

# 複数データセット間の構造の違いを 探る等質性分析

——「日本人の国民性調査」データへの適用——

統計数理研究所 土 屋 隆 裕

(1995 年 1 月 受付)

## 1. はじめに

同一の質問項目群に対して複数のデータセットがあったとき、それらのデータセット間の構造の違いを探りたいことがある。例えば、世代によって回答パターンが異なるのか、時代によってももの考え方が違ってくるのか、といった場合である。そのとき、項目間の構造、あるいは考えの筋道(林(1993a))を分析する手法の一つである数量化3類を各データセットに対して適用する、という方法が考えられる。そして、各データセットから得られた結果を比較し、どのような差異が見られるのか、検討するのである。

例えば、「日本人の国民性調査」では、表1に示した“義理人情”に関する質問項目について、7回分のデータセットが得られている。そこで、各データセットに対して数量化3類を適用した結果を図1に示す。選択肢の散布図は、どの年も同じような形をしており、義理人情的なものとしてそうでないものを対比させる構造は安定している、と言える(林(1992))。

また、別の例として、穂山(1993)による対人行動の調査がある。これは、同一の質問項目について、毎年データを収集し、各年のデータセットに対する数量化3類の結果を3次元散布図にして見比べることで、項目の構造の安定性を検討しているものである。

目により散布図を比較するこのような方法は、構造が異なるのか否か、という大局的な様子を知るには、直観に訴えるため、有効である。しかし、よりつっこんだ分析を行なうには、不十分であろう。例えば、図1に示される各年のパターンは、全く同じわけではない。それらの差異のなかに系統的な違いがある可能性も残るが、目視による比較では、それは明らかではない。また、非常に多数のデータセットがある場合には、それらを見比べて的確な知見を得ることは難しい。どのデータセット間の構造が似ているのか、という判断もあいまいになる。

林(1982, 1993b)は、数量化3類の結果得られた布置を基に、データセットの構造間の距離を求め、それをMDA-ORで分析している。この方法を用いれば、データセット間の構造の類似性を把握することは容易となる。しかし、構造に差異があることが分かった場合、具体的な違いの様子を知るためには、やはり、数量化3類の個々の結果を見比べるという方法に頼ることになる。

そこで、本論文では、複数データセット間の構造の違いを探索的に分析するためのモデルを提案する。そして、パラメタの推定法について述べた後、人工データと「日本人の国民性調査」データを分析した結果について考察する。

表1. “義理人情”に関する質問項目・選択肢の  $L$ .

	$r=1$	$r=2$	$r=3$
<b># 4.4 ‘先生が悪いことをした’</b>			
「先生が悪いことをした」というような話を、子供が聞いてきて、親にたずねたとき、親はそれがほんとうであることを知っている場合、子供には「そんなことはない」といった方がよいと思いますか、それとも「それはほんとうだ」といった方がよいと思いますか？			
1 そんなことはないという	-.194	-.306	.178
2 ほんとうだという	.104	.164	-.226
<b># 5.1 ‘恩人がキトクするとき’</b>			
[絵を見せながら] 南山さんという人は、小さいときに両親に死に別れ、となりの親切な西木野さんに育てられ、大学まで卒業させてもらいました。そして南山さんはある会社の社長にまで出世しました。ところが故郷の、育ててくれた、西木野さんが「キトクだからスグカエレ」という電報を受けとったとき、南山さんの会社がつぶれるか、つぶれないか、ということがきまってしまう大事な会議があります。			
[ここでリストを見せる] あなたはつぎのどちらの態度をとるのがよいと思いますか、よいと思う方を1つだけえらんで下さい？			
1 なにをおいても、すぐ故郷へ帰る	.543	-.618	.016
2 故郷のことが気になっても、大事な会議に出席する	-.624	.704	.056
<b># 5.1b ‘親がキトクするとき’</b>			
[質問 a と同じ絵, 同じリストで] いまの質問では、恩人が死にそうなきを、うかがいましたが、もしキトクなのが恩人ではなくて、南山さんの親だったら、どうしたらよいと思いますか、どちらかえらんで下さい？			
1 なにをおいても、すぐ故郷へ帰る	.554	-.650	.012
2 故郷のことが気になっても、大事な会議に出席する	-.586	.682	.051
<b># 5.1c1 ‘入社試験 (親戚)’</b>			
[リスト] あなたが、ある会社の社長だったとします。その会社で、新しく職員を1人採用するために試験をしました。入社試験をまかせておいた課長が、「社長のご親戚の方は2番でした。しかし、私としましては、1番の人でも、ご親戚の方でも、どちらでもよいと思いますがどうでしょうか」と社長のあなたに報告しました。あなたはどちらをとれ (採用しろ) といいますか？			
1 1番の人を採用するよういいう	.330	.285	.287
2 親戚を採用するよういいう	-1.227	-1.120	.653
<b># 5.1c2 ‘入社試験 (恩人の子)’</b>			
[つぎのリスト] それでは、このばあい、2番になったのがあなたの親戚の子供ではなくて、あなたの恩人の子供だったとしたら、あなたはどうしますか？(どちらをとれといいますか？)			
1 1番の人を採用するよういいう	.567	.519	.465
2 恩人の子供を採用するよういいう	-.695	-.665	.353
<b># 5.6 ‘めんどうをみる課長’</b>			
[リスト] ある会社に次のような2人の課長がいます。もしあなたが使われるとしたら、どちらの課長につかわれる方がよいと思いますか、どちらか1つあげてください？			
1 規則をまげてまで、無理な仕事をさせることはありませんが、仕事以外のことで人のめんどうを見ません	.130	.426	-.227
2 時には規則をまげて、無理な仕事をさせることもありますが、仕事のこと以外でも人のめんどうをよく見ます	-.019	-.059	.069

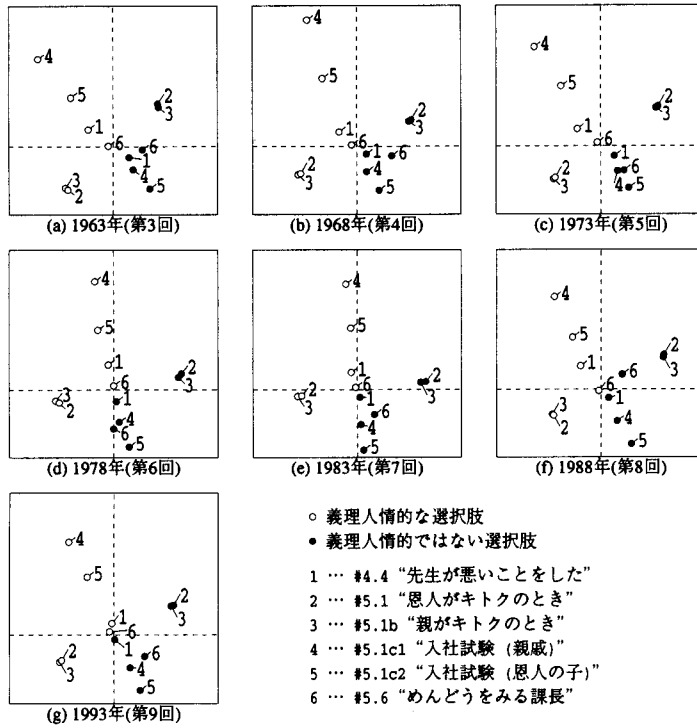


図1. “義理人情”に関する質問項目群の数量化3類の結果。

## 2. モデル

### 2.1 等質性分析

数量化3類と同じ結果を導く手法として、等質性分析 (Gifi (1990)) がある (Tenenhaus and Young (1985), 柳井 (1994))。本論文では、等質性分析の考え方にそって方法を述べる。

第  $i$  番目のデータセットにおける項目  $j$  のアイテムカテゴリデータを、 $N_i \times C_j$  行列  $X_{ij}$  ( $i=1, \dots, I; j=1, \dots, J$ ) によって表わす。 $X_{ij}$  の各行は、ある要素のみ1で、残りの要素は0である。等質性分析では、各項目は基本的には同じ特性を測っていると仮定する (Meulman (1982))。すなわち、各項目ごとに構成する合成変数  $X_{ij}W_{ij}$  ( $j=1, \dots, J$ ) を、ある特性に対する得点行列  $M_i$  におきかえても情報の損失は少ないであろうと考える。そこで、 $X_{ij}W_{ij}$  と  $M_i$  との差の二乗  $\sigma$  を最小にする  $W_{ij}, M_i$  を求めるのである。

$$(2.1) \quad \sigma = \sum_i \sum_j \text{tr}(X_{ij}W_{ij} - M_i)'(X_{ij}W_{ij} - M_i) = \sum_i \sum_j \|X_{ij}W_{ij} - M_i\|^2$$

ただし、 $M_i$  の原点や尺度は任意であるため、

$$M_i' \mathbf{1} = 0, \quad M_i' M_i = N_i I \quad (i=1, \dots, I)$$

とする。

(2.1) を  $W_{ij}$  について最小化すれば、 $P_{ij} = X_{ij}(X_{ij}'X_{ij})^{-1}X_{ij}'$  として、

$$\begin{aligned}\sigma &\geq \sum_i \sum_j \text{tr} \mathbf{M}_i' (\mathbf{I} - \mathbf{P}_{ij}) \mathbf{M}_i = \sum_i \sum_j (\text{tr} \mathbf{M}_i' \mathbf{M}_i - \text{tr} \mathbf{M}_i' \mathbf{P}_{ij} \mathbf{M}_i) \\ &= J \sum_i \text{tr} \mathbf{M}_i' \mathbf{M}_i - \sum_i \text{tr} \mathbf{M}_i' \left( \sum_j \mathbf{P}_{ij} \right) \mathbf{M}_i\end{aligned}$$

が成り立つ。すなわち、 $\text{tr} \mathbf{M}_i' \mathbf{M}_i$  を人に与える数量の全分散、 $\text{tr} \mathbf{M}_i' \mathbf{P}_{ij} \mathbf{M}_i$  を、選んだ選択肢が同じ人々のグループの間の分散と考えれば、(2.1) の基準は、グループ内分散を最小にするよう  $\mathbf{W}_{ij}$ 、 $\mathbf{M}_i$  を求めることと等価となり、 $\mathbf{W}_{ij}$  は多重対応分析 (大隅 他 (1994)) あるいは数量化3類 (林 (1993b)) において項目に与えられる数量に等しくなる。具体的には、 $\mathbf{M}_i$  は  $\sum_j \mathbf{P}_{ij}$  の二番目以降に大きい固有値に対応する固有ベクトルによって与えられ、各  $\mathbf{W}_{ij}$  は、 $\mathbf{W}_{ij} = (\mathbf{X}_{ij}' \mathbf{X}_{ij})^{-1} \mathbf{X}_{ij}' \mathbf{M}_i$  として求められる (柳井 (1994))。

ところで、一つの項目に対して、データセットごとに異なる重み行列  $\mathbf{W}_{ij}$  を仮定したのは、構造は異なることが予想されるからである。しかし、(2.1) の基準では、各データセットごとに分析を行なうことと同じになる。 $\mathbf{W}_{1j}, \dots, \mathbf{W}_{ij}$  は同じ項目に対する重みであるため、それらの間に何らかの関係を持たせてもよいであろう。そこで、 $\mathbf{W}_{ij}$  を構造化することを考える。

## 2.2 射影による構造化

まず、項目間の構造が異なるということが何を意味するのか考察する。ここで言う項目間の構造とは、数量化3類、あるいは等質性分析における  $\mathbf{W}_{ij}$  を指す。

構造に違いがあるときには、二通りの考え方をすることができよう。

第一の考え方は、考えの筋道が異なっている、というものである。例えば、「日本人の国民性調査」における伝統対近代の質問群で、回答者がある選択肢を選ぶ場合を考えよう。数量化3類の結果、伝統対近代という構図がはっきりしている (林 (1992)) ことから、1950年代では、それが伝統的な意見なのか、あるいは近代的な意見なのか、という観点から選択肢をながめた上で、自分の意見に近い選択肢を選んでいることになる。極端に単純化してしまえば、ある選択肢が選ばれるのは、それが伝統的な意見だから、ということである。しかし、1980年代では、伝統対近代という構図が崩れている (林 (1992))。すなわち、1950年代と同じ選択肢が選ばれるとしても、それは、その選択肢が伝統的な意見だからではない。なぜなら、もし、伝統的な意見であるという理由から選ばれたのなら、散布図の上で伝統的な選択肢のかたまりは崩れないはずだからである。1980年代では、伝統対近代とは別の観点から選択肢をながめ、その上で回答している、ということになる。

質問項目や選択肢の字句が全く同じだとしても、ある一つの選択肢を選ぶに至る考えの筋道はいろいろである。別の言い方をすれば、一つの質問項目をどうとらえるか、という観点・視点はさまざまである。ある考えの筋道をたどれば、選択肢 a を選んだ人は、選択肢 b を選ぶが、別の考えの筋道をたどれば、選択肢 b を選ばないかもしれない。すなわち、ある視点から見れば、選択肢 a と選択肢 b は近いところにあるが、別の視点から見ると、それらは遠いかもしれないということである。

第二の考え方としては、意見分布の違いが構造にあらわれる、ということが挙げられる。例えば、表2に示した二つのデータセットに対する数量化3類の結果は異なる。二つのデータセットの間で、考えの筋道、あるいは項目に対する視点に変化したと考えることもできよう。しかし、単に意見分布が変わっただけである、という解釈の方が自然である。第一の選択肢と他の選択肢を対立させる考えの筋道は変わらず、選ぶ意見が変化しただけということである。

第一の考え方、すなわち、質問項目に対する考えの筋道の違い、視点の相違が構造の違いにあらわれる、という考え方に従えば、 $\mathbf{W}_{ij}$  を以下のように表わすことができる。

表 2. 仮想的な二つのデータセットのクロス集計.

第 1 データセット				第 2 データセット			
項目 B	項目 A			項目 B	項目 A		
	1	2	3		1	2	3
1	100	0	0	1	40	0	0
2	0	100	50	2	0	120	60
3	0	50	100	3	0	60	120

まず、全体で  $C = \sum_i C_i$  個の選択肢が  $R$  次元空間上に布置を持つと仮定する。これを、 $C \times R$  行列  $L = (L'_1 \cdots L'_C)'$  によって表わす。 $L$  はデータセットによらず常に一定である。なぜなら、選択肢の字句は常に全く同一だからである。そして、第  $i$  番目のデータセットにおける考えの筋道、視点を、ある  $D$  次元部分空間  $S_i (D < R)$  とすると、 $L$  の  $S_i$  上への射影が、そのデータセットの構造となる。すなわち、 $S_i$  の正規直交基底を  $R \times D$  行列  $F_i$  とすると、

$$(2.2) \quad W_i = (W'_{i1} \cdots W'_{ij})' = LF_i$$

となる。データセット間の項目の構造変化は、 $F_i$  の変化によって示される。構造が変わらなければ、 $F_i$  はほとんど変わらない。 $F_i$  が大きく異なるときには、構造に何らかの違いがあるといえる。また、 $L$  は全てのデータセットをまとめあげた上での、項目の構造である。具体的にどの項目がどう異なり、どの項目が変わらないか、ということは、 $L$  を調べることによって分かる。

ところで、第二の考え方、すなわち単なる意見分布の違いが構造の違いにあらわれる、という考え方に基づけば、考えの筋道  $F_i$  というものを導入した (2.2) は不適當なように思われる。しかし、 $W_i$  の違いがいかにして起こったかにかかわらず、 $W_1, \dots, W_I$  がほとんど同じならば  $R$  を  $D$  に近づけて、 $F_1, \dots, F_I$  を等しくすればよく、逆に  $W_1, \dots, W_I$  が非常に異なるなら、 $R$  を大きくすることでその差異を表現することができる。すなわち、 $F_i$  を、第  $i$  番目のデータセットにおける考えの筋道と考えず、 $L$  から  $W_i$  を導くための射影子であり、構造変化を示すための指標と考えれば、第二の考え方に基づいても、(2.2) を仮定することはできるであろう。 $F_i$  を導入するために考えの筋道という概念を用いたが、以後は、 $F_i$  をデータ縮約のための単なるパラメタとみなす、ということである。

以上の議論から、最小とすべき基準  $\sigma$  は、

$$(2.3) \quad \sigma = \sum_i \sum_j \|X_{ij} L_j F_i - M_i\|^2$$

ただし、 $F_i' F_i = I, M_i' 1 = 0, M_i' M_i = N_i I \quad (i=1, \dots, I)$

となる。

### 3. パラメタの推定

(2.3) の  $\sigma$  を最小とする  $L, F_i, M_i$  を代数的に同時に求めることは難しいので、それぞれについて交互に最小化する。すなわち、

1.  $F_i^{(t)}, M_i^{(t)}$  を固定して、 $L$  について最小化し、それを  $L^{(t+1)}$  とおく。
2.  $L^{(t+1)}, M_i^{(t)}$  を固定して、 $F_i$  について最小化し、それを  $F_i^{(t+1)}$  とおく。

3.  $L^{(t+1)}, F_i^{(t+1)}$  を固定して,  $M_i$  について最小化し, それを  $M_i^{(t+1)}$  とおく.

以上を値が収束するまで繰り返す. 以下では, 各パラメタについての最小化の方法を順に述べる.

### 3.1 $L$ についての最小化

(2.3) を次のように書き直す.

$$\begin{aligned} \sigma &= \sum_i \sum_j \|\text{vec}(X_{ij}L_jF_i) - \text{vec}M_i\|^2 \\ &= \sum_j \left\| \begin{pmatrix} F_1' \otimes X_{1j} \\ \vdots \\ F_I' \otimes X_{Ij} \end{pmatrix} \text{vec}L_j - \begin{pmatrix} \text{vec}M_1 \\ \vdots \\ \text{vec}M_I \end{pmatrix} \right\|^2 \end{aligned}$$

これを  $\text{vec}L_j$  の各成分で偏微分し, その結果を  $\mathbf{O}$  とおいて整理すると,

$$(3.1) \quad \text{vec}L_j = \left( \sum_i F_i F_i' \otimes X_{ij}' X_{ij} \right)^{-1} \text{vec} \left( \sum_i X_{ij}' M_i F_i' \right)$$

が得られる. 特に,  $X_{ij}$  がアイテムカテゴリ型の行列である場合には,  $X_{ij}' X_{ij}$  は対角行列となるため,

$$L_{jk} = y'_{jk} \left( \sum_i x_{ijk} F_i F_i' \right)^{-1}$$

となる. ただし,  $l_{jk}, y'_{jk}$  は, それぞれ  $L_j, \sum_i X_{ij}' M_i F_i'$  の第  $k$  行ベクトル,  $x_{ijk}$  は  $X_{ij}' X_{ij}$  の第  $k$  対角要素である.

### 3.2 $F_i$ についての最小化

(2.3) では,  $L$  と  $F_i$  の間,  $F_i$  と  $M_i$  の間に回転に対する不定性がある. すなわち, ある正規直交行列を  $R, T$  とすると,

$$\begin{aligned} LF_i &= LR'RF_i = L^*F_i^*, \quad \text{ただし, } L^* = LR, F_i^* = RF_i \\ \|X_{ij}L_jF_i - M_i\|^2 &= \|X_{ij}L_jF_iT - M_iT\|^2 = \|X_{ij}L_jF_i^{\dagger} - M_i^{\dagger}\|^2 \\ &\quad \text{ただし, } F_i^{\dagger} = F_iT, M_i^{\dagger} = M_iT \end{aligned}$$

が成り立つ. 各パラメタについて交互に最小化するため, 計算を行なう時点ではこのことは特に問題とはならない. しかし, 結果の見やすさを考え, 第1番目のデータセットの  $F_1$  については,

$$F_1 = \begin{pmatrix} I_D \\ \mathbf{O} \end{pmatrix}$$

とする.

次に, 他の  $F_i (i=2, \dots, I)$  については, これらの列ベクトルごとにさらに交互最小化を繰り返す. すなわち,  $F_i$  から列ベクトル  $f_{id}$  を除いた行列を  $F_{i(-d)}$  とすると,  $F_{i(-d)}$  を固定した上で,  $f_{id}$  について順に最小化を行なう. 最小とすべき基準は, ラグランジュの未定乗数  $\lambda, \xi$  を用いて,

$$(3.2) \quad \sigma(f_{id}) = \sum_j \|X_{ij}L_jf_{id} - m_{id}\|^2 + \lambda(f'_{id}f_{id} - 1) + f'_{id}F_{i(-d)}\xi$$

となる。ただし、 $\mathbf{m}_{id}$  は  $\mathbf{M}_i$  の第  $d$  列ベクトルである。

$\sigma(\mathbf{f}_{id})$  を  $\mathbf{f}_{id}$  について微分してその結果を  $\mathbf{0}$  とおき、 $\xi$  を消去すると、 $\mathbf{X}_i = (\mathbf{X}_{i1} \cdots \mathbf{X}_{id})'$  として、

$$(3.3) \quad \begin{aligned} (\mathbf{Q}_{i(-d)} \mathbf{L}' \mathbf{X}_i' \mathbf{X}_i \mathbf{L} \mathbf{Q}_{i(-d)} + \lambda \mathbf{Q}_{i(-d)}) \mathbf{f}_{id} &= \mathbf{Q}_{i(-d)} \mathbf{L}' \mathbf{X}_i' \mathbf{m}_{id} = \mathbf{z}_{id} \\ \mathbf{Q}_{i(-d)} &= \mathbf{I} - \mathbf{F}_{i(-d)} \mathbf{F}_{i(-d)}' \end{aligned}$$

が得られる。

ところで、 $\mathbf{Q}_{i(-d)} \mathbf{F}_{i(-d)} = \mathbf{0}$  が成り立つため、 $\mathbf{F}_{i(-d)}$  は、 $\mathbf{Q}_{i(-d)} \mathbf{L}' \mathbf{X}_i' \mathbf{X}_i \mathbf{L} \mathbf{Q}_{i(-d)}$  の固有値 0 に対応する固有ベクトルのある一組である。従って、ある  $R \times (R-D+1)$  次の行列  $\mathbf{U}$  を用いて、

$$\mathbf{Q}_{i(-d)} \mathbf{L}' \mathbf{X}_i' \mathbf{X}_i \mathbf{L} \mathbf{Q}_{i(-d)} = (\mathbf{U} \mathbf{F}_{i(-d)}) \begin{pmatrix} \Phi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} (\mathbf{U} \mathbf{F}_{i(-d)})' = \mathbf{V} \Phi \mathbf{V}'$$

と分解することができる。ただし、 $\mathbf{V} \mathbf{V}' = \mathbf{V}' \mathbf{V} = \mathbf{I}$  であり、 $\Phi$  は対角行列である。そこで、(3.3) は、

$$\mathbf{V} \begin{pmatrix} \Phi + \lambda \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}' \mathbf{f}_{id} = \mathbf{U} \mathbf{D}_{(\lambda)} \mathbf{U}' \mathbf{f}_{id} = \mathbf{z}_{id}$$

と表わされる。 $\mathbf{F}_{i(-d)}' \mathbf{f}_{id} = \mathbf{F}_{i(-d)}' \mathbf{z}_{id} = \mathbf{0}$  であることから、

$$\mathbf{f}_{id} = \mathbf{U} \mathbf{D}_{(\lambda)}^{-1} \mathbf{U}' \mathbf{z}_{id}$$

が得られる。Gander (1981) にならって、

$$(3.4) \quad \mathbf{f}_{id}' \mathbf{f}_{id} = \mathbf{z}_{id}' \mathbf{U} \mathbf{D}_{(\lambda)}^{-2} \mathbf{U}' \mathbf{z}_{id} = 1$$

を満たす  $\lambda$  の中で最大の  $\lambda_{\max}$  により、(3.2) を最小とする  $\mathbf{f}_{id}$  が求まる。

(3.4) を満たす  $\lambda$  はニュートン・ラフソン法により求められる。 $\Phi$  の対角要素  $\phi_{rr}$  の符号は正であり、その中で最小のものを  $\phi_{\min}$  とすると、

$$\begin{aligned} \lim_{\lambda \rightarrow -\phi_{\min}} \mathbf{z}_{id}' \mathbf{U} \mathbf{D}_{(\lambda)}^{-2} \mathbf{U}' \mathbf{z}_{id} &\rightarrow \infty \\ \lim_{\lambda \rightarrow \infty} \mathbf{z}_{id}' \mathbf{U} \mathbf{D}_{(\lambda)}^{-2} \mathbf{U}' \mathbf{z}_{id} &\rightarrow 0 \end{aligned}$$

が成り立つ。そこで、 $\varepsilon$  をある正の小さな数としたとき、 $\lambda^{(0)} = -\phi_{\min} + \varepsilon$  をニュートン・ラフソン法の初期値とする。

### 3.3 $\mathbf{M}_i$ についての最小化

$\mathbf{F}_i$  のときと同様に、 $\mathbf{M}_i$  についても、各列ベクトルごとに交互最小化を繰り返す。 $\mathbf{M}_i$  から列ベクトル  $\mathbf{m}_{id}$  を除いた行列を  $\mathbf{M}_{i(-d)}$  とすると、ラグランジュの未定乗数  $\alpha, \beta, \gamma$  を用いて、

$$(3.5) \quad \sigma(\mathbf{m}_{id}) = \sum_j \|\mathbf{X}_{ij} \mathbf{L}_j' \mathbf{f}_{id} - \mathbf{m}_{id}\|^2 + \alpha (\mathbf{m}_{id}' \mathbf{m}_{id} - N_i) + \beta \mathbf{m}_{id}' \mathbf{1} + \mathbf{m}_{id}' \mathbf{M}_{i(-d)} \gamma$$

が最小にすべき基準となる。(3.5) を  $\mathbf{m}_{id}$  で微分し、その結果を  $\mathbf{0}$  とおいたものを整理すると、

$$\begin{aligned} \mathbf{m}_{id} &= c^{-1/2} \mathbf{R}_{i(-d)} \mathbf{X}_i \mathbf{L} \mathbf{f}_{id} \\ c &= \frac{1}{N_i} \mathbf{f}_{id}' \mathbf{L}' \mathbf{X}_i' \mathbf{R}_{i(-d)} \mathbf{X}_i \mathbf{L} \mathbf{f}_{id} \\ \mathbf{R}_{i(-d)} &= \mathbf{I} - \mathbf{1}(\mathbf{1}' \mathbf{1})^{-1} \mathbf{1}' - \frac{1}{N_i} \mathbf{M}_{i(-d)} \mathbf{M}_{i(-d)}' \end{aligned}$$

が得られる。ただし、「日本人の国民性調査」データに適用した場合には、 $m_{id}$ の次数がかなり大きくなる上、 $m_{id}$ の値を結果の解釈に利用することはほとんどないと思われるため、実際の計算では、

$$m'_{id}X_i = c^{-1/2}f'_{id}L'X_iR_{i(-d)}X_i$$

を求める。

### 3.4 局所解について

全てのデータセットに共通の構造  $L$  を、 $F_i$  によって変換することで、各データセットの構造が得られる、という考え方は、同時主成分分析 (Kiers and Ten Berge (1989))、因子分析法における PARAFAC モデル (Mayekawa (1987))、多次元尺度構成法における Horan モデル (Horan (1969), 齋藤 (1980)) の考え方に似ている。すなわち、同時主成分分析では、データセットごとの成分得点  $F_i$  を求めるための重み  $W$  は、各データセットに対して一定であると仮定する。また、PARAFAC モデルでは、共通の因子負荷行列  $A$  に重み  $W_i$  を乗じたものが、各データセットの因子負荷行列になると仮定する。さらに、INDSCAL (Carroll and Chang (1970)) や ALSCAL (Takane et al. (1977)) に代表される Horan モデルでは、全ての個体に共通の刺激空間  $X$  を  $W_i$  というパラメタで重みづけることにより、個体  $i$  の刺激空間が得られる、という仮定をおく。これらのモデルでは、本章で述べたのと同じように、最小化基準を各パラメタについて交互最小化することにより、パラメタ推定を行なう。

ところで、このようなアルゴリズムには、局所最小値が存在する (Takane et al. (1977))。次章以降で述べる分析を行なった際にも、初期値の選び方により、得られる結果が異なる場合があった。本論文では、初期値を様々に変えて、 $\sigma$  の値が最も小さくなったときの結果を示している。ALSCAL では、初期値を工夫することにより、局所解の問題を克服しようとしているが、本論文で述べたモデルにおいても、今後、局所解を避ける工夫が必要であろう。

## 4. 人工データの分析例

まず、人工データに対してモデルを適用することによって、モデルの性質や結果の見方について述べる。

用いるデータは、表3に示す三つのデータセット群である。三つの項目は、それぞれ三つの選択肢を持つ。第1のデータセットに対して数量化3類あるいは等質性分析を適用し、2番目の解まで求めたとき、選択肢の散布図は放物線を描くことが知られている。また、第3のデータセットを分析すると、選択肢は円を描く。すなわち、表3に示したのは、項目Aの第1選択肢(以後、選択肢A-1と略記)と項目Cの第3選択肢(選択肢C-3と略記)との間の関連が次第に強くなるようなデータセット群である。このデータセット群に対して、図示を容易にするため、 $R=3$ ,  $D=2$ としてモデルを適用した。

計算の結果得られた  $F_i$  の値は、各データセットの構造が似ているか否かを表わすものである。ただし、3.2節で述べたように、パラメタは回転に対する不定性がある。したがって、各  $F_i$  の値を直接比較することはできない。しかし、 $R-D=1$  の場合には、

$$F'_i g_i = 0, \quad g'_i g_i = 1 \quad (i=1, \dots, I)$$

を満たす法線ベクトル  $g_i$  が、符号を除いて一意に定まる。そのため、 $g_i$  の値を比較することで、構造変化の様子を知ることができる。なお、 $R-D>1$  の場合には、例えば、各データセットの



表3. 人工的な三つのデータセット.

個 体	データセット								
	1			2			3		
	項 目	項 目	項 目	項 目	項 目	項 目	項 目	項 目	項 目
A	B	C	A	B	C	A	B	C	
1	1	1	1	1	1	1	1	1	1
2	2	1	1	2	1	1	2	1	1
3	2	2	1	2	2	1	2	2	1
4	2	2	2	2	2	2	2	2	2
5	3	2	2	3	2	2	3	2	2
6	3	3	2	3	3	2	3	3	2
7	3	3	3	3	3	3	3	3	3
8	—	—	—	1	3	3	1	3	3
9	—	—	—	—	—	—	1	1	3

表中の数字は、各項目で選ばれた選択肢を示す。  
また、“—”はデータがないことを表わす。

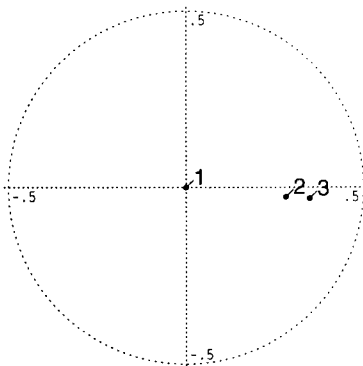


図2. 三つの人工データセットの  $g_i$  の散布図.

$h_i = F_i w_i$  になるべく似るような  $w_i$  を求め、この  $h_i$  を比較するといった方法も考えられるかもしれないが、この点に関しては今後の課題である。

$g_i$  は数値のままでは分かりにくいので、 $g_i$  の第1要素を横軸座標、第2要素を縦軸座標として、各データセットに対する  $g_i$  をプロットした。その際、 $g_i$  の第3要素の符号が正になるよう、各要素の符号を定めた。その結果を図2に示す。

$F_i$  に課した制約から、 $g_1$  の位置は原点となる。仮に、項目間の構造が変化していなければ、各  $g_i$  は原点に一致する。しかし、実際には  $g_1$  と  $g_2$  は大きく離れており、 $g_2$  と  $g_3$  は接近している。すなわち、第1データセットと第2データセットの間に大きな構造変化があったことが示唆されている。ただし、ここで気をつけなければならないことは、 $g_i$  間の差の絶対的な大きさが、構造の変化の大きさに対応するわけではない、という点である。「日本人の国民性調査」データの分析結果でも見るように、 $L$  の値によっては、 $g_i$  の変化が小さくとも、大きな構造変化が見られることがある。 $g_i$  の相対的な位置関係のみが重要、ということである。

具体的にどのような変化があったのかを知るためには、 $L$  の値と、それを各  $F_i$  上に射影した  $LF_i$  の値を調べる必要がある。そこで、 $L$  を3次元上にプロットしたものを図3、 $LF_i$  を2次元上にプロットしたものを図4に示す。図2に示された各データセットの  $g_i$  の位置は、図3のグレーで表わされた平面上に、全ての選択肢を覆う透明の大きな半球を置いたとき、その半球上のどの位置から選択肢群を見ているか、ということを表わすものである。図2と図3を見比べてみれば分かるように、第1のデータセットでは、選択肢群を真上からながめていたことになる。すなわち、図3の各点のグレーの平面上への射影が第1のデータセットの構造となる。射

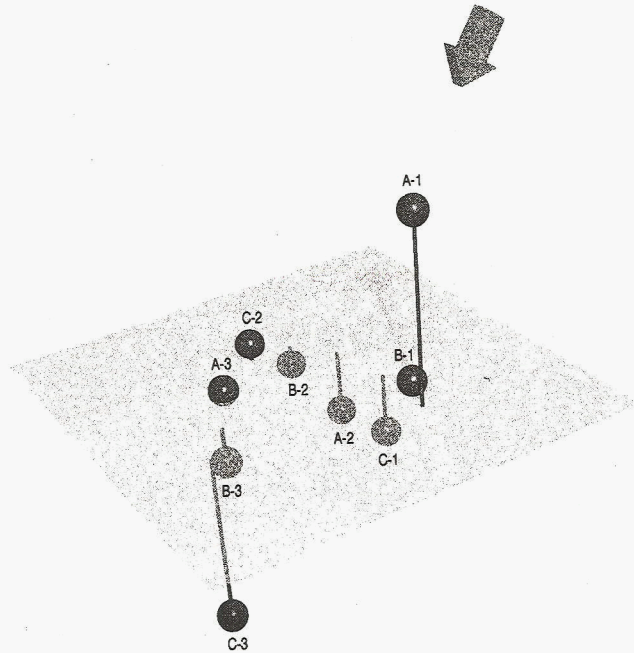


図3. 三つの人工データセットの  $L$  の3次元散布図.

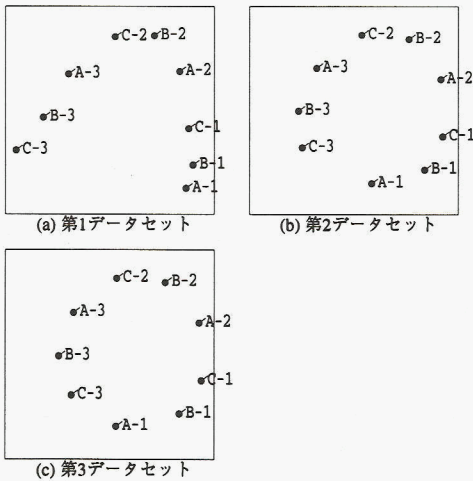


図4. 三つの人工データセットにおける  $L$  の  $F_1$  への射影図.

影された図は、図4(a)に見られるように放物線を描き、よく知られている結果に一致する。放物線が右上に凸となっているが、これは  $L$  が回転に対して不定性を持つため、たまたまそのような形になっただけである。数量化3類で得られる第1軸、第2軸というような方向はこのモデルには存在しない。第3のデータセットでは、選択枝群を図3の矢印の方向からながめていたことになる。ただし、図の矢印は説明のため便宜的に示したもので、その位置や角度は厳密には正確ではない。射影された結果は、図4(c)の円となり、このデータセットの構造が正しく表現される。図4(b)に示される第2データセットの構造は、図2において、 $g_2$  が  $g_3$  に近いことから分かるように、第3データセットの構造に近い。以上のように、 $L$  と  $F_i$  を用いることで、

異なる構造を持つデータセット群の構造の様子を統一的に表現できることが、本モデルの特徴である。

ところで、図3においてグレーの平面からの偏差の大きい選択枝が射影される位置は、 $g_i$  が図2の原点からわずかにずれただけでも第1データセットのときの位置とは大きく異なる。すなわち、偏差の大きい選択枝が大きな変化を示した、ということが出来る。図3では、選択枝

A-1 や選択肢 A-3 がそれに当たる。したがって、それらの選択肢の動きに着目して、どのような変化があったのかを探ることになる。この章の例では、図 4 を見れば明らかであるが、第 3 データセットでは、選択肢 A-1 と選択肢 A-3 の関連が、第 1 データセットに比べて強い、ということが出来る。そして、この解釈は人工データセット群を作成したときの方針に一致している。

## 5. 「日本人の国民性調査」データの分析

“義理人情”，“日本の水準”，“伝統対近代” に関してそれぞれ継続的に用いられている質問項目群の分析結果を示す。ただし、分析に使用する質問項目に対して、一つでも‘その他’，‘D.K.’ を選んだ回答者は分析から除外した。ところで、あまり多くの次元を求めても解釈が難しくなる上、結果も見にくくなる。また、林 (1992) の結果から、 $R=D+1$  でも変化を十分にとらえきれると考えられる。そこで、以下の分析は全て、 $R=3$ 、 $D=2$  として行なった。

### 5.1 “義理人情”

まず始めに、冒頭で述べた“義理人情”に関する質問項目群を取り上げる。林 (1992) は、これらの項目群の構造がほぼ安定していることを見出しているが、本論文のモデルによってさらに付加すべき結果を得たため、ここでその結果を述べる。

質問項目は表 1 に示すとおりである。林 (1992) では、# 5.1d “大切な道徳” も含めているが、この質問項目は複数回答を要求しており、データはアイテムカテゴリ型であるという本論文のモデル仮定から外れるため、分析には用いなかった。1963 年 (第 3 回) から 1993 年 (第 9 回) までの 7 回のデータを分析し、得られた  $L$  の値を表 1 に示した。

まず、質問項目間の構造変化の様子を見るために、 $F_i$  の値、すなわち法線ベクトル  $g_i$  の値を調べる。 $g_i$  をプロットした結果を図 5 に示す。

調査年の順に図 5 の各点を追ってみると、1963 年には中心にあった位置が、1968 年には右上の方向へ動いている。しかし、1973 年には、また原点に戻るような動きを示し、1978 年には、逆に左下の方向へ大きく位置を変えている。1983 年、1988 年と、変化は小さいものの、さらに左下へ動き、1993 年になって動きが止まったように見える。すなわち、1963 年から 1968 年にかけて、ある方向へ構造が変化していった。しかし、1973 年からは、それとは逆の方向への大

きな変化が始まり、その変化は 1980 年代に至るまで続いていった、と言えるであろう。本論文の最初に示した図 1 は少しずつ異なっている。それらの中に、系統的な変化があったということになる。特に、1970 年は公害が深刻化した年であり、1973 年は第 1 次オイルショックの年である。この時期は、これらの出来事に象徴されるように、戦後の高度成長のひずみが見えはじめた時期である。このような時代背景が、義理人情の考え方にも影響を与えたのではないかと想像される。

具体的な変化の様子を調べるために、表 1 の  $L$  の値を 3 次元にプロットしたものを図 6 に示す。グレーの平面からの偏差が大き

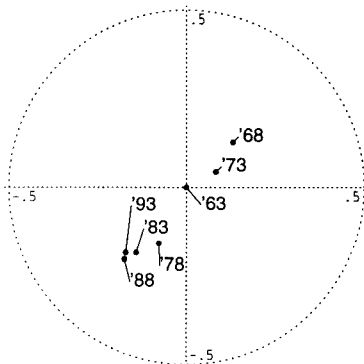


図 5. 7 回の調査時期の  $g_i$  の散布図 (“義理人情”に関する質問項目群)。

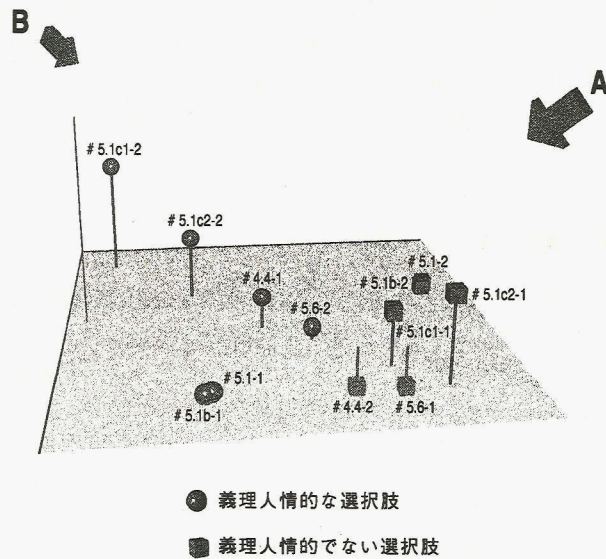


図6.  $L$  の3次元散布図 (“義理人情”に関する質問項目群)。

い選択肢は、#5.1c1 や#5.1c2 の選択肢である。特にこれらの選択肢の動きに注目することになる。各年の射影図  $LF_i$  については数が多くなる上、図1で数量化3類の結果を示しているの、ここでは省略する。

図5と見比べると、1968年の時点では図6の矢印Aの方向から選択肢群をとらえていたことが分かる。この方向から見ると、義理人情的でない選択肢どうしは近くに見える。例えば、#5.1c2-1と#5.1c1-1を見ればよい。逆に、#5.1c1-2と#5.1c2-2を見れば分かるように、義理人情的な選択肢どうしは離れる。すなわち、1968年には、義理人情的でない選択肢間の関連が強かった、と言うことができる。

1978年以降は、図6の矢印Bの方向へ移動する。この方向から見ると、Aの方向とは逆に、義理人情的でない選択肢どうしは離れ、義理人情的な選択肢どうしが近づく。すなわち、1978年以降は、義理人情的な選択肢間の関連が強まってきた、ということになる。

実際に、図1の数量化3類の結果を見ると、1978年以降の義理人情的な選択肢のバラツキ具合は、1968年のものよりも小さい。

また、義理人情的な選択肢間の関連が強いということは、それらが同時に選ばれることが多い、ということの意味する。したがって、義理人情的な選択肢をいくつ選んだか、という義理人情スケール(林(1992))をつくれれば、1978年以降は、その値が大きくなることが予想される。値の分布を調べると、確かに、“1963年から10年間はやや義理人情的でない方に動くかに見えたが、1978年からまたもとの分布の形にもどり、今日まで続いているのである(林(1992))。”

以上の結果をまとめると、義理人情的なものとそうでないものとを対比させる、という大局的な考えの筋道に変化はない。しかし、詳細に見ると、1963年から1968年にかけては、義理人情的でない選択肢の関連が強まった。1973年には、それとは逆の動きが起こり、1978年からは義理人情的な選択肢の関連が強くなってきている、と言える。

## 5.2 “日本の水準”

次に、 $L$ の値を調べることにより、項目群のグループ分けを行なうことができた結果を述べ

表 4. “日本の水準”に関する質問項目・選択肢の  $L$ .

	$r=1$	$r=2$	$r=3$
<u>#.9.12 ‘日本の「科学技術の水準」</u>			
a 現在の日本は「科学技術の水準」という点では、つぎの4つのどれに当たると思いますか？			
1 非常によい	-.181	.728	.657
2 ややよい	-.115	-.559	.038
3 ややわるい	1.429	-.480	1.203
4 非常にわるい	2.697	.620	11.908
<u>#.9.12b ‘日本の「芸術」</u>			
b では、「芸術」という点では、日本はどれでしょうか？			
1 非常によい	-.100	1.247	-.155
2 ややよい	-.245	-.265	-.357
3 ややわるい	.646	-.337	-.770
4 非常にわるい	2.531	-.206	-1.899
<u>#.9.12c ‘日本の「経済力」</u>			
c では、「経済力」という点では、日本はどれでしょうか？			
1 非常によい	-.313	.977	1.227
2 ややよい	-.306	-.466	.090
3 ややわるい	.361	-.478	2.951
4 非常にわるい	2.258	.395	5.525
<u>#.9.12d ‘日本の「生活水準」</u>			
d では、「生活水準」という点では、日本はどれでしょうか？			
1 非常によい	-.408	1.701	1.568
2 ややよい	-.470	-.185	.357
3 ややわるい	.302	-.345	2.645
4 非常にわるい	1.993	.291	7.004
<u>#.9.12e ‘日本の「心の豊かさ」</u>			
e では、「心の豊かさ」という点では、日本はどれでしょうか？			
1 非常によい	-.215	2.047	1.487
2 ややよい	-.413	-.082	-.605
3 ややわるい	-.031	-.198	-.175
4 非常にわるい	1.397	.149	1.152

る。

1973年(第5回), 1978年(第6回), 1988年(第8回), 1993年(第9回)の4回にわたって, “日本の水準”に関して, 表4に示す質問がなされた。表4には, この4回の調査データから求められた  $L$  の値も示されている。

図7は, 4回の  $g_i$  の値をプロットしたものである。1973年と1978年, 1988年と1993年の間ではそれぞれ項目間の構造は似ているが, 両者の間には違いがあることが読み取れる。すなわち, 1980年代に“日本の水準”に関する質問項目群の構造に何らかの変化があったことがうかがえる。1983年にこの質問項目を実施していれば, より細かい動きを知ることができたかもしれない。

具体的な変化の様子を調べるために、 $L$ の値をプロットしたものを図8に示す。ただし、表4に示された $L$ の $r=3$ に対する値に非常に大きなものがあるため、図8でグレーの平面に直交する方向の長さを3分の1の縮尺にしてある。また、図8では、同じ質問項目に属する選択肢を線で結んでいる。線で結ばれた選択肢がほぼ放物線を描いていることから、ここで取り上げた質問項目の選択肢は順序をなしていることが分かる。また、‘ややわるい’、‘非常にわるい’といった選択肢の動きが大きかったことがよく分かる。

ところで、図8で近くにある選択肢どうしは、どの方向へ射影しても常に近くにある。それらの選択肢は、ある特定の年だけでなく、常に強く関連していたということになる。そこで、特に‘ややわるい’や‘非常にわるい’といった選択肢の位置関係に注目し、5つの項目を見てみると、それらをほぼ3つのグループに分けることができる。すなわち、#9.12c‘日本の「経済力」’と#9.12d‘日本の「生活水準」’、#9.12‘日本の「科学技術の水準」’、#9.12b‘日本の「芸術」’と#9.12e‘日本の「心の豊かさ」’である。これらの3つのグループは、それぞれ1980年代に同じような変化をした項目のグループということになる。

“経済力”と“生活水準”の‘非常にわるい’、‘ややわるい’という選択肢どうしは近くにあ

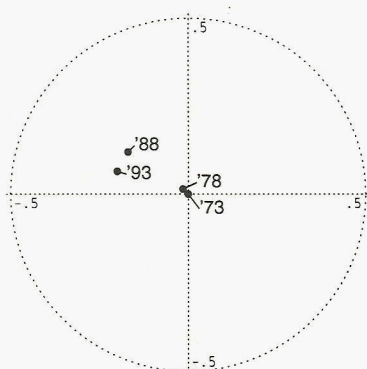


図7. 4回の調査時期の $g_i$ の散布図(“日本の水準”に関する質問項目群)。

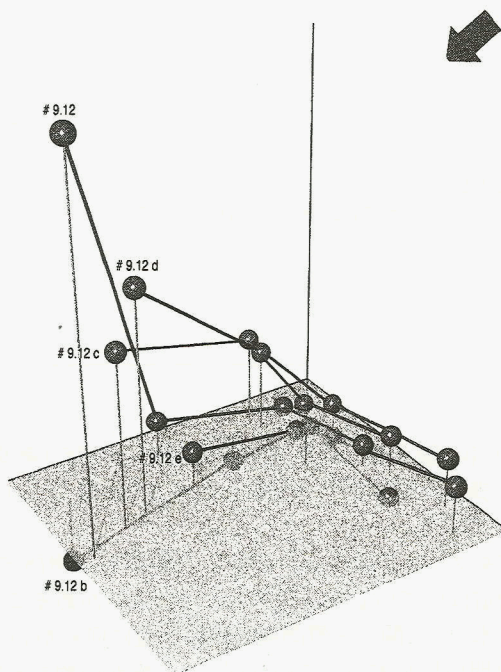


図8.  $L$ の3次元散布図(“日本の水準”に関する質問項目群)。



る。20年間、“経済力”がわるいと思っている人々にとっては、“経済力”と“生活水準”は密接な関連がある、ということである。ただし、‘非常によい’という選択肢どうしはやや離れているため、“経済力”がよいと思っている人々が、必ずしも“生活水準”がよいと考えているわけではない、ということが分かる。

“科学技術の水準”の‘非常にわるい’という選択肢の変化は非常に大きく、この項目を特徴づけている。また、各選択肢の位置が、図8で言えば全体的に左寄りになっている。1980年代は、

表5. “伝統対近代”に関する質問項目・選択肢のL.

	r=1	r=2	r=3
<b># 2.1 ‘しきたりに従うか’</b>			
あなたは、自分が正しいと思えば世のしきたりに反しても、それをおし通すべきだと思いますか、それとも世間のしきたりに、従った方がまちがいないと思いますか？			
1 おし通せ	-.049	-.692	1.016
2 従え	-.738	.377	-.934
3 場合による	.977	.494	-.316
<b># 2.5 ‘自然と人間の関係’</b>			
[リスト] 自然と人間との関係について、つぎのような意見があります。あなたがこのうち真実に近い（ほんとうのことに近い）と思うものを、1つだけえらんで下さい？			
1 人間が幸福になるためには、 自然に従わなければならない	-.865	.541	-2.521
2 人間が幸福になるためには、 自然を利用しなければならない	.361	-.072	-.268
2 人間が幸福になるためには、 自然を征服してゆかななければならない	.270	-.418	4.351
<b># 4.5 ‘子供に「金は大切」と教える’</b>			
小学校に行っているくらいの子供をそだてるのに、つぎのような意見があります。「小さいときから、お金は人にとって、いちばん大切なものだ」と教えるのがよい」というのです。あなたはこの意見に賛成ですか、それとも反対ですか？			
1 賛成	-.402	.161	1.595
2 反対	.457	-.939	.789
3 いちがいいはいえない	1.414	.919	1.184
<b># 4.10 ‘他人の子供を養子にするか’</b>			
子供がないときは、たとえ血のつながりのない他人の子供でも、養子にもらって家をつがせた方がよいと思いますか、それとも、つがせる必要はないと思いますか？			
1 つがせた方がよい	-.292	.157	2.014
2 つがせないでもよい、意味がない	.350	-.801	1.280
3 場合による	1.169	1.072	-.157
<b># 7.4 ‘日本と個人の幸福’</b>			
[リスト] あなたはつぎの意見の、どちらに賛成ですか。1つだけあげてください？			
1 個人が幸福になって、はじめて日本全体がよくなる	.273	-.324	-.441
2 日本がよくなって、はじめて個人が幸福になる	-.397	.317	.876
3 日本がよくなることも、個人が幸福になることも同じである	.183	-.063	.243

‘日本の「科学技術の水準」の高さを実感させるような時期であったということであろう。

“芸術”と“心の豊かさ”はやや離れているように見える。しかし、他の3項目に比べると、グレーの平面からの偏差が小さく、変化が小さかった、という意味において似ていると言える。1988年、1993年は図8の矢印の方向から項目をとらえているため、‘非常にわるい’という選択肢どうしの距離が近くなり、関連は強まっている。

### 5.3 “伝統対近代”

次に、林 (1992) が分析した“伝統対近代”に関する質問項目群を分析する。これらの項目群の構造は、変化してきたことが分かっているが (林 (1992)),  $g_i$  を可視化することで、その変化がいつ起きてきたのかが、分かりやすくなったため、ここでその結果を述べる。

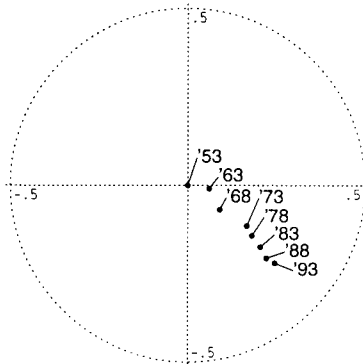


図9. 8回の調査時期の  $g_i$  の散布図 (“伝統対近代”に関する質問項目群).

質問項目は表5に示すとおりである。#8.1 ‘政治家にまかせるか’は、1978年から選択肢が変更されたので、今回の分析では用いなかった。1953年 (第1回), 1963年 (第3回) から1993年 (第9回) の8回のデータを分析した結果を同じく表5に示す。また、今までと同様に、 $g_i$  をプロットしたものを図9に示す。

図9には、“伝統対近代”という質問項目群の構造が、年を経るごとに徐々に変化してきている様子をはっきりと表われている。その変化の形は、最初に示した“義理人情”のようなUターンを示すものではなく、一方向的である。林 (1992) では、1973年、1978年から“伝統対近代”という構図が崩れ出したと述べられている。確かに、1968年と1973年の間は、他の年の間よりも離れているが、図9では、1963年、1968年の時点で、既に右下の方向へ向かって動き出しており、構造の変化する徴候が見られていたことが分かる。そして、1973年になってその動きが大きくはっきりと現れてきたのである。

具体的にどのような変化があったのかを見るために、 $L$  をプロットしたものを図10に示す。図10の矢印は今までと同様に、 $g_i$  の変化した方向を便宜的に示したものである。この矢印の方向から散布図を見ると、伝統的な選択肢と近代的な選択肢が混ざりあって見える。すなわち、“伝統対近代”という構図が崩れてきている、という林 (1992) の主張を支持する結果が得られた。

図10を見ると、特に変化の大きかった選択

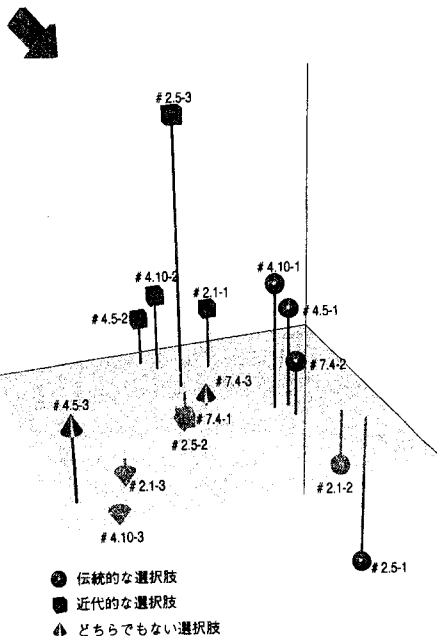


図10.  $L$  の3次元散布図 (“伝統対近代”に関する質問項目群).



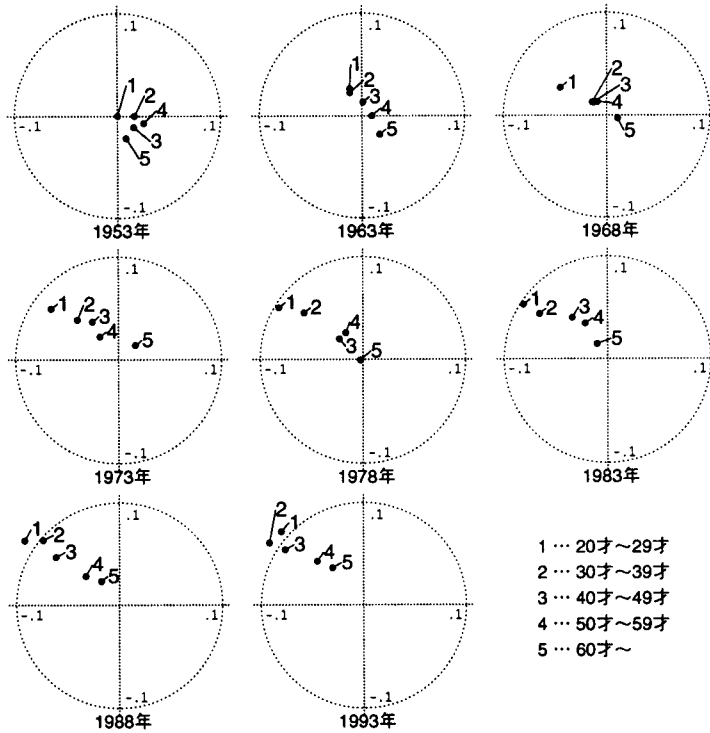


図 11. 年齢層ごとの 8 回の調査時期の  $g_i$  の散布図 (“伝統対近代”に関する質問項目群)。

肢は、# 2.5 ‘自然と人間との関係’という質問項目の‘自然に従わなければならない’、‘自然を征服してゆかなければならない’である。

# 4.5 ‘子供に「金は大切」と教える’と # 4.10 ‘他人の子供を養子にするか’という質問項目の‘賛成’と‘つがせた方がよい’、‘反対’と‘つがせないでもよい、意味がない’という選択肢どうしは互いに近い位置にある。“伝統対近代”という構図が崩れている中で、この二つの質問項目に対しては、共通する考えの筋道がある、ということが示唆される。

#### 5.4 年齢層別“伝統対近代”

データセットの数が非常に多いときに、本論文のモデルを用いた  $g_i$  の可視化が、構造変化の様子を捉えるために有効なことを示す別の分析結果を最後に述べる。

前節と同じ“伝統対近代”に関する質問項目群において、各年のデータをさらに年齢層別に分割したデータセットを作成した。その際、各年齢層に属する回答者数があまり少なくならないよう、回答者を全部で五つの年齢層に分類した。すなわち、20 歳～29 歳、30 歳～39 歳、40 歳～49 歳、50 歳～59 歳、60 歳以上、である。各調査年の各年齢層が一つのデータセットとなるため、分析に用いるデータセットの総数  $I$  は、5 年齢層 × 8 調査回数 = 40 となる。40 個の  $g_i$  を調査年ごとにプロットしたものを図 11 に示す。また、40 個のデータセットに対して共通の  $L$  が一つ求められるが、前節で示したものとほぼ同じ形が得られたので、その図示については割愛する。

1953 年では、年齢層間に特にはっきりとした傾向は見られない。ところが、1963 年からは、五つの年齢層が一列となって、図の左上の方向へ向かって動き出している様子がよく分かる。す

なわち、20歳～29歳を先頭に、30歳～39歳、40歳～49歳、50歳～59歳、60歳以上と年齢層順に続いて動いている。まず若い年齢層の人々の考えの筋道が変化し、それが徐々に高年齢層にも伝わっていった、ということである。回答分布の形が高年齢層へ伝わっていく様子を調べることは難しくない。しかし、構造が伝わる様子は、本論文のモデルを使ってはじめて明確になったと言える。

構造の伝わり方には、コウホートによるものや加齢、時代の流れによるものなどを考えることができる。それらを分離するためには、コウホート分析的な考え方をモデルに取り入れる必要がある。

ところで、1953年では各年齢層の位置がバラバラに見える。また、図の円の中心部と辺縁部とでは見かけの距離がやや変わる、ということも考慮しても、1988年、1993年では若い年齢層の動きが少なくなってきている。これらの原因として、次の二つが考えられる。

まず第一には、 $R=3$ では、1953年や1993年の考えの筋道の変化を表現するには不十分である、ということが挙げられる。変化は、1953年に始まり1993年に終わるわけではない。1953年以前から続いており、今後も続いていくと考えられる。調査を行なった40年間の変化は $R=3$ でとらえられるが、それ以外の時期の変化をとらえるには、 $R=4$ や $R=5$ とすることが必要なのではないか、ということである。

第二には、特に、動きが少なくなった原因として、各年齢層における考えの筋道の変化をとらえるためには、“伝統対近代”に関する質問項目群では不十分になってきている、ということが挙げられる。従来の質問項目群は、“伝統対近代”を対比させるか、させないか、ということ調べるには適している。しかし、既に対比させなくなってきている以上、再び“伝統対近代”を対比させるように変化していかない限り、今後生じる別の変化を探るには新たな質問項目群が必要となる。

## 6. ま と め

本論文では、複数のデータセット間の構造の違いを探るモデルを提案した。このモデルでは、 $F_i$ により構造の類似性を表現することができた。また、 $L$ を図示することで、どのような構造の違いがあるのか容易に把握することが可能となった。そして、このモデルを「日本人の国民性調査」データに適用することで、質問項目の構造変化の様子が明らかとなった。今後、国際比較のために収集されたデータ等にも適用し、モデルの有効性をさらに検証することが求められる。 $D < R$ としたことが、本モデルの特徴の一つであるが、国際比較のような、構造が大きく異なると予想されるデータセット群に対しては、 $R-D$ を大きくする必要がある。その際、3章で述べた局所解、 $R$ や $D$ の値の選択、結果の図示の方法といったことが問題になると考えられる。

## 謝 辞

本論文に対して有益な御助言を頂いた統計数理研究所 林知己夫名誉教授、駒澤 勉教授、坂元慶行教授、中村 隆助教授に感謝いたします。

## 参 考 文 献

穂山貞登 (1993). 『数量化のグラフィックス』, 行動計量学シリーズ, 朝倉書店, 東京.

- Carroll, J.D. and Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of "Eckart-Young" decomposition, *Psychometrika*, **35**, 283-319.
- Gander, W. (1981). Least squares with a quadratic constraint, *Numer. Math.*, **36**, 291-307.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*, Wiley, Chichester.
- 林知己夫 (1982). 科学文明観の位置づけ, 『第4日本人の国民性』(統計数理研究所国民性調査委員会 編), 208-214, 出光書店, 東京.
- 林知己夫 (1992). 日本人の国民性について, 『第5日本人の国民性』(統計数理研究所国民性調査委員会 編), 187-210, 出光書店, 東京.
- 林知己夫 (1993a). 『行動計量学序説』, 行動計量学シリーズ, 朝倉書店, 東京.
- 林知己夫 (1993b). 『数量化——理論と方法——』, 朝倉書店, 東京.
- Horan, C.B. (1969). Multidimensional scaling: combining observations when individual have different perceptual structures, *Psychometrika*, **34**, 139-165.
- Kiers, H.A.L. and Ten Berge, J.M.F. (1989). Alternating least squares algorithms for simultaneous component analysis with equal component weight matrices in two or more populations, *Psychometrika*, **54**, 467-473.
- Mayekawa, S. (1987). Maximum likelihood solution to the PARAFAC model, *Behaviormetrika*, **21**, 45-63.
- Meulman, J. (1982). *Homogeneity Analysis of Incomplete Data*, DSWO Press, Leiden.
- 大隅 昇, L. ルバル, A. モリノウ, K.M. ワーウィック, 馬場康維 (1994). 『記述的多変量解析法』, 日科  
技連出版社, 東京.
- 齋藤堯幸 (1980). 『多次元尺度構成法』, 朝倉書店, 東京.
- Takane, Y., Young, F.W. and de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features, *Psychometrika*, **42**, 7-67.
- Tenenhaus, M. and Young, F.W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data, *Psychometrika*, **50**, 91-119.
- 柳井晴夫 (1994). 『多変量データ解析法——理論と応用——』, 行動計量学シリーズ, 朝倉書店, 東京.

## A Homogeneity Analysis for Exploring Structural Differences among Several Data Sets : Its Application to Data from the Surveys of the Japanese National Character

Takahiro Tsuchiya

(The Institute of Statistical Mathematics)

In this paper, the author proposed a new model to explore structural differences among several categorical data sets. The model is an extension of a homogeneity analysis and produces two kinds of figures. One illustrates structural similarities among the data sets and the other gives detailed information on the differences among those data sets.

The performance of this model is demonstrated by its successful application to three examples obtained from the data of the study on the Japanese national character. From the results of the analysis of the first example data, we found that attitudes toward *giri* had changed around 1973. The analysis of the second example showed that the rating of people for the level of Japan had structurally changed in 1980's. Finally, from the analysis of the third example, it was revealed that the distinction between traditional vs. modern attitudes was gradually becoming ambiguous. These results coincide with findings obtained from our intuitive analysis.