

# 社会システム分析のための統合化プログラム 1.2

## — 共分散構造分析（中間報告） —

福井正康，陳文龍\*，王嘉琦\*

福山平成大学経営学部経営学科

\*福山平成大学大学院経営学研究科経営情報学専攻

### 概要

我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は統計の分野でよく利用されている共分散構造分析について新しくプログラムを作成した。しかしこのプログラムはアルゴリズムやユーザーインターフェイスなど取り組むべき課題が多く、この論文は中間段階までの報告である。

### キーワード

College Analysis, 社会システム分析, 統計, OR, 意思決定, 共分散構造分析

URL: <http://www.heisei-u.ac.jp/ba/fukui/>

## 1. はじめに

共分散構造分析はこれまでの多変量解析の手法を包含する優れた分析手法であり、第2世代の多変量解析と呼ばれることがある。利用者は観測される変数や内部に潜在する直接観測されない変数間の関係を記述するネットワーク型の統計モデルを作成し、そのモデルと観測値とで各変数間の直接的な影響力を推測する。統計モデルはこれまでの多変量解析に比べて複雑な構造を記述可能で、その中に重回帰分析や因子分析などの構造を複数含めることができる<sup>1)</sup>。

我々は、社会システム分析教育用ソフトウェア College Analysis の機能拡張のため、新たに共分散構造分析のプログラムを追加することにした。College Analysis には、集計や検定を扱う基本統計や多変量解析のプログラムが含まれているが、共分散構造分析の重要性を考えるとこの分析手法は避けて通ることができないものと思われる。しかしこの分析のプログラムは分量が多く、グラフィックでの構造図入力や複雑なアルゴリズムなど取り組むべき課題も多い。そこで今回はグラフィック部分を除き、アルゴリズムも従来のものを用いて、中間段階までを公表することにした。

共分散構造分析は変数間の関係を構造方程式と呼ばれる線形の式で与え、変数間の影響の強さを表すパラメータの値は観測変数の共分散行列から推定する。その際一般にパラメータ数は共分散行列の独立な成分数と異なるため、パラメータの値は厳密には決まらない。パラメータの推定にはある評価関数を用いて、これを最小化するような方法を考える。この評価関数の選び方によって、推定値の導出にはいくつかの方法がある。その中で最もよく利用されるのが最小2乗法や最尤法である。

我々のプログラムの最大の問題はこの最小化のアルゴリズムにある。最小2乗法では評価関数はパラメータについて高次の多項式となり、最尤法では非線形の長大な数式となる。これらの数式の最小化問題は非常に繊細で、これまでの Newton-Raphson 法では限界があるし、計算の手順によっては時間が膨大にかかる場合もある。今回のプログラムではこのアルゴリズムに Levenberg-Marquart 法を応用したものを採用し、計算の方法もできる限り時間的な無駄を省くように考え、簡単なモデルであれば何とか辛抱できる時間で計算できるところまで来た。しかし、Amosなどのプログラムでは 1985 年以降発展してきたマルコフ連鎖モンテカルロ法などが採用されており、短時間で比較的安定な解を求めるができるようになってきている。我々も今後このようなアルゴリズムを使ったプログラムに変更していく必要があるが、現段階では 2 つのアルゴリズムの違いを実感しておくのも今後のための教訓となる。

この論文では非常に簡単なモデルから、多少複雑な（まだ実用モデルの段階ではないが）モデルまで我々のプログラムと Amos の結果とを比較してみた。その中で我々のプログラムだけでなく、Amos の利用上の注意点も少しだけ見えてきた。これらの問題についても例を見ながら考えて行く。

## 2. 共分散構造分析の基礎理論

### 2.1 モデルの構造と方程式

ここでは図 2.1 の構造モデルを例として共分散構造分析の理論の説明をする。

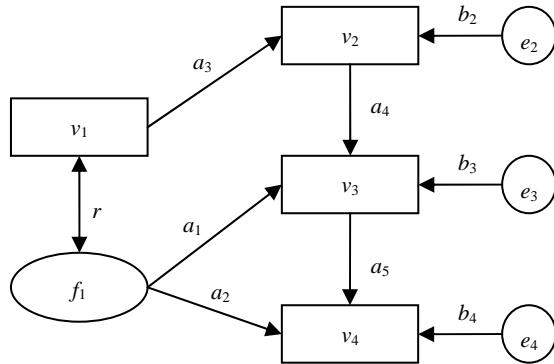


図 2.1 構造モデル

四角や楕円や円で表される量はモデルに含まれる変数で、形によりその意味するところが異なり、それぞれラベルが付けられている。矢印は因果関係を表すパラメータで、これにもラベルが付けられている。また、双方向の矢印は相関を表すパラメータである。

このモデルをよく利用される影響行列の形で表現すると表 2.1 のようになる。左側の変数が始点、上側の変数が終点である。

表 2.1 構造モデルの影響行列

	$f_1$	$v_1$	$v_2$	$v_3$	$v_4$	$e_2$	$e_3$	$e_4$
$f_1$		$r$		$a_1$	$a_2$			
$v_1$	$r$		$a_3$					
$v_2$				$a_4$				
$v_3$					$a_5$			
$v_4$								
$e_2$			$b_2$					
$e_3$				$b_3$				
$e_4$					$b_4$			

変数は通常、いくつかの視点から以下のように分けられる。

#### 観測変数と潜在変数

観測変数とは実測値の分かっている変数であり、図 2.1 の構造モデルでは  $v_1, v_2, v_3, v_4$  などの変数が

これに相当し、構造図では四角形で表現される。潜在変数とは直接には観測されない変数で、因子分析の因子や誤差などがこれに当り、構造図では楕円や円で表現される。図 2.1 の例では  $f_1, e_2, e_3, e_4$  などの変数である。ここでは  $f_1$  が因子変数、 $e_2, e_3, e_4$  が誤差変数である。特に因子変数は楕円、誤差変数は円（または円なし）で表現される場合がある。

### 外生変数と内生変数

外生変数は構造モデルで相関を除いてどこからも影響を受けない（片側矢印が入らない）変数で、図 2.1 の構造モデルでは  $v_1, f_1, e_2, e_3, e_4$  がこれに当る。内生変数はそれ以外の変数で  $v_2, v_3, v_4$  などである。

### 構造変数と誤差変数

構造変数とは後に述べるモデルの構成要素に使われる変数で、図 2.1 の構造モデルでは  $f_1, v_1, v_2, v_3, v_4$  などがこれに当る。誤差変数とはモデルでは説明できないゆらぎの成分を表すもので  $e_2, e_3, e_4$  がこれに当る。

これらの変数の関係は構造方程式と呼ばれる式で表現される。図 2.1 の構造モデルでは以下となる。

$$\begin{aligned}v_2 &= a_3 v_1 + b_2 e_2 \\v_3 &= a_1 f_1 + a_4 v_2 + b_3 e_3 \\v_4 &= a_2 f_1 + a_5 v_3 + b_4 e_4\end{aligned}$$

この方程式の左辺を構造変数に拡張し、以下のような式を考える

$$\begin{aligned}f_1 &= f_1 \\v_1 &= v_1 \\v_2 &= a_3 v_1 + b_2 e_2 \\v_3 &= a_1 f_1 + a_4 v_2 + b_3 e_3 \\v_4 &= a_2 f_1 + a_5 v_3 + b_4 e_4\end{aligned}$$

構造方程式の左辺には構造変数と呼ばれる変数を取るが、そのうちの内生変数は必ず誤差変数からの影響を受けるようにする。上の構造方程式を行列表示すると以下のようになる。

$$\begin{pmatrix} f_1 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a_3 & 0 & 0 & 0 \\ a_1 & 0 & a_4 & 0 & 0 \\ a_2 & 0 & 0 & a_5 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & b_2 & 0 & 0 \\ 0 & 0 & 0 & b_3 & 0 \\ 0 & 0 & 0 & 0 & b_4 \end{pmatrix} \begin{pmatrix} f_1 \\ v_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

今、以下のように定義すると、

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & a_3 & 0 & 0 & 0 \\ a_1 & 0 & a_4 & 0 & 0 \\ a_2 & 0 & 0 & a_5 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & b_2 & 0 & 0 \\ 0 & 0 & 0 & b_3 & 0 \\ 0 & 0 & 0 & 0 & b_4 \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} f_1 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} f_1 \\ v_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

構造方程式は(1)式のように表すことができる。

$$\mathbf{t} = \mathbf{At} + \mathbf{Bh} \quad (1)$$

ここに  $\mathbf{t}$  は構造変数からなるベクトル、 $\mathbf{h}$  は外生変数からなるベクトルである。またパラメータは行列  $\mathbf{A}$  と  $\mathbf{B}$  に含まれる。

構造方程式は以下のように変形できる。

$$\mathbf{t} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{Bh} \quad (2)$$

ここでベクトル  $\mathbf{t}$  のうち観測変数に注目し、観測変数で作られたベクトル  $\mathbf{v}$  とそれを取り出す行列  $\mathbf{G}$  を以下のように定義する。

$$\mathbf{v} = \mathbf{Gt}, \quad \text{ここに} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

この関係を用いると、上式は(3)式のように変形される。

$$\mathbf{v} = \mathbf{G}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{Bh} \quad (3)$$

次に観測変数  $\mathbf{v}$  および外生変数  $\mathbf{h}$  の共分散行列を考える。簡単のため潜在変数は平均が 0、分散が 1 になるように標準化されているものとする。変数  $\mathbf{v}$  の共分散行列を  $E(\mathbf{vv}')$ 、変数  $\mathbf{h}$  の共分散行列を  $E(\mathbf{hh}')$  とするとそれらの関係は(4)式のようになる。

$$E(\mathbf{vv}') = \mathbf{G}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} E(\mathbf{hh}') \mathbf{B}' (\mathbf{I} - \mathbf{A})'^{-1} \mathbf{G}' \quad (4)$$

実際の計算では  $E(\mathbf{vv}')$  を標本から得られた不偏共分散行列(共分散行列の不偏推定量)で置き換え、 $E(\mathbf{hh}')$  についても観測変数部分は不偏共分散行列、潜在変数部分は分散を 1、共分散には必要に応じて共分散を表すパラメータを設定する。図 2.1 の構造モデルの場合は、潜在変数間または外生の観測変数と潜在変数間で、 $f_1$  と  $v_1$  の間だけに共分散  $r$  を仮定しているので、以下の形となる。

$$E(\mathbf{v}\mathbf{v}') \Rightarrow \mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{23} & u_{33} & u_{34} \\ u_{41} & u_{24} & u_{34} & u_{44} \end{pmatrix}, \quad E(\mathbf{h}\mathbf{h}') \Rightarrow \mathbf{H} = \begin{pmatrix} 1 & r & 0 & 0 & 0 \\ r & u_{11} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

ここで  $\mathbf{U}$  は不偏共分散行列であるが、標準化したデータの場合には相關行列となる。これを用いて(4)式を書き換えると以下のようになる。

$$\mathbf{G}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} \mathbf{H} \mathbf{B}' (\mathbf{I} - \mathbf{A})'^{-1} \mathbf{G}' = \mathbf{U} \quad (5)$$

これは観測値とパラメータを結びつける方程式である。この方程式を丁度方程式と呼び、一意的な解が存在する場合、その解を丁度解と呼ぶ。しかし丁度解が存在する場合はまれで、一般には解が不定になっていたり、不能になっていたりする。解が不定になっている場合をパラメータは識別不能という。不能になっている場合は最適近似解を求める。最適近似解を求める方法はいくつかあるが、ここでは主に利用される 2 つの方法について紹介する。

## 2.2 パラメータの推定

パラメータの推定は方程式の近似解を求めるための評価関数を作り、それを最小化する方法が採られるが、この節ではよく利用される 2 つの評価関数について説明する。

### 最小 2 乗法

方程式(5)の左辺と右辺の差の 2 乗和を最小化するために以下の評価関数を考える。

$$f_{MS}(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=i}^n (\Sigma(\boldsymbol{\theta})_{ij} - u_{ij})^2$$

ここに  $\boldsymbol{\theta}$  はパラメータを総称したものであり、 $n$  は観測変数の数、 $\Sigma(\boldsymbol{\theta})$  は以下のように(5)式の左辺を表す。

$$\Sigma(\boldsymbol{\theta}) = \mathbf{G}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} \mathbf{H} \mathbf{B}' (\mathbf{I} - \mathbf{A})'^{-1} \mathbf{G}'$$

丁度解の場合  $f(\boldsymbol{\theta})$  の値は 0 である。

### 最尤法

我々はまず観測値を与える確率変数  $\mathbf{x}_\lambda$  ( $\lambda = 1, 2, \dots, N$ ) がそれぞれ独立に  $n$  変量正規分布に従うと考える。共分散行列を  $\Sigma(\boldsymbol{\theta})$  とすると、 $\mathbf{x}_\lambda$  の確率密度関数は以下で与えられる。

$$f(\mathbf{x}_\lambda | \boldsymbol{\theta}) = (2\pi)^{-p/2} |\Sigma(\boldsymbol{\theta})|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{x}_\lambda - \boldsymbol{\mu})' \Sigma(\boldsymbol{\theta})^{-1} (\mathbf{x}_\lambda - \boldsymbol{\mu}) \right]$$

$N$  回の独立な観測に関する確率密度関数は以下で与えられる。

$$f(\mathbf{x} | \boldsymbol{\theta}) = \prod_{\lambda=1}^N f(\boldsymbol{\theta} | \mathbf{x}_\lambda)$$

最尤法ではこの確率密度関数に実測値を代入した尤度関数  $f(\boldsymbol{\theta})$  を最大化するようにパラメータを決定する。実際には計算の簡単化のため、尤度関数を対数変換した対数尤度関数の符号を変えたものを最小化する。符号を変えた対数尤度関数は以下で与えられる。

$$\begin{aligned} -\log f(\boldsymbol{\theta}) &= -\sum_{\lambda=1}^N \log f(\boldsymbol{\theta} | \mathbf{x}_\lambda) \\ &= \frac{1}{2} \sum_{\lambda=1}^N (\mathbf{x}_\lambda - \boldsymbol{\mu})' \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\mathbf{x}_\lambda - \boldsymbol{\mu}) - \frac{N}{2} \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}| + const. \\ &= \frac{N-1}{2} \left( \text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{U}) - \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}| \right) + const. \end{aligned}$$

但し、

$$\mathbf{U} = \frac{1}{N-1} \sum_{\lambda=1}^N (\mathbf{x}_\lambda - \boldsymbol{\mu})(\mathbf{x}_\lambda - \boldsymbol{\mu})'$$

通常最尤法の評価関数としては、上の対数尤度関数に定数を加えた以下の式が用いられることが多い。

$$f_{ML}(\boldsymbol{\theta}) = \text{tr}(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{U}) - \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{U}| - n$$

これらの評価関数の最小化法には様々な方法が用いられるが、現在我々は最小 2 乗法では、Levenberg-Marquart 法、最尤法では最初のパラメータ設定に最小 2 乗法を用い、求められた値を初期値として Levenberg-Marquart 法を応用した Newton-Raphson 法を用いている。

## 2.3 モデルの評価

ここではモデルの良し悪しを評価するいくつかの指標とその性質についてまとめておく。

### 解の検定

帰無仮説  $H_0$  : 構成されたモデルは正しい。

対立仮説  $H_1$  : 構成されたモデルは正しくない。

$$\chi^2 = (N-1)f_{ML} \sim \chi^2_{df}, \quad df = \frac{1}{2}n(n+1) - p$$

ここに  $N$  はデータ数、 $n$  は観測変数の数、 $p$  は自由パラメータ数（外生観測変数数+パス係数数+誤差変数数+共分散（相関）数）であり、 $df$  は  $\chi^2$  分布の自由度である。この検定はデータ数を増やして精度を上げるほど対立仮説である「モデルは正しくない」という結果が出やすくなるという矛

盾を含んでいる。

### 適合度指標

#### GFI (Goodness of Fit Index)

これは実測値による共分散行列とパラメータで表された共分散行列の類似の程度を見る指標で以下のよう与えられる。

$$GFI = 1 - \frac{\text{tr}((\Sigma(\hat{\theta})^{-1} \mathbf{U} - \mathbf{I})^2)}{\text{tr}((\Sigma(\hat{\theta})^{-1} \mathbf{U})^2)}$$

ここに  $\text{tr}(\mathbf{A}^2) = \text{tr}(\mathbf{A}\mathbf{A}')$

この指標の値は 0.9 以上が良いとされるが、モデルの自由度が大きくなると値を大きくすることが難しくなる。

#### AGFI (Adjusted Goodness of Fit Index)

これは GFI の自由度の問題を改善した指標で、相関を加えて自由度を見かけ上小さくしても値が改善されるとは限らない指標である。

$$AGFI = 1 - \frac{n(n+1)}{2df}(1 - GFI)$$

一般に  $AGFI \leq GFI$  の関係がある。

### 情報量基準

#### AIC (Akaike's Information Criterion)

これは一般的統計モデルの評価指標として有名であり、以下で定義される。

$$AIC = \chi^2 - 2df$$

この値が小さいほど良いモデルとされる。この指標には、標本数が多い場合、自由度が小さい（パラメータ数が多い）モデルが良いモデルと判断される傾向がある。

#### CAIC (Consistent Akaike's Information Criterion)

これは AIC の標本数の影響を抑えた指標である。

$$CAIC = \chi^2 - (\log(N) + 1)df$$

### パラメータの検定

最尤法の推定値  $\hat{\theta}$  を用いると、以下のようになることが知られている。

$$z_i = \frac{\hat{\theta}_i}{\sigma_{\hat{\theta}_i}} \sim N(0,1) \quad \text{ここに、} \sigma_{\hat{\theta}_i} \equiv \frac{N-1}{2} \frac{\partial^2}{\partial \theta_i^2} f_{ML}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

これを用いてパラメータの値を 0 と比較する検定を行うことができる。

### 3. プログラムの動作

ここでは 2 章で述べた例を用いてプログラムの動作を説明する。プログラムを起動すると図 3.1a のような初期メニューが表示される。これは授業用にできるだけ簡易化したメニューである。この中で拡張メニュー ボタンをクリックすると図 3.1b のような拡張メニューが表示される。

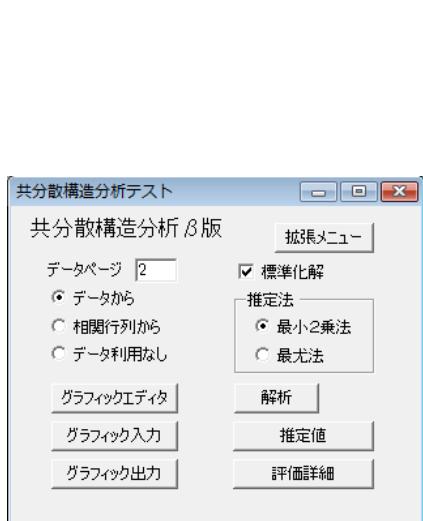


図 3.1a 初期メニュー画面

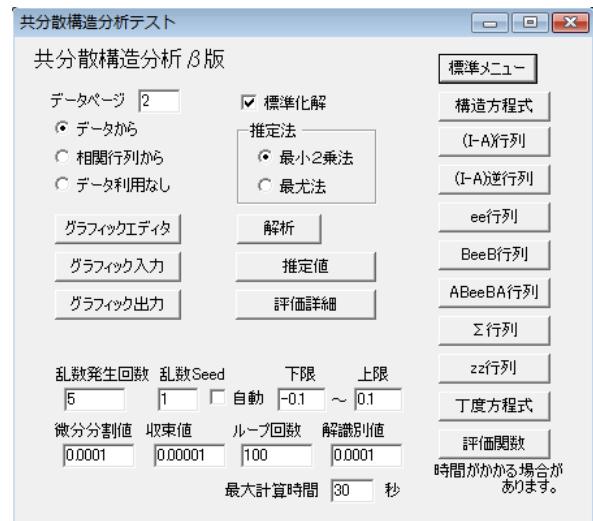


図 3.1b 拡張メニュー画面

拡張メニューには細かな設定や、数式表示のためのボタンが含まれている。以後すべての機能が揃った拡張メニューをもとに説明していく。これらのメニューの中の「グラフィックエディタ」、「グラフィック入力」、「グラフィック出力」ボタンについては他の分析との関係でまだ十分に検討しておらず、ここでは触れない。

共分散構造分析のデータは基本的にデータ構造を記述したページと観測変数のデータ値を表すページに分かれる。前者を図 3.2 に示す。後者については通常の統計データの画面である。

データ編集 semorg2.txt

	f1	v1	v2	v3	v4	e2	e3	e4
f1		r		a1	a2			
v1				a3				
v2					a4			
v3						a5		
v4				b2				
e2					b3			
e3						b4		
e4								

1/2 (1,1) 分析: [ ] 備考: [ ]

図 3.2 構造データ

分析は、メインメニュー左上の「データページ」テキストボックスに観測値のページ番号を記入し、図 3.2 の構造データを表示して実行する。1 つの観測データに複数のモデルを考える場合は、データを 1 ページ目にして、2 ページ目以降を構造データにするのがよい。

最初に共分散構造分析の基礎となる数式について、表示結果を説明する。図 3.1b のメニュー画面の「構造方程式」ボタンをクリックすると構造方程式が図 3.3 のように表示される。

構造方程式

	f1	v1	v2	v3	v4	f1	v1	e2	e3	e4
f1						1				f1
v1							1			v1
v2	=			a3			v2	+	b2	e2
v3		a1		a4			v3		b3	e3
v4		a2			a5		v4			b4 e4

図 3.3 構造方程式

ここに 2 章 (1) 式中の行列  $\mathbf{A}$  は図の四角形で囲まれた部分である。メニューの「(I-A) 行列」ボタンをクリックすると図 3.4 のように  $\mathbf{I}-\mathbf{A}$  行列の結果が表示される。

(I-A) 行列

1				
	1			
		-a3	1	
		-a1	-a4	1
		-a2		-a5 1

図 3.4 (I-A) 行列

メニューの「(I-A) 逆行列」ボタンをクリックすると  $\mathbf{I}-\mathbf{A}$  逆行列の結果が表示される。

(I-A) 逆行列

1					分母
	1				
		a3	1		
	a1	a3*a4	a4	1	
	a1*a5+a2	a3*a4*a5	a4*a5	a5	1 1

図 3.5 (I-A) 逆行列

分母の列の最下行では  $\mathbf{I}-\mathbf{A}$  行列の行列式を表す。メニューの「ee 行列」ボタンをクリックすると図

3.6 のように行列  $\mathbf{H}$  が表示される。

	f1	v1	e2	e3	e4
▶ f1	1	r			
v1	r	1			
e2			1		
e3				1	
e4					1

図 3.6 ee 行列

モデルで相関を仮定した部分はここにそのパラメータが残る。後はすべて無相関と仮定される。

メニューの「BeeB 行列」ボタンクリックすると図 3.7 のように行列  $\mathbf{BHB}'$  が表示される。

	1	r			
▶	1	r			
r		1			
			b2*b2		
				b3*b3	
					b4*b4

図 3.7  $\mathbf{BHB}'$  行列

「ABeeBA 行列」ボタンクリックすると図 3.8 のように行列  $(\mathbf{I} - \mathbf{A})^{-1} \mathbf{BHB}' (\mathbf{I} - \mathbf{A})'^{-1}$  が表示される。

	1	r	r*a3	a1+r*a3*a4	a1*a5+a2+r...	分母
▶	1	r	a3	r*a1+a3*a4	r*a1*a5+r*a...	
r		1	a3	r*a1+a3*a4	r*a1*a5+r*a...	
a3*r	a3	a3*a3+b2*b2	a3*a1+a3*a...	a3*r*a1*a5...		
a1+a3*a4*r	a1+r*a3*a4	a1*r*a3+a3...	a1*a1+a3*a...	a1*a1*a5+a...		
a1*a5+a2+a...	a1*a5*r+a2...	a1*a5*r*a3...	a1*a5*a1+a...	a1*a5*a1*a...		1

図 3.8  $(\mathbf{I} - \mathbf{A})^{-1} \mathbf{BHB}' (\mathbf{I} - \mathbf{A})'^{-1}$  行列

「Σ 行列」ボタンクリックすると図 3.9 のように丁度方程式左辺の  $\Sigma(\theta)$  が表示される。

	1	a3	r*a1+a3*a4	r*a1*a5+r*a...	分母
▶	1	a3	a3*a3+b2*b2	a3*r*a1+a3...	a3*r*a1*a5...
a3	a3	a3*a3+b2*b2	a3*r*a1+a3...	a3*r*a1*a5...	
a1+r*a3*a4	a1+r*a3*a4	a1*a1+a3*a...	a1*a1*a5+a...		
a1*a5*r+a2...	a1*a5*r+a2...	a1*a5*r*a3...	a1*a5*a1+a...	a1*a5*a1*a...	1

図 3.9 Σ 行列

「zz 行列」ボタンクリックすると図 3.10 のように行列  $\mathbf{U}$  が表示される。これは観測変数の共分散行列（標準化解の場合は相関行列）である。

zz 行列				
	v1	v2	v3	v4
► v1		1	0.834553	0.895450
v2	0.834553		1	0.879707
v3	0.895450	0.879707		1
v4	0.887599	0.826780	0.953547	

図 3.10 観測変数の相関行列

メニューの「丁度方程式」ボタンクリックすると図 3.11 のように丁度方程式が表示される。

```

■ 丁度方程式

 $\frac{5}{3} - 0.834553140280362 = 0$ 
 $rka1 + 3ka4 - 0.895449827584345 = 0$ 
 $rka1 + 3ka3 + rk2 + 3*a4*a5 - 0.887599355784528 = 0$ 
 $a3 + 3*a2 + 2*b2 - 1 = 0$ 
 $a3 + 3*a2 + 3*a3 + 4*b2 + 2*a4 - 0.879707324211929 = 0$ 
 $a3 + 3*rk1 + 5 + 3*a3 + 3*rk2 + 2 + 3*a3 + 3*a4 + 5 + 2*b2 + 2*a4 + 5 - 0.826779612393259 = 0$ 
 $a1 + a1 + 3*a4 + 2*a5 + 1 + 1*rk1 + 3*a3 + 4*a4 + 3*a5 + 4*a4 + 4*b2 + 2*b2 + 2*a4 + b3 + b4 - 1 = 0$ 
 $a1 + a1 + 5 + 1 + a2 + 3*a4 + 4*rk1 + 1 + 5 + 3*a4 + 4*rk2 + 2 + 1 + rk1 + a3 + a4 + a5 + a3 + a4 + a3 + a4 + a5 + a5 + a4 + 2*b2 + b2 + a4 + a5 + b3 + b3 - 0.953546824928684 = 0$ 
 $a1 + a5 + 1 + a6 + 1 + a5 + 2*a2 + 1 + a5 + 2*a2 + 2*a3 + a4 + 5 + rka1 + a5 + a4 + 4*a5 + rka2 + a1 + a5 + 2*b2 + b2 + a4 + a5 + a5 + a5 + a3 + b3 + a5 + b4 + b4 - 1 = 0$ 
#パラメータ数： 9
#方程式数 : 9

```

図 3.11 丁度方程式

メニューの「評価関数」ボタンをクリックすると図 3.12 のように評価関数が表示される。

```
(a-0.834553140280362)^2+(r*a1+a3*a4-0.895449827584345)^2+(r*a1*a5+r*a2+a3*a4*a5-0.8875989355784528)^2+  
(a3*a3+b2*b2-1)^2+(a3*r*a1+a3*a3*a4+b2*b2*a4-0.879707234211923)^2+  
(a3*r*a1+a5*a3*r*a2+a3*a3*a4*a5+b2*b2*a4*a5-0.8267798612393259)^2+  
(a1*a1+a3*a4*r*a1+a1*r*a3*a4*a3*a4+a4*b2*b2*a4+b3*b3-1)^2+  
(a1*a1+a2*a2+a3*a4*r*a1*a3+a3*a4*a4*r*a2+a1*r*a3*a4*a4+b3*a3*a4*a5+a4*b2*b2*a4*a5+b3*b3*a5-  
0.953516294928664)^2+  
(a1*a5*a1*a5+a1*a5*a2*a2*a1*a5+a2*a2+a3*a4*a5*a5*a1*a5+a3*a4*a5*r*a2+1*a5*r*a3*a4*a5+a2*r*a3*a4*a5+b4*b4-1)^2
```

図 3.12 評価関数

これは最小2乗法における評価関数で、これを最小化するようにパラメータは選ばれる。最尤法の場合、表示が膨大になるのでかなり時間がかかる場合がある。

推定値については、「推定法」のグループの「最尤法」ラジオボタンを選択して、最初に「解析」ボタンクリックし、それから「推定値」をクリックすると図 3.13 のように表示される。

	f1	v1	v2	v3	v4	e2	e3	e4
▶ f1		0.9192		0.5781	0.3326			
v1	0.9192		0.8345					
v2				0.4361				
v3					0.6499			
v4								
e2			0.5509					
e3				0.2981				
e4					0.2692			
分散	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
評価閾値	0.0000							

図 3.13 最尤法の推定値

これは最尤法の推定値であるが、工度方程式の解でもある。グラフィックエディタを用いて構造図を

作成した場合は、構造図中にも推定値が表示されるようにしたい。さらに「評価詳細」ボタンをクリックすると、異なった形式の推定値と評価値が図 3.14a と図 3.14b のように表される。

パラメータ推定値						
	変数		変数	推定値	標準誤差	検定値
▶ r	v1	<->	f1	0.9192	0.0620	14.8289
a1	v3	<-	f1	0.5781	0.1353	4.2719
a2	v4	<-		0.3326	0.2424	1.3724
a3	v2	<-	v1	0.8345	0.1023	8.1566
a4	v3	<-	v2	0.4361	0.1265	3.4490
a5	v4	<-	v3	0.6499	0.2319	2.8030
b2	v2	<-	e2	0.5509	0.0723	7.6156
b3	v3	<-	e3	0.2981	0.0849	3.5097
b4	v4	<-	e4	0.2692	0.0514	5.2398
評価閾値	0.0000					

図 3.14a 推定値の詳細表示

モデルの評価	
推定値の評価 推定値の検定（データ数が増えれば「正しくない」と結論され易いことに注意） 自由度の値が不適切で計算できません。	
適合度指標 自由度 0 GFI (>= 0.9 が良いが、自由度が小さくなると改善されることに注意 1.0000) AGFI (AGFI < GFI 自由度を小さくしても必ずしも改善されない) 自由度の値が不適切で計算できません。	
情報量基準, $\chi^2 - 2 \times df$ AIC (モデル比較の代表的指標、値が小さいほど良いモデル 標本数が多くなると自由度が小さいほど「良く」なる特徴を持つ) 0.0002 CAIC (AICより標本の影響を抑えた指標) 0.0002	
自由度当りのモデルの分布と最尤モデルとの乖離を表す指標 RMSEA (<= 0.05: 当てはまりが良い, >= 0.1: 当てはまりが悪い) 自由度の値が不適切で計算できません。	

図 3.14a モデルの評価（表示の後半部分）

最小 2 乗法の場合、推定値の検定部分や評価指標の適合度指標以外の部分は表示されない。

## 4. Amos との比較

我々はプログラムの評価のために、我々の結果と Amos の結果とを以下の構造図の場合について比較した<sup>1)</sup>。まだ我々の計算のアルゴリズムが不十分なため、ごく小さなモデルについてのみの比較に限られている。なお名称は参考文献に名前がある場合はその名前を使用し、名前がない場合は我々が与えた。また結果の符号については、潜在変数の符号の任意性に起因すると思われる場合は、結果が同一のものと判断した。

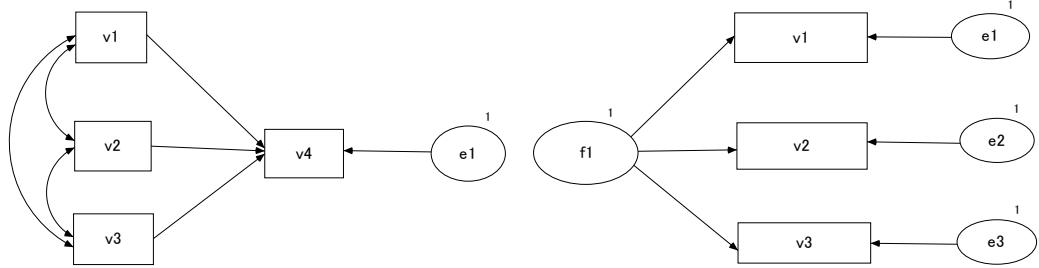


図 4.1 回帰分析モデル

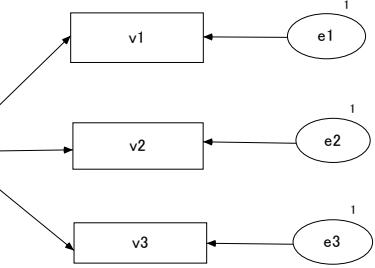


図 4.2 因子分析モデル

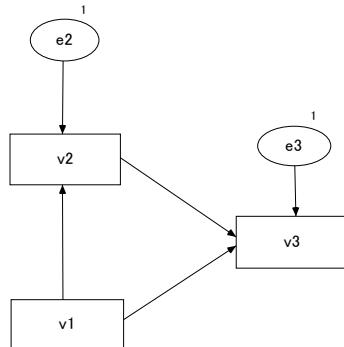


図 4.3 回帰分析の複合モデル 1

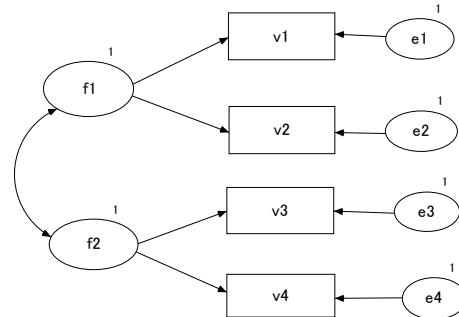


図 4.4 因子分析の複合モデル 1

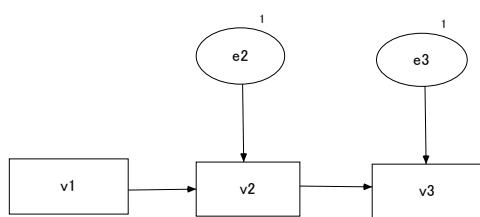


図 4.5 連結モデル

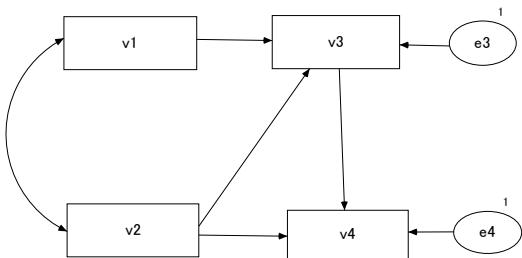


図 4.6 逐次モデル

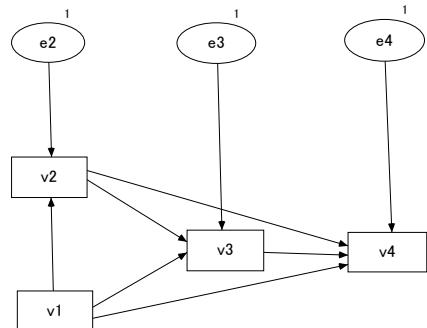


図 4.7 回帰分析の複合モデル 2

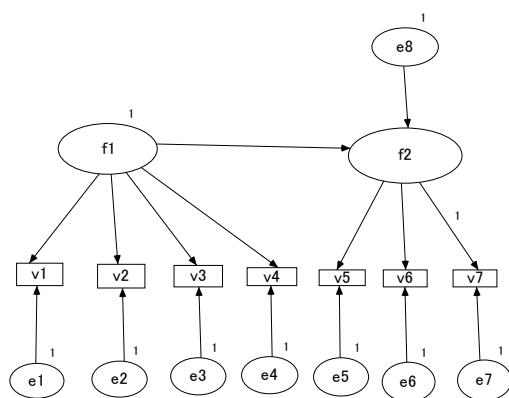


図 4.8 因子分析の複合モデル 2

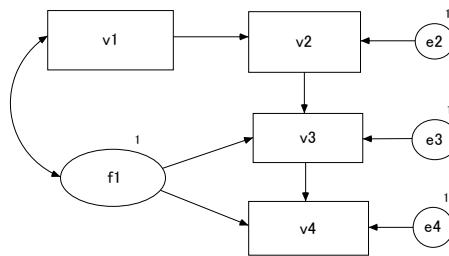


図 4.9 回帰分析と因子分析の複合モデル

いずれの場合もモデルがデータをよく表す場合は Amos の結果と我々のプログラムの結果は一致する。しかしモデルがデータに適合しない場合（我々は乱数を用いてデータを作成して試した）、興味深い結果が出たので紹介する。

表 4.1 は図 4.4 の場合の最尤法による両者の比較である。

表 4.1 図 4.4 の結果の比較

	変数		変数	Amos	CAnalysis	CAnalysis 別解
r	f2	$\leftarrow\rightleftharpoons$	f1	-0.253	0.2532	45.8391
a1	v1	$\leftarrow$	f1	1.000	1.0000	0.0069
a2	v2	$\leftarrow$	f1	-0.244	-0.2444	0.0060
a3	v3	$\leftarrow$	f2	1.000	-0.9998	-1.0218
a4	v4	$\leftarrow$	f2	0.133	-0.1333	-0.1361
b1	v1	$\leftarrow$	e1	0.000	0.0000	1.0000
b2	v2	$\leftarrow$	e2	0.970	0.9697	1.0000
b3	v3	$\leftarrow$	e3	0.000	0.0197	0.0001
b4	v4	$\leftarrow$	e4	0.991	-0.9911	-0.9911
評価関数値				0.087	0.0867	0.0664

我々の結果はパラメータの初期値の与え方によって何種類かの異なる結果が得られ、その中で Amos の結果と一致する解の他に、例えば CAnalysis 別解のような解が得られた。別解では相関係数が 1 以上の値になるが、評価関数値は Amos の値より小さくなる。Amos ではこのような非現実な解は排除しているように見える。さらに我々のプログラムでパラメータを一部固定してみると評価関数値が Amos の値より小さい現実的な推定値を求めることが可能となる。我々はこれまで評価関数が極値となる推定値を求めようとしてきたが、パラメータの値が 1 の近傍になる場合には境界を持つ最小化問題となっているように思われる。Amos であってもこのような場合には注意する必要がある。

次に表 4.2 は図 4.8 の結果の比較である。

表 4.2 図 4.8 の結果の比較

	変数		変数	AMOS	CAnalysis
a0	f2	<-	f1	0.173	0.8608
a1	v1	<-	f1	-0.157	0.0983
a2	v2	<-	f1	0.692	0.1063
a3	v3	<-	f1	0.139	0.4869
a4	v4	<-	f1	0.270	-0.0344
a5	v5	<-	f2	-0.235	-0.6597
a6	v6	<-	f2	0.106	0.0874
a7	v7	<-	f2	1.000	0.5206
c1	v1	<-	e1	0.988	0.9952
c2	v2	<-	e2	0.722	0.9944
c3	v3	<-	e3	0.990	-0.8736
c4	v4	<-	e4	0.963	0.9994
c5	v5	<-	e5	0.972	0.8233
c6	v6	<-	e6	0.994	-0.9972
c7	v7	<-	e7	0.000	0.8940
c8	f2	<-	e8	0.986	0.0001
評価関数値				0.190	0.1451

ここでは現実的な値の範囲で Amos より良い解が得られている。この場合にも Amos の推定値の中に境界値 1 が含まれている。またこのような場合でも Amos での GFI の値が 0.951 と高いことにも注意を要する。

さらに図 4.9 については Amos と College Analysis で同じ解が得られ、いずれも標準化解のパラメータの推定値が非現実な値となる場合も見られた。これを見ると Amos でも完全に非現実なパラメータを除外しているわけではなさそうである。

実際の分析ではパラメータの推定値が現実的な値となるようなモデルを考えるため、ここで述べたようなことは起こらないが、分析に不慣れな利用者は十分注意する必要がある。特に非標準化解の場合はそれに気が付かない可能性があるので、結果の検討が必要である。

## 5. おわりに

我々は共分散構造分析についてプログラムの開発を進め、中間段階にまで到達した。殆ど知識のない状態から始めたので、計算の手順の失敗やアルゴリズムの問題から計算時間の短縮にかなり回り道をした。しかしこれらの問題を考える過程で知識を得ることもできた。特に計算時間については実際にプログラムを作成しなければ分らない部分も多い。対話的に処理を行う場合、著者らは人がストレスなく待てる計算時間の上限を 10 秒程度に考えているが、これまでに College Analysis の中で開発してきたプログラムでは特に気になることはなかった。しかし共分散構造分析のプログラムでは、今

のままのアルゴリズムでは、4章で試したモデル程度が限界である。この意味でも Amos で採用されているマルコフ連鎖モンテカルロ法は優れている。我々のプログラムを実用的なものにするためにはどうしても取り入れなくてはならない。

構造図について我々のプログラムでは行列形式で入力するが、紙に書かれたものを入力する場合はかなり効率良く行える。しかし、頭の中でモデルを考える場合、この行列形式の入力は有効とは言えない。このため我々は新しくグラフィック入力用のエディタを開発しなければならない。これは共分散構造分析のメニューから呼び出し、構造図を作成して、結果を表形式のエディタに戻すものにする。また他の分析でも使用するため、汎用的なものにすることも必要である。現在、その大部分は開発が終わっており、その実行画面は図 5.1 のようになる予定である。このエディタの機能やデータ構造については他の分析との関係もあるので、次の論文で詳しく説明する。

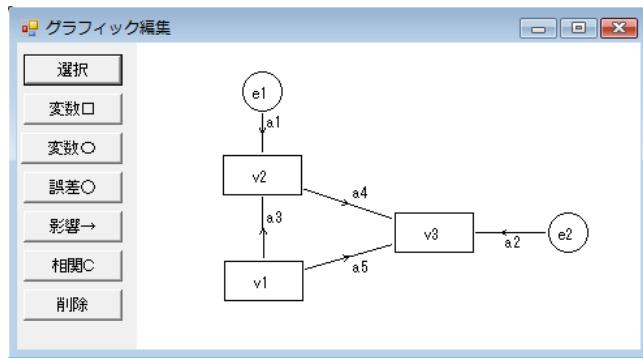


図 5.1 グラフィックエディタ画面

## 参考文献

- 1) 共分散構造分析 [入門編] –構造方程式モデリング–, 豊田秀樹, 朝倉書店, 1998.
- 2) はじめての共分散構造分析 Amos によるパス解析, 小塩真司, 東京図書, 2008.
- 3) 共分散構造分析 Amos 編 –構造方程式モデリング–, 豊田秀樹, 東京図書, 2007.

# **Multi-purpose Program for Social System Analysis 12**

## **- Analysis of Covariance Structures (Interim Report) -**

Masayasu FUKUI, Wenlong CHEN\* and Jiaqi WANG\*

Department of Business Administration, Faculty of Business Administration,  
Fukuyama Heisei University

\* Graduate School of Business Administration,  
Fukuyama Heisei University

### **Abstract**

We have been constructing a unified program on the social system analysis for the purpose of education. This time we created a new program of analysis of covariance structures. This program has many issues such as algorithms and user interface. Therefore, this paper is to report the intermediate stages.

### **Keywords**

College Analysis, social system analysis, OR, statistics, analysis of covariance structures

URL: <http://www.heisei-u.ac.jp/ba/fukui/>