を適用することにより推定する．

3 Group Lasso に基づく因子分析モデリング

(1) 式で与えられる因子分析モデルの第 i 番目の観測変数を成分表示すると，

$$X_i = \sum_{j=1}^{k} \ell_{ir} F_j + \varepsilon_i$$

となる．ここで，$\ell_{i1}, \cdots, \ell_{ik}$ がすべて 0 であるという状況を考える．このとき，

$$X_i = \varepsilon_i$$

となり，第 i 変数が独自因子のみの変数となる．すなわち，第 i 変数は共通因子で表されないと考えることができ，因子分析を行う上で不要な変数とみなすことができる．

本報告では，因子分析を行う上で不要な観測変数 X_i が含まれるとき，$\ell_{i1}, \cdots, \ell_{ik}$ のすべてを 0 に縮小推定できるモデリング手法を提案する．

3.1 罰則項付き最尤法

Group Lasso (Yuan and Lin, 2006) は，線形回帰モデルにおいて，回帰係数ベクトルをグループに分けたとき，そのグループに含まれるパラメータの推定値をすべて 0 に縮小推定することのできる手法である．本報告では，Group Lasso の基本的な考え方に基づいて，対数尤度関数に因子負荷行列 L の各列に L_1 ノルムの罰則項を加えた罰則付き対数尤度関数

$$l_\lambda(L,\Psi) = l(L,\Psi) - N\lambda \sum_{i=1}^{p} \|\boldsymbol{\ell}_i\| \tag{7}$$

の最大化によりモデルを推定することを考える．ただし，$\boldsymbol{\ell}_i$ は因子負荷行列 L の第 i 行ベクトル，$\lambda\ (>0)$ は平滑化パラメータ，$\|\boldsymbol{\ell}_i\|$ は

$$\|\boldsymbol{\ell}_i\| = \sqrt{\sum_{j=1}^{k} \ell_{ij}^2}$$

とする．この罰則項は，$\ell_{i1},\cdots,\ell_{ik}$ のすべてを 0 に縮小推定することができるという特徴を持つ．

しかし，(7) 式を用いてパラメータを推定すると，平滑化パラメータの値が小さい時は 0 に推定されず，ある程度大きな値になると同時にほとんど 0 に推定するという現象が起こった．この問題に対処するため，次の罰則項に重みをつけた罰則付き対数尤度関数の最大化を考える．

$$l_\lambda(L,\Psi) = l(L,\Psi) - N\lambda \sum_{i=1}^{p} w_i \|\boldsymbol{\ell}_i\|. \tag{8}$$

ただし，w_i は正数とする．これは，Zou (2006) や，Tateishi, Matsui and Konishi (2009) の提案した重み付きの Lasso 推定に基づいて構成されている．Tateishi, Matsui and Konishi (2009) は，ガウス型基底関数に基づく非線形回帰モデルにおいて，重みをつけた Lasso を適用することにより，より安定した推定量を求めることを可能にした．

ここで，どのような重み w_i を用いるかが重要となる．本報告では，変数間の相関に着目し，次のような重みを用いる．

$$w_i = \left(\frac{1}{p}\sum_{h=1}^{p} s^{hh}\right)\cdot\frac{1}{s^{ii}}. \tag{9}$$

ただし，s^{ii} は，分散共分散行列の逆行列の第 (i,i) 成分とする．s^{ii} は分散拡大係数とよばれ，変数相互間の相関の高さを表し，この値が大きいほど他の変数との相関が高いと考えられる．ゆえに，この重みを採用することにより他の変数と相関の低い変数の共通性が 0 に推定されるという特徴を持つ．

3.2 推定

因子分析モデルを最尤法で推定するアルゴリズムは数多く提案されており，よく用いられるアルゴリズムとして，行列 $\Psi^{-1/2}S\Psi^{-1/2}$ または $S^{-1/2}\Psi S^{-1/2}$ の固有値問題を解くことにより L と Ψ を交互に推定するという方法がある (たとえば，Jennrich and Robinson, 1982)．

しかし，この方法を罰則付き最尤法の推定に拡張することは困難である．そこで，EM アルゴリズム (Rubin and Thayer, 1982) を用いて罰則付き対数尤度関数の最大化を行うことを考える．

まず，潜在変数 $\boldsymbol{F}$ を不観測変数とみなし，観測データ $\boldsymbol{X}=\boldsymbol{x}$ が与えられたもとでの $\boldsymbol{F}$ の事後分布 $\boldsymbol{F}|\boldsymbol{x}$ を計算する．次に，完全対数尤度関数に罰則項を加えた罰則付き完全対数尤度関数を計算し，その最大化を行う．しかし，罰則項をそのまま用いると M ステップでの因子負荷行列 L の更新式を陽に表すことが困難となる．そこで，Fan and Li (2001) に基づいて，罰則項を次のように 2 次近似することを考える．

$$\|\boldsymbol{\ell}_i\| \approx \|\boldsymbol{\ell}_i^{(\mathrm{old})}\| + \frac{1}{2\|\boldsymbol{\ell}_i^{(\mathrm{old})}\|}(\boldsymbol{\ell}_i'\boldsymbol{\ell}_i - \boldsymbol{\ell}_i^{(\mathrm{old})\prime}\boldsymbol{\ell}_i^{(\mathrm{old})}) \qquad \text{for } \boldsymbol{\ell}_i \neq \mathbf{0}. \tag{10}$$

ただし，$\boldsymbol{\ell}_i^{(\mathrm{old})}$ はパラメータを更新する前の値とする．この 2 次近似を適用することにより，L, Ψ の更新式を求めることができ，次で与えられる．

$$\hat{\boldsymbol{\ell}}_i = \left\{(M^{-1}+M^{-1}L'\Psi^{-1}S\Psi^{-1}LM^{-1})/\psi_i^2 + (\lambda w_i/\|\boldsymbol{\ell}_i\|)I_k\right\}^{-1}\left\{\frac{1}{\psi_i^2}M^{-1}L'\Psi^{-1}S_{i.}\right\},$$

$$\hat{\Psi} = \mathrm{diag}\left[S - 2S\Psi^{-1}LM^{-1}\hat{L}' + \hat{L}M^{-1}\hat{L}' + \hat{L}M^{-1}L'\Psi^{-1}S\Psi^{-1}LM^{-1}\hat{L}'\right].$$

ただし，$S_{i.}$ は行列 S の第 i 列，$M = L'\Psi^{-1}L + I_k$ とする．なお，EM アルゴリズムにおいてステップを更新している途中に $\|\boldsymbol{\ell}_i\|$ の値が十分小さな値に達したとき，$\boldsymbol{\ell}_i = \mathbf{0}$ と推定する．

3.3 モデル評価基準

Group Lasso に基づく因子分析モデルにおいて，どの観測変数を不要な変数と見なすかということが本質的となる．この不要な変数の選択は，平滑化パラメータ λ の選択と対応する．また，因子分析モデルにおけるもう一つの重要な問題として，適切な因子数をどのように選択するかということがある．本報告では，モデル評価基準 GBIC (Konishi *et al.*, 2004) を Group Lasso に基づく因子分析モデルの枠組みで導出し，因子数と平滑化パラメータの値を同時に選択する方法を提案する．

まず，(8) 式の最大化は，パラメータ L に次の事前分布を仮定して，事後分布のモードで推定することと同等である．

$$\pi_\lambda(L) = C\prod_{i=1}^{p}\exp\left(-N\lambda w_i\|\boldsymbol{\ell}_i\|\right). \tag{11}$$

ただし，C は基準化定数

$$C = (C_k)^{p-k}\prod_{j=1}^{k-1}C_j, \quad C_j = \frac{(N\lambda w_j)^j}{2^j\pi^{\frac{j-1}{2}}\Gamma\left(\frac{j+1}{2}\right)} \tag{12}$$

とする．ここで，$\Gamma(\cdot)$ はガンマ関数とする．

一般化ベイズ型モデル評価基準 GBIC は，ベイズ型モデル評価基準 BIC (Schwarz, 1978) を拡張することによって導出された評価基準である．BIC は最尤法で推定されたモデルを評価する基準であるが，GBIC は罰則付き最尤法で推定されたモデルを評価することのできる基準である．GBIC の基本的な考え方は，モデルが生起する確からしさを表す周辺尤度を最大とするようなモデルを選ぶというものである．一般に，周辺尤度の積分計算は困難であるため，ラプラス近似を適用することにより次の GBIC の一般式を得る．

$$\mathrm{GBIC} = -2\log\left\{\int f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\lambda)d\boldsymbol{\theta}\right\} \tag{13}$$

$$\approx -p^*\log(2\pi) + p^*\log N + \log|J_\lambda(\hat{\boldsymbol{\theta}}_{p^*})| - 2\left\{\log f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\hat{\boldsymbol{\theta}}_{p^*}) + \log\pi(\hat{\boldsymbol{\theta}}_{p^*}|\lambda)\right\}. \tag{14}$$

ただし，p^* はパラメータ数

$$p^* = p(k+1) - \frac{k(k-1)}{2} \tag{15}$$

とし，$\hat{\boldsymbol{\theta}}_{p^*}$ は事後分布のモード，$J_\lambda(\hat{\boldsymbol{\theta}}_{p^*})$ は

$$J_\lambda(\hat{\boldsymbol{\theta}}_{p^*}) = -\frac{1}{N}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left\{\log f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta}) + \log\pi(\boldsymbol{\theta}|\lambda)\right\}\Bigg|_{\hat{\boldsymbol{\theta}}_{p^*}}\right] \tag{16}$$

とする．

(11) 式の事前分布を (14) 式に代入することにより，提案した Group Lasso に基づく因子分析モデルの枠組みで GBIC を導出できる．

$$\begin{aligned}\mathrm{GBIC} \approx& -p^*\log(2\pi) + p^*\log N + \log|J_\lambda(\hat{\boldsymbol{\theta}}_{p^*})| + N\left\{p\log(2\pi) + \log|\hat{\Sigma}| + \mathrm{tr}(\hat{\Sigma}^{-1}S_N)\right\}\\ &-2\log C + 2N\lambda\sum_{i=1}^{p} w_i\|\boldsymbol{\ell}_i\|.\end{aligned} \tag{17}$$

ただし，$\hat{\Sigma}, \hat{\psi}_i^2$ は Σ, ψ_i^2 の事後分布のモードとする．なお，$J_\lambda(\hat{\boldsymbol{\theta}}_{p^*})$ に含まれる罰則付き対数尤度関数の 2 階微分は次で与えられる．

$$\begin{aligned}\frac{\partial^2 l_\lambda(\boldsymbol{\theta})}{\partial\ell_{ab}\partial\ell_{cd}} = N\Bigg\{& (\Sigma^{-1})_{(a,c)}(L'\Sigma^{-1}L)_{(b,d)} + (\Sigma^{-1}L)_{(a,d)}(\Sigma^{-1}L)_{(c,b)}\\ &-(\Sigma^{-1}S\Sigma^{-1})_{(a,c)}(L'\Sigma^{-1}L)_{(b,d)} - (\Sigma^{-1}L)_{(a,d)}(\Sigma^{-1}S\Sigma^{-1}L)_{(c,b)}\\ &-(\Sigma^{-1})_{(a,c)}(L'\Sigma^{-1}S\Sigma^{-1}L)_{(b,d)} - (\Sigma^{-1}S\Sigma^{-1}L)_{(a,d)}(\Sigma^{-1}L)_{(c,b)}\\ &-(\Sigma^{-1})_{(a,c)}(I_k)_{(b,d)} + (\Sigma^{-1}S\Sigma^{-1})_{(a,c)}(I_k)_{(b,d)}\\ &+\lambda w_c\frac{(L)_{(a,b)}(L)_{(c,d)}(I_p)_{(a,c)}}{\sqrt{(LL')_{(c,c)}}^3} - \lambda w_c\frac{(I_p)_{(a,c)}(I_k)_{(b,d)}}{\sqrt{(LL')_{(c,c)}}}\Bigg\},\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{\partial^2 l_\lambda(\boldsymbol{\theta})}{\partial\psi_i^2\partial\ell_{cd}} = N\Bigg\{& (\Sigma^{-1})_{(c,i)}(\Sigma^{-1}L)_{(i,d)} - (\Sigma^{-1}S\Sigma^{-1})_{(c,i)}(\Sigma^{-1}L)_{(i,d)}\\ &-(\Sigma^{-1})_{(c,i)}(\Sigma^{-1}S\Sigma^{-1}L)_{(i,d)}\Bigg\},\end{aligned} \tag{19}$$

$$\frac{\partial^2 l_\lambda(\boldsymbol{\theta})}{\partial\psi_i^2\partial\psi_j^2} = \frac{N}{2}\left\{(\Sigma^{-1})^2_{(i,j)} - 2(\Sigma^{-1})_{(i,j)}(\Sigma^{-1}S\Sigma^{-1})_{(i,j)}\right\}. \tag{20}$$

ただし，$A_{(\alpha,\beta)}$ は，行列 A の (α,β) 成分とする．

ここで，共通性が 0 に推定されると，GBIC をそのまま計算することが困難となるという問題が生じる．実際，共通性が 0 に推定されると，(18) 式は発散する．この問題に対処するため，共通性が 0 となった観測変数に対して，推定する前に共通性が 0 であるという制約のもとで推定したと解釈する．すなわち，第 i 番目の変数の共通性が 0 と縮小推定されたら，$\ell_{i1},\cdots,\ell_{ik}$ はあらかじめ 0 という条件のもとで推定されたと見なす．また，それに伴い，パラメータ数も変わる．一般に，u 個の観測変数の共通性が 0 に推定されたら，パラメータ数は

$$p^* = p(k+1) - \frac{k(k-1)}{2} - uk \tag{21}$$

となる．

なお，因子負荷行列の回転の不変性があるため，GBIC を導出する際に，一意性の条件が必要となる．本報告では，因子負荷行列の一意性の条件である $\ell_{ij} = 0$（$i < j$）という制約 (たとえば，Anderson and Rubin, 1956 参照) のもとで GBIC を導出する．

表 1: 因子数が 1 ～ 6 のときの AIC，BIC，GBIC と平滑化パラメータ λ の値．AIC，BIC は最尤法によって推定されたモデルを評価しており，GBIC は Group Lasso に基づく因子分析モデルを評価している．

因子数	AIC	BIC	不適解の発生	提案手法（GBIC）	λ
1	1660.54	1716.68	発生しなかった	1718.50	0.0363
2	1605.39	1687.72	発生しなかった	1705.18	0.0513
3	1562.46	1669.11	発生しなかった	1699.33	0.0829
4	1532.43	1661.54	発生した	1708.89	0.0796
5	1525.34	1675.04	発生した	1725.95	0.1287
6	1521.25	1689.66	発生した	1748.31	0.1286

4 実データへの適用

提案したモデリング手法の有効性を検証するため，Kendall のデータ（Kendall, 1980）に適用した．このデータは，ある求人に対する 48 人の応募者についての 15 変数スコア，すなわち，$N = 48,\ p = 15$ である．その具体的なスコアは以下のとおりである．

1. 履歴書	2. 容姿	3. 学力	4. 好感度	5. 自信
6. 明晰さ	7. 誠実さ	8. 技量	9. 経験	10. 志望動機
11. 抱負	12. 理解力	13. 将来性	14. 志望程度	15. 会社向き

このデータに対し，因子数が 1 ～ 6 のとき，最尤法および提案手法を適用した．まず，最尤法で推定し，推定されたモデルを AIC，BIC で評価した．また，提案した Group Lasso に基づく因子分析モデルを推定し，推定されたモデルを GBIC で評価した．なお，不適解が発生したときは，独自分散が 0.005 以上という制約のもとで推定した．これらの結果を表 1 にまとめた．表 1 より，AIC，BIC はともに不適解の発生したモデルを選択している．一方，提案手法は因子数 3 を選択し，かつ「学力」という変数を不要な変数と見なしている（表 3 参照）．

図 1 は，因子数を 3，4 としたときの横軸を平滑化パラメータ λ，縦軸を GBIC としたときのグラフである．これらのグラフから，GBIC は凸になっておらず，所々急激に小さな値を取るところがある．この境目が，不要な変数を取り除いた部分であり，たとえば，因子数 3 のときは，最初の変化点が学力という変数を取り除いたときとなっている．

因子数が 4 の時の最尤推定値と，提案したモデリング手法が選択したモデルを比較してみる．因子数が 4 の時の最尤推定値を表 2 に，提案手法が選択したモデルを表 3 に示す． 表 2 の 4 因子モデルでは不適解が発生したが，第 4 因子をみると，因子負荷量は「学力」と「志望程度」のみ大きな値を取っており，そのほかの値は 0 に近い．よって，第 4 因子を「学力」と解釈することはできる．

一方，表 3 の結果を見ると,「学力」という変数を不要と見なしている．4 因子モデルのときには,「学力」という変数が重要な役割を果たしていたが，3 因子モデルでは不要であると見なされるのは興味深い．Group Lasso に基づいて推定された共通因子は，第一因子が「やる気と能力」，第 2 因子が「経歴と適性」．第三因子が「人柄」と適切な解釈ができる．

また，3 因子モデルで，提案手法が学力を「不要」とみなしたのは，$\lambda = 0.0478$ の時であった．一方，罰則項に重みをつけない (7) 式の最大化によりモデルを推定すると，学力を「不要」

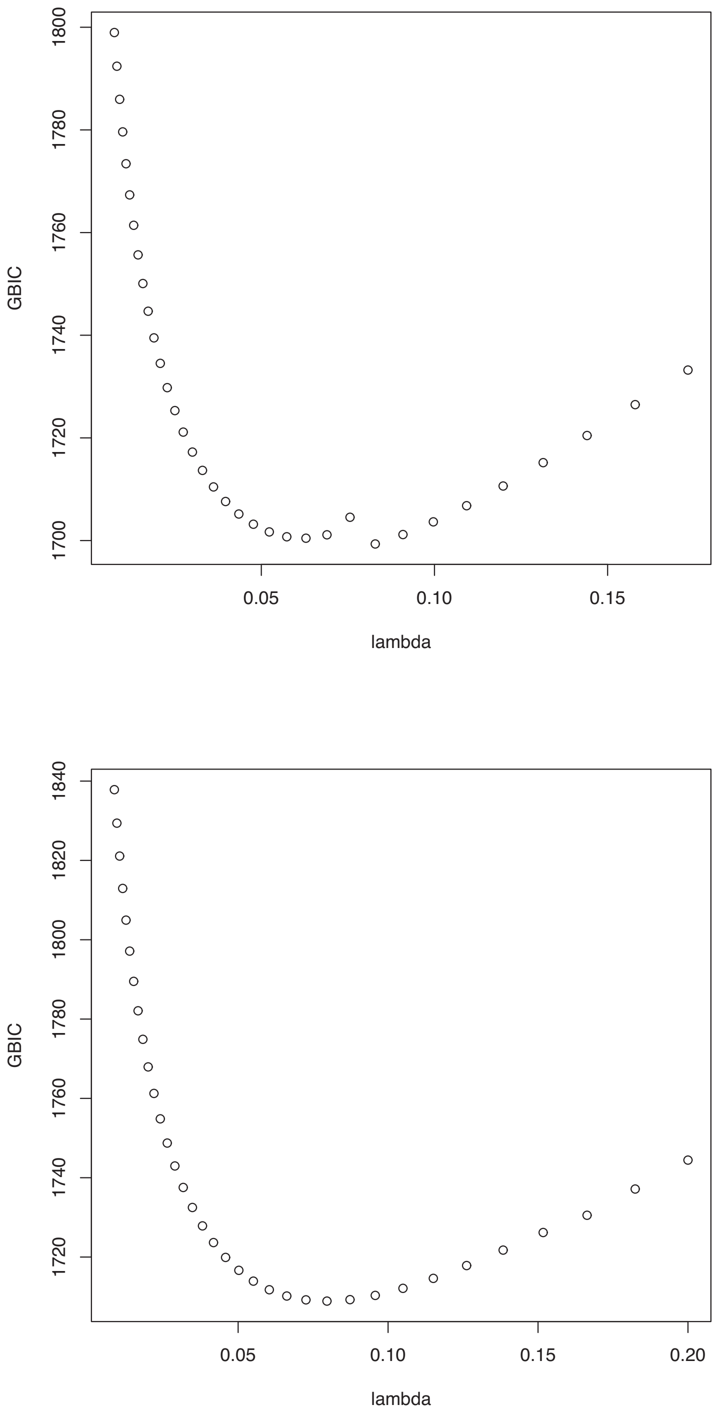

図 1: 因子数 3（上図），4（下図）のときの，平滑化パラメータを動かしたときのグラフ．横軸が平滑化パラメータ λ，縦軸が GBIC の値を表す．

表 2: 4 因子モデルの最尤推定値

	Factor1	Factor2	Factor3	Factor4	uniqueness
履歴書	0.13	0.72	0.11	-0.12	0.44
容姿	0.45	0.14	0.24	0.16	0.69
学力	0.07	0.12	0.00	0.68	0.52
好感度	0.23	0.24	0.83	-0.05	0.20
自信	0.92	-0.10	0.15	-0.09	0.11
明晰さ	0.84	0.12	0.30	0.06	0.19
誠実さ	0.25	-0.22	0.74	-0.02	0.34
技量	0.89	0.24	0.08	-0.07	0.13
経験	0.09	0.78	-0.05	0.17	0.36
志望動機	0.77	0.39	0.18	-0.06	0.22
抱負	0.90	0.18	0.11	-0.06	0.14
理解力	0.78	0.28	0.36	0.16	0.15
将来性	0.73	0.35	0.44	0.25	0.09
志望程度	0.42	0.39	0.56	-0.59	0.00
会社向き	0.36	0.77	0.05	0.14	0.25

表 3: 提案手法の選択したモデル（$\lambda = 0.0478$）

	Factor1	Factor2	Factor3	uniqueness
履歴書	0.08	0.45	0.10	0.57
容姿	0.27	0.12	0.17	0.72
学力	**0.00**	**0.00**	**0.00**	1.00
好感度	0.10	0.17	0.86	0.01
自信	0.79	-0.12	0.11	0.11
明晰さ	0.69	0.08	0.25	0.19
誠実さ	0.18	-0.17	0.56	0.45
技量	0.74	0.18	0.08	0.14
経験	0.05	0.60	-0.05	0.41
志望動機	0.63	0.30	0.12	0.23
抱負	0.75	0.10	0.10	0.16
理解力	0.65	0.21	0.24	0.21
将来性	0.58	0.28	0.34	0.19
志望程度	0.32	0.19	0.48	0.42
会社向き	0.26	0.70	0.05	0.16

とみなしたのは，$\lambda = 0.2$ の時であった．すなわち，提案手法は，平滑化パラメータの値が小さくても不要な因子を削除することができるという特徴を持つことが分かった．

5　まとめと今後の課題

本報告では，安定した推定量を求め， 因子分析を行う上で不要な観測変数が存在した時にその変数を削除し，さらに適切な因子数を選択することのできる新たな因子分析モデリング手法を提案した．まず，推定の不安定性という問題に対処するために，対数尤度関数に罰則項を加えた罰則付き対数尤度関数の最大化によりモデルを推定した．ここで，どのような罰則項を用いるかということが重要となるが，Group Lasso (Yuan and Lin, 2006) の基本的な考え方に基づいて，因子負荷行列の各行ベクトルに L_1 ノルムの重み付き罰則項を導入した．このモデルを推定するため，EM アルゴリズムと 2 次近似を組み合わせるアルゴリズムを用いた．さらに，このモデルに含まれる平滑化パラメータの値および因子数を同時に選択するために，モデル評価基準 (Konishi *et al.*, 2004) を導出した．

また，提案したモデリング手法を Kendall のデータに適用し，その有効性を検証した．まず，最尤法で推定し，推定されたモデルを BIC で評価すると，因子数 4 が選択された．しかし，因子数 4 では不適解が発生しており，推定されたモデル解釈が困難となる．一方，提案手法は，「学力」という変数を取り除いた因子数 3 のモデルが選択された．このようなモデルは，最尤法では決して推定できないモデルであり，提案手法の有用性を実証できた．

参考文献

[1] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis, In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **5**, (pp. 111–150). Berkeley: University of California Press.

[2] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). *Ann. Statist.*, **32**, 407–499.

[3] Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, **96**, 1348–1360.

[4] Jennrich, R. I. and Robinson, S. M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis. *Psychometrika*, **34**, 111–123.

[5] 狩野裕 (1998)．不適解の原因と処理．探索的因子分析．大阪大学人間科学部紀要．

[6] Kendall, M. G. (1980). *Multivariate Analysis* (2nd.ed.). Charles Griffin.

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

[8] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.

[9] Rubin, D. B. and Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, **47**, 69–76.

[10] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.

[11] Shimamura, T., Imoto, S., Yamaguchi, R. and Miyano, S. (2007). Weighted lasso in graphical Gaussian modeling for large gene network estimation based on microarray data. *Genome Inform.* **19**, 142–153.

[12] Tateishi, S., Matsui, H. and Konishi, S. (2009). Nonlinear regression modeling via the lasso-type regularization. Preprint, MI 2009-8, Kyushu University.

[13] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, **58**, 267–288.

[14] Yuan, M and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. Ser. B*, **68**, 49–67.

[15] Zou, H. (2006). The adaptive Lasso and its oracle properties. *J. Amer. Statist. Assoc.*, **101**, 1418–1429.