382　教育心理学研究, 2007, 55, 382—392

構造方程式モデリングにおける検定統計量の Bartlett 補正

岡　田　謙　介*　　星　野　崇　宏**　　繁　桝　算　男***

構造方程式モデリング（SEM）では柔軟なモデル構成が可能であるために，モデルの評価・選択が重要になる。統計的モデル選択において基本となる統計量に尤度比検定統計量Tがあり，SEM でもこれが漸近的に χ^2分布にしたがう性質を利用した検定が可能である。しかし，標本サイズが小さいとき検定統計量Tの標本分布は χ^2分布から正方向に逸脱する。逸脱の度合いは１因子あたりの観測変数数が大きいとき，とくに大きくなる。通常得られる教育心理学データの標本数程度では，この逸脱のため，適切なモデル選択が行えなくなってしまう。また，適合度指標の大部分は χ^2分布にしたがうTの関数として構成されるので，Tの分布のゆがみが直接的に波及する。そこで，我々は今回Tに Bartlett 補正を適用し，その χ^2分布への近似精度を向上させる方法を提案する。モンテカルロ実験により，提案した方法がTの標本分布の χ^2分布に対する近似精度を大幅に改善していることを確認する。

キーワード：構造方程式モデリング，尤度比検定，χ^2値，Bartlett 補正，適合度

問　　　題

構造方程式モデリング（Structural Equation Modeling, SEM）は，教育心理学をはじめ心理学・行動科学の研究においてなくてはならない多変量解析法である。SEM の大きな特徴のひとつとして，しばしばその柔軟なモデル構成力が挙げられる。回帰分析や正準相関分析といった古典的な多変量解析では，統計モデルはデータを得る前に固定されており，研究者はモデルに合わせた形で観測データを準備する必要があった。しかし SEM の枠組みでは，研究者は事前知識とデータの性質とを考慮して，状況に応じたモデルを構成し分析することができる。このように，研究者のアイディアを自在に統計モデルの形で表現できることが SEM の長所である。

このことは，SEM を用いた実践研究では可能なモデルが数多くありえる，ということでもある。したがって，観測データに対するモデルの当てはまりの良さを測ったり，競合するモデル間の優劣を比較したりといった「モデルの比較・検証」が，実質科学的観点からも重要となる。

*　東京大学大学院総合文化研究科・日本学術振興会
〒153-8902　東京都目黒区駒場 3-8-1　東京大学大学院
総合文化研究科　広域科学専攻
生命環境科学系　認知行動科学大講座
ken@bayes.c.u-tokyo.ac.jp
**　東京大学教養学部／大学院総合文化研究科
***　東京大学大学院総合文化研究科

SEM における母数推定は，母数θによって構造化されたモデル上の分散共分散行列$\Sigma(\theta)$と標本数Nの観測データ[1]とから計算した分散共分散行列Sとの乖離の度合いを表す関数

$$F=F(\Sigma(\theta),\ S) \quad (1)$$

をθについて最小化することにより達成される。このFを目的関数（objective function），もしくは乖離度関数（discrepancy function）という。Fの具体的な形は，母数推定にどのような方法を用いるかによって決まる。ここでの母数推定法としては最尤法，最小二乗法，ADF 法などいくつかの手法が提案されているが，点推定量の一致性と漸近正規性とが成り立つことから，一般に最尤法が基本とされる（狩野・三浦, 2002 ; Pp.137-138）。ほとんどの SEM ソフトウェアでも，最尤法が標準となっている。したがって，本論文でも最尤法を用いた母数推定を考える。

いま，観測変数の数をp，推定する母数の数をq，潜在変数の数をrで表す。正規性の仮定が満たされる場合の，最尤法による目的関数は

$$F=\log|\Sigma(\theta)|+tr[S\Sigma(\theta)^{-1}]-\log|S|-p \quad (2)$$

となる。各種の最適化手法を用いてθに関してFを最小化し，このモデルの母数推定を行うことができる。そうして推定された目的関数を$\hat{F}=F(\Sigma(\hat{\theta}),S)$で表そう。$\Sigma(\hat{\theta})$は，真の母数の値$\theta$が未知であるため，その構造方程式モデルによる推定値$\hat{\theta}$を用いて導いたモデル分散共分散行列の推定値である。

[1]　記述の便宜上，本論文では以下$n=N-1$とおく。

SEM を含め，複雑な統計モデルにおけるモデル選択には尤度比検定が広く用いられる。(既存の) SEM における尤度比検定統計量Tは

$$T=n\hat{F} \tag{3}$$

によって構成される。そして，これが帰無仮説の下で漸近的に自由度dfのカイ二乗分布（以下，$\chi^2{}_{df}$分布と表記する）にしたがう性質を利用してモデルの比較•検証を行うことができる(尤度比検定)。自由度dfは，帰無仮説下での母数空間と対立仮説下での母数空間との次元の差に等しくなる。

尤度比検定における帰無仮説と対立仮説はそれぞれ，

H_0：いま考えているモデルが真のモデルである

H_1：母数に制約のないモデル（飽和モデル）が真のモデルである

となる。なお，(3)式のTは，カイ二乗検定統計量，もしくはカイ二乗値と呼ばれることも多い。尤度比検定統計量はしばしば記号χ^2を使って表されるが，本論文では観測データから既存の方法で計算される標本尤度比検定統計量を記号Tで記述し，漸近論の仮定が満たされる場合の，$\chi^2{}_{df}$分布にしたがう検定統計量と区別する。

心理学の有名学術誌 13 誌に掲載された 41 の論文を調べた McDonald & Ho (2002) のレビューにおいて，尤度比検定統計量はモデルのよさを測る統計量の中で唯一，全ての研究において報告されていた。このように，検定統計量Tは SEM においてモデルのよさを測り，モデルの比較・検証を行う上で最も一般的と言える統計量である。

しかし，既存のシミュレーション研究の結果を見ると，たとえ考えているモデルが真であっても，標本検定統計量Tはその期待値から逸脱した値が報告されている (Curran, West & Finch, 1996: p.22 Table 1 ; Hu, Bentler, & Kano, 1992 : p.356 Table 3 ; Marsh, Balla, & McDonald, 1988 : p.400 Table 3 ; ほか多数)。先行研究の結果からは，一般に標本数が小さいほど逸脱の度合いはより大きい（もっとも，ここで引用した 3 論文を見ればわかるように，この標本数とTの乖離の関係は一貫したものではなかった；この原因についてはシミュレーション(1)の目的の項で述べる)。

尤度比検定は検定統計量Tが$\chi^2{}_{df}$分布にしたがうことを前提とした検定である。この前提が満たされない場合，尤度比検定の結果は信用できない。とくに教育心理学における SEM を用いた研究では，100 未満〜数百程度という決して多くない標本数で分析を行わざるを得ない場合が多い。このような小標本下でも不適切なふるまいをしない検定統計量をつくることは，実際的な観点からも非常に重要である。

そこで本論文では，第 1 に，Bartlett 補正を用いて検定統計量Tの分布の$\chi^2{}_{df}$分布からの逸脱を補正することを提案する。また第 2 に，先行研究の欠点を改善した大規模なシミュレーションにより，既存の検定統計量Tと，補正済み検定統計量 T^*の挙動を比較し，提案された補正がたしかに尤度比検定統計量の性質を改善していることを調べる。

方　　法

本論文で適用する Bartlett 補正は，尤度比検定が行える任意の統計モデルにおいて，検定統計量の$\chi^2{}_{df}$分布に対する近似精度を向上させるための補正である。

Tを$\chi^2{}_{df}$分布で近似したときの誤差のオーダーは通常$O(n^{-1})$である。Bartlett (1937) は，標本統計量Tの$\chi^2{}_{df}$分布に対する近似を改善する方法を提案した。いまTの期待値が，aを帰無仮説のもとで一致推定量である定数として

$$E(T)=df\left\{1+\frac{a}{n}+O(n^{-2})\right\} \tag{4}$$

と展開できるとしよう。Bartlett は，このとき尤度比検定統計量Tを

$$T^*=\left(1-\frac{a}{n}\right)T \tag{5}$$

と補正することによって，補正済み統計量T^*の期待値が

$$E(T^*)=df+O(n^{-2}) \tag{6}$$

を満たすように，近似を改善できることを示した。

(5)式で補正に用いた係数

$$m=\left(1-\frac{a}{n}\right) \tag{7}$$

を，Bartlett 補正因子 (Bartlett correction factor) と呼ぶ。

nは常に 0 以上なので，(7)式でもしもaが正であればmは 1 より小さくなる。このとき補正済み統計量$T^*=mT$ の分布は，元のTの分布を零方向に引き戻した形になる。逆に$a<0$のときには$m>1$となり，T^*の分布は元のTの分布をより正方向に押し広げた形の分布となる。

実際に Bartlett 補正因子mを求めるには，(4)式の漸近展開を導かなければならない。これは多変量の複雑な統計モデルでは必ずしも容易ではない。最近の結果としては，例えば Moulton, Weissfeld, & Laurent (1993)はロジスティック回帰モデルについて，Zucker, Liberman, & Manor (2000) は混合線形モデルについ

て，また Lagos & Morettin (2004) は ARMA モデルについてそれぞれ Bartlett 補正因子mを導出したことを報告している。Bartlett 補正に関する網羅的なレビューとしては，Cribari-Neto & Cordeiro (1996) を参照いただきたい。

Wakaki, Eguchi, & Fujikoshi (1990) は，構造を持った分散共分散行列間の距離関数ρと標本数nとの積

$$T_w=n\rho(S,\ \Sigma(\hat{\theta})) \tag{8}$$

で書ける統計量T_wが Bartlett 補正可能であることを示し，その Bartlett 補正因子を導いた。ρ は S と$\Sigma(\hat{\theta})$との近さを測る関数でいくつかの条件をみたす必要があるが，SEM での母数推定の目的関数Fはこれらの条件をみたす。このとき Wakaki et al. (1990) の結果にしたがえば，Bartlett 補正因子mは

$$m=1+\frac{2c}{dfn} \tag{9}$$

で与えられる。ここでcはT_wの特性関数を漸近展開したときに現れる係数であり，

$$c=\frac{1}{72df}\{-3p(2p^2+3p-1)$$
$$-9g^{abcd}K_{abcd}+g^{abcdef}K_{abc,def}\}+\frac{1}{16df}g^{ab}g^{cd}$$
$$\{4K_{[ab]cd}-K_{[ab][cd]}+2K_{[ac][bd]}\} \tag{10}$$

と書ける。ここでの各g，Kは，Einstein の規約 (Einstein's convention) を用いて

$$g^{abcdef}=g^{ab}g^{cdef}+g^{ac}g^{bdef}+g^{ad}g^{bcef}+g^{ae}g^{bcdf}$$
$$+g^{af}g^{bcde} \tag{11}$$

$$g^{abcd}=g^{ab}g^{cd}+g^{ac}g^{bd}+g^{ad}g^{bc} \tag{12}$$

$$K_{abc,def}=K_{abc}K_{def} \tag{13}$$

$$K_{[ab][cd]}=\mathrm{tr}(J_{[ab][cd]}) \tag{14}$$

$$K_{[ab]cd}=\mathrm{tr}(J_{[ab]}J_cJ_d) \tag{15}$$

$$K_{abc\cdots}=\mathrm{tr}(J_aJ_bJ_c\cdots) \tag{16}$$

$$J_{[ab]}=J_{ab}-\frac{1}{2}J_cg^{cd}\mathrm{tr}J_aJ_{ab} \tag{17}$$

$$J_{ab}=\Sigma^{1/2}\left[\frac{\partial}{\partial\xi^a}\frac{\partial}{\partial\xi^b}\Sigma^{-1}\right]\Sigma^{1/2} \tag{18}$$

$$J_a=\Sigma^{1/2}\left[\frac{\partial}{\partial\xi^a}\Sigma^{-1}\right]\Sigma^{1/2} \tag{19}$$

のように表現される。ただし Einstein の規約とは，一つの項の中で上・下の双方に登場する添え字については，それで和をとっていることを表すとするものである。例えば，J_cg^{cd} は通常のシグマ記号を用いた記法での $\sum_{c=1}^{q}J_cg^{cd}$ にあたる。Einstein の規約を用いることで，シグマ記号がいくつも重なる見苦しい表記を避けることができる。

(11)～(19)式をよく見ると，J_aとJ_{ab}さえ求めれば，SEM における尤度比検定統計量Tを Bartlett 補正できることがわかる。J_aとJ_{ab}は，それぞれ共分散行列を各母数について1階微分・2階微分可能であれば，求めることができる。したがって，母数をその推定値で置き換え，共分散行列の未知母数に関する1階微分および2階微分を計算すれば，SEM での標本尤度比検定統計量Tに Bartlett 補正を適用できる。

シミュレーション(1)

目的

シミュレーション1の目的は，Tの標本分布が特定の条件下でχ^2_{df}分布から逸脱することを確認し，またその逸脱に影響する要因を調べることである。

代表的な SEM でのシミュレーション研究における繰り返し数を TABLE 1 にまとめた。先行研究には，1実験水準あたりの繰り返し数が十分大きいとは言えない，という問題があった。繰り返し数は，多くても 500，通常は 200 前後であったことがわかる。上述のように，先行研究における標本サイズとTの平均値の大きさとの関係は，必ずしも単調減少せず，しばしば変動していたが，これは単に標本誤差による可能性がある。

そこで今回，我々は繰り返し数を1実験水準あたり 10,000 回に増やして実験を行った。先行研究の 50～100 倍の繰り返し数を採用することで，Tの標本誤差の影響をおよそ無視できる程度に抑えることができる。また，単にTの平均値にとどまらず，有意水準を超えた標本の割合やヒストグラムをも考慮に入れたより精緻な議論が可能になる。

方法

モデル Kenny & McCoach (2003) にならった確認的因子分析モデルを用いた。Hu & Bentler (1998) が

TABLE 1 先行研究と本研究とにおける，1実験水準あたりの繰り返し数

出典	繰り返し数
Anderson & Gerbing (1984)	100
Hu, Bentler, & Kano (1992)	200
Ding, Velicer, & Harlow (1995)	100
Hu & Bentler (1998)	200
Fan, Thompson, & Wang (1999)	200
Jackson (2001)	200
Curran, Bollen, Paxton, Kirby, & Chen (2002)	500
Kenny & McCoach (2003)	100
Jackson (2003)	200
本研究	10,000

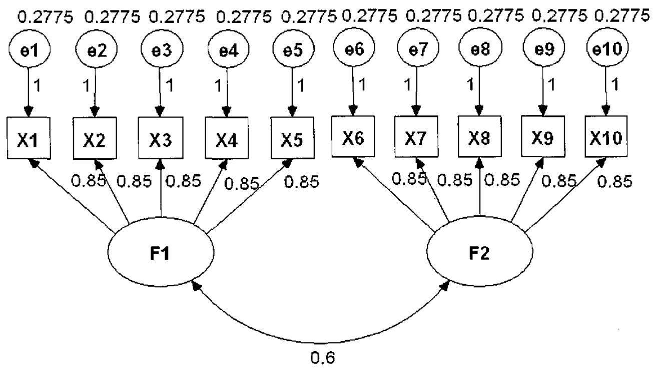

FIGURE 1 シミュレーションに用いた確認的因子分析モデル

モデル 3（10 観測変数，2 潜在変数）の場合を示す。モデル 1，モデル 2 も同様で，1 因子あたりの観測変数がそれぞれ 3 変数，4 変数となったものである。

指摘するように，確認的因子分析モデルは最も代表的でよく利用される SEM のモデルである。また同値モデル（適合度が等しくなるモデル）として因子間にパスがあるモデルなどが存在するので，結果を因子分析モデルのみに限定することなく，より SEM に一般的ものとして解釈することが可能である。

因子数は 2 とし，観測変数数については 6 変数，8 変数，10 変数の 3 条件を用意した。順にモデル 1，モデル 2，モデル 3 と呼ぶことにする。この設定は，因子あたり観測変数数が適合度指標に影響を及ぼすという報告（Anderson & Gerbing, 1984 ; Ding, Velicer, & Harlow, 1995）をふまえ，これを操作することを意図したものである。真値は FIGURE 1 のように設定した。

データ生成と推定　観測変数のシミュレーションデータセットは，次のように生成された。まず，モデル母数の真値を用いて真の分散共分散行列を計算した。次に，平均が零ベクトル・分散共分散行列がいま計算した真値である多変量正規分布から，乱数発生により N 標本の「観測データ」を発生させた。乱数発生法には，さまざまな良い性質を持つとされる Mersenne Twister を用いた（Matsumoto & Nishimura, 1998）。そして，真のモデルに対する尤度比検定統計量 T を計算した。

繰り返し　上記のデータセット生成を，1 水準につき 10,000 回繰り返した。

標本サイズ　生成する 1 データセットあたりの標本数 N は，行動科学学分野の研究を代表することを念頭に，N=50, 100, 200, 300 の 4 条件を用意した。

実装　Windows XP の動作するパーソナルコンピュータ上で，統計ソフトウェア SAS バージョン 8 の PROC CALIS・PROC IML を利用して分析を行った。

結果

各条件における 10,000 回の標本平均を TABLE 2 (a) に，期待値 df で割って標準化した値を同(b)に示す。また，χ^2_{df} 分布の有意水準 α 別に見た棄却率（有意水準 α で棄却される標本の割合）を TABLE 3 に示す。さらに，3 条件を代表して，モデル 3 における 10,000 回サンプリングした T のヒストグラムを，モデル 1～3 のそれぞれについて FIGURE 2 に示す。図中の曲線は，T が尤度比検定理論上漸近的にしたがう χ^2_{df} 分布である。

標本平均および棄却率の期待値からの逸脱は，N が小さいほど大きくなった。また標準化した平均値で比較すると，N が小さい条件どうしでは因子あたり観測変数数が大きいときの方がより乖離が大きくなった。さらに，ヒストグラムに見る T の分布の χ^2_{df} 分布からの逸脱も，標本数 N が小さい条件ほど大きくなった。全実験条件中で標本サイズが最小・因子あたり観測変数数が最大であるモデル 3（観測変数数=10）・N=50 条件のヒストグラムを見ると（FIGURE 2 左上），顕著な χ^2_{df} 分布からの逸脱が見てとれる。

上記の結果は主要な先行研究の流れとも一致するものである（先行研究について詳しくは星野・岡田・前田（2005）のレビューを参照）。本研究では繰り返し回数を先行研究よりも大幅に増やすことによって，より一貫性が高く信頼できる結果を得ることができた。また標本ヒスト

TABLE 2　標本 T の平均値

(a)　補正前

	N=50	N=100	N=200	N=300	期待値
V6	8.510 (0.043)	8.233 (0.041)	8.138 (0.041)	8.116 (0.041)	8.000
V8	20.905 (0.069)	19.911 (0.065)	19.428 (0.063)	19.222 (0.062)	19.000
V10	37.929 (0.092)	35.871 (0.087)	34.882 (0.086)	34.471 (0.083)	34.000

(b)　補正前［標準化］

	N=50	N=100	N=200	N=300	期待値
V6	1.064 (0.011)	1.029 (0.010)	1.017 (0.010)	1.015 (0.010)	1.000
V8	1.100 (0.011)	1.048 (0.011)	1.023 (0.010)	1.012 (0.010)	1.000
V10	1.116 (0.011)	1.055 (0.011)	1.026 (0.010)	1.014 (0.010)	1.000

(a)は標本値，(b)は期待値で割って標準化した値についての結果である。各セルで上段は平均値，下段カッコ内は標準誤差である。

TABLE 3 有意水準αで棄却域に入った標本Tの割合

		$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.20$
V6	$N=50$	0.016	0.069	0.127	0.238
	$N=100$	0.012	0.058	0.114	0.217
	$N=200$	0.012	0.057	0.108	0.207
	$N=300$	0.010	0.056	0.109	0.212
V8	$N=50$	0.023	0.106	0.184	0.315
	$N=100$	0.016	0.082	0.137	0.282
	$N=200$	0.010	0.059	0.116	0.210
	$N=300$	0.010	0.049	0.095	0.207
V10	$N=50$	0.035	0.121	0.212	0.357
	$N=100$	0.020	0.083	0.153	0.266
	$N=200$	0.014	0.063	0.123	0.233
	$N=300$	0.011	0.057	0.113	0.212

グラムを示したことにより，Tの標本分布が$\chi^2{}_{df}$分布から乖離していることを確認できた。

シミュレーション(2)

目的

シミュレーション2の目的は，本論文で提案したSEMにおける検定統計量のBartlett補正の有用性を，シミュレーション実験によって確かめることである。前項では小標本数などの条件下でTの標本分布が$\chi^2{}_{df}$分布から逸脱することがわかったが，本項では我々の提案した補正を適用し，こうした条件下でもT^*の$\chi^2{}_{df}$分布に対する当てはまりが改善されることを確認する。

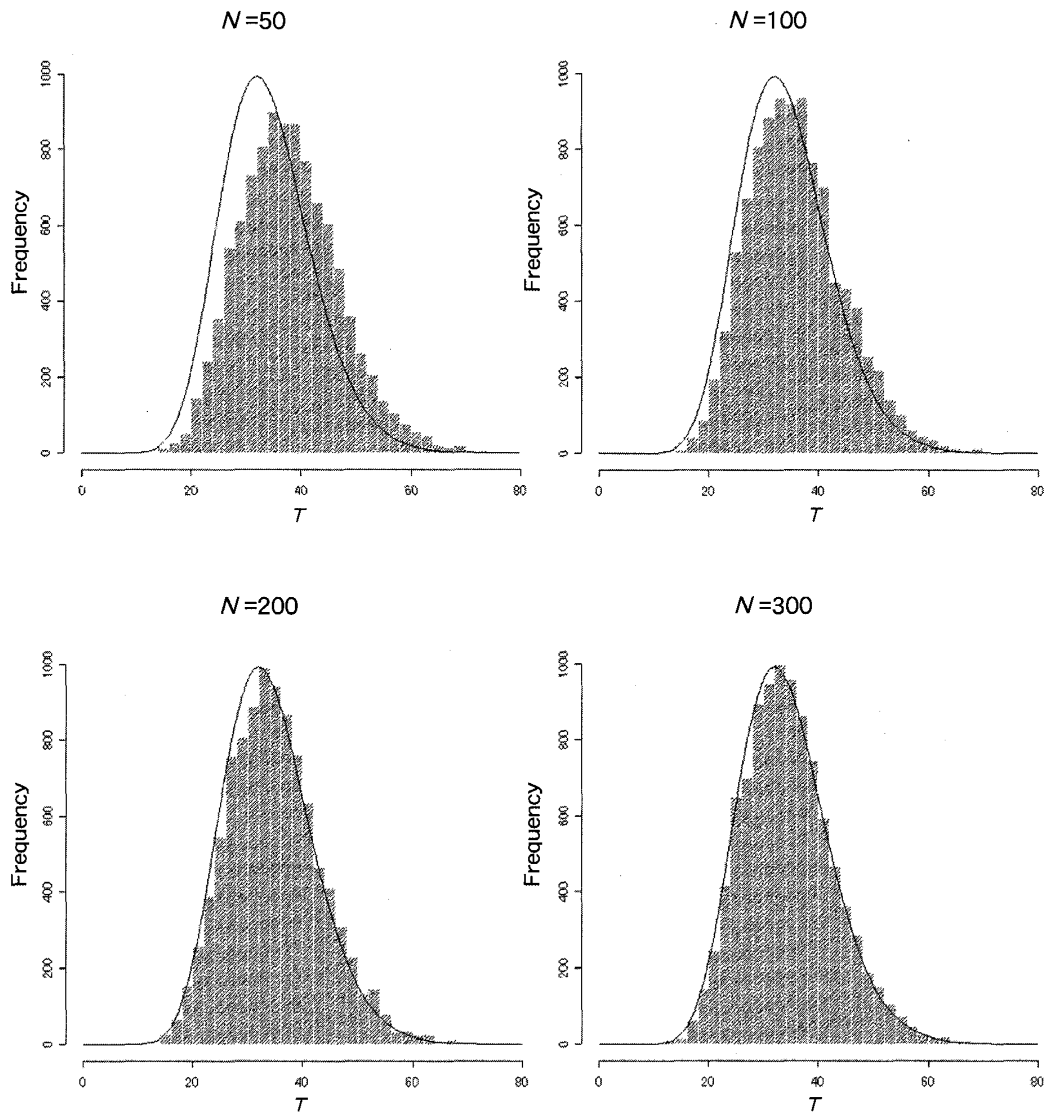

FIGURE 2 標本Tのヒストグラムと，漸近理論によるχ^2分布
モデル3（10観測変数，2潜在変数）の場合を示す。とくに小標本（*e.g.* $N=50$）の条件で，標本Tが漸近理論によるχ^2分布（曲線）から大幅に逸脱してしまっている。

方法

モデル，データ生成，繰り返し，標本サイズ，実装等の仕様については，シミュレーション1と同様とした。各々のデータについて，提案した方法で Bartlett 補正因子m，補正済み尤度比検定統計量T^*を計算した。この算出過程は，煩雑になるので付録に示す。

結果

シミュレーション1と同様に結果を示す。まず，各条件の平均値と標準誤差とを TABLE 4 に，有意水準αごとの棄却率を TABLE 5 に示す。また，モデル3でのT^*のヒストグラムを，FIGURE 3 に示す。

同一実験条件で補正前と補正後とを比較すると，ヒストグラムに見る小標本時のTの分布のχ^2_{df}分布からの逸脱が Bartlett 補正によって劇的に改善された。また，平均値・棄却率もそれぞれ，Bartlett 補正によって大きくχ^2_{df}分布での期待値に近づいたことがわかる。

さらに，モデル3をとりあげて，補正の効果を詳しく見る。補正前T・補正後T^*各々10,000 回の平均値をグラフ化し，FIGURE 4 に示した。ただし期待値$T=34$に太線が引いてあり，エラーバーは標準誤差である。Nのどの水準においても，Bartlett 補正がT^*の分布をχ^2_{df}分布へと近づけている様子がわかる。また，有意水準を$\alpha=0.01$, 0.05, 0.10, 0.20 に各々設定したときの，棄却域に入る標本の割合を FIGURE 5 にグラフ化した。Bartlett 補正によってχ^2_{df}分布への近似精度が改善されている様子が理解できる。

TABLE 4　標本T^*の平均値

(a)　補正後

	$N=50$	$N=100$	$N=200$	$N=300$	期待値
V6	7.961	7.966	8.006	8.029	8.000
	(0.040)	(0.040)	(0.040)	(0.040)	
V8	19.277	19.133	19.011	18.971	19.000
	(0.063)	(0.063)	(0.061)	(0.061)	
V10	34.500	34.246	34.092	33.949	34.000
	(0.083)	(0.083)	(0.084)	(0.082)	

(b)　補正後［標準化］

	$N=50$	$N=100$	$N=200$	$N=300$	期待値
V6	0.995	0.996	1.001	1.004	1.000
	(0.010)	(0.010)	(0.010)	(0.010)	
V8	1.015	1.007	1.001	0.998	1.000
	(0.010)	(0.010)	(0.010)	(0.010)	
V10	1.015	1.007	1.003	0.999	1.000
	(0.010)	(0.010)	(0.010)	(0.010)	

(a)は標本値，(b)は期待値で割って標準化した値についての結果である。各セルで上段は平均値，下段カッコ内は標準誤差である。TABLE 2と見比べると，どの条件についても補正後の結果がχ^2 分布での期待値へと近づいていることがわかる。

TABLE 5　有意水準αで棄却域に入った標本T^*の割合

		$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.20$
V6	$N=50$	0.010	0.050	0.097	0.195
	$N=100$	0.009	0.049	0.099	0.201
	$N=200$	0.010	0.053	0.102	0.198
	$N=300$	0.009	0.053	0.105	0.205
V8	$N=50$	0.013	0.064	0.115	0.220
	$N=100$	0.009	0.065	0.111	0.240
	$N=200$	0.006	0.045	0.102	0.194
	$N=300$	0.009	0.046	0.084	0.200
V10	$N=50$	0.013	0.054	0.107	0.217
	$N=100$	0.012	0.055	0.109	0.205
	$N=200$	0.012	0.052	0.104	0.204
	$N=300$	0.009	0.050	0.099	0.195

TABLE 4と見比べると，どの条件についても補正後の結果がχ^2分布での期待値へと近づいていることがわかる。

シミュレーション(1)と(2)の結果をまとめてTおよびT^*から主要な適合度指標を計算すると，その平均値は TABLE 6 に示すようになった。いずれの適合度指標についても標本数Nに対して指標値が単調に増加もしくは減少する関係が確認できたので，各指標ごとに$N=$50, 100, 200, 300 の4条件中，最大値と最小値との比を計算した。

考　察

本論文では，とくに小標本下で標本尤度比検定統計量がχ^2_{df}分布から正方向へ逸脱するという問題を指摘・確認し，Bartlett 補正を用いたその解決法を提案した。

これまで同問題が意識されることは少なく，SEM を用いた応用研究では補正しない検定統計量Tやそのp値が報告されてきた。また，SEM で発展している各種適合度指標はその多くがTの関数であるが，適合度指標の算出においても補正しないTが使われてきた。

しかし，補正しないTのχ^2_{df}分布からの逸脱の程度は，決して無視できるものではない。シミュレーション(1)におけるモデル3，$N=50$の条件では，5％水準で有意な値をとる標本Tがχ^2_{df}分布での期待値の2倍をゆうに超えて12.1%も出現した(FIGURE 5 右上)。小標本下でTをχ^2_{df}分布で近似することは，単に精度が悪いというだけでなく，もはや誤りである。これに対し，本研究のシミュレーション実験では，提案された

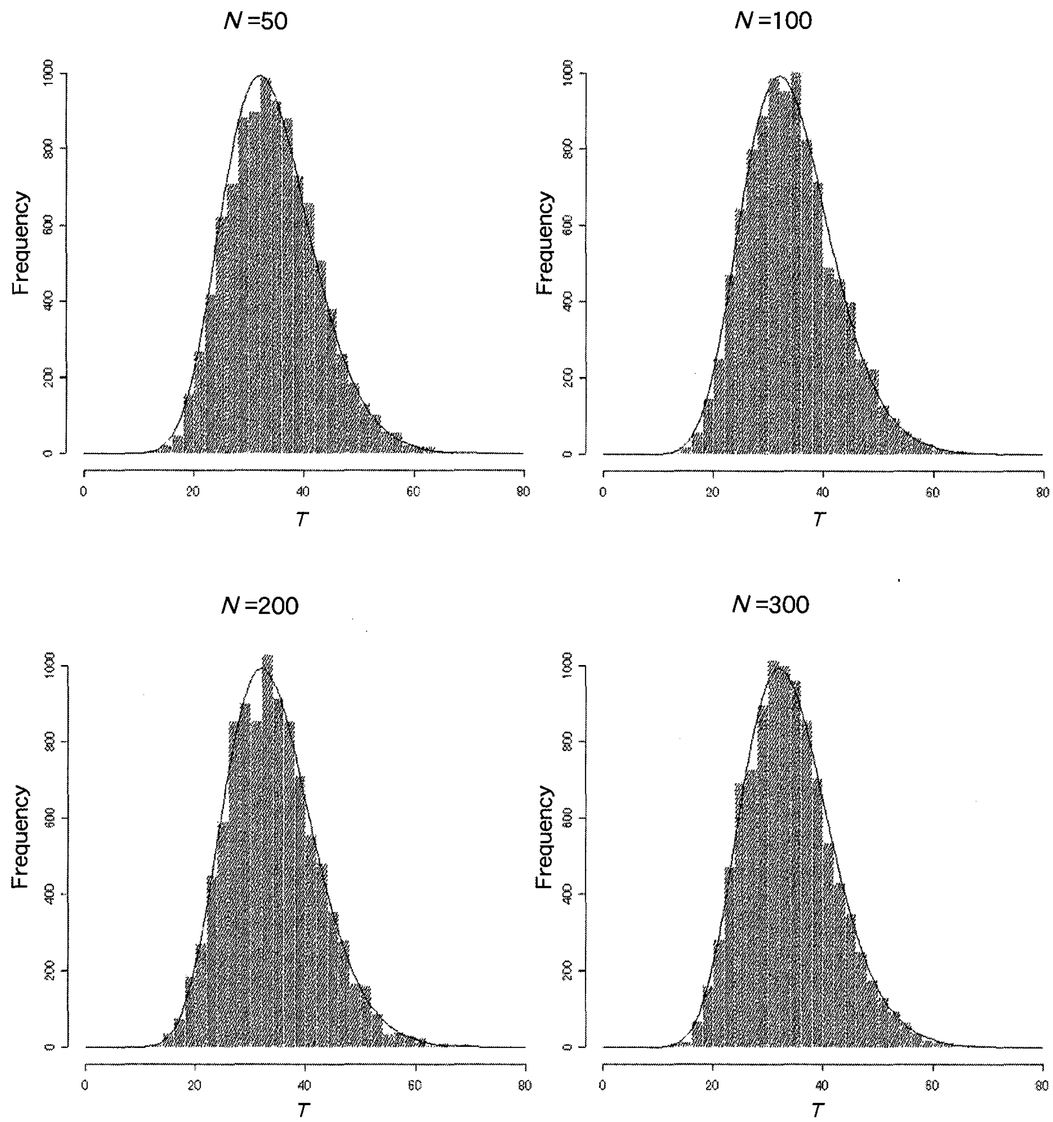

FIGURE 3 標本T^*のヒストグラムと，漸近理論によるχ^2分布
モデル3（10観測変数，2潜在変数）の場合を示す。FIGURE 2と見比べると，各条件で標本分布がχ^2分布へと近づいていることがわかる。

方法によってTをT^*に置き換えることで，カイ二乗分布への近似精度が大きく向上した（分布・分布の統計量の双方で確認）。

TABLE 6に示した標本数を変化させたときの各適合度指標ごとの最大値と最小値との比は、本来同じ真のモデルから発生したデータについて計算した指標であるので，小さい（1に近い）ほど指標がNに依存しないことを示し，良いことになる。同一のモデル・指標について最大値-最小値比を比較すると，どの適合度指標についても補正済みT^*を用いて計算した場合の方が，標本サイズに起因する変動が大幅に抑えられていた。このように，今回の方法はより広くSEMの適合度指標の補正のために応用できることが示唆された。

Paxton, Curran, Bollen, Kirby, & Chen（2001）が主張するとおり，心理学や社会学，政治学などにおけるSEMの応用では100以下の標本サイズで解析を行わなければならない事態もしばしば生じる。したがって，小標本下でも正しいモデル選択ができることは重要である。今回提案した方法は，こうした事態への有効な処方箋となりうる。

ただし，本研究の結果だけからBartlett補正の利用が正当化されるわけではない。まず，今回のシミュレーションの設定は真のモデルに限られたものであり，モデルが誤設定された場合のT^*の挙動については明ら

TABLE 6 適合度指標のばらつき（平均値）

(a) 補正前

	N=50	N=100	N=200	N=300	MAX/MIN
T	37.92893	35.87104	34.88246	34.47066	1.10032
AIC[1]	-26.07107	-28.12896	-29.11754	-29.52934	1.13265
BIC[2]	-131.61723	-133.67512	-134.66369	-135.07549	1.02628
CAIC[3]	-163.61723	-165.67512	-166.66369	-167.07549	1.02114
CAK[4]	0.42175	0.41141	0.40644	0.40438	1.04298
CK[4]	0.43528	0.42494	0.41997	0.41790	1.04158
RMSEA[5]	0.02656	0.02184	0.01995	0.01890	1.40529
Gamma hat[6]	0.99416	0.99620	0.99718	0.99759	1.00345
Mc[7]	0.98554	0.99060	0.99305	0.99406	1.00864

(b) 補正後

	N=50	N=100	N=200	N=300	MAX/MIN
T^*	34.49976	34.24599	34.09160	33.94949	1.01621
AIC	-29.50024	-29.75401	-29.90840	-30.05051	1.01865
BIC	-135.04640	-135.30016	-135.45455	-135.59667	1.00407
CAIC	-167.04640	-167.30016	-167.45455	-167.59667	1.00329
CAK	0.40452	0.40325	0.40247	0.40176	1.00688
CK	0.41805	0.41677	0.41600	0.41528	1.00666
RMSEA	0.01909	0.01832	0.01823	0.01776	1.07499
Gamma hat	0.99756	0.99782	0.99797	0.99811	1.00055
Mc	0.99398	0.99462	0.99500	0.99535	1.00137

モデル 3（10 観測変数，2 潜在変数）の場合を示す。MAX/MINは各適合度指標の最大値と最小値との比であり，これが小さいほど適合度指標がNに依存しないことを示す。
[1]Akaike (1973) [2]Schwarz (1978) [3]Bozdogan (1987) [4]Browne & Cudeck (1989)
[5]Steiger & Lind (1980) [6]Steiger (1989) [7]McDonald & Marsh (1990)

かではない。また，滑らかでないTの関数に対して，TをT^*で置き換える効果があるのかも不明である。こうした点について，今後研究を進める必要がある。

尤度比検定統計量Tに関する補正としては，Satorra-Bentler の補正がよく知られている (Satorra & Bentler, 1988)。これは観測変数が多変量正規分布にしたがっていないときの補正であり，実際にシミュレーションによってその非正規分布に対する頑健性が確かめられている (Hu et al., 1992)。一方今回我々が提案した補正は，多変量正規分布が仮定できた下でも依然存在する小標本時のゆがみの補正であり，Satorra-Bentler の補正と相補的な役割を果たすことが期待される。

また，ブートストラップ法を利用したアプローチも近年注目を集めている。

適合度指標をブートストラップで求める方法には，パラメトリック・ブートストラップとノンパラメトリック・ブートストラップの 2 つの方法がある。本研究で行ったシミュレーション実験は，シミュレーション 1 では既存の尤度比検定統計量の，またシミュレーション 2 では提案した Bartlett 補正済み尤度比検定統計量の，それぞれパラメトリック・ブートストラップによる経験分布の導出を行ったことになる。しかし，パラメトリック・ブートストラップは通常 SEM ソフトウェアでは実装されておらず，応用研究で利用することは困難である。

一方，ノンパラメトリック・ブートストラップ法は標本サイズが小さいときにはうまく機能しないことが指摘されている (Ichikawa & Konishi, 1995 ; Yung & Bentler, 1994)。さらに，ブートストラップ法に共通した難点として，乱数発生に伴う再現性が確実でないことが挙げられる。

本研究では 1 実験条件について 10,000 回の試行を行った。この繰り返し数は先行研究の標準から考えると非常に大きな数である。しかし計算機の発達した現代では，この程度の繰り返し数は計算時間的にも十分許容できる。標本変動を抑えられることや，ヒストグラムを描いて理論上の分布と比較できることといった利点も指摘できる。今回の我々の設定は，今後のシミュレーション研究における標準となっていくと考える。

なお，尤度比検定統計量が標本サイズに依存する，ということに関して議論するとき，2 つの異なった問題があることに注意が必要である。すなわち，一つは

χ^2検定の検定力がNに依存するという問題，もう一つはTの分布のχ^2_{df}分布からの乖離の大きさがNに依存するという問題である。前者の方は，小標本では検定力が小さいために誤ったモデルをも検出できなく，逆に大標本では検定力が過大になるためにわずかな標本ゆらぎをもモデルを棄却する要因として検出してしまう，ということである。これはそもそも統計的仮説検定一般が宿命的に有する問題であり，古くからよく認識されてきた (Satorra & Saris, 1985)。しかし，後者の問題，すなわちそもそも小標本ではTの標本統計量がχ^2_{df}分布にしたがわないという問題については，これまで認識が十分であったとは言えないだろう。

尤度比検定統計量にとどまらず，多くの適合度指標は，非心度やベースラインモデルとの差などの漸近理論に基づいて導出されている。したがって，標本サイズの小さいときには今回提案したような方法によって，漸近理論による分布からの乖離を補正する必要がある。既存のTを使ったブートストラップ法では，このような事態に対処することはできなかった。

本論文で提案したBartlett補正に関する研究が今後さらに発展し，SEMでのモデルの当てはまりを議論する際に幅広く使われるようになることを期待する。

引用文献

Akaike, H. 1973 Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*. Budapest : Akademiai Kiado. Pp. 267-281.

Anderson, J. C., & Gerbing, D. W. 1984 The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, **49**, 155-173.

Bartlett, M. S. 1937 Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London, Series A*, **160**, 268-282.

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., & Cudeck, R. 1989 Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, **24**, 445-455.

Cribari-Neto, F., & Cordeiro, G. M. 1996 On Bartlett and Bartlett-type corrections. *Econometric Reviews*, **15**, 339-367.

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. 2002 The noncentral chi-square distribution in misspecified structural equation models : Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research*, **37**, 1-36.

Curran, P. J., West, S. G., & Finch, J. F. 1996 The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, **1**, 16-29.

Ding, L., Velicer, F., & Harlow, L. L. 1995 Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, **2**, 119-144.

Fan, X., Thompson, B., & Wang, L. 1999 Effects of sample size, estimation methods, and model specification on structural equation modeling. *Structural Equation Modeling*, **6**, 56-83.

星野崇宏・岡田謙介・前田忠彦 2005 構造方程式モデリングにおける適合度指標とモデル改善について：展望とシミュレーション研究による新たな知見 行動計量学, **32**, 209-235. (Hoshino, T., Okada, K., & Maeda, T. 2005 Fit indices and model modification in structural equation modeling : A review and new findings. *Japanese Journal of Behaviormetrics*, **32**, 209-235.)

Hu, L., & Bentler, P. M. 1998 Fit indices in covariance structure modeling : Sensitivity to underparameterized model-misspecification. *Psychological Methods*, **3**, 424-453.

Hu, L., Bentler, P. M., & Kano, Y. 1992 Can test statistics in covariance structure analysis be trusted ? *Psychological Bulletin*, **112**, 351-362.

Ichikawa, M., & Konishi, S. 1995 Application of the bootstrap methods in factor analysis. *Psychometrika*, **60**, 77-93.

Jackson, D. L. 2001 Sample size and number of parameter estimation in maximum likelihood confirmatory factor analysis : A Monte Carlo investigation. *Structural Equation Modeling*, **8**, 205-223.

Jackson, D. L. 2003 Revisiting sample size and number of parameter estimates : Some support for the N : q hypothesis. *Structural Equation Modeling*, **10**, 128-141.

狩野 裕・三浦麻子 2002 グラフィカル多変量解析 増補版—目で見る共分散構造分析— 現代数学社

Kenny, D. A., & McCoach, D. B. 2003 Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, **10**, 333-351.

Lagos, B. M., & Morettin, P. A. 2004 Improvement of the likelihood ratio test statistic in ARMA models. *Journal of Time Series Analysis*, **25**, 83-101.

Marsh, H. W., Balla, J., & McDonald, R. P. 1988 Goodness-of-fit indexes in confirmatory factor analysis : The effect of sample size. *Psychological Bulletin*, **103**, 391-410.

Matsumoto, M., & Nishimura, T. 1998 Mersenne Twister : A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**, 3-30.

McDonald, R. P., & Ho, M-H. R. 2002 Principles and practice in reporting structural equation analyses. *Psychological Methods*, **7**, 64-82.

McDonald, R. P., & Marsh, H. W. 1990 Choosing a multivariate model : Noncentrality and goodness of fit. *Psychological Bulletin*, **107**, 247-255.

Moulton, L. H., Weissfeld, L. A., & Laurent, R. T. S. 1993 Bartlett correction factors in logistic regression models. *Computational Statistics and Data Analysis*, **15**, 1-11.

Mulaik, S. A. 1971 A note on some equations of confirmatory factor analysis. *Psychometrika*, **36**, 63-70.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. 2001 Monte Carlo experiments : Design and implementation. *Structural Equation Modeling*, **8**, 287-312.

Satorra, A., & Bentler, P. M. 1988 Scaling corrections for chi-square statistics in covariance structure analysis. In *1988 Proceedings of Business and Economics sections of the American Statistical Association*, 308-313.

Satorra, A., & Saris, W. 1985 Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, **50**, 83-90.

Schwarz, G. 1978 Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Steiger, J. H. 1989 A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, **5**, 173-180.

Steiger, J. H., & Lind, J. 1980 *Statistically-based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Wakaki, H., Eguchi, S., & Fujikoshi, Y. 1990 A class of tests for a general covariance structure. *Journal of Multivariate Analysis*, **32**, 313-325.

Yung, Y-F., & Bentler, P. M. 1994 Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, **47**, 63-84.

Zucker, D. M., Liberman, O., & Manor, O. 2000 Improved small sample inference in the mixed liner model : Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society, Series B*, **62**, 827-838.

(2005.9.16 受稿, '07.1.11 受理)

付　　録

ここでは，今回のシミュレーションで用いた確認的因子分析モデルにおける Bartlett 補正因子 m の具体的な計算手順を述べる。モデルの共分散構造を

$$\Sigma=\Lambda\Phi\Lambda^t+\Psi \tag{20}$$

で表現しよう。ただし，因子負荷量行列 Λ は母数ベクトル ξ によって $\Lambda(\xi)$ と構造化され，いくつかの要素が 0 に固定されている。

因子分析モデルでの推定すべき母数とは，各 Λ, Φ, Ψ の非零要素である。それらを一列に並べると，たとえばモデル 3 (FIGURE 1) の場合には

$$\xi=\{\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{41}, \lambda_{51}, \lambda_{62}, \lambda_{72}, \lambda_{82}, \lambda_{92}, \lambda_{10\,2}, \phi_{12}, \psi_{11}, \psi_{22}, \psi_{33}, \psi_{44}, \psi_{55}, \psi_{66}, \psi_{77}, \psi_{88}, \psi_{99}, \psi_{10\,10}\}^t \tag{21}$$

のようにベクトル表現できる（モデル 1, 2 でも要素の数が異なる以外は同様である）。

このとき Mulaik (1971) の結果にしたがうと，Σ の母数による 1 階微分はそれぞれ

$$\frac{\partial\Sigma}{\partial\lambda_{ij}}=\frac{\partial(\Lambda\Phi\Lambda^t+\Psi)}{\partial\lambda_{ij}}$$

$$= \Lambda\Phi 1_{ji} + 1_{ji}\Phi\Lambda^t (i=1, \ldots, p, j=1, \ldots, r) \tag{22}$$

$$\frac{\partial\Sigma}{\partial\phi_{gh}} = \Lambda(1_{gh}+1_{hg}-1_{gh}I1_{gh})\Lambda^t (g, h=1, \ldots, r) \tag{23}$$

となる。同様に2階微分は,

$$\frac{\partial\Sigma}{\partial\lambda_{ij}\partial\lambda_{ts}} = 1_{ti}\phi_{sj} + 1_{it}\phi_{js} (i, t=1, \ldots, p, \quad j, s=1, \ldots, r) \tag{24}$$

$$\frac{\partial^2\Sigma}{\partial\lambda_{ij}\partial\phi_{gh}} = \Lambda(1_{gh}+1_{hg}-1_{gh}I1_{gh})1_{ji} + 1_{ij}(1_{gh}+1_{hg}-1_{gh}I1_{gh})\Lambda^t \quad (i=1, \ldots, p, \quad j, g, h=1, \ldots, r) \tag{25}$$

$$\frac{\partial^2\Sigma}{\partial\lambda_{ij}\partial\psi_{tt}} = 0 (i, t=1, \ldots, p, \quad j=1, \ldots, r) \tag{26}$$

$$\frac{\partial^2\Sigma}{\partial\phi_{gh}\partial\phi_{uv}} = 0 (g, h, u, v=1, \ldots, r) \tag{27}$$

$$\frac{\partial^2\Sigma}{\partial\phi_{gh}\partial\psi_{tt}} = 0 (t=1, \ldots, p, \quad g, h=1, \ldots, r) \tag{28}$$

$$\frac{\partial^2\Sigma}{\partial\psi_{ii}\partial\psi_{tt}} = 0 (i, t=1, \ldots, p) \tag{29}$$

となる。ただし,1_{ij}はi,j要素だけが1で他の要素はすべて0の行列を表すこととする。

実際の(18),(19)式におけるJ_a, J_{ab}は,Σ^{-1}の1階・2階微分である。これは,逆行列の微分の性質である

$$\frac{\partial\Sigma^{-1}}{\partial\xi^a} = -\Sigma^{-1}\frac{\partial\Sigma}{\partial\xi^a}\Sigma^{-1} \tag{30}$$

を用いると,それぞれ

$$J_a = -\Sigma^{-1/2}\frac{\partial\Sigma}{\partial\xi^a}\Sigma^{-1/2} \tag{31}$$

$$J_{ab} = \Sigma^{-1/2}\left[\frac{\partial\Sigma}{\partial\xi^a}\Sigma^{-1}\frac{\partial\Sigma}{\partial\xi^b} - \frac{\partial^2\Sigma}{\partial\xi^a\partial\xi^b} + \frac{\partial\Sigma}{\partial\xi^b}\Sigma^{-1}\frac{\partial\Sigma}{\partial\xi^a}\right]\Sigma^{-1/2} \tag{32}$$

によって計算できる。J_a, J_{ab}が得られれば,あとは(9)〜(19)式を用いてBartlett補正因子mを得ることができる。以上のプロセスは,今回用いたSAS/IML等の行列計算言語を用いれば容易に実装できる。

Bartlett Correction of Test Statistics in Structural Equation Modeling

Kensuke Okada (Graduate School of Arts and Sciences, University of Tokyo ; Japan Society for the Promotion of Science), Takahiro Hoshino (Graduate School of Arts and Sciences, University of Tokyo ; Komaba Educational Development, School of Arts and Sciences, University of Tokyo) and Kazuo Shigemasu (Graduate School of Arts and Sciences, University of Tokyo) Japanese Journal of Educational Psychology, 2007, 55, 382−392

Model selection is one of the most important steps in the application of structural equation modeling (SEM). In this process, the likelihood ratio test statistic, T, is a commonly employed index. Hypothesis testing can be performed on the basis that T asymptotically follows the χ^2 distribution. Various fit indices have been proposed, most of which are based on the assumption that T asymptotically follows the χ^2 distribution. When the size of the sample is small, however, the distribution of T deviates considerably from the χ^2 distribution. This problem, especially pronounced when there is a large number of indicators per factor, is serious because it violates the theoretical justification of utilizing for model selection not only T, but also many other fit indices. In the present article, we propose a Bartlett correction of T to improve its approximation to the χ^2 distribution. When the efficacy of our method was evaluated by Monte Carlo simulation, the results showed that this method was superior to the current standard.

Key Words : structural equation modeling, likelihood ratio test, χ^2 value, Bartlett correction, goodness of fit