資料

相関構造分析におけるフィッシャーの z 変換を 利用したパラメータ推定法

豊田 秀樹*

A LEAST-SQUARES SOLUTION FOR CORRELATION STRUCTURE ANALYSIS WITH FISHER'S Z-TRANSFORMATION

Hideki TOYODA

The purpose of this study is to propose a least-squares method for parameter estimation using Fisher's z -transformation in correlation structure analysis. An advantage of this method in comparison with the unweighted least squares method (ULS) is that the residuals are normally distributed and their variances are homogeneous. A numerical example is given for factor analysis by using the data of Spearman (1904). Through the results of computer simulations in higher-order factor analysis, this method is shown to have less significant errors of parameter estimates than the ULS.

Key words: Fisher's z -transformation, least-squares method, correlation structure analysis, computer simulation, higher-order factor analysis.

目 的

本研究の目的は、相関構造分析におけるフィッシャーの z 変換を利用した新しいパラメータの推定法を提案することである。

相関構造分析 (Jöreskog, 1978, 小笠原, 1981) におけるパラメータ推定は最小2乗法・一般化最小2乗法・最尤推定法のうち何れかを用いるのが一般的である。標本数 N が大きい場合は漸近的に最良である最尤推定法を用いることが多く、標本数 N が小さい場合は(たとえば $N \leq 100$)計算が容易で、不適解・非収束の生じる確率が低く、推定誤差も小さい最小2乗法 (Unweighted Least Squares method, 以下 ULS と略記) を用いることが多い(丘本, 1986)。

ULS を用いた相関構造分析のパラメータ推定とは、変量 i と変量 j の母相関係数 ρ_{ij} がパラメータ θ の関数

$$\rho_{ij} = f(\theta) \quad (1)$$

で表わされているとき、その制限の下で、

$$FULS = \sum_i \sum_j^n (r_{ij} - \rho_{ij})^2 \quad (2)$$

$$= t_r(R - \Sigma)^2 \quad (3)$$

を最小にする θ を推定することである。ここで n は分析対象となっている変数の数、 r_{ij} は標本相関係数、 R は標本相関行列、 Σ は母相関行列を表わす。ところが分析対象となる変数が多変量正規分布に従っていると仮定すると、 r_{ij} に対する ρ_{ij} の最小2乗法は $\langle 1 \rangle \rho$ が0から遠ざかるにつれて r の標本分布は歪みが大きくなり誤差の分布が正規分布に従わない。 $\langle 2 \rangle \rho$ が0から遠い場合(± 1 に近い場合)の r の標本分布の分散は0に近い場合の分散より小さいので誤差の分布の分散が等しくないという性質がある。

この問題を解決するために、岡太・今泉(1977)は因子分析における因子負荷行列を推定する際に r_{ij} と ρ_{ij} に Hotelling (1953) の z 変換を施し、変換後の両者の差の2乗和を最小にする方法を提案している。本研究では、パラメータ推定の対象を因子分析を含む全ての相関構造分析モデルに拡張し、 r_{ij} と ρ_{ij} にフィッシャーの z 変換を施し、

$$\zeta_{ij} = \frac{1}{2} \log \frac{1 + \rho_{ij}}{1 - \rho_{ij}} \quad (i \neq j) \quad (4)$$

* 東京大学 (The University of Tokyo)

$$z_{ij} = \frac{1}{2} \log \frac{1+r_{ij}}{1-r_{ij}} \quad (i \neq j) \quad (5)$$

z_{ij} に対する ζ_{ij} の最小2乗法, つまり

$$F_{FLS} = \sum_{i \neq j}^n \sum_{j \neq i}^n (z_{ij} - \zeta_{ij})^2 \quad (6)$$

を最小化することによって θ を推定する方法を提案する。 z の標本分布は $N > 50$ で ρ の値によらず (± 1 に近くとも) 正規分布によく近似し, 等分散 ($(N-3)^{-1}$) とみなすことができるので, z 変換を行うことによって比較的小さい標本数で近似的に誤差分布が正規分布に従い, その分散が等しい状態で θ の最小2乗推定値を得ることができる。本研究では本推定法の数値例を示すと共に, 数値シミュレーションを行い ULS との推定精度の比較を行う。

方 法

フィッシャーの z 変換を利用した最小2乗推定法 (Fisher's z -transformed Least Squares method, 以下 FLS と略記) を行う際に必要となる (6) 式の θ による偏導関数の一般式は

$$\frac{\partial F_{FLS}}{\partial \theta} = T' \left[T \frac{\partial \Sigma}{\partial \theta} \right] \quad (7)$$

である。 T' は

$$t_{ij} \begin{cases} i=j \text{ のときは } 0 \\ i \neq j \text{ のときは } \frac{\zeta_{ij} - z_{ij}}{1 - \rho_{ij}^2} \end{cases}$$

を要素とする行列である。推定値を得るには (7) 式の値を 0 とおき, θ についての同時方程式を解く。

従来の方法との形式的な比較

(7) 式における T は目的関数を Σ で微分したものであるから, T を

$$T = (\Sigma - R)$$

$$T = R^{-1}(\Sigma - R)R^{-1}$$

$$T = \Sigma^{-1}(\Sigma - R)\Sigma^{-1}$$

に置き換え, θ を推定すると, それぞれ ULS・一般化最小2乗法・最尤法による解となる。換言するならば, 従来の推定法による相関構造分析の計算プログラムの T に相当する箇所のみを書き直すことにより本推定法を実行することができる。

数値例

Spearman (1904) の 6 変量のデータに対して FLS を用いて因子分析を行い, 具体的な数値例を示す。推定すべきパラメータは因子負荷行列 A であり, 母相関係数は

$$\rho_{ij} = \sum_{k=1}^m a_{ik} a_{jk} \quad (i \neq j) \quad (8)$$

$$\rho_{ii} = \sum_{k=1}^m a_{ik}^2 + d_i^2 = 1 \quad (9)$$

の様に A の要素で表わされる (1) 式に相当)。 m は因子数であり, ここでは $m=1$ とする。独自性 d_i は A の関数として得られるので自由パラメータではない。そこで (8) 式の制約の下で (6) 式を偏微分すると

$$\frac{\partial F_{FLS}}{\partial A} = TA \quad (10)$$

となる (1) 式に相当)。推定値を得るためには最急降下法による以下の計算手順を用いた。

[1] R に対して主成分分析を行い, 第1主成分を A の初期値 A_1 とする。 $l=1$

[2] (10) 式を用い A_l における目的関数の勾配ベクトル g_l を求める。

[3] $E = \text{MAX}(\text{ABS}(g_l)) < .0001$ なら収束と判定し, [6] へ。

[4] $A_{l+1} = A_l - \beta g_l$ ただし $\beta = 1/300$

[5] $l = l+1$ として [2] へ戻る。

[6] 計算終了。

以上の計算により得られた FLS による推定値は

$$\hat{A}'_{FLS} = \{.955 .873 .805 .750 .683 .649\}$$

であった。ULS による推定値は

$$\hat{A}'_{ULS} = \{.958 .882 .803 .750 .673 .646\}$$

である。FLS は, 値の高い推定値は ULS より低めに, 値の低い推定値は高めに推定している。

精度の検討

真の構造がわかっている人工的なデータに高次因子分析を行い, FLS と ULS の推定精度を比較検討する。

TABLE 1 データの発生に用いる計画行列

	η_1	η_2	η_3	η_4
y_1	.9	.0	.0	.0
y_2	.9	.0	.0	.0
y_3	.9	.0	.0	.0
y_4	.9	.0	.0	.0
y_5	.0	.9	.0	.0
y_6	.0	.9	.0	.0
y_7	.0	.9	.0	.0
y_8	.0	.9	.0	.0
y_9	.0	.0	.9	.0
y_{10}	.0	.0	.9	.0
y_{11}	.0	.0	.9	.0
y_{12}	.0	.0	.9	.0
y_{13}	.0	.0	.0	.9
y_{14}	.0	.0	.0	.9
y_{15}	.0	.0	.0	.9
y_{16}	.0	.0	.0	.9

データの発生

分析対象となる変量数 $n=16$ のデータ行列 $Y(16 \times N)$ を

$$Y = A\eta + \sigma \quad (11)$$

によって発生させる。因子負荷行列 $A(16 \times 4)$ は TABLE 1 で示される計画行列、誤差変量 $\sigma(16 \times N)$ は第 i 行が平均 0・分散 $1 - \sum_{j=1}^4 a_{ij}^2$ に従う正規乱数である。内生的な潜在変量 $\eta(4 \times N)$ は

$$\eta = \Gamma \xi + \varepsilon \quad (12)$$

によって発生させる。 $\Gamma' = \{.7 \ .7 \ .7 \ .7\}$ は 2 次因子の因子負荷行列、外生的な潜在変量 $\xi(1 \times N)$ は平均 0・分散 1 に従う正規乱数、誤差変量 $\varepsilon(4 \times N)$ は平均 0・分散 .51 (= $1 - .7^2$) に従う正規乱数である。

本データは、16 の下位テストから構成される知能検査に対して 2 次因子分析を行い、単純構造を示す 4 つの知能因子（1 次因子、たとえば計算因子・記憶因子・言語因子・推理因子）を仮定し、知能因子が下位テストに与える影響の大きさを調べると共に、それ等の知能因子の背後に 1 つの一般知能因子（2 次因子）を仮定し、一般知能因子が各知能因子に与える影響の大きさを調べる場面を想定して設定した。TABLE 1 は、第 1 の知能因子が変量 1~4 に、第 2 の知能因子が変量 5~8 に、第 3 の知能因子が変量 9~12 に、第 4 の知能因子が変量 13~16 に影響を与えることを示し、 Γ は一般知能因子が 4 つの知能因子に影響を与えることを示している。

パラメータの推定

推定すべきパラメータはベクトル Γ の要素の全て、つまり $\gamma_1 \cdot \gamma_2 \cdot \gamma_3 \cdot \gamma_4$ の 4 個と、行列 A の要素 $a_{11} \cdot a_{21} \cdot a_{31} \cdot a_{41} \cdot a_{52} \cdot a_{62} \cdot a_{72} \cdot a_{82} \cdot a_{93} \cdot a_{103} \cdot a_{113} \cdot a_{123} \cdot a_{134} \cdot a_{144} \cdot a_{154} \cdot a_{164}$ の 16 個の合計 20 個であり、これを順に $\Theta = \{\theta_1, \theta_2 \sim \theta_{20}\}$ と表わす。全て自由パラメータとする。行列 A のその他の要素は固定パラメータとし、推定せず、値は 0 とする。母相関行列は

$$\Sigma = A(\Gamma\Gamma' + \psi)A' + \theta\sigma \quad (13)$$

と制約される (11 式に相当)。 ψ と $\theta\sigma$ は

$$\psi = \text{diag}(I - \Gamma\Gamma') \quad (14)$$

$$\theta\sigma = \text{diag}(I - A(\Gamma\Gamma' + \psi)A') \quad (15)$$

の様に Γ と A の関数として得られ、自由パラメータではないから (13 式) の制約の下で (6 式) を Γ と A で偏微分すると

$$\frac{\partial F_{\text{FLS}}}{\partial \Gamma} = A'TA\Gamma \quad (16)$$

$$\frac{\partial F_{\text{FLS}}}{\partial A} = TA(\Gamma\Gamma' + \psi) \quad (17)$$

となる (17 式に相当)。 Θ における勾配は (16)・(17 式) の対応する要素を順に並べることによって得る。推定値を得るためには最急降下法による以下の計算手順を用いた。

[1] Γ と A の初期値には真値を用いる。 Γ の要素である $\theta_1 \sim \theta_4$ は .7, A の要素である $\theta_5 \sim \theta_{20}$ は .9 である。 $l=1$

[2] $l \geq 1500$ なら非収束と判定し、[8] へ。

[3] (16)・(17 式) を用い θ_l における目的関数の勾配ベクトル g_l を求める。

[4] $E = \text{MAX}(\text{ABS}(g_l)) < .0001$ なら収束と判定し、[8] へ。

[5] $\theta_{l+1} = \theta_l - \beta g_l$ ただし $\beta = 1/100$

[6] θ_{l+1} の要素の中で $-\alpha$ 以下のものは $-\alpha$ に α 以上のものは α にする。 $\alpha = .9999$

[7] $l = l + 1$ [2] へ戻る。

[8] 計算終了

ULS によってパラメータを推定する場合は (16)・(17 式) 中の T を

$$T = (\Sigma - R) \quad (18)$$

とする。

シミュレーション計画

人工データを $N=50 \cdot 100 \cdot 200 \cdot 400$ についてそれぞれ 100 回ずつ発生させ、ULS・FLS で Θ の推定を行う。2 次因子の因子負荷の真値が .7 である推定量を M とし、1 次因子の因子負荷の真値が .9 である推定量を H とすると 1 回の収束計算で M の推定値は 4 個 ($\theta_1 \sim \theta_4$)、 H の推定精度は 16 個 ($\theta_5 \sim \theta_{20}$) 得ることができる。推定精度は M と H について真値の回りの 2 次のモーメントの平方根で評価し、収束したケース数 c と共に TABLE 2 に示

TABLE 2 推定精度と収束したケース数

N	推 定 精 度				収束したケース数	
	H		M		MAX=100	
	ULS	FLS	ULS	FLS	ULS	FLS
50	.050	.037	.104	.106	79	98
100	.036	.024	.071	.071	99	100
200	.025	.017	.055	.055	100	100
400	.018	.012	.037	.037	100	100

す。推定精度の計算のために使われる推定値の数は M については $c \times 4$ 個であり、 H については $c \times 16$ 個である。たとえば非収束のケースが 1 つもない場合は、それぞれ 400 個・1600 個となる。非収束のケース数は $100 - c$ 個である。

結果と考察

H の推定精度は $N=50 \cdot 100 \cdot 200 \cdot 400$ の全ての条件で FLS が成績が良く、ULS の推定精度に対する比は順に 74%・67%・68%・66% であり、26%~34% 推定誤差が小さくなっている。また FLS の推定精度は 2 倍の被

験者数の ULS の推定精度にほぼ等しい。

M の推定精度は $N=50$ の条件では ULS の成績が良く、 $N=100 \cdot 200 \cdot 400$ の条件では差がない。しかし $N=50$ の条件での FLS の ULS に対する推定精度の比は 102% であり、 H と比べてその差は小さい。

非収束のケースは、 $N=50$ の条件では ULS に 21 ケース、FLS に 2 ケースあり、 $N=100$ の条件では ULS に 1 ケースある。収束ケース数は FLS が成績が良い。

結 論

以上のことから、前節で設定した分析条件では $N=100 \cdot 200 \cdot 400$ の条件では FLS を用いるべきであり、 $N=50$ の条件でも H の推定精度の改善に比べて M の推定精度の損失はわずかであるので FLS を用いるのが望ましいといえる。

本研究で提案したパラメータ推定法は因子分析・高因子分析に限らず LISREL (Jöreskog & Sörbom, 1984) や COSAN (McDonald, 1978) で扱われる相関構造分析と呼ばれるモデル群(たとえば、パス解析・シンプレックス構造分析・多重特性多重法行列の分析・多重指標分析など)に適用することが可能な一般的な方法である。

付 記

本研究の計算は東京大学大型計算機センターを利用して行った。

引用文献

Jöreskog, K.G. 1978 Structural Analysis of Covariance and Correlation Matrices. *Psychometrika*,

43, 443—447.

Jöreskog, K.G. & Sörbom, D. 1984 LISREL VI User's Guide: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood, Scientific Software, Inc.

Hotelling, H. 1953 New Light on the Correlation Coefficient and its Transforms. *Journal of the Royal Statistical Society., Series B*, 15, 193—225.

McDonald, R.P. 1978 A Simple Comprehensive Model for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, 31, 59—72.

小笠原春彦 1981 検査再検査事態の共分散および相関構造分析 鉄道労働科学, 35, 21—31.

岡太彬訓・今泉 忠 1977 因子分析の一方法 行動計量学会・大会論集, 134—137.

丘本 正 1986 因子分析の基礎 日科技連

Spearman, C. 1904 "General Intelligence," Objectively Determined and Measured. *American Journal of Psychology*. 15, 201—293.

謝 辞

暖かく見守りながら的確な指導をして下さる芝 祐順先生(東京大学教育学部教授)、有益な助言を下さった村上隆先生(名古屋大学教育学部助教授)、南風原朝和先生(新潟大学教育学部助教授)、アブストラクトの書き方を指導して下さい渡部 洋先生(東京大学教育学部助教授)に深謝します。

(1988年5月2日受稿)