

非正規データにおける情報量規準を用いた 共分散構造モデルの選択問題

—「日本人の国民性調査」データへの適用—

柳原 宏和[†]

(受付 2004年7月30日; 改訂 2005年2月4日)

要 旨

データに正規性を仮定した下での共分散構造モデルの選択問題に関する赤池情報量規準(AIC)は、実際に観測値が従う分布である真の分布が正規分布でなく未知の非正規分布であった場合、真の分布の尖度に依存した定数バイアスを持つ。本論文の目的は、部分的に交差確認法(クロス・バリデーション)を適用することでバイアスを補正した、修正 AIC を選択規準に用いることにより真の分布の非正規性の影響が小さくなるような選択法を提案することである。この修正 AIC のバイアスは AIC や TIC のバイアスよりも小さくなることを漸近展開式とシミュレーション実験により確かめる。また、実解析への適用例として国民性調査第11回全国調査のデータに「日本人の満足感」に関する因子分析モデルをあてはめ、その最適モデルを探索する。

キーワード：赤池情報量規準，因子分析モデル，交差確認法，非正規性，モデル選択。

1. はじめに

日本人の国民性調査データに関する解析における重要な目的の一つに、類似した複数の質問項目の背後にあってそれらの質問項目に対する共通の回答傾向を生ぜしめる要因を探索するということがある。しかしながら、このような要因は直接観測されないことが多い。そのため、意味のある因果関係を、観測変数間の関係としてよりもむしろ、潜在変数(直接観測されない変数)間の関係として表現し、その潜在変数を確率変数と置く。そうすると、潜在変数間の因果関係の探索問題は、観測変数の共分散構造分析などに代表される、共分散行列の構造に関する推定問題に帰着され、目的とする因果関係の推論が可能となる。以上のことから、国民性調査データに関する解析において共分散構造モデルは非常に重要なモデルであると言え、前田(1995)においてもその有用性が論じられている。共分散構造モデルにおいて、仮定した因果関係にはそれに対応した共分散構造があり、それらのどの構造を選べば良いかということは非常に重要な問題となる。

観測値が本当に従う分布である真の分布はほとんどの場合未知であり、一般的な統計的解析手法では観測値が正規分布に従っているという仮定の下で議論を進めているものが多い。他の分布に較べ最尤推定法が簡便であり、また従来のソフトウェアでは正規性の仮定の下で提案さ

[†] 筑波大学大学院 システム情報工学研究科：〒305-8573 茨城県つくば市天王台 1-1-1；
yanagi@sk.tsukuba.ac.jp

れた統計的手法が主流になっていることを考えると、解析の初期段階で取敢えず正規性を仮定することは、理論的には問題はあるかもしれないが、実用面から考えると解析の妥当な第一歩であると考えられる。正規性の仮定の下で最も一般的なモデル選択法は、Akaike(1973)により提案された赤池情報量規準(AIC)が最も小さくなるモデルを最適なモデルとする手法である。AICは予測に関するカルバック=ライブラーの距離に基づくリスクの推定量の一つであり、標本対数尤度関数の -2 倍にリスクに対するバイアス補正項である「独立パラメータ数の2倍」を足すことにより定義され、その式の簡便性から様々な分野で応用されている。ところが、AICはそのバイアス補正項のシンプルさから様々な状況で無視できないバイアスを持つことがある。そのため、Sugiura(1978)、Bozdogan(1987)、Hurvich and Tsai(1989)、Fujikoshi and Satoh(1997)等の多くの研究者により、そのようなバイアスを補正した情報量規準が多く提案されている。

しかしながら、これらのバイアスを補正した情報量規準は仮定したモデルの分布と真の分布が同じであるという条件の下で提案されたもので、もしそれらの分布間に違いが生じればまた異なるバイアスが発生する。国民性調査データなど社会学、心理学の分野において興味のあるデータはカテゴリカルデータであることが多いため、正規性の仮定を満たしていないものが多い。そのため、仮定した分布と真の分布の違いの影響を受けやすい手法でモデル選択を行えば、解析結果にバイアスが生じることも十分考えられる。

竹内(1976)によって提案されたTICは、候補のモデルの分布と真の分布が異なるという仮定の下で導出された情報量規準であり、真の分布が独立に同一な分布であればAICが持つような定数バイアスを補正しているということが知られている。このTICでは、最尤推定量の漸近正規性を用いてバイアス補正項を計算している。ところが多変量モデルでは、観測値の次元が標本の大きさに較べ大きくなれば推定量の正規分布への収束が遅くなり、結果的にTICのバイアスが大きくなることもある。またTICのバイアス補正項は真の分布の尖度に依存しているため、標本数が巨大でなければうまくバイアスを補正することができないということもFujikoshi et al.(2005)で報告されている。

情報量規準を用いる以外に仮説検定を行うことで共分散行列の構造を選択する方法もある。観測値が正規分布に従っているという仮定の下で導出された検定統計量の帰無分布は、真の分布が正規分布であった場合、漸近的に χ^2 分布に従うことが知られており、 χ^2 近似により検定を行うことがほとんどである。しかしながら、その漸近帰無分布は、正規性は仮定したが真の分布は非正規であるといった状況の下では真の分布の尖度に基づく重みを持った χ^2 分布の重み付き和の分布になる(Satorra and Bentler(1988)、Yuan and Bentler(1999)、Yanagihara et al.(2005)等参照)。そのため、真の分布が正規分布ではないときに χ^2 近似を用いて検定を行うと、パーセント点はずれ、解析者が望む有意水準と実際の有意水準が大きく異なってしまい、検定結果に誤りを与えてしまうことがある。特に、正規性の下での尤度比検定では、真の分布の尖度が大きくなるほど検定統計量の帰無仮説の下での漸近平均と漸近分散は χ^2 分布の平均と分散よりも大きくなるので、 χ^2 分布を使った検定は非保守的な検定になる傾向があることがYanagihara et al.(2005)で述べられている。真の分布の非正規性の影響を小さくするような修正をほどこした検定統計量もBrowne(1984)、Satorra and Bentler(1988)、Kano et al.(1990)、Kano(1992)等により提案されている。しかしながら、それらの多くは尖度の推定量に依存した補正法に基づくものであり、その補正がうまく機能するためには非常に大きな標本数が必要となる(Hu et al.(1992)、Yuan and Bentler(1998)等参照)。以上のことから、仮説検定で共分散構造を選択する場合、大きな危険性が存在していると考えられる。

本論文では、Yanagihara(2004a)でのアイデアを基に部分的に交差確認法(クロス・バリデーション)を用いることでAICのバイアスを補正し、その新しい情報量規準を使うことで非正

規性の影響を受けにくい選択法を提案する。真の分布がどのような分布であってもバイアスを補正しているこの情報量規準を使う選択法では、真の分布の非正規性の影響が小さくなっている。以下、第2章では、本論文で用いたモデルと今までに提案されている情報量規準 AIC と TIC を紹介し、新しいバイアス修正情報量規準である修正 AIC (Corrected AIC : AIC_c) を提案する。第3章では、 AIC_c は AIC や TIC よりも小さいバイアスを持ち、またリスクが最も小さいモデルを選ぶ頻度も改善していることをシミュレーションにより確かめる。第4章では、第11回国民性調査データ(坂元 他(2004))に「日本人の満足感」の構造(前田(1995))に関する因子分析モデルをあてはめ、その最適モデルの探索に新しい情報量規準を用いた解析例を紹介する。

2. 共分散構造モデルとモデル選択規準

本章では、共分散構造モデルとその最適モデルの選択のための従来の情報量規準を紹介し、更に従来の情報量規準の問題点を改善した新しい情報量規準を提案する。

2.1 カルバック=ライブラーの距離に基づくリスク

第 i 番目の個体 ($i = 1, \dots, n$) に対する $p \times 1$ 観測値ベクトルを $y_i = (y_{i1}, \dots, y_{ip})'$ とする。共分散構造モデルでは、観測値ベクトル y_i に $q \times 1$ 未知パラメータベクトル $\theta = (\theta_1, \dots, \theta_q)'$ で表現できる共分散行列 $\Sigma(\theta)$ を持つ候補のモデル、

$$(2.1) \quad M : y_i \sim \text{i.i.d. } N_p(\mu, \Sigma(\theta)), \quad (i = 1, \dots, n),$$

をあてはめる。ただし $\mu = (\mu_1, \dots, \mu_p)'$ は $p \times 1$ 未知パラメータベクトルである。本論文では、候補のモデルに正規性を仮定するが、真の分布は正規分布に従っているかどうか分からないという状況を考える。すなわち、真のモデルを

$$(2.2) \quad M_* : y_i = \mu + \Sigma_*^{1/2} \varepsilon_i, \quad (i = 1, \dots, n),$$

とする。ただし、それぞれの誤差ベクトル $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{ip})'$ は独立に同一な平均 $E[\varepsilon_i] = 0$ 、共分散行列 $\text{Cov}[\varepsilon_i] = I_p$ をもつ未知の分布に従うとし、 Σ_* は未知の真の共分散行列であるとする。

今、 $\xi = (\mu', \theta')'$ を全ての未知パラメータを並べた $(p+q) \times 1$ ベクトルとし、 $f(y_i|\xi)$ を平均 μ 共分散行列 $\Sigma(\theta)$ を持つ p 変量正規分布の密度関数、

$$(2.3) \quad f(y_i|\xi) = \left(\frac{1}{2\pi}\right)^{p/2} |\Sigma(\theta)|^{-1/2} \exp\left\{-\frac{1}{2}(y_i - \mu)' \Sigma(\theta)^{-1} (y_i - \mu)\right\},$$

とする。このとき、候補のモデル M (2.1) の下での対数尤度関数は、

$$(2.4) \quad \begin{aligned} \ell(\xi|Y) &= \sum_{i=1}^n \log\{f(y_i|\xi)\} \\ &= -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma(\theta)| - \frac{1}{2} \sum_{i=1}^n (y_i - \mu)' \Sigma(\theta)^{-1} (y_i - \mu), \end{aligned}$$

となる。ただし $Y = (y_1, \dots, y_n)'$ である。これにより、候補のモデル M の下での μ の最尤推定量は、

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

となり, θ の最尤推定量は,

$$(2.5) \quad \hat{\theta} = \arg \min_{\theta} \{ \log |\Sigma(\theta)| + \text{tr}(S\Sigma(\theta)^{-1}) \},$$

となる. ただし S は標本共分散行列で

$$S = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})',$$

である.

複数の候補のモデルに対してモデルの適合度を比較するためには, 何かしらのリスクを定義する必要がある. 最終的には, それぞれのモデルにおけるリスクを比較し, リスクが最小になるようなモデルを最適なモデルとみなす. 本論文では, そのようなリスクとして予測に基づくカルバック=ライブラーの距離を用いる. 今, $p \times 1$ ベクトル u_i ($i = 1, \dots, n$) を互いに独立に, また y_i とは独立に同一の分布に従うような将来の観測値ベクトルとすると, 候補のモデル M の下でのリスクは,

$$(2.6) \quad \begin{aligned} R_{KL} &= np \log(2\pi) + nE_Y^* \left[\log |\Sigma(\hat{\theta})| \right] \\ &\quad + \sum_{i=1}^n E_U^* E_Y^* \left[(u_i - \bar{y})' \Sigma(\hat{\theta})^{-1} (u_i - \bar{y}) \right] \\ &= np \log(2\pi) + nE_Y^* \left[\log |\Sigma(\hat{\theta})| \right] \\ &\quad + nE_Y^* \left[\text{tr}(\Sigma_* \Sigma(\hat{\theta})^{-1}) \right] + nE_Y^* \left[(\bar{y} - \mu)' \Sigma(\hat{\theta})^{-1} (\bar{y} - \mu) \right], \end{aligned}$$

と定義される. ただし, E^* は真のモデル M_* (2.2) の下での期待値を表すとし, $U = (u_1, \dots, u_n)'$ である.

2.2 AIC 規準

Akaike (1973) により提案された AIC はリスク (2.6) の推定量の 1 つである. その AIC は, $\hat{\xi} = (\bar{y}', \hat{\theta}')'$ とすると (2.4) 式 の対数尤度関数に $\xi = \hat{\xi}$ を代入した標本対数尤度関数 $\ell(\hat{\xi}|Y)$ を用いて,

$$(2.7) \quad \begin{aligned} AIC &= -2\ell(\hat{\xi}|Y) + 2(p+q) \\ &= np \log(2\pi) + n \log |\Sigma(\hat{\theta})| + \text{tr}(S\Sigma(\hat{\theta})^{-1}) + 2(p+q), \end{aligned}$$

と定義される. 共分散構造モデルの下位モデルである因子分析モデルでの AIC に関しては, Akaike (1987) により詳しく議論されているのでそちらを参照して欲しい. 式 (2.7) からわかるように, AIC は標本対数尤度の -2 倍, すなわち, $-2\ell(\hat{\xi}|Y)$ でリスク R_{KL} (2.6) を推定したときに生じるバイアスを $2(p+q)$ で補正した推定量となっている. つまり AIC では,

$$\begin{aligned} R_{KL} - E_Y^*[-2\ell(\hat{\xi}|Y)] &= nE_Y^*[\text{tr}(\Sigma_* \Sigma(\hat{\theta})^{-1})] \\ &\quad + nE_Y^*[(\bar{y} - \mu)' \Sigma(\hat{\theta})^{-1} (\bar{y} - \mu)] - nE_Y^*[\text{tr}(S\Sigma(\hat{\theta})^{-1})] \\ &\approx 2(p+q), \end{aligned}$$

と近似していることになる. しかしながら上記の式からわかるように, バイアスは真の分布に依存しているにも拘らず, AIC のバイアス補正項はどのような分布に対しても一定である. その結果, AIC はリスク R_{KL} に対し真の分布に依存するバイアスを持つことになる.

AIC のバイアスを計算すると以下ようになる. 今, 候補のモデル M (2.1) の下で標本数 n を大きくしたときに $\hat{\theta}$ (2.5) が収束する値を θ_0 , つまり, $\theta_0 = \lim_{n \rightarrow \infty} \hat{\theta}$ とし, $\Omega = \Sigma_*^{1/2} \Sigma(\theta_0)^{-1/2}$

と置く．このとき， AIC のリスク R_{KL} に対するバイアス $B_{AIC} = R_{KL} - E_Y^*[AIC]$ は，

$$(2.8) \quad B_{AIC} = 2\text{tr}(\Omega'\Omega) + 2\text{tr}\left(\{\Omega'\Omega \otimes \Omega'\Omega\} D(\theta_0|\Sigma_*)\right) \\ + \text{tr}\left(\Gamma_\kappa \{\Omega \otimes \Omega\} D(\theta_0|\Sigma_*) \{\Omega' \otimes \Omega'\}\right) - 2(p+q) + O(n^{-1}),$$

となる(導出に関しては Appendix A.3 を参照)．ただし， $D(\cdot|\cdot)$ は Appendix A.2 の(A.6)式で定義される $p^2 \times p^2$ 行列である．また， δ_{ab} を $a = b$ のとき 1 で $a \neq b$ であれば 0 となるクロネッカーデルタ， $e_j^{(p)}$ を第 j 番目の成分が 1 で残りの成分が 0 となる $p \times 1$ ベクトル， $\sum_{a_1, a_2, \dots}^p$ は $\sum_{a_1=1}^p \sum_{a_2=1}^p \dots$ を示す表記とすると， Γ_κ は Yanagihara et al. (2005) で定義される，

$$(2.9) \quad \Gamma_\kappa = \sum_{a,b,c,d}^p \kappa_{abcd} \left\{ e_a^{(p)} e_b^{(p)'} \otimes e_c^{(p)} e_d^{(p)'} \right\},$$

であるような真の分布の尖度に依存する $p^2 \times p^2$ 行列である．ここでの κ_{abcd} は ε_i の四次キュムラントであり， $\varepsilon = (\varepsilon_1, \dots, \varepsilon_p)'$ を ε_i と同一の分布に従う確率変数とすると，

$$(2.10) \quad \kappa_{abcd} = E_Y^*[\varepsilon_a \varepsilon_b \varepsilon_c \varepsilon_d] - \delta_{ab} \delta_{cd} - \delta_{ac} \delta_{bd} - \delta_{ad} \delta_{bc},$$

である．今 $\kappa_4^{(1)}$ を

$$(2.11) \quad \kappa_4^{(1)} = \sum_{a,b}^p \kappa_{aabb} = E_Y^*[(\varepsilon' \varepsilon)^2] - p(p+2),$$

のように定義される多変量尖度(Mardia (1970), Isogai (1983)等)とすると(2.9)と(2.11)は $\kappa_4^{(1)} = \text{tr}(\Gamma_\kappa)$ という関係式を満たす．このことから， AIC のバイアスは，真の分布の尖度に依存していること，更に候補のモデル M の下での共分散行列の期待値と真のモデルの共分散行列の構造の違いにも依存していることがわかる．ここで更に特別な形を考える．今， P_A を行列 A の列ベクトルで張る空間への射影行列 $P_A = A(A'A)^{-1}A'$ ，また $\Delta(\cdot)$ を Appendix A.1 の(A.1)式で定義される $\Sigma(\theta)$ の一次の偏微分に関する $p^2 \times q$ 行列であるとする．このとき， $\Sigma_* = \Sigma(\theta_0)$ であれば， $\Omega = I_p$ ， $D(\theta_0|\Sigma_*) = P_{\Delta(\theta_0)}$ となるのがわかる．このことにより， $\Sigma_* = \Sigma(\theta_0)$ であれば，式(2.8)での AIC のバイアスは以下ようになる．

$$B_{AIC} = \text{tr}(\Gamma_\kappa P_{\Delta(\theta_0)}) + O(n^{-1}).$$

もし ε_i が正規分布に従えば， $\kappa_{abcd} = 0$ である．よって，式(2.7)で与えられる AIC は，真のモデルと候補のモデルが一致する，すなわち， ε_i が正規分布でかつ $\Sigma(\theta_0) = \Sigma_*$ であれば定数バイアスを持たないことがわかる．しかしながら，真のモデルと候補のモデルが異なれば定数バイアスを持つ．さらに真の分布の尖度が大きくなれば AIC のバイアスも大きくなる傾向があることもわかる．

2.3 TIC 規準

竹内(1976)により提案された TIC は， AIC が持つようなモデルの不一致によるバイアスが修正された情報量規準である．この TIC では， $\sqrt{n}(\hat{\theta} - \theta_0)$ が漸近的に $(p+q)$ 変数正規分布に従うことを利用して， AIC のバイアス補正項である $2(p+q)$ を再評価している．今， $(p+q) \times (p+q)$ 行列 $J(\hat{\xi})$ と $I(\hat{\xi})$ を(2.3)式の p 変数正規分布の密度関数 $f(y_i|\xi)$ と(2.4)式対数尤度関数 $\ell(\xi|Y)$ を用いて以下のように定義する．

$$(2.12) \quad J(\hat{\xi}) = -\frac{1}{n} \left(\frac{\partial^2}{\partial \xi \partial \xi'} \ell(\xi|Y) \Bigg|_{\xi=\hat{\xi}} \right),$$

$$(2.13) \quad I(\hat{\xi}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \xi} \log\{f(y_i|\xi)\} \frac{\partial}{\partial \xi'} \log\{f(y_i|\xi)\} \Bigg|_{\xi=\hat{\xi}}.$$

TIC の一般公式(竹内(1976))から, 候補のモデル M (2.1)における TIC は,

$$(2.14) \quad TIC = -2\ell(\hat{\xi}|Y) + 2\text{tr}(I(\hat{\xi})J(\hat{\xi})^{-1}),$$

と定義される. 上記のトレース部分を更に計算すると,

$$(2.15) \quad TIC = np \log(2\pi) + n \log |\Sigma(\hat{\theta})| + (n+2)\text{tr}(S\Sigma(\hat{\theta})^{-1}) \\ + \frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}'_i \otimes \hat{\varepsilon}'_i) D(\hat{\theta}|S) (\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i) \\ - \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2})' D(\hat{\theta}|S) \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2}),$$

となる(導出に関する詳細は Appendix A.2 を参照). ただし $\hat{\varepsilon}_i$ は候補のモデル M の下での標準化された残差ベクトルの推定量,

$$(2.16) \quad \hat{\varepsilon}_i = \Sigma(\hat{\theta})^{-1/2} (y_i - \bar{y}),$$

である. 式(2.15)から, TIC は, AIC とは異なり, 多変量尖度の推定量に依存し真の分布と仮定した分布の違いを考慮に入れた情報量規準となっていることがわかる. 実際に ε_i が独立に同一な分布に従えば, TIC は AIC が持つような定数バイアスを修正しているということが知られている. つまり, B_{TIC} を TIC のバイアスとすると,

$$B_{TIC} = R_{KL} - E_Y^*[TIC] = O(n^{-1}),$$

となる.

ところが, 理論上ではバイアスが補正されていると雖も, 実際の状況において TIC のバイアス補正項がうまく機能しないことがある. そのような TIC の欠点は, 多変量線形モデルにおける変数選択の場合ではあるが, Yanagihara (2004a), Fujikoshi et al. (2005)にも述べられている. 一般的に, 多変量分布の尖度は次元 p が大きくなればなるほど大きくなる. しかしながら, Mardia (1970)に代表される従来の最小二乗残差に基づく推定量により大きい尖度の推定を行えば, 巨大な標本数が無ければ極端な過小推定をしてしまい, その結果非常に大きなバイアスが生じることが知られている(Yanagihara (2004b)). 以上のようなことが(2.15)式の TIC でも起き, 適度な標本数であっても非正規性に対する補正がうまく行われないことがある. また TIC は, 標本数 n に関してリスク R_{KL} (2.6)を $\sqrt{n}(\hat{\theta} - \theta_0)$ でテーラー展開し, その第一項目による近似を基にした推定量である. もし p が大きくなれば, 無視した第二項目以降が大きくなり, このテーラー展開による近似の精度は悪くなる. そのため, 次元が大きいときに TIC のバイアスが大きくなるという現象が起こる. 一般的に, 共分散構造モデルの次元は大きくならざるをえない. 以上のことから, 通常の場合であれば TIC と AIC は結果的にほとんど変わらない状況が往々に起こりうる.

2.4 クロス・バリデーションを用いた情報量規準

2.2 と 2.3 章からわかるように, AIC では真の分布によりバイアスが変動し, TIC では多変量尖度の推定がうまくいかなかったり次元の増大により $\hat{\theta}$ の正規近似が悪くなったりして, バイアスがうまく補正されないという状況が起こることがあり問題がある. そのため, 候補のモデルにおける分布と真の分布との相違に関する影響がより小さくなるような情報量規準が必要となる. 本論文では, Yanagihara (2004a)でのアイデアを用い, リスク R_{KL} (2.6)における未来の観測値ベクトルに関する部分だけをクロス・バリデーションによる予測残差に置き換える

ことで新しい情報量規準を構成する．クロス・バリデーションを用いた情報量規準の構成は多くの研究者により行われている．特に Stone(1974)が有名であり，その規準量は漸近的に AIC と同値であることが Stone(1977)により示されている．また，正規多変量線形モデルと正規一般化多変量分散分析(GMANOVA)モデルにおいて，Fujikoshi et al(2003)によりクロス・バリデーションを用いた情報量規準のバイアスを補正した新しい規準量も提案されている．

新しい情報量規準である修正 AIC (Corrected AIC : AIC_c) を以下のように定義する．まず， $\bar{y}_{[-i]}$ と $S_{[-i]}$ を第 i 番目の個体に関する観測値ベクトル y_i を取り除いたデータにより得た以下のような推定量とする．

$$(2.17) \quad \bar{y}_{[-i]} = \frac{1}{n-1} \sum_{j \neq i}^n y_j = \frac{n}{n-1} \bar{y} - \frac{1}{n-1} y_i,$$

$$(2.18) \quad S_{[-i]} = \frac{1}{n-1} \sum_{j \neq i}^n (y_j - \bar{y}_{[-i]})(y_j - \bar{y}_{[-i]})' \\ = \frac{n}{n-1} \left\{ S - \frac{1}{n-1} (y_i - \bar{y})(y_i - \bar{y})' \right\}.$$

上記の式変形については Fujikoshi et al(2003)等を参照してほしい．今， $\hat{\theta}_{[-i]}$ を第 i 番目の個体に関する観測値ベクトル y_i を取り除いたデータにより得た θ の最尤推定量とする．つまり，

$$(2.19) \quad \hat{\theta}_{[-i]} = \arg \min_{\theta} \{ \log |\Sigma(\theta)| + \text{tr}(S_{[-i]} \Sigma(\theta)^{-1}) \}.$$

このとき AIC_c を，

$$(2.20) \quad AIC_c = np \log(2\pi) + n \log |\Sigma(\hat{\theta})| \\ + \sum_{i=1}^n (y_i - \bar{y}_{[-i]})' \Sigma(\hat{\theta}_{[-i]})^{-1} (y_i - \bar{y}_{[-i]}) \\ = np \log(2\pi) + n \log |\Sigma(\hat{\theta})| \\ + \left(\frac{n}{n-1} \right)^2 \sum_{i=1}^n (y_i - \bar{y})' \Sigma(\hat{\theta}_{[-i]})^{-1} (y_i - \bar{y})$$

と定義する． AIC_c のバイアスは TIC と同じ

$$B_{AIC_c} = R_{KL} - E_Y^*[AIC_c] = O(n^{-1}),$$

である(証明は Appendix A.5 を参照)．しかし， TIC と異なって個別に尖度の推定をする必要がないため， AIC_c では尖度の推定量による過小推定の問題は発生しない．更に，テラー展開を使っていないため，次元の増大に対するバイアス項への影響は TIC に較べ小さくなる．

共分散構造の選択問題においては，Cudeck and Browne(1983)，Browne and Cudeck(1989)，De Gooijer(1995)等でクロス・バリデーションを用いたモデル選択の指標が提案されている．これらの指標では，データを二つ(もしくは複数)の calibration 標本と validation 標本に分けそれらの標本から指標値を計算している．しかしながら， AIC_c ではそのような手順を踏む必要はない．

式(2.20)からもわかるように， AIC_c は第 i 番目の個体に関する観測値ベクトル y_i を取り除いたデータにより得た θ の最尤推定量である $\hat{\theta}_{[-i]}$ (2.19) に依存している．そのため， AIC_c を得るためには n 回の繰り返しが必要となる．もともと $\hat{\theta}$ を得るためには，いくらかの繰り返し計算が必要であり，それだけで多くの計算時間を必要とする．そのため，標本数 n が大きい時には，何らかの手法を用いて計算時間の短縮をはかったほうが良い．標本数 n が十分大き

い時には, Fujisawa and Eguchi (2003) のような近似公式を使い, $\hat{\theta}_{[-i]}$ の計算時間を短縮することもできる. その近似公式は,

$$(2.21) \quad \hat{\theta}_{[-i]} \approx \hat{\theta} + \frac{1}{n-1} H(\hat{\theta}|S)^{-1} \Delta(\hat{\theta})' \text{vec}(\Sigma(\hat{\theta})^{-1/2} W_i \Sigma(\hat{\theta})^{-1/2}),$$

である(導出に関する詳細は Appendix A.4 を参照). ただし $H(\cdot|S)$ は Appendix A.1 での (A.5) 式で定義される尤度関数の二次偏微分に関する行列であり, W_i は,

$$(2.22) \quad W_i = S - \frac{n}{n-1} (y_i - \bar{y})(y_i - \bar{y})',$$

である. しかしながら, 近似公式 (2.21) において, 観測値の次元 p が標本数 n に比べ大きいときは, 第二項目以降の次元の影響が大きくなり, 近似の精度が悪くなることがある. またこれは TIC でも同じことが言えるのだが, $H(\hat{\theta}|S)$ が正則でない場合に近似公式 (2.21) は使えないものとなる. 更に, 正則の場合でも $H(\hat{\theta}|S)^{-1}$ が不安定な値になることもある. 以上のことから, この近似公式 (2.21) を使用するときは十分に注意しなければならない.

3. 数値実験

第 2 章では, 共分散構造モデルの選択問題に関する従来の情報量規準のバイアスに関する問題点を考え, それらを改善するような新しい情報量規準を提案した. 本章では, 新しく提案した AIC_c のバイアスが実際に AIC や TIC のバイアスと比べ小さくなっていることをシミュレーション実験により確かめる.

3.1 実験に用いたモデル

本論文の数値実験では, 真のモデル M_* (2.2) の共分散行列として Hu et al. (1992) で使用している以下のような 15 次元の因子分析モデルを使用した.

$$(3.1) \quad \Sigma_* = \Lambda_* \Lambda_*' + \Psi_*$$

ただし, 因子負荷量行列 Λ_* と 独自因子分散行列 Ψ_* は,

$$\Lambda_* = (I_3 \otimes l_*) Q_*^{1/2}, \quad \Psi_* = \text{diag}(1_3 \otimes w_*),$$

である. ここでの 1_k は全ての成分が 1 である $k \times 1$ ベクトルであり, l_* , Q_* , w_* はそれぞれ

$$l_* = \begin{pmatrix} 0.70 \\ 0.70 \\ 0.75 \\ 0.80 \\ 0.80 \end{pmatrix}, \quad Q_* = \begin{pmatrix} 1.0 & 0.3 & 0.4 \\ 0.3 & 1.0 & 0.5 \\ 0.4 & 0.5 & 1.0 \end{pmatrix}, \quad w_* = \begin{pmatrix} 0.5100 \\ 0.5100 \\ 0.4375 \\ 0.3600 \\ 0.3600 \end{pmatrix},$$

である. 非正規誤差は, Yuan and Bentler (1997) で提案されている, 以下のようなモデルにより発生させた.

データモデル x_1, \dots, x_k ($k \geq p$) はそれぞれ独立に同一な平均 0 分散 1 を持つ分布に従う確率変数とし, $x = (x_1, \dots, x_k)'$ とする. また r は, x と独立で, 分散 1 を持つ分布に従う確率変数とする. このとき, データを

$$(3.2) \quad \varepsilon = r A' x,$$

で発生させる．ただし $A = (a_1, \dots, a_k)'$ は $\text{rank}(A) = p$ である $k \times p$ 行列で $A'A = I_p$ を満たすものとする．このとき，式(2.11)で定義される ε の多変量尖度は，

$$\kappa_4^{(1)} = \beta(\eta - 3) \sum_{i=1}^k (a_i' a_i)^2 + (\beta - 1)p(p + 2),$$

となる．ただし $E[x_i^4] = \eta$ であり $E[r^4] = \beta$ である．

χ_f^2 を自由度 f の χ^2 分布に従う確率変数とし， 16×15 行列 B を

$$B = \begin{pmatrix} I_{15} \\ \mathbf{1}'_{15} \end{pmatrix} (I_{15} + \mathbf{1}_{15} \mathbf{1}'_{15})^{-1/2},$$

とする．本論文では(3.2)式のデータモデルに対して，以下のような x, A, r により誤差分布を発生させた．この設定は Yanagihara (2004b)での数値実験と同じものである．

誤差分布 1 (正規分布) $x_i \sim \text{i.i.d. } N(0, 1)$, $r = 1$, $A = I_{15}$ ($\kappa_4^{(1)} = 0$).

誤差分布 2 (t -分布) $x_i \sim \text{i.i.d. } N(0, 1)$, $r = \sqrt{6/\chi_8^2}$, $A = I_{15}$ ($\kappa_4^{(1)} = 127.5$).

誤差分布 3 (一様分布) $x_i = w_i/(5/\sqrt{3})$, $w_i \sim \text{i.i.d. } U(-5, 5)$, $r = 1$, $A = B$ ($\kappa_4^{(1)} = -16.875$).

誤差分布 4 (χ^2 分布) $x_i = (w_i - 4)/2\sqrt{2}$, $w_i = \chi_4^2$, $r = \sqrt{6/\chi_8^2}$, $A = B$ ($\kappa_4^{(1)} = 190.781$).

誤差分布 5 (対数正規分布) $x_i = (w_i - e^{1/8})/(e^{1/2}\sqrt{e^{1/4} - 1})$, $\log w_i \sim \text{i.i.d. } N(0, 1/4)$, $r = \sqrt{6/\chi_8^2}$, $A = B$ ($\kappa_4^{(1)} = 251.920$).

3.2 実験結果

3.1 章での式(3.1)と(3.2)を用いて発生させた擬似データ $y_i = \mathbf{1}_{15} + \Sigma_*^{1/2} \varepsilon_i$ ($i = 1, \dots, 200$) に対して，因子分析モデル，

$$\Sigma(\theta) = \Lambda \Lambda' + \Psi, \quad (\Lambda = (\lambda_1, \dots, \lambda_m), \Psi = \text{diag}(\psi_1, \dots, \psi_p)),$$

を当てはめ，その因子数 m を AIC, TIC, AIC_c を用いて決定する実験を 500 回繰り返し行った．このモデルでは $m = 3$ のときが真のモデルである． Λ と Ψ の実際の推定アルゴリズムは，柳井 他(1990), pp. 49–92等を参考にさせていただきたい．因子数 m の探索範囲は $1 \leq m \leq 6$ とした．図 1 には，それぞれの誤差モデルで得たリスクと AIC, TIC, AIC_c の平均を示している．図中の横軸の目盛りは候補のモデルでの因子数に対応している．更に，リスクは実線で， AIC の平均は○で， TIC の平均は□で， AIC_c の平均は△で表している．また，表 1 はそれぞれの誤差モデルにおけるリスクとバイアス，モデルの選択確率(度数(%))を示している．

図と表から， AIC のバイアスは真の分布の尖度が大きくなれば大きくなっていることがわかる．またそれに伴い，尖度が大きくなればリスクが最小になるモデル(ここでは真のモデル($m = 3$))を選択する確率が低くなっている． TIC は AIC に較べバイアスが小さくなっていることが見てとれるが，その大きさはまだまだ大きく，十分にバイアスを補正しているとは言い難い．このような現象は，誤差分布 2, 4, 5 のように尖度が大きい分布において顕著に現れている．その原因は尖度の過小推定にあると考えられる．また， TIC のバイアスは AIC のバイアスよりも小さくなっているとしても，モデル選択確率は逆に AIC よりも悪くなっている．一方， AIC_c のバイアスはほとんどの場合に最も小さくなっている．それは特に誤差分布 5 で顕著な結果として現れている．更に，モデルの選択確率は，尖度が小さい分布である誤差分布 1 と 3 でこそ AIC よりも低い確率になっているがその差は小さいものであり，尖度が大きい分布では AIC よりもはるかに高い確率となっている．また，すべての分布において真のモデル

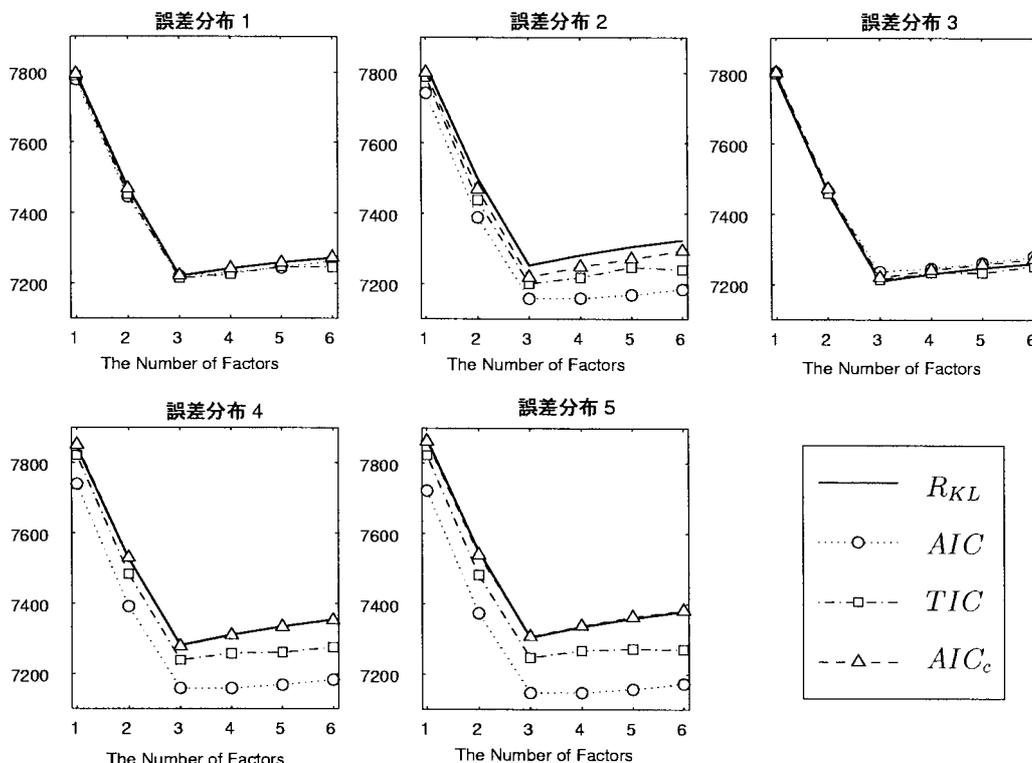


図 1. リスク R_{KL} と AIC , TIC , AIC_c の平均.

を選ぶ確率が 90% を超えているため、 AIC_c を用いたモデル選択法では、 AIC や TIC を用いた手法と異なり選択確率における分布間の違いはあまり無い。

以上のことから、 AIC を使った選択法は、真の分布が正規分布、または真の分布の尖度が小さい場合にはほとんど問題ないが、真の分布の尖度が大きくなればそのバイアスも大きくなり、リスクが最小になるモデルを選ぶ確率が下がる。一方、 AIC_c はすべて誤差分布においてバイアスを最も良く補正しており、それを用いた選択法は、リスクが最小になるモデルを選ぶ確率もどの分布においてもあまり変わらない。そのため、 AIC_c を用いた選択手法は、候補のモデルと真のモデルの分布間の相違に関する影響を最も受けにくい手法であると言える。

4. 国民性データへの適用

本章では、国民性調査第 11 回全国調査 M 型調査票のデータ(坂元 他(2004))に「日本人の満足感(前田(1995))に関する因子分析モデルを当てはめ、2.4 章で提案した新しい情報量規準 AIC_c を用いて潜在因子の最適構造の探索を行う。

4.1 データと候補のモデル

本論文の解析では、前田(1995)と同様の表 2 で示した 14 項目を変数として使用した。具体的な質問項目については、坂元 他(2004)を参照して欲しい。それぞれの項目において「その他」と「D.K.」回答が一つでもある個体を取り除いた 835 名のデータを用いて解析を行った。

表 1. 数値実験の結果(リスク, バイアスの平均, モデル選択確率).

誤差 分布	情報量 規準	— 潜在因子の数 —					
		1	2	3	4	5	6
1	リスク	7798.99	7471.85	7219.99	7241.33	7258.78	7272.20
	AIC バイアス (度数 (%))	21.00 (0.0)	26.79 (0.0)	0.10 (94.8)	12.34 (5.2)	14.99 (0.0)	9.68 (0.0)
	TIC バイアス (度数 (%))	9.58 (0.0)	18.62 (0.0)	6.47 (75.0)	16.08 (12.2)	12.32 (6.4)	26.47 (6.4)
	AIC_c バイアス (度数 (%))	5.18 (0.0)	3.46 (0.0)	-1.35 (90.6)	-0.38 (7.6)	0.99 (1.8)	-0.90 (0.0)
2	リスク	7821.15	7500.16	7252.40	7281.58	7305.45	7323.99
	AIC バイアス (度数 (%))	79.40 (0.0)	111.58 (0.0)	94.43 (63.0)	122.59 (33.0)	136.69 (4.0)	139.72 (0.0)
	TIC バイアス (度数 (%))	34.24 (0.0)	62.48 (0.0)	51.78 (63.6)	63.89 (18.0)	57.57 (9.0)	83.98 (9.4)
	AIC_c バイアス (度数 (%))	21.43 (0.0)	31.54 (0.0)	33.23 (93.2)	32.69 (6.0)	34.60 (0.8)	30.03 (0.0)
3	リスク	7789.96	7462.24	7209.69	7230.35	7247.14	7260.55
	AIC バイアス (度数 (%))	-11.92 (0.0)	-3.02 (0.0)	-26.55 (96.4)	-15.88 (3.6)	-14.16 (0.0)	-19.65 (0.0)
	TIC バイアス (度数 (%))	-9.57 (0.0)	3.74 (0.0)	-5.01 (76.4)	-3.57 (11.0)	13.91 (7.0)	8.11 (5.6)
	AIC_c バイアス (度数 (%))	-11.60 (0.0)	-9.47 (0.0)	-9.96 (91.4)	-11.02 (7.0)	-10.55 (1.6)	-12.08 (0.0)
4	リスク	7844.40	7527.22	7280.21	7309.99	7334.04	7353.85
	AIC バイアス (度数 (%))	105.03 (0.0)	136.02 (0.0)	122.23 (59.0)	152.11 (36.8)	166.43 (4.2)	171.24 (0.0)
	TIC バイアス (度数 (%))	23.03 (0.0)	43.49 (0.0)	42.07 (59.0)	52.73 (15.8)	73.59 (13.6)	78.68 (11.6)
	AIC_c バイアス (度数 (%))	-6.66 (0.0)	-1.57 (0.0)	2.66 (92.0)	1.13 (7.4)	0.88 (0.4)	2.05 (0.2)
5	リスク	7868.89	7550.01	7304.70	7334.82	7358.81	7378.63
	AIC バイアス (度数 (%))	146.84 (0.0)	176.37 (0.0)	157.98 (60.0)	187.52 (36.8)	201.49 (3.0)	206.11 (0.2)
	TIC バイアス (度数 (%))	46.01 (0.0)	67.82 (0.0)	58.02 (55.4)	67.46 (16.0)	87.16 (15.6)	108.45 (13.0)
	AIC_c バイアス (度数 (%))	5.89 (0.0)	10.09 (0.0)	-2.11 (94.4)	-2.42 (5.2)	-4.13 (0.4)	-2.51 (0.0)

また, “健康状態 (#2.80)と記したものは, 前田(1995)と同様に, 五つの下位項目, すなわち, “頭痛・偏頭痛 (#2.80a), “背中の痛み (#2.80b), “いらいら (#2.80c), “うつ状態” (#2.80d), “不眠症 (#2.80e), に対してここ一ヶ月の間に ‘かかったことがない’ と回答した場合に 1 ポイントを与えて, その合計点を新たな変数と定義したものである.

分析の目的は, 上記のデータに対して因子分析モデル,

$$(4.1) \quad \Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi,$$

をあてはめ, 「日本人の満足感」に関する潜在因子の最適構造を探索することである. ただし, Λ は成分に部分的に 0 を持つ $14 \times m$ 因子負荷量行列, Φ は対角成分が 1 であるような $m \times m$

表 2. 分析に使用した項目と関連する構成概念の内容.

構成概念の内容	No.	項目番号	項目内容
健康状態	1	#2.80	5 項目の病的症状の有無に関する解答
生活水準	2	#1.8	帰属階級
	3	#7.29	くらしむき
	4	#7.30a	生活水準 10 年の変化
日本の物質的豊かさへの評価	5	#9.12	日本の「科学技術の水準」
	6	#9.12c	日本の「経済力」
	7	#9.12d	日本の「生活水準」
個人生活への満足	8	# 2.3c	家庭に満足か
	9	# 2.3i	仕事や職場に満足か
	10	#2.3j	余暇(レジャー)に満足か
	11	#2.3k	健康状態に満足か
	12	#2.3l	生活全体に満足か
社会への満足	13	#2.3d	社会に満足か
	14	#7.40	社会は公平か

因子間相関行列, また, Ψ は $\Psi = \text{diag}(\psi_1, \dots, \psi_{14})'$ の 14×14 独自因子分散行列である. 前田(1995)では, 「日本人の満足感」というものを「個人生活への満足感」と「社会への満足感」という二つの構成概念によって検討している. 解析で取り扱ったデータでは, 「個人生活への満足感」が潜在因子になっていると考えられる変数は 1, 8, 9, 10, 11, 12, であり, 「社会への満足感」が潜在因子となっていると考えられる変数は 2, 3, 4, 5, 6, 7, 13, 14 である. 本論文では, 「日本人の満足感」は「個人の満足感」と「社会への満足感」の二つの構成概念によって表現されていると捉え, 表 2 における変数の構成概念から「個人の満足感」と「社会への満足感」を複数に分割した潜在因子によって「日本人の満足感」を説明することを試みる. 実際のあてはめに用いたモデルは以下の六つのモデルである.

モデル 1: 潜在因子は, 「個人の満足感」と「社会への満足感」($m = 2$).

モデル 2: 潜在因子は, 「個人の満足感」と「帰属意識」と「日本の物質的豊かさへの評価」($m = 3$).

モデル 3: 潜在因子は, 「個人の満足感」と「豊かさへの評価」と「社会への満足」($m = 3$).

モデル 4: 潜在因子は, 「個人の満足感」と「生活水準」と「社会への満足」($m = 3$).

モデル 5: 潜在因子は, 「個人の満足感」と「生活水準」と「日本の物質的豊かさへの評価」と「社会への満足」($m = 4$).

モデル 6: 潜在因子は, 「健康状態」と「生活水準」と「日本の物質的豊かさへの評価」と「個人生活への満足」と「社会への満足」($m = 5$).

それぞれのモデルにおけるパス図を図 2 に, 因子負荷量行列 Λ (4.1) のパターンを Appendix A.6 に示している.

4.2 解析結果

表 3 は, 候補のモデル 1~6 を当てはめたときのそれぞれの情報量規準 AIC , TIC , AIC_c とモデルの適合度 GFI の結果である. また, 図 3 はそれぞれのモデルにおける情報量規準の変動を示している. 図 3 において, AIC は \circ , TIC は \square , AIC_c は \triangle で表されている. また, 図中の横軸の目盛りは使用したモデルの番号に対応している. 表 3 と図 3 から, 候補のモデル

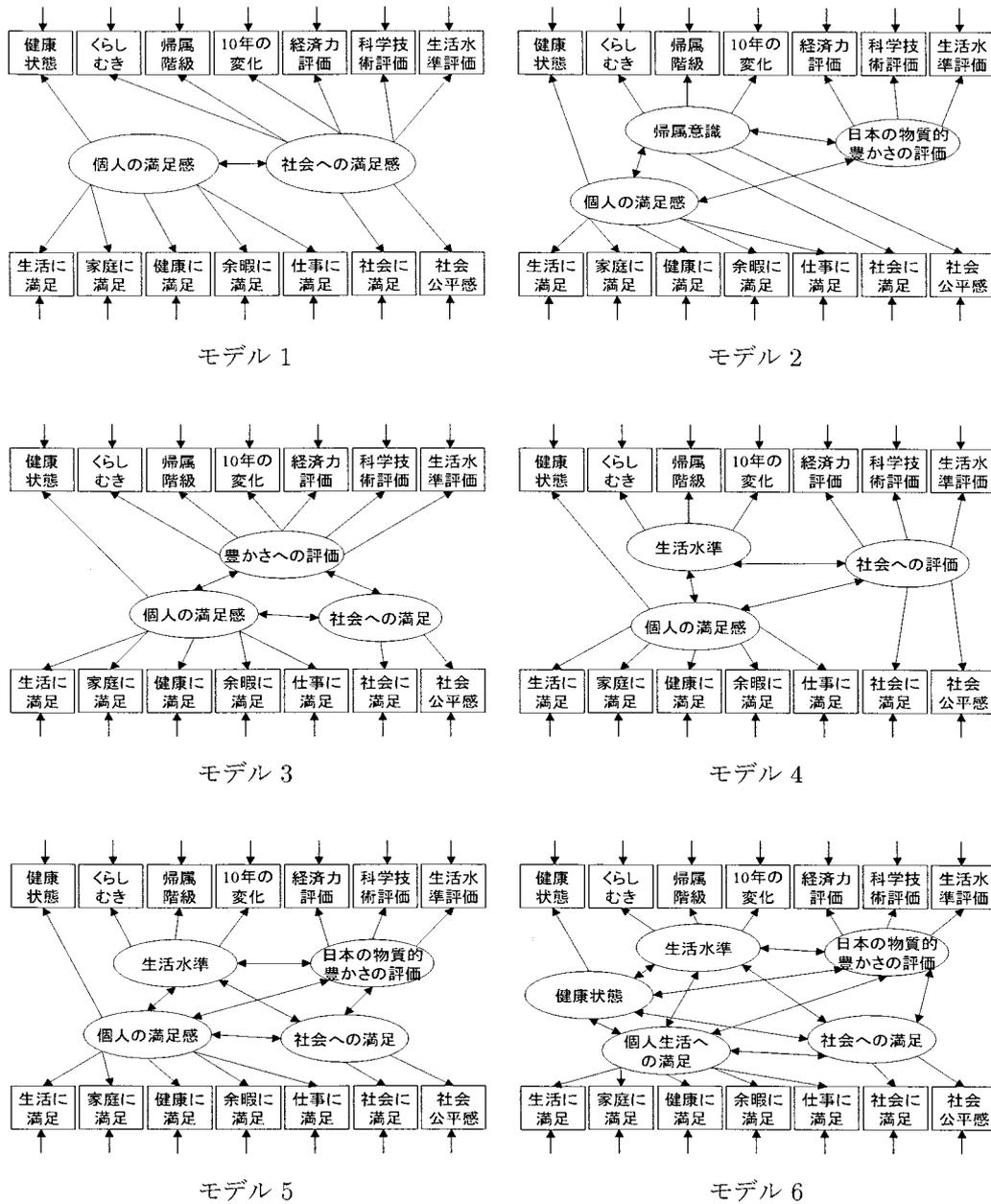


図 2. それぞれのモデルでのパス図.

表 3. それぞれの候補のモデルでの情報量規準の比較.

モデル	自由母数数	AIC	TIC	AIC_c	GFI
1	29	25328.58	25338.40	25339.02	0.917
2	31	25148.91	25159.20	25159.92	0.946
3	31	25289.05	25302.71	25303.40	0.922
4	31	25157.81	25168.90	25169.65	0.946
5	34	25059.54	25070.80	25071.64	0.964
6	37	25078.96	25090.51	25091.47	0.962

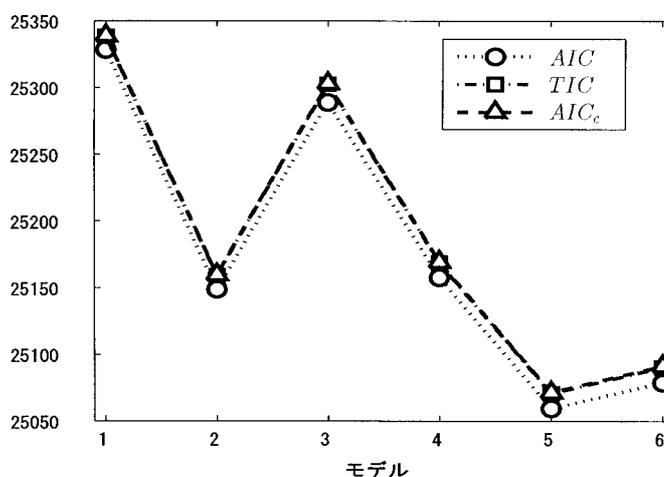


図 3. それぞれのモデルでの情報量規準.

の中で最適なモデルは, すべてのモデルの中で最も GFI が大きいモデル 5 であることがわかる.

表 3 と図 3 から, それぞれのモデルにおいて, 情報量規準 AIC , TIC , AIC_c の差があまり無いことに気づく. この原因には, データの尖度があまり大きくないということが関係していると予想される. 使用したデータの変数別のヒストグラムを図 4 に示す. この図から, 変数 1 は大きく正規分布から逸脱しているが, それ以外の変数ではそれほど正規分布から逸脱していないことが見てとれる. 実際に真の分布が正規分布のときに不偏推定量になるような多変量尖度の推定量 (Mardia (1970), Browne (1984) 等参照),

$$(4.2) \quad \hat{\kappa}_4^{(1)} = \frac{(n+1)}{n(n-1)} \sum_{i=1}^n (\tilde{\varepsilon}_i' \tilde{\varepsilon}_i)^2 - p(p+2),$$

を計算してみると, $\hat{\kappa}_4^{(1)} = 26.67$ となる. ただし $\tilde{\varepsilon}_i = S^{-1/2}(y_i - \bar{y})$ である. この数値からはデータが正規分布から大きく逸脱しているとは言えない. しかしながら, 上記の推定量のような従来の最小二乗残差に基づく推定量では, データの非正規性に対してそれほど敏感に反応しない (Yanagihara (2004b)) ため, データの非正規性を過小に評価していることが考えられる. そのため, Yanagihara (2004b) で提案されている (4.2) よりも非正規性に対して敏感に反応する多変量尖度の推定量でもデータの非正規性を評価してみる. Yanagihara (2004b) ではチューニングパラメータ α によって残差ベクトルの長さを調節した多変量尖度の推定量 $\tilde{\kappa}_4^{(1)}(\alpha)$ を提案している. この推定量は $\alpha = 0$ のとき式 (4.2) での推定量と一致し, $\alpha = 0.5, 1$ のときはそれ

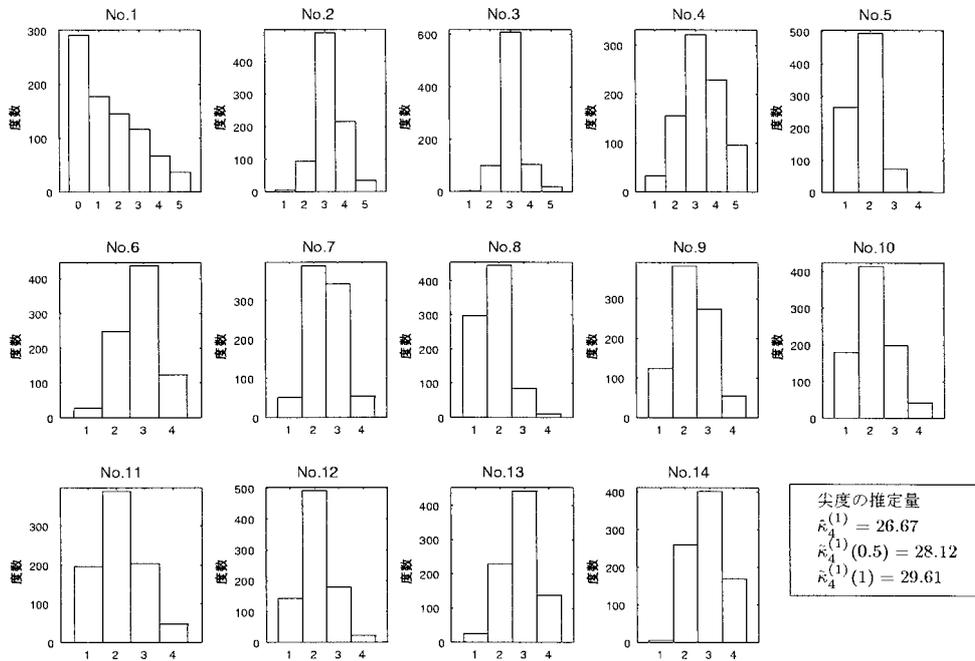


図 4. それぞれの変数のヒストグラム .

それ,

$$(4.3) \quad \hat{\kappa}_4^{(1)}(0.5) = \frac{(n-p-3)}{n(n-1)} \sum_{i=1}^n \frac{(\tilde{\varepsilon}'_i \tilde{\varepsilon}_i)^2}{1 - \tilde{\varepsilon}'_i \tilde{\varepsilon}_i / (n-1)} - p(p+2),$$

$$(4.4) \quad \hat{\kappa}_4^{(1)}(1) = \frac{(n-p-3)(n-p-5)}{n(n-1)^2} \sum_{i=1}^n \left\{ \frac{\tilde{\varepsilon}'_i \tilde{\varepsilon}_i}{1 - \tilde{\varepsilon}'_i \tilde{\varepsilon}_i / (n-1)} \right\}^2 - p(p+2),$$

となる . それぞれ計算すると, $\hat{\kappa}_4^{(1)}(0.5) = 28.12$, $\hat{\kappa}_4^{(1)}(1) = 29.61$ となる (4.2)(4.3)(4.4)において, $p(p+2) = 224$ であることに注意すると, データの尖度は大きくなく, ほとんど正規分布とみなせることがわかる(厳密性を求めるならここで仮説検定を行うべきであろうが, 本論文ではそれを省略する). データが正規分布からそれほど逸脱していないので, それぞれの情報量規準の結果がさほど変わらないものになったと推測できる .

最後に, 情報量規準により最適なモデルであると選択されたモデル₅での解析結果を図5に示す . この結果から, 本論文で取り扱ったデータでは, 「日本人の満足感」は, 「個人の満足感」と「社会への満足感」を三つに分割した「生活水準」, 「日本の物質的豊かさへの評価」, 「社会への満足」という因子で説明されていることになる . また, それぞれの潜在変数間の相関はそれほど高くないという結果も得た . このモデルでの GFI は 0.964 であり, 現在求められている, “ GFI が 0.95 付近 ” という規準(狩野・三浦(2002), p. 141)を満たしている . しかしながら, それぞれの因子負荷には 0.26 等の値が低いものも存在し, さらに説明力が高いモデルを探索する必要があることを指摘したい .

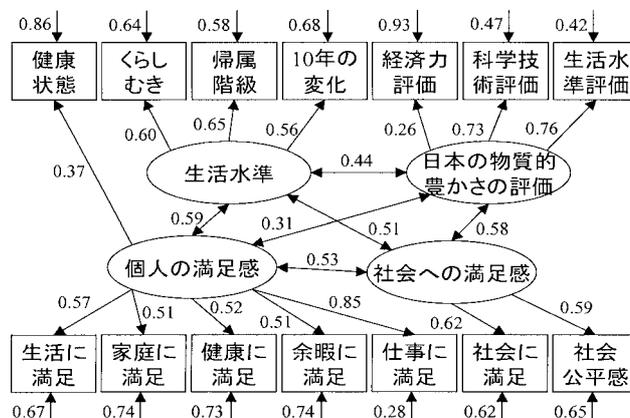


図5. 最適なモデルにおける解析結果.

5. おわりに

従来の統計手法の多くは正規性の仮定の下で提案されたものである。統計解析のソフトウェアでは正規性の下で提案された手法が一般的であることや真の分布は未知であることを考えると、解析の初期段階においてこのような手法を用いることはある意味妥当ではある。しかしながら、国民性全国調査によるデータはカテゴリカルデータであり、明らかに正規分布に従っていない。そのため、従来の正規性の仮定に基づく手法をそのままデータに適用したときに、仮定した正規分布とデータが従う真の分布の相違が解析結果にどのような影響を与えるかを調べ、もし必要であればその非正規性の影響を小さくするような修正を施す必要がある。本論文では、分布に正規性を仮定したときの共分散構造モデルの選択問題に関する AIC のバイアスを漸近展開式から評価し、非正規性の影響が小さくなるような新しいバイアス補正情報量規準である AIC_c を提案した。実際のシミュレーション実験により、この AIC_c のバイアスは AIC や TIC のバイアスよりも小さくなり、更に AIC_c を用いたモデル選択ではリスクを最小にするモデルを選ぶ確率がどのような分布に関してもほぼ同じくらいになることが確かめられた。

国民性調査データを使った解析例においては、正規分布からさほど外れていないデータであったため、 AIC , TIC , AIC_c のどの情報量規準もほぼ同じ値になり、すべての情報量規準が同じモデルを最適なモデルと選んだ。この結果は、本論文で提案した情報量規準をモデルの最適化に用いた場合、データが正規分布に“近い”場合でも、分布に正規性を仮定して導出された AIC を用いた結果と同じような結果となることを示している。これは非常に重要な結果である。もしデータが正規分布に“近い”場合、非正規性の下でバイアスを補正した情報量規準が正規性の仮定の下での情報量規準とまったく違う結果を与えてしまうのであれば、バイアスを補正した情報量規準を安心して使用することができない。その点を考えれば、4章での解析結果は非常に重要な結果である。

しかしながら、前田(1995)で述べているように、4章で選ばれたモデルは候補のモデルの中で最適なモデルであり、本当の意味での真のモデルを示しているのではない。また、4章の最後にも述べたが、因子負荷の問題も残っており、さらに説明力が高いモデルを探索する必要がある。また、カテゴリカルデータに正規分布の仮定をしても良いか? という問題もある。カテゴリー数が大きければさほど問題はないが、厳密な意味で言えば、因子分析モデルもカテゴリカ

ルデータにおける因子分析モデル(柳井 他(2000), pp. 129–145)等)を用いる必要がある。カテゴリカルデータでは、観測値ベクトル y_i に $E[y_i] = \mu$, $\text{Cov}[y_i] = \Sigma(\theta)$ である離散分布を仮定する方法, y_i は平均 μ , 共分散行列 $\Sigma(\theta)$ を持つ連続分布からの潜在変数をカテゴリ化した変数であると仮定する方法等が考えられる。前者の場合は問題無いが、後者の場合 $E[y_i] \neq \mu$, $\text{Cov}[y_i] \neq \Sigma(\theta)$ であるため、本論文で用いた推定法では推定結果にバイアスが生じ、 AIC_c でも正確にバイアスを補正できない。このようなモデルで推定を行った結果と、正規性の仮定の下で推定を行った結果との違いも調べる必要がある。また、本論文での数値実験では、モデル選択法での非正規性の影響を調べるということを中心に行ったが、その他にも標本数、因子負荷量の大きさ等も影響を与えていると考えられる(Bandalos (1993))。更に、仮定した分布と真の分布が異なった場合に情報量規準のバイアスを補正するその他の手法として、ブートストラップ法を用いた方法があり、Ishiguro et al. (1997)ではブートストラップ法でバイアスを補正した情報量規準である EIC を提案している。本論文では EIC との比較を行っていないが、多変量モデルに関していえば、標本数が次元に対して十分大きくなければブートストラップ法は良い近似を与えないということが Wakaki et al. (2002)や Fujikoshi et al. (2005)等で報告されている。これらが検証できる数値実験も行う必要がある。以上のような点がこれからの課題になる。

謝 辞

因子分析に関して有益なコメントを頂いた 筑波大学 上市秀雄先生、ハワイ大学 林健太郎先生、本論文の内容に対して適切なコメントを頂いた 統計数理研究所 坂元慶行先生に厚く御礼申し上げます。またこの特集号の編集委員、さらに二人の査読者の方々には大変丁寧に論文を読んで戴き、改訂のための助言を戴きました。ここに深く感謝の意を表します。

Appendix

A.1 $\Sigma(\theta)$ の偏微分に関する表記

共分散行列 $\Sigma(\theta)$ に関する標準化された j 次偏微分行列を

$$\dot{\Sigma}_{a_1 \dots a_j}(\theta_*) = \Sigma(\theta)^{-1/2} \left\{ \frac{\partial^j}{\partial \theta_{a_1} \dots \partial \theta_{a_j}} \Sigma(\theta) \right\} \Sigma(\theta)^{-1/2} \Big|_{\theta=\theta_*},$$

とする。この行列を用い、以下のような $p^2 \times q$ の一次偏微分行列を定義する。

$$(A.1) \quad \begin{aligned} \Delta(\theta_*) &= (\text{vec}(\dot{\Sigma}_1(\theta_*)), \dots, \text{vec}(\dot{\Sigma}_q(\theta_*))) \\ &= \left\{ \Sigma(\theta)^{-1/2} \otimes \Sigma(\theta)^{-1/2} \right\} \frac{\partial}{\partial \theta'} \text{vec}(\Sigma(\theta)) \Big|_{\theta=\theta_*}. \end{aligned}$$

また θ に関する対数尤度関数を

$$(A.2) \quad l_n(\theta|S) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma(\theta)| - \frac{n}{2} \text{tr}(S\Sigma(\theta)^{-1}),$$

とし、この対数尤度に対し、一次と二次の偏微分に基づく $q \times 1$ ベクトル、 $q \times q$ 行列を、

$$g(\theta_*|S) = -\frac{2}{n} \left(\frac{\partial}{\partial \theta} l_n(\theta|S) \Big|_{\theta=\theta_*} \right), \quad H(\theta_*|S) = -\frac{2}{n} \left(\frac{\partial^2}{\partial \theta \partial \theta'} l_n(\theta|S) \Big|_{\theta=\theta_*} \right),$$

と定義する。今、候補のモデル $M(2.1)$ の下での共分散行列で標準化した残差行列を

$$(A.3) \quad R(\theta|S) = \Sigma(\theta)^{-1/2} \{ \Sigma(\theta) - S \} \Sigma(\theta)^{-1/2},$$

とくと、共分散構造モデルにおける $g(\theta|S)$ と $H(\theta|S)$ は以下ようになる。

$$(A.4) \quad g(\theta|S) = \Delta(\theta)' \text{vec}(R(\theta|S)),$$

$$(A.5) \quad H(\theta|S) = \Delta(\theta)' \Delta(\theta) - 2\Delta(\theta)' \{I_p \otimes R(\theta|S)\} \Delta(\theta) \\ + \sum_{a,b}^q \text{tr} \left(R(\theta|S) \dot{\Sigma}_{ab}(\theta) \right) e_a^{(q)} e_b^{(q)'}$$

ただし、 $e_j^{(q)}$ は第 j 番目の成分が 1 で残りの成分が 0 となる $q \times 1$ ベクトルである。

A.2 TIC の導出

竹内(1976)によって提案された TIC の一般公式は、式(2.14)で定義される。共分散構造モデルにおける $J(\hat{\xi})$ (2.12) と $I(\hat{\xi})$ (2.13) は、

$$J(\hat{\xi}) = \begin{pmatrix} \Sigma(\hat{\theta})^{-1} & O \\ O & \frac{1}{2}H(\hat{\theta}|S) \end{pmatrix}, \quad I(\hat{\xi}) = \begin{pmatrix} \hat{I}_{11} & \hat{I}_{12} \\ \hat{I}_{12}' & \hat{I}_{22} \end{pmatrix},$$

である。ただし O は全ての成分が 0 であるような行列であり、

$$\hat{I}_{11} = \Sigma(\hat{\theta})^{-1} S \Sigma(\hat{\theta})^{-1}, \\ \hat{I}_{12} = \frac{1}{2} \Sigma(\hat{\theta})^{-1/2} \left\{ \frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}_i' \otimes \hat{\varepsilon}_i \hat{\varepsilon}_i') \right\} \Delta(\hat{\theta}), \\ \hat{I}_{22} = \frac{1}{4} \Delta(\hat{\theta})' \left\{ \text{vec}(R(\hat{\theta}|S)) \text{vec}(R(\theta|S))' + \frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}_i \hat{\varepsilon}_i' \otimes \hat{\varepsilon}_i \hat{\varepsilon}_i') \right. \\ \left. - \text{vec}(I_p - R(\hat{\theta}|S)) \text{vec}(I_p - R(\hat{\theta}|S))' \right\} \Delta(\hat{\theta}),$$

である。ここでの $\hat{\varepsilon}_i$ は式(2.16)で定義される候補のモデル $M(2.1)$ の下での残差ベクトルの推定量であり、 $R(\cdot)$ は式(A.3)で定義した候補のモデル $M(2.1)$ の下での共分散行列で標準化した残差行列とする。今、 $\hat{\theta}$ は最尤推定量であるので、 $g(\hat{\theta}|S) = \Delta(\hat{\theta})' \text{vec}(R(\hat{\theta}|S)) = 0$ を満たすことを利用すると \hat{I}_{22} は、

$$\hat{I}_{22} = \frac{1}{4} \Delta(\hat{\theta})' \left\{ \frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}_i \hat{\varepsilon}_i' \otimes \hat{\varepsilon}_i \hat{\varepsilon}_i') \right. \\ \left. - \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2}) \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2})' \right\} \Delta(\hat{\theta}),$$

となる。ここで、

$$(A.6) \quad D(\theta|S) = \Delta(\theta) H(\theta|S)^{-1} \Delta(\theta)',$$

という $p^2 \times p^2$ 行列関数を定義すると、求める式(2.14)中のトレースは、

$$2\text{tr}(I(\hat{\xi})J(\hat{\xi})^{-1}) \\ = 2\text{tr}(S\Sigma(\hat{\theta})^{-1}) + \frac{1}{n} \sum_{i=1}^n (\hat{\varepsilon}_i' \otimes \hat{\varepsilon}_i') D(\hat{\theta}|S) (\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i) \\ - \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2})' D(\hat{\theta}|S) \text{vec}(\Sigma(\hat{\theta})^{-1/2} S \Sigma(\hat{\theta})^{-1/2}),$$

となる。よって、式(2.15)を得ることができる。

A.3 AIC のバイアス

観測値ベクトル y_i が独立同一分布に従う場合, TIC のリスク R_{KL} (2.6) に対するバイアスは $R_{KL} - E_Y^*[TIC] = O(n^{-1})$ となる(竹内(1976)). この関係式を使うと AIC のバイアスは,

$$\begin{aligned}
 (A.7) \quad B_{AIC} &= R_{KL} - E_Y^*[AIC] \\
 &= E_Y^*[TIC] - E_Y^*[AIC] + O(n^{-1}) \\
 &= 2E_Y^*[\text{tr}(\mathbf{I}(\hat{\xi})\mathbf{J}(\hat{\xi})^{-1})] - 2(p+q) + O(n^{-1}) \\
 &= B_1 - B_2 + B_3 - 2(p+q) + O(n^{-1}),
 \end{aligned}$$

となる. ただし,

$$\begin{aligned}
 B_1 &= 2E_Y^* \left[\text{tr}(\mathbf{S}\Sigma(\hat{\theta})^{-1}) \right], \\
 B_2 &= E_Y^* \left[\text{vec}(\Sigma(\hat{\theta})^{-1/2}\mathbf{S}\Sigma(\hat{\theta})^{-1/2})' \mathbf{D}(\hat{\theta}|\mathbf{S}) \text{vec}(\Sigma(\hat{\theta})^{-1/2}\mathbf{S}\Sigma(\hat{\theta})^{-1/2}) \right], \\
 B_3 &= \frac{1}{n} \sum_{i=1}^n E_Y^* \left[(\hat{\varepsilon}'_i \otimes \hat{\varepsilon}'_i) \mathbf{D}(\hat{\theta}|\mathbf{S})(\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i) \right],
 \end{aligned}$$

である. 今, $E_Y^*[\hat{\theta}] \rightarrow \theta_0$, $E_Y^*[\mathbf{S}] \rightarrow \Sigma_*$ ($n \rightarrow \infty$) であることを利用すると,

$$(A.8) \quad E_Y^* \left[\text{tr}(\mathbf{S}\Sigma(\hat{\theta})^{-1}) \right] = \text{tr}(\mathbf{\Omega}'\mathbf{\Omega}) + O(n^{-1}),$$

$$\begin{aligned}
 (A.9) \quad E_Y^* \left[\text{vec}(\Sigma(\hat{\theta})^{-1/2}\mathbf{S}\Sigma(\hat{\theta})^{-1/2})' \mathbf{D}(\hat{\theta}|\mathbf{S}) \text{vec}(\Sigma(\hat{\theta})^{-1/2}\mathbf{S}\Sigma(\hat{\theta})^{-1/2}) \right] \\
 = \text{vec}(\mathbf{\Omega}'\mathbf{\Omega})' \mathbf{D}(\theta_0|\Sigma_*) \text{vec}(\mathbf{\Omega}'\mathbf{\Omega}) + O(n^{-1}),
 \end{aligned}$$

となることがわかる. ただし $\mathbf{\Omega} = \Sigma_*^{1/2}\Sigma(\theta_0)^{-1/2}$ である. また同様に, $\hat{\varepsilon}_i = \mathbf{\Omega}'\varepsilon_i + O_p(n^{-1/2})$ を利用すると,

$$\begin{aligned}
 (A.10) \quad &\frac{1}{n} \sum_{i=1}^n E_Y^* \left[(\hat{\varepsilon}'_i \otimes \hat{\varepsilon}'_i) \mathbf{D}(\hat{\theta}|\mathbf{S})(\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i) \right] \\
 &= \text{tr} \left(\left\{ \mathbf{\Omega} \otimes \mathbf{\Omega} \right\} \mathbf{D}(\theta_0|\Sigma_*) \left\{ \mathbf{\Omega}' \otimes \mathbf{\Omega}' \right\} \right. \\
 &\quad \left. \times \left[\sum_{a,b,c,d}^p E_Y^*[\varepsilon_a \varepsilon_b \varepsilon_c \varepsilon_d] \left\{ e_a^{(p)} e_b^{(p)'} \otimes e_c^{(p)} e_d^{(p)'} \right\} \right] \right) + O(n^{-1}),
 \end{aligned}$$

を得ることができる. ここで, K_p を Magnus and Neudecker ((1999), pp. 46–47) で定義されている, 任意の $p \times p$ 行列 C に対して $K_p \text{vec}(C) = \text{vec}(C')$ を満たす $p^2 \times p^2$ の交換行列 (commutation matrix) とする. Yanagihara et al. (2005) より, この交換行列 K_p は,

$$K_p = \sum_{a,b}^p \left\{ e_a^{(p)} e_b^{(p)'} \otimes e_b^{(p)} e_a^{(p)'} \right\},$$

となる. これと四次キュムラント(2.10)の関係式より,

$$\sum_{a,b,c,d}^p E_Y^*[\varepsilon_a \varepsilon_b \varepsilon_c \varepsilon_d] \left\{ e_a^{(p)} e_b^{(p)'} \otimes e_c^{(p)} e_d^{(p)'} \right\} = \Gamma_\kappa + \mathbf{I}_{p^2} + K_p + \text{vec}(\mathbf{I}_p) \text{vec}(\mathbf{I}_p)',$$

を得る. ただし Γ_κ は式(2.9)で定義される真の分布の尖度に依存する $p^2 \times p^2$ 行列である. 上記の期待値を式(A.10)に代入すると以下ようになる.

$$\text{tr} \left(\left\{ \mathbf{\Omega}' \otimes \mathbf{\Omega}' \right\} \mathbf{D}(\theta_0|\Sigma_*) \left\{ \mathbf{\Omega} \otimes \mathbf{\Omega} \right\} \left\{ \Gamma_\kappa + \mathbf{I}_{p^2} + K_p + \text{vec}(\mathbf{I}_p) \text{vec}(\mathbf{I}_p)' \right\} \right) + O(n^{-1}).$$

ここで, Magnus and Neudecker ((1999), p. 47) より, 任意の $p \times p$ 行列 C に対して, $K_p(C \otimes C) = (C \otimes C)K_p$ であり, また, もし C が対称行列であれば $K_p \text{vec}(C) = \text{vec}(C)$ を満たすということを利用すると, $\Delta(\theta)$ の定義式(A.1)より, $\dot{\Sigma}_\alpha(\theta)$ ($\alpha = 1, \dots, q$) は対称行列であるので,

$$K_p(\Omega \otimes \Omega)\Delta(\theta_0) = (\Omega \otimes \Omega)\Delta(\theta_0),$$

となることがわかる. $D(\theta_0|\Sigma_*) = \Delta(\theta_0)H(\theta_0|\Sigma_*)^{-1}\Delta(\theta_0)'$ であるので, 上記の式より,

$$K_p(\Omega \otimes \Omega)D(\theta_0|\Sigma_*) = (\Omega \otimes \Omega)D(\theta_0|\Sigma_*),$$

である. よって,

$$\begin{aligned} (A.11) \quad & \frac{1}{n} \sum_{i=1}^n E_{\mathbf{Y}}^* \left[(\hat{\varepsilon}'_i \otimes \hat{\varepsilon}'_i) D(\hat{\theta}|S) (\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i) \right] \\ & = \text{tr}(\Gamma_\kappa \{ \Omega \otimes \Omega \} D(\theta_0|\Sigma_*) \{ \Omega' \otimes \Omega' \}) \\ & \quad + 2\text{tr}(\{ \Omega' \Omega \otimes \Omega' \Omega \} D(\theta_0|\Sigma_*)) \\ & \quad + \text{vec}(\Omega' \Omega)' D(\theta_0|\Sigma_*) \text{vec}(\Omega' \Omega) + O(n^{-1}), \end{aligned}$$

である. 式(A.8)(A.9)(A.11)を(A.7)に代入すれば, AIC のバイアスは(2.8)式のようになることがわかる.

A.4 クロス・バリデーションに関する近似公式

第 i 番目の個体の観測値ベクトル y_i ($i = 1, \dots, n$) を除いたデータで定義される θ の対数尤度関数は,

$$(A.12) \quad l_{n-1}(\theta|S_{[-i]}) = -\frac{n-1}{2} \{ \log(2\pi) + \log|\Sigma(\theta)| + \text{tr}(S_{[-i]}\Sigma(\theta)^{-1}) \},$$

である. ここで, 式(2.18)と W_i (2.22) から $S_{[-i]} = S + W_i/(n-1)$ であることがわかるので, 式(A.12)は式(A.2)でのすべてのデータを使った対数尤度関数 $l_n(\theta|S)$ と W_i を使って,

$$l_{n-1}(\theta|S_{[-i]}) = \frac{n-1}{n} l_n(\theta|S) - \frac{1}{2} \text{tr}(W_i \Sigma(\theta)^{-1}),$$

と書き換えることができる. ここで, $\hat{\theta}_{[-i]}$ は $l_{n-1}(\theta|S_{[-i]})$ を最大にする解であることを利用すると,

$$\begin{aligned} & \left. \frac{\partial}{\partial \theta} l_{n-1}(\theta|S_{[-i]}) \right|_{\theta=\hat{\theta}_{[-i]}} \\ & = -\frac{n-1}{2} g(\hat{\theta}_{[-i]}|S) + \frac{1}{2} \Delta(\hat{\theta}_{[-i]})' \text{vec}(\Sigma(\hat{\theta}_{[-i]})^{-1/2} W_i \Sigma(\hat{\theta}_{[-i]})^{-1/2}) = 0, \end{aligned}$$

である. ただし $g(\cdot|S)$ は式(A.4)で与えられるすべてのデータに基づく対数尤度関数 $l_n(\cdot|S)$ の一次微分に関するベクトルであり, 0 はすべての成分が 0 であるようなベクトルである. 今, $z_i = (z_{i1}, \dots, z_{iq})' = \hat{\theta}_{[-i]} - \hat{\theta}$ とすると, テーラー展開により上記の式は,

$$-\frac{n-1}{2} \{ g(\hat{\theta}|S) + H(\hat{\theta}|S) z_i \} + \frac{1}{2} \Delta(\hat{\theta})' \text{vec}(\Sigma(\hat{\theta})^{-1/2} W_i \Sigma(\hat{\theta})^{-1/2}) + O(\|z_i\|) = 0,$$

($\|z_i\| \rightarrow 0$) と展開できる. ここで, $\hat{\theta}$ は最尤推定量であるので $g(\hat{\theta}|S) = 0$ となることに注意すれば,

$$z_i = \frac{1}{n-1} \{ H(\hat{\theta}|S)^{-1} \Delta(\hat{\theta})' \text{vec}(\Sigma(\hat{\theta})^{-1/2} W_i \Sigma(\hat{\theta})^{-1/2}) + O(\|z_i\|) \}.$$

$W_i = O_p(1)$, $S = \Sigma_* + O_p(n^{-1/2})$, $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$ であるので, 上記の式から $z_i = O_p(n^{-1})$ であることがわかる. よって,

$$(A.13) \quad \hat{\theta}_{[-i]} = \hat{\theta} + \frac{1}{n-1} H(\hat{\theta}|S)^{-1} \Delta(\hat{\theta})' \text{vec}(\Sigma(\hat{\theta})^{-1/2} W_i \Sigma(\hat{\theta})^{-1/2}) + O_p(n^{-2}),$$

となり, この式の第二項目まで考慮に入れたものがクロス・バリデーシヨンの近似公式(2.21)となる.

A.5 AIC_c のバイアス

式(2.17)から $\bar{y}_{[-i]} - \mu$ は,

$$\bar{y}_{[-i]} - \mu = \frac{n}{n-1} (\bar{y} - \mu) - \frac{1}{n-1} (y_i - \mu),$$

と変形できる. この式より AIC_c の期待値は,

$$\begin{aligned} E_Y^*[AIC_c] &= np \log(2\pi) + n E_Y^* [\log |\Sigma(\hat{\theta})|] + \sum_{i=1}^n E_Y^* [\text{tr}(\Sigma_* \Sigma(\hat{\theta}_{[-i]})^{-1})] \\ &\quad + \left(\frac{n}{n-1} \right)^2 \sum_{i=1}^n E_Y^* [(\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1} (\bar{y} - \mu)] \\ &\quad - \frac{2n}{(n-1)^2} \sum_{i=1}^n E_Y^* [(\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1} (y_i - \mu)] + O(n^{-1}), \end{aligned}$$

となることがわかる. 今, 式(A.13)を,

$$\hat{\theta}_{[-i]} - \hat{\theta} = \frac{1}{n} d_i + O_p(n^{-2}),$$

と変形する. ただし,

$$d_i = H(\hat{\theta}|S)^{-1} \Delta(\hat{\theta})' \text{vec}(\Sigma(\hat{\theta})^{-1/2} \{S - (y_i - \bar{y})(y_i - \bar{y})'\} \Sigma(\hat{\theta})^{-1/2}),$$

である. ここで, $nS = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})'$ であることに気がつけば, $\sum_{i=1}^n d_i = 0$ となる. これを利用すると, テーラー展開により任意の $p \times p$ 行列 C に関して,

$$\begin{aligned} \text{tr}(C \Sigma(\hat{\theta}_{[-i]})^{-1}) &= \text{tr}(C \Sigma(\hat{\theta})^{-1}) \\ &\quad + \frac{1}{n} \text{vec}(\Sigma(\hat{\theta})^{-1/2} C \Sigma(\hat{\theta})^{-1/2})' \Delta(\hat{\theta})' d_i + O_p(n^{-2}), \end{aligned}$$

であるので,

$$\sum_{i=1}^n \text{tr}(C \Sigma(\hat{\theta}_{[-i]})^{-1}) = n \text{tr}(C \Sigma(\hat{\theta})^{-1}) + O_p(n^{-1}),$$

となることがわかる. 今, $(\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1} (\bar{y} - \mu) = \text{tr}((\bar{y} - \mu)(\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1})$ であるので,

$$\begin{aligned} \sum_{i=1}^n \text{tr}(\Sigma_* \Sigma(\hat{\theta}_{[-i]})^{-1}) &= n \text{tr}(\Sigma_* \Sigma(\hat{\theta})^{-1}) + O_p(n^{-1}), \\ \sum_{i=1}^n (\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1} (\bar{y} - \mu) &= n (\bar{y} - \mu)' \Sigma(\hat{\theta})^{-1} (\bar{y} - \mu) + O_p(n^{-1}), \\ \sum_{i=1}^n (\bar{y} - \mu)' \Sigma(\hat{\theta}_{[-i]})^{-1} (y_i - \mu) &= n (\bar{y} - \mu)' \Sigma(\hat{\theta})^{-1} (\bar{y} - \mu) + O_p(n^{-1}), \end{aligned}$$

である．以上より， $n^2/(n-1)^2 = 1 + 2/n + O(n^{-2})$ ， $-2n/(n-1)^2 = -2/n + O(n^{-2})$ であることを考慮すると，

$$\begin{aligned} E_Y^*[AIC_c] &= np \log(2\pi) + nE_Y^*[\log |\Sigma(\hat{\theta})|] \\ &\quad + nE_Y^*[\text{tr}(\Sigma_* \Sigma(\hat{\theta})^{-1})] + nE_Y^*[(\bar{y} - \mu)' \Sigma(\hat{\theta})^{-1} (\bar{y} - \mu)] + O(n^{-1}). \end{aligned}$$

よって式(2.6)のリスク R_{KL} と比較すれば， AIC_c のバイアスのオーダーは $O(n^{-1})$ となることがわかる．

A.6 実解析での因子負荷量行列のパターン

4章での実解析で用いた候補のモデル(4.1)における因子負荷量行列のパターンは以下のようなものである．ただし因子負荷量行列 Λ_j の添え字は，それぞれのモデルの番号に対応するものである．

$$\begin{aligned} \Lambda_1 &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \\ 0 & \lambda_7 \\ \lambda_8 & 0 \\ \lambda_9 & 0 \\ \lambda_{10} & 0 \\ \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ 0 & \lambda_{13} \\ 0 & \lambda_{14} \end{pmatrix}, & \Lambda_2 &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & \lambda_3 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & 0 & \lambda_5 \\ 0 & 0 & \lambda_6 \\ 0 & 0 & \lambda_7 \\ \lambda_8 & 0 & 0 \\ \lambda_9 & 0 & 0 \\ \lambda_{10} & 0 & 0 \\ \lambda_{11} & 0 & 0 \\ \lambda_{12} & 0 & 0 \\ 0 & \lambda_{13} & 0 \\ 0 & \lambda_{14} & 0 \end{pmatrix}, & \Lambda_3 &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & \lambda_3 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & \lambda_6 & 0 \\ 0 & \lambda_7 & 0 \\ \lambda_8 & 0 & 0 \\ \lambda_9 & 0 & 0 \\ \lambda_{10} & 0 & 0 \\ \lambda_{11} & 0 & 0 \\ \lambda_{12} & 0 & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & \lambda_{14} \end{pmatrix}, \\ \\ \Lambda_4 &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & \lambda_3 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & 0 & \lambda_5 \\ 0 & 0 & \lambda_6 \\ 0 & 0 & \lambda_7 \\ \lambda_8 & 0 & 0 \\ \lambda_9 & 0 & 0 \\ \lambda_{10} & 0 & 0 \\ \lambda_{11} & 0 & 0 \\ \lambda_{12} & 0 & 0 \\ 0 & 0 & \lambda_{13} \\ 0 & 0 & \lambda_{14} \end{pmatrix}, & \Lambda_5 &= \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 \\ 0 & 0 & \lambda_5 & 0 \\ 0 & 0 & \lambda_6 & 0 \\ 0 & 0 & \lambda_7 & 0 \\ \lambda_8 & 0 & 0 & 0 \\ \lambda_9 & 0 & 0 & 0 \\ \lambda_{10} & 0 & 0 & 0 \\ \lambda_{11} & 0 & 0 & 0 \\ \lambda_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 & \lambda_{14} \end{pmatrix}, & \Lambda_6 &= \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 & 0 \\ 0 & 0 & \lambda_5 & 0 & 0 \\ 0 & 0 & \lambda_6 & 0 & 0 \\ 0 & 0 & \lambda_7 & 0 & 0 \\ 0 & 0 & 0 & \lambda_8 & 0 \\ 0 & 0 & 0 & \lambda_9 & 0 \\ 0 & 0 & 0 & \lambda_{10} & 0 \\ 0 & 0 & 0 & \lambda_{11} & 0 \\ 0 & 0 & 0 & \lambda_{12} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{13} \\ 0 & 0 & 0 & 0 & \lambda_{14} \end{pmatrix}. \end{aligned}$$

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiado, Budapest.
- Akaike, H. (1987). Factor analysis and AIC, *Psychometrika*, **52**, 317–332.
- Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models, *Multivariate Behavioral Research*, **28**, 351–342.

- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**, 345–370.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.
- Browne, M. W. and Cudeck, R. (1989). Single sample cross-validation indices for covariance structures, *Multivariate Behavioral Research*, **24**, 445–455.
- Cudeck, R. and Browne, M. W. (1983). Cross-validation of covariance structures, *Multivariate Behavioral Research*, **18**, 147–167.
- De Gooijer, J. G. (1995). Cross-validation criteria for covariance structures, *Communications in Statistics, Simulation and Computation*, **24**, 111–147.
- Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression, *Biometrika*, **84**, 707–716.
- Fujikoshi, Y., Noguchi, T., Ohtaki, M. and Yanagihara, H. (2003). Corrected versions of cross-validation criteria for selecting multivariate regression and growth curve models, *Annals of the Institute of Statistical Mathematics*, **55**, 537–553.
- Fujikoshi, Y., Yanagihara, H. and Wakaki, H. (2005). Bias corrections of some criteria for selection multivariate linear regression models in a general case, *American Journal of Mathematical and Management Sciences* (in press).
- Fujisawa, H. and Eguchi, S. (2003). Robust estimation in the normal mixture model, Research Memo., No. 867, The Institute of Statistical Mathematics, Tokyo.
- Hu, L., Bentler, P. M. and Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted?, *Psychological Bulletin*, **112**, 351–362.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.
- Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC, *Annals of the Institute of Statistical Mathematics*, **49**, 411–434.
- Isogai, T. (1983). On measures of multivariate skewness and kurtosis, *Mathematica Japonica*, **28**, 251–261.
- Kano, Y. (1992). Robust statistics for test-of-independence and related structural models, *Statistics & Probability Letters*, **15**, 21–26.
- 狩野 裕, 三浦麻子 (2002). 『AMOS, EQS, CALIS によるグラフィカル多変量解析—目でみる共分散構造分析—(増補版)』, 現代数学社, 京都.
- Kano, Y., Berkane, M. and Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters, *Biometrika*, **77**, 575–585.
- 前田忠彦 (1995). 日本人の満足感の構造とその規定因に関する因果モデル—共分散構造分析の「日本人の国民性調査」への適用—, *統計数理*, **43**, 141–160.
- Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, John Wiley & Sons, New York.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519–530.
- 坂元慶行, 中村 隆, 前田忠彦, 土屋隆裕 (2004). 国民性調査第 11 回全国調査—2003 年全国調査—, *統計数理研究所研究レポート*, No. 92.
- Satorra, A. and Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis, *Proceedings of the Business and Economic Statistics Section*, 308–313, American Statistical Association, Alexandria.
- Stone, M. (1974). Cross-validated choice and assessment of Statistical predictions, *Journal of the Royal Statistical Society, Series B*, **36**, 111–147.

- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *Journal of the Royal Statistical Society, Series B*, **39**, 44–47.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Communications in Statistics, Theory and Methods*, **A7**, 13–26.
- 竹内 啓 (1976). 情報統計量の分布とモデルの適切さの規準, *数理科学*, **153**, 12–18.
- Wakaki, H., Yanagihara, H. and Fujikoshi, Y. (2002). Asymptotic expansion of the null distributions of test statistics for multivariate linear hypothesis under nonnormality, *Hiroshima Mathematical Journal*, **32**, 17–50.
- Yanagihara, H. (2004a). Corrected version of *AIC* for selecting multivariate normal linear regression models in a general nonnormal case, Tech. Report, No. 04–04, Statistical Research Group, Hiroshima University, Hiroshima.
- Yanagihara, H. (2004b). A family of estimators for multivariate kurtosis in a nonnormal linear regression model, Tech. Report, No. 04–08, Statistical Research Group, Hiroshima University, Hiroshima.
- Yanagihara, H., Tonda, T. and Matsumoto, C. (2005). The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption, *Journal of Multivariate Analysis* (in press).
- 柳井晴夫, 繁樹算男, 前川眞一, 市川雅数 (1990). 『因子分析—その理論と方法—』, 朝倉出版, 東京.
- Yuan, K.-H. and Bentler, P. M. (1997). Generating multivariate distributions with specified marginal skewness and kurtosis, *SoftStat' 97—Advances in Statistical Software 6—* (eds. W. Bandilla and F. Faulbaum), 385–391, Lucius & Lucius, Stuttgart.
- Yuan, K.-H. and Bentler, P. M. (1998). Normal theory based test statistics in structural equation modelling, *British Journal of Mathematical and Statistical Psychology*, **51**, 289–309.
- Yuan, K.-H. and Bentler, P. M. (1999). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions, *Statistica Sinica*, **9**, 831–853.

Selection of Covariance Structure Models in Nonnormal Data
by Using Information Criterion: An Application to Data
from the Survey on the Japanese National Character

Hirokazu Yanagihara

University of Tsukuba

This paper deals with a bias correction of Akaike's Information Criterion (AIC) for a selection of covariance structure models under a general nonnormal case, when we assume that an observation is distributed according to the normal distribution. The AIC under the normal assumption has a constant bias that depends on a kurtosis of the true distribution. We correct this bias by partly using the cross-validation method. We verify that our criterion is better than the AIC and TIC by calculating a theoretical bias and conducting a numerical experiment. Further, an illustrative analysis applying factor models with respect to Japanese people's level of satisfaction to an actual data set from the eleventh nationwide survey on the Japanese national character is also demonstrated. Using our new criterion, we search the best model from the candidate ones.