完全単純構造・主クラスター成分分析・resampling による確認：心理測定尺度構成のための単純な分析方法試論

村　上　　　隆

『中京大学現代社会学部紀要』　第８巻　第１号　抜　刷

２０１４年９月　　PP. 47~90

完全単純構造・主クラスター成分分析・resampling による確認：心理測定尺度構成のための単純な分析方法試論

村　上　　　隆

1．心理測定のための因子分析の利用

複数の質問項目からなる inventory を多数の対象者に対して実施し，その質問項目への回答を何らかの形で合成して個人の得点を求める心理学的尺度構成のために，主成分分析を含む広い意味の因子分析を用いることは，社会心理学，性格心理学，教育心理学といった領域ではごく普通の手続きである。特に，質問項目に対して，「賛成」，「やや賛成」，「どちらでもない」，「やや反対」，「反対」といった５段階前後の（順序のついた）カテゴリカルな判断を求め，それをそのまま 1~5（あるいは 1~7）の整数値として複数の項目にわたって加算して得られる合計点（sum scores），それを尺度得点とする方法は，態度測定の分野では Likert 法と呼ばれているが，実際，項目反応の形だけでなく，それらにもとづく個人差測定尺度の作成のための基本的手順は Likert(1932) にまとめられている。事実，この方法で作られ，使用されている尺度は膨大に存在する（たとえば，Robinson, Shaver, & Wrightsman, 1991)。

１群の項目から尺度を複数構成する場合には，各項目反応をそのまま数値（心理測定用語では間隔尺度）とみなして行われる（広義の）探索的因子分析が，項目を相互相関にもとづいて幾つかのクラスターに分類する手

段として用いられている（Carmine & Zeller, 1979）。すなわち，単純構造化された因子負荷行列の列ごとに，高い（salient）負荷量をもつ項目を選び出し，項目を相互に背反なグループに分類して，それらの項目反応の合計点によって個人差測定尺度を定義するわけである。

このような手続きは，従来，必ずしも厳密な理論的考察を経ないまま，一種の heuristics として活用されてきた。それにもかかわらず，パーソナリティの 5 因子理論のような個人差査定の重要な枠組みを，試行錯誤を経ながらほぼ確立してきた（Digman, 1996）ように，因子分析は，種々の心理学研究の中で重要な役割を果たしてきた（たとえば，Streiner & Norman, 2003）。ただし，近年の確認的因子分析の手法の台頭により，研究の進行とともに，探索的因子分析の守備範囲は徐々に狭まっており，主に尺度構成の初期の段階に限定されつつある。しかしながら，明確な概念の定義も仮説も存在しない研究の初期の段階では，他に替わりうる方法がなく，探査的分析の必要性は失われていない。

構成された尺度の質は，しばしば信頼性係数の下限である Cronbach の α 係数（Cronbach, 1951）によって評価される。これは項目間相関係数がほぼ等しいという条件のもとに，次で算出される。

$$\alpha_l = \frac{p_l \overline{r_l}}{1+(p_l-1)\overline{r_l}} \tag{1}$$

ここで，p_l は l 番目のクラスターに属する項目数，$\overline{r_l}$ はそれらの項目間相関係数の平均値である[i]。

実際，個々のクラスター内部で平行測定（parallel measurement），すなわち，すべての項目が同じ真の得点と同じ分散をもつランダム誤差であるという仮定が成立するならば，すべての項目間相関係数は等しいことになる（たとえば，Carmine & Zeller, 1979, p.33）。その相関係数を r_l とすると，そうした条件の下で α 係数は，次のように書けることになる。

$$\rho_l = \frac{p_l r_l}{1+(p_l-1)r_l} \tag{2}$$

この ρ_l は，尺度 l の（下限ではなく）正確な信頼性係数となる（Spearman−Brown の公式，(1)の方が一般形とされる，Nunnally, 1978）。

このように考えてくると，研究者は項目反応データに因子分析を適用にあたって，だいたい次のようなことを期待していることになる。ここでは，仮に，質問項目が q 個のクラスターに分かれており，それぞれのクラスターは，それぞれ，$p_1, p_2, \cdots, p_q$ 個の項目からなっているものとしよう。このとき，$p = p_1 + p_2 + \cdots + p_q$ である。さらに，項目を適当に並べ替えることによって，同一のクラスターに属する項目を隣接する位置にもってくることができたとする。そうすると，項目間相関行列は次のように分割できる。

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1q} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{q1} & \mathbf{R}_{q2} & \cdots & \mathbf{R}_{qq} \end{bmatrix}$$

個々の下位行列のうち主対角要素となる $\mathbf{R}_{ll}$ は，平行測定の仮定にしたがってすべての項目間相関が等しく，次の形をとる。

$$\mathbf{R}_{ll} = \begin{bmatrix} 1 & r_l & \cdots & r_l \\ r_l & 1 & \cdots & r_l \\ \vdots & \vdots & \ddots & \vdots \\ r_l & r_l & \cdots & 1 \end{bmatrix}$$

もし，$\mathbf{R}_{ll'} = \mathbf{O}\left(l \neq l'\right)$ ならば，対応する負荷行列は次のようになるであろう（ここでは，共通性推定を伴う狭義の因子分析を想定している）。なお，$\mathbf{O}$ はゼロ行列である。

$$\mathbf{A}=\begin{bmatrix} \sqrt{r_1} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \sqrt{r_1} & 0 & \cdots & 0 \\ 0 & \sqrt{r_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \sqrt{r_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{r_q} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{r_q} \end{bmatrix}$$

負荷行列の各行の非ゼロ要素が最大1個であるならば，完全単純構造（perfect simple structure）という（Bernaard & Jennrich, 2003）。上の $\mathbf{A}$ は完全単純構造となっている。なお，$\mathbf{R}_{ll'}=\mathbf{O}\left(l\neq l'\right)$ でなくても，斜交回転によって完全単純構造に近づけることができる場合もあろう。その際，因子間相関係数は0にならないが，多くの適用例において，項目全体は何らかの単一の構成概念に対応しており，因子はその複数の下位概念に対応すると解される場合が多いから，このことはむしろ好ましいこととも考えられる。

なお, 4節で述べるように, 項目の一部は逆転項目である場合も多いが, ここではすべての逆転項目の反応は反転され，項目間の相関係数はすべて正であるものとしている。

こうした状況が実現していれば，項目を q 個のクラスターに分類し，それらの単純加算による q 次元の尺度を構成することができる。また，(2) によって（実データでは近似的に（1）によって），それぞれの尺度の信頼性を評価することもできる。

しかしながら，もちろんこれは理論の側から見た理想状態にすぎない。実データの分析では，いろいろと判断しなければならない問題が生じてくることになる。

2. 探索的因子分析利用上の問題点

1節で述べたような理念が，現実のデータ分析においてどの程度満たされるかを見るために，1つの例を挙げてみよう。表1は，Mehrabian & Epstein（1972）による情緒的共感性尺度（Emotional Empathy Scale：EES）の日本語版を1645名の大学生に実施した結果（中村，2000）を，主成分分析（いわゆる因子分析主成分解）し，プロマックス回転を行って得られた因子パターン行列である。なおこの表の右端の列，R^2は，通常の因子分析では共通性と呼ばれるが，主成分解では個々の変数の全因子への重回帰による重相関係数の2乗である。パターン行列の係数は，標準偏回帰係数なので，その2乗和とR^2は一致しない。また，斜交回転では全変数による因子は，説明力を因子ごとに分解することができず，表の「因子寄与」の欄は空欄となっている。他方，全体としての説明力は計算でき，これはR^2の和と一致する。

しばしば行われるように，高い負荷量（ここでは，絶対値が0.4以上）をゴシックで示している。負荷量の符号は，因子（得点）が大きいほど（常識的な意味で）共感性が高くなる方向に変換している。すべての因子間相関係数は正だから，これらの因子が全体として共感性と解釈される潜在的な属性と関連していると考えられよう。項目の配列順は，後に本研究で提案する方法による4因子（クラスター）解に合わせてある。因子は順に，「感情の伝染」，「共感の理解」，「感情移入」，「道徳的共感」と名づけることができよう[ii]。

この表では，22, 15など複数の項目の負荷量が低く，どの尺度にも加えることができないと判断されるであろう。信頼性にこだわる限り，一部の項目の取捨選択は必要である。

一般論として。この手続きには，次のような問題点が指摘されるであろう。

1）項目の分類だけが目的であるとすれば，手続きが複雑である。因子抽出と回転の2つの段階に分かれ，それぞれに，多数の選択肢があっ

表１　感情的共感性尺度のパターン行列

質問項目	Ⅰ	Ⅱ	Ⅲ	Ⅳ	R^2
13 友人が動揺しているからといって，自分も動揺するようなことはない。	**-0.73**	-0.04	0.04	0.03	0.52
25 まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	**0.63**	0.01	-0.06	0.22	0.51
10 まわの人の気分によって自分の気分も大きく左右される。	**0.65**	-0.21	0.06	0.03	0.47
9 人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	**0.58**	-0.08	0.14	0.06	0.42
7 友人の悩みに感情的に巻き込まれてしまう傾向がある。	**0.57**	0.07	0.11	0.06	0.41
32 周囲が興奮していても，自分は冷静でいられることが多い。	**-0.68**	-0.18	-0.03	0.19	0.48
16 泣いている人を見ると自分も動揺してしまう。	**0.52**	0.12	0.05	0.21	0.43
20 まわりの人たちがいらいらしていても，自分は平静でいられる。	**-0.68**	0.06	0.09	0.12	0.42
24 人の気持ちに左右されずに判断を下すことができる。	**-0.53**	-0.05	0.07	0.08	0.26
5 まわりに神経質な人がいると，自分も神経質になる。	0.28	**-0.47**	0.01	0.18	0.31
30 他人の涙を見ると，同情するよりいらいらする。	-0.10	**-0.66**	-0.10	-0.01	0.52
6 幸せのあまり泣くというのは，馬鹿げている。	-0.03	**-0.40**	-0.27	-0.13	0.36
21 友人が悩み事を話し始めると，別の事柄に話を変えようと努める。	0.11	**-0.42**	-0.18	-0.13	0.30
23 映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。	-0.05	**-0.39**	-0.31	0.04	0.33
4 不幸な人がすっかりしょげているのを見ると，いらだたしくなる。	0.00	**-0.61**	0.19	-0.15	0.37
26 人がどうしてそんなに動揺しているのかをなかなか理解できない。	-0.13	**-0.49**	0.01	-0.01	0.26
22 他の人の笑いに引きずられて笑ったりはしない。	-0.24	-0.18	-0.25	0.04	0.21
15 孤独な人は多分友情を求めていないのであろう。	0.03	-0.27	-0.09	-0.21	0.16
3 感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。	-0.06	**-0.58**	0.12	0.15	0.32
2 人は動物の感情や感受性について思い入れすぎである。	0.03	-0.37	0.04	-0.06	0.14
11 自分がであった外国人の多くは，クールで，感情的ではないように見えた。	0.22	-0.23	-0.23	0.00	0.16
33 小さい子どもはときどき，はっきりとした理由もなく泣くものである。	-0.05	-0.25	0.20	-0.07	0.07
31 映画を観ると強く引き込まれてしまう。	0.02	0.02	**0.68**	0.04	0.50
18 小説の登場人物の気持ちになりきってしまう。	0.00	-0.19	**0.74**	0.02	0.50

17 歌によって幸せな気分になることがある。	0.01	-0.13	**0.68**	0.00	0.42
28 本や映画の世界に浸るのは，少し馬鹿げている。	0.11	-0.22	**-0.63**	0.09	0.48
8 ときどきラブソングの歌詞に深く感動することがある。	0.24	-0.02	**0.57**	-0.07	0.41
14 人がプレゼントを開けるところを見るのが好きである。	-0.01	0.07	0.31	0.14	0.16
1 グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。	0.00	-0.02	0.01	**0.67**	0.45
27 動物が苦しんでいるのを見るとひどく動揺してしまう。	-0.05	0.11	0.04	**0.61**	0.41
29 無力な老人を見ると動揺してしまう。	0.08	-0.06	-0.10	**0.68**	0.45
19 ひどい目にあっている人を見ると，強い憤りを感じる。	-0.08	0.00	0.23	**0.52**	0.38
12 職業訓練所で働くよりは社会福祉関係の仕事をしたい。	0.02	0.11	-0.06	**0.48**	0.25
主成分寄与					11.84

主成分間相関				
Ⅰ	1.00	0.09	0.23	0.28
Ⅱ	0.09	1.00	0.32	0.15
Ⅲ	0.23	0.32	1.00	0.32
Ⅳ	0.28	0.15	0.32	1.00

て煩雑である。因子抽出と回転が別のプロセスになることの理由の説明も簡単ではない。また，それぞれの段階で必要となる数理的知識は不必要に多く，教育上の隘路となる。特に，回転に関しては，因子分析の弱い環，あるいは「因子分析における最も古くて最も困難の多い問題」と言われることもあり（Harshman & Lundy,1984；p.122），できれば回避したい手続きである。

2）回転に対応して，出力も複数あり，かつ相互の関係は単純でない。特に，プロマックス回転のような斜交回転では，SPSS のような統計ソフトウェアにおいて，回転前の負荷行列，因子を独立変数とし個別変数を従属変数とする標準偏回帰係数の行列であるパターン行列，因子と個別変数との相関係数である因子構造行列がまったく同じ形式で出力され，ときに誤りの原因となる。通常，パターン行列が解釈されるが，これが偏回帰係数であることは，それらの大きさを評価するのに困難にする（あまり指摘されていないが）。

3）因子数の決定が難しい。因子数を１だけ増減することは，１つの因子が２つに分離したり，２つの因子が癒合したりするような結果になる，すなわち，因子構造は階層構造をもっていると見られることが多いが，場合によって，因子数の変化にともなって大きな「組み替え」が生起することがある。この現象の説明は難しく，ほとんどの解説書にも書かれていないが，それに気づいたユーザーを戸惑わせることになる。

4）その一方で，探索的因子分析が出力する負荷行列が説得力を持ちすぎることも問題になる。すなわち，あらかじめ緩やかな仮説を持っていたとしても，因子分析の結果はそれを忘れさせ，因子分析によって得られた項目クラスターがまったく疑う余地のないものとして受け入れられてしまう可能性も高い。これは，探索的因子分析の概念形成の方法としての強力さ（村上，2008）を示すものであるが，こうした複雑な手続きの結果得られた負荷行列には当然標本誤差が加

わっており，それを評価しないまま唯一の数値を信じ続けることは継続した研究の一貫性という観点からは問題になる。

5）しかしながら，現実問題として，ユーザーがまったく何の仮説も持たないで分析にあたることも少ない。すべての項目の所属クラスターが決定できなくても，部分的仮説を分析に反映させることは適切であろう。ここで述べる方法は，簡単にこの問題に対処できるが，これについては稿を改めることにしたい。

6）もう1つ，比較的頻繁に問題にされるのが，Likert 尺度の反応をそのまま数値(間隔尺度)とみなして分析することの正当性である。これについて，筆者は，村上（2013）で1つの対処方法を提案しているが，ここでは，とりあえず間隔尺度とみなす取り扱いを承認した上で，先に進むことにしたい。

以下においては，上記1）～4）の問題点の解消を目指した単純な方法を提案する。

3. 方法の概要

主成分分析　主成分分析の定式化は，大きく分けて2通りある。1つは，データ行列に対して，因子スコア行列と因子負荷行列の積で定義される因子分析モデルを最小2乗法的に当てはめるもの（たとえば，足立・村上，2011）であり，もう1つは，データ行列の各変数の1次結合（合成変量）を分散（の和）の最大化，あるいは，素点と合成変量との相関係数（因子構造に相当する）の2乗（和）の最大化という条件で定義するものである。ここでは，その両方を含めた定式化から出発しよう。

まず，n，p，q をそれぞれ，個体，項目，因子の数とする。$\mathbf{Z}$ を標準化された Likert 型の質問項目への反応の $n \times p$ の行列とする（以下，標準化されたデータ行列と呼ぶ)。すなわち，次の条件を満たす。

$$\mathbf{Z}'\mathbf{1} = \mathbf{0} \tag{3}$$

$$\mathrm{diag}\left(n^{-1}\mathbf{Z}'\mathbf{Z}\right) = \mathbf{I} \tag{4}$$

ここで，**1** はすべての要素が 1 であるような（列）ベクトル，**0** はゼロベクトル，**I** は単位行列である。その結果，次の行列は項目間の相関行列となる

$$\mathbf{R} = n^{-1}\mathbf{Z}'\mathbf{Z} \tag{5}$$

ここで，最小化基準を次のように定義する。

$$\varphi(\mathbf{A},\mathbf{V} \mid \mathbf{Z}) = n^{-1} \| \mathbf{Z} - \mathbf{ZVA}' \|^2 \tag{6}$$

ここで, **V** は $p{\times}q$ の重み行列であり, 以下の制約条件を満たすものとする。

$$\mathrm{diag}\left(\mathbf{V}'\mathbf{RV}\right) = \mathbf{I} \tag{7}$$

すなわち，次の $n{\times}q$ の行列，

$$\mathbf{F} = \mathbf{ZV} \tag{8}$$

の要素である主成分得点（項目反応の 1 次合成変量）は（平均値が 0 で標準偏差が 1 の）標準得点となる。**A** は $p \times q$ の負荷行列であり，行に対応する変数の主成分への標準偏回帰係数の行列である。この基準は，$n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}$ を制約条件とする次の因子モデルの当てはめと実質的に等価であることは，簡単に示される。

$$\varphi(\mathbf{A},\mathbf{F} \mid \mathbf{Z}) = n^{-1} \| \mathbf{Z} - \mathbf{FA}' \|^2 \tag{9}$$

以上の定式化をパス図（path diagram）の形で示したのが図 1 である。

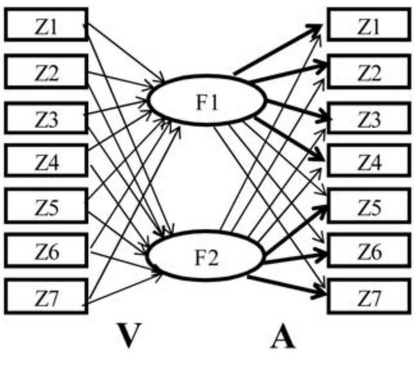

図 1　主成分分析のパス図表現。７つの変数の２つの因子。
太い矢印は高い負荷量を示す。

制約条件つき主成分分析　ここで，$\mathbf{A}$ には斜交回転が行われ，可能な限り完全単純構造に近づけることが目指されるが，一般にそれは可能でない。式（6）であらわされる基準の値を多少とも犠牲にしない限り，完全単純構造を達成することはできない。その方法は幾つかあると思われるが，ここでは，最も単純と思われるやり方，すなわち，主成分得点（に相当する合成変量）への各変数の回帰を重回帰から単回帰に変えるという方法を考えよう。すなわち，行列 $\mathbf{A}$ の各行に1つだけ非ゼロの値を残し，他はすべて0にするということである。これは，完全単純構造の定義そのものである（図2）。

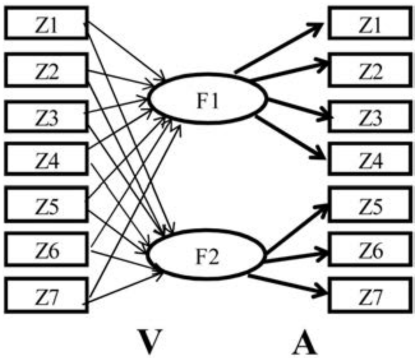

図２　主成分分析の負荷量を１項目１因子に限定する

この結果，最小化基準は次のように書き換えられることになる。

$$\varphi(a_1, a_2, \ldots, a_m, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_q \mid \mathbf{Z}) = n^{-1} \sum_{l=1}^{q} \sum_{j \in J_l} \| \mathbf{z}_j - \mathbf{Z}\mathbf{v}_l a_j \|^2 \tag{10}$$

ここで，$J_1, J_2, \ldots, J_q$ は，非ゼロの負荷量，$a_1, a_2, \ldots, a_p$ が，それぞれ，$l, \ldots, q$ のどの列に属するかを示す添え字の集合であり，添え字（項目番号）$l, \ldots, p$ は，$J_1, J_2, \ldots, J_q$ のどれか1つにだけ所属する（この段階では，集合 $J_1, J_2, \ldots, J_q$ のうちに空のものはないとする。このことは4節で証明される）。また，$\mathbf{z}_j$ は $\mathbf{Z}$ の第 j 列である n 次元ベクトル，$\mathbf{v}_l$ は $\mathbf{V}$ の第 l 列の p

次元ベクトルである。

これは，負荷行列の各行の $q-1$ 個の値を 0 とするという制約条件つきの主成分分析である（constrained PCA)。ただし，探索的分析では，それぞれの項目がどの集合（主成分）に非ゼロの負荷量をもつかが，当然，事前には明らかでない。したがって，問題の中心は（10）を最小化するような項目の分類（クラスタリング，clustering）ということになる。

項目のクラスター分析　ここで，方法をさらに単純化することができる次の命題が成り立つ。

命題 1　式（10）の極小値において，$k \notin J_l$ であるすべての項目 k に対応する $\mathbf{v}_l$ の要素 v_{kl} は 0 である。その結果，この方法の最小化基準は，

$$\varphi(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_q, \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_q \mid \mathbf{Z}) = n^{-1} \sum_{l=1}^{q} \| \mathbf{Z}_l - \mathbf{Z}_l \mathbf{v}_l \mathbf{a}_l' \|^2 \qquad (11)$$

となる。ここで，$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 & \cdots & \mathbf{Z}_q \end{bmatrix}$ で，$\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_q$ はそれぞれ，$J_1, J_2, ..., J_q$ に属する項目に対応する $\mathbf{Z}$ の列からなる $n \times |J_1|, n \times |J_2|, \ldots, n \times |J_q|$ の行列である。ただし，$|J_l|$ は J_l に含まれる項目数を示すものとする。なお，すべての項目は相互に 1 次独立，すなわち，項目間相関行列 $\mathbf{R}$ のランクは p とする。

この命題の意味するところは，図 3 に示されている。すなわち，(10)を最小化する 1 次合成変量は，項目が所属するクラスターだけの範囲で算出される。

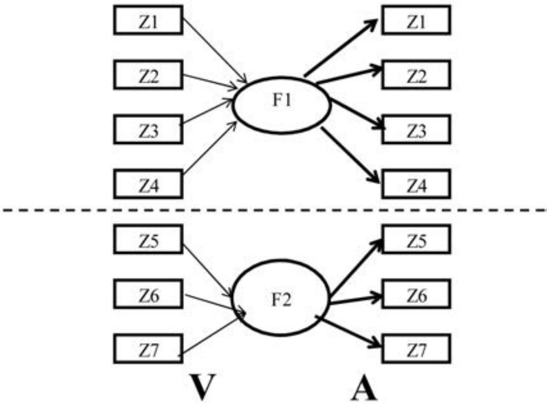

図３　主クラスター成分分析のパス図

証明　まず，(10) 式の右辺は，

$$n^{-1}\sum_{l=1}^{q}\sum_{j\in J_l}\|\mathbf{z}_j-\mathbf{Z}\mathbf{v}_l a_j\|^2=p-\sum_{l=1}^{q}\sum_{j\in J_l}a_j^2$$

したがって，問題は，

$$\sigma^2=\sum_{l=1}^{q}\sum_{j\in J_l}a_j^2$$

の最大化に置き換えられる。

命題は，１つの変数クラスターについて証明すれば十分である。J を当該クラスターに含まれる変数の添字の集合，K を含まれない変数の添字の集合とする。全変数間の相関行列 $\mathbf{R}$ は，（適当な変数の並べ替えによって，）次のように書ける。

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_{JJ} & \mathbf{R}_{JK}\\ \mathbf{R}_{KJ} & \mathbf{R}_{KK}\end{bmatrix}$$

J および K の変数に対する重みベクトルを $\mathbf{v}_L$，$\mathbf{v}_K$ とすると，制約条件 (7) から，

$$[\mathbf{v}'_J\quad \mathbf{v}'_K]\begin{bmatrix}\mathbf{R}_{JJ} & \mathbf{R}_{JK}\\ \mathbf{R}_{KJ} & \mathbf{R}_{KK}\end{bmatrix}\begin{bmatrix}\mathbf{v}_J\\ \mathbf{v}_K\end{bmatrix}=1=\mathbf{v}'\mathbf{R}\mathbf{v}$$

である。

ここで，クラスター J に含まれる変数 j の1次合成変量への（単）回帰係数は，

$$a_j = \frac{\mathbf{v}'_J \mathbf{r}_{jJ} + \mathbf{v}'_K \mathbf{r}_{jK}}{\mathbf{v}'\mathbf{R}\mathbf{v}} = \mathbf{v}'_J \mathbf{r}_{jJ} + \mathbf{v}'_K \mathbf{r}_{jK}$$

したがって，最大化基準は，

$$\sum_{j\in J} a_j^2 = \mathbf{v}'_J \left(\sum_{j\in J} \mathbf{r}_{Jj} \mathbf{r}'_{Jj} \right) \mathbf{v}_J + 2\mathbf{v}'_J \left(\sum_{j\in J} \mathbf{r}_{Jj} \mathbf{r}'_{Kj} \right) \mathbf{v}_K + \mathbf{v}'_K \left(\sum_{j\in J} \mathbf{r}_{Kj} \mathbf{r}'_{Kj} \right) \mathbf{v}_K$$

ここで，

$$\sum_{j\in J} \mathbf{r}_{Jj} \mathbf{r}'_{Kj} = \mathbf{R}_{JJ} \mathbf{R}_{JK}$$

等を用いて，

$$\sum_{j\in J} a_j^2 = \mathbf{v}'_J \mathbf{R}^2_{JJ} \mathbf{v}_J + 2\mathbf{v}'_J \mathbf{R}_{JJ} \mathbf{R}_{JK} \mathbf{v}_K + \mathbf{v}'_K \mathbf{R}_{KJ} \mathbf{R}_{JK} \mathbf{v}_K \tag{12}$$

また，定義により，

$$\mathbf{v}'\mathbf{R}\mathbf{v} = \mathbf{v}'_J \mathbf{R}_{JJ} \mathbf{v}_J + 2\mathbf{v}'_J \mathbf{R}_{JK} \mathbf{v}_K + \mathbf{v}'_K \mathbf{R}_{KK} \mathbf{v}_K = 1$$

そこで，λ を Lagrange の乗数として，次の関数を考える。

$$\begin{aligned} f(\mathbf{v}_J, \mathbf{v}_K) = {} & \mathbf{v}'_J \mathbf{R}^2_{JJ} \mathbf{v}_J + 2\mathbf{v}'_J \mathbf{R}_{JJ} \mathbf{R}_{JK} \mathbf{v}_K + \mathbf{v}'_K \mathbf{R}_{KJ} \mathbf{R}_{JK} \mathbf{v}_K \\ & - \lambda \left(\mathbf{v}'_J \mathbf{R}_{JJ} \mathbf{v}_J + 2\mathbf{v}'_J \mathbf{R}_{JK} \mathbf{v}_K + \mathbf{v}'_K \mathbf{R}_{KK} \mathbf{v}_K - 1 \right) \end{aligned}$$

これを，$\mathbf{v}_J$，$\mathbf{v}_K$ で偏微分して0と置くことにより，

$$\mathbf{R}^2_{JJ} \mathbf{v}_J + \mathbf{R}_{JJ} \mathbf{R}_{JK} \mathbf{v}_K - \lambda \left(\mathbf{R}_{JJ} \mathbf{v}_J + \mathbf{R}_{JK} \mathbf{v}_K \right) = \mathbf{0} \tag{13}$$

$$\mathbf{R}_{KJ} \mathbf{R}_{JJ} \mathbf{v}_J + \mathbf{R}_{KJ} \mathbf{R}_{JK} \mathbf{v}_K - \lambda \left(\mathbf{R}_{KJ} \mathbf{v}_J + \mathbf{R}_{KK} \mathbf{v}_K \right) = \mathbf{0} \tag{14}$$

式（13）の左から $\mathbf{R}_{KJ} \mathbf{R}^{-1}_{JJ}$ を掛けて，そこから（14）を引くことにより，

$$\lambda\left(\mathbf{R}_{KK}-\mathbf{R}_{KJ}\mathbf{R}_{JJ}^{-1}\mathbf{R}_{JK}\right)\mathbf{v}_K=\mathbf{0}$$

が得られる。なお，すべての項目は１次独立と仮定したので $\mathbf{R}_{JJ}^{-1}$ は必ず存在する。

ここで，(13) と (14) を加算し，(12) を用いると，

$$\lambda=\sum_{j\in J}a_j^2$$

したがって，$\lambda\neq 0$ かつ，項目の１次独立性の仮定により $\left|\mathbf{R}_{KK}-\mathbf{R}_{KJ}\mathbf{R}_{JJ}^{-1}\mathbf{R}_{JK}\right|=0$ だから，

$$\mathbf{v}_K=\mathbf{0}$$

である。これが証明すべき事実であった。すなわち，$\mathbf{v}_l$ の非ゼロ要素は，集合 J_l に含まれる項目に対応する部分だけである。したがって，最小化基準は (10) から (11) に書き換えられる。

主クラスター成分分析　$\mathbf{v}_k=\mathbf{0}$ を (13) に代入して，

$$\mathbf{R}_{JJ}^2\mathbf{v}_J=\lambda\mathbf{R}_{JJ}\mathbf{v}_J$$

ここで，項目の１次独立性の仮定により，$|\mathbf{R}_{JJ}|\neq 0$ が成り立つから，

$$\mathbf{R}_{JJ}\mathbf{v}_J=\lambda\mathbf{v}_J \tag{15}$$

である。この式は，１主成分の主成分分析の式に他ならない。つまり，解は，クラスター内の変数のみを対象とした１因子の主成分解であり，重みは，相関行列の最大固有値に対応する固有ベクトルを，$\mathbf{v}_l'\mathbf{v}_l=1/\lambda_l$ と基準化したものである。いずれにせよ，これによって解くべき問題は，最大固有値の和 σ^2 が最大になるような項目のクラスターを見出すことに帰着する。

ところで，結果的にこの方法は，Braverman (1970)，Escoufier (1988)

によって独立に提案され，Kiers(1993）によって主クラスター成分分析（principal cluster component analysis；PCCA）と呼ばれた方法と一致することになる。ただし，この方法は完全単純構造を目指したものではなく，因子分析の「弱い環」である回転を回避する点に大きな目標があり，実際，Braverman も Kiers も負荷行列については，通常の因子構造（個別変数と因子との相関係数の行列）を採用している。すなわち，ここで提案したパターン行列の非ゼロ要素を各行 1 つに制限して導出される主成分分析は，PCCA とアルゴリズム的に一致することになるのである。

アルゴリズム　この方法は，基本的に p 個の項目を q 個の背反なクラスターに分類するという離散的なものであり，そのためのアルゴリズムはいろいろと考えられる（たとえば，Givens & Hoetings, 2013）が，本研究では以下のような単純な手続きを，数値計算プログラム MATLAB でコーディングして用いた。

(0) 初期化

クラスターの数を q とする。すべての項目を q 番目のクラスターに所属させる。

全項目間の相関行列を計算，その最大固有値を λ とする。$s_0 = \lambda$ とする。1, ..., p の整数をランダムに並べ替え，$c(1), ..., c(p)$ とする。これを試行順序として固定する。t=1 とする。

(1) j=1 とする。

(2) 項目 $c(j)$ を選択し，その項目が現在所属しているクラスター l から暫定的に取り除き，残りの変数による相関行列の最大固有値を計算，λ_l とする。

(3) 項目 $c(j)$ を，順次，クラスター $l' = 1, \ldots, q\ (l' \neq l)$ に移し，その最大固有値 $\lambda_{l'}$ を計算する。

(4) $\max \lambda_{l'} > \lambda_l$ なら，項目 $c(j)$ を， $\max \lambda_{c'}$ が実現しているクラ

スター l' に移す。そうでなければ，クラスター l にもどす。

(5) $j+1 \rightarrow j$ とする。$j<p$ なら（2）に戻る。$j=p$ なら（6）に進む。

(6) $t+1 \rightarrow t$ とし，$s_t = \sum_{l=1}^{q} \lambda_l$ とする。$s_t > s_{t-1}$ なら（1）に戻る。そうでなければ，収束したと判断し，$\sigma^2 = s_t$ として終わり。

次の4節で論じるように，このアルゴリズムは必ず収束する。しかしながら，極大（local maxima）の問題を回避することはできず，少なくとも，初期値（項目の移動順序）を変えて複数回反復する必要がある。この点については5節で検討しよう。

4. ここで提案した方法の幾つかの性質

負荷行列と重み行列の意味　2節で述べたように，通常の斜交回転を伴う因子分析では，パターン行列と構造行列がそれぞれ，標準化回帰係数と相関係数という異なる意味をもち，その間の関係も単純ではない。また，主成分スコアを求めるための重み行列も因子負荷行列との関係は単純でない。

しかしながら，ここで提案した方法では，負荷行列の要素は標準単回帰係数であると同時に（当然ながら）相関係数でもある。また，重み行列もそれらを列ごとに見れば比例している。すなわち，負荷量自体は相関係数という単純な解釈ができるだけでなく，因子スコアも（標準化された）項目反応に負荷量を掛けて加算し，結果を標準化したものであると解釈できる。こうした単純性の実用上の効果は決して過小評価されるべきでない。

変数のクラスター分析との相違　変数間の相関係数を次の式によって距離の2乗に変換し，通常の系統的，あるいは非系統的クラスター分析を適用することは，従来から行われてきた（たとえば，Izenman, 2008；pp.439–443）。

$$d_{jk}^2 = 1 - r_{jk} \tag{16}$$

しかしながら，この単純なクラスター分析の適用と，ここで提案した方法の結果には大きな違いが生じる可能性がある。それは，通常の Likert 型の項目の加算にあたっては反応を逆転する必要がある場合が生じるからである。

たとえば，表１の最初の２つの項目，「友人が動揺しているからといって，自分も動揺するようなことはない」と「まわりに落ち込んでいる人がいると，自分も平静のままではいられない」は，ともに「感情の伝染」という次元に関連しているが，方向は逆であり，この２つの項目の間には高い逆相関がある。式（16）を単純に用いれば，この２つの項目は別のクラスターに含まれることになる。図４に示すように，こうしたクラスター分析からは，余分な次元が生じてしまうことになり，不適切な解釈が行われる可能性も高い。

項目の意味にもとづいて，項目反応をあらかじめ逆転しておくという措置も考えられるが，その手順は誤りが発生しやすく，かつ，現実には５節で見るように，逆転すべきかどうかの判断が難しいケースも少なくない。

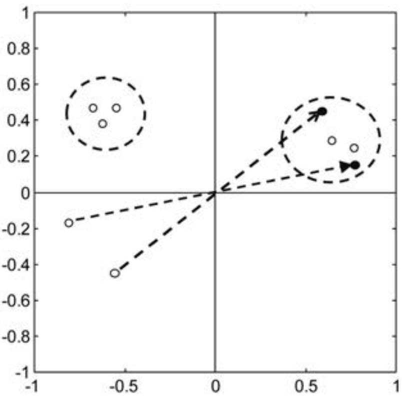

図４　２つの項目の反転によって３つのクラスターの数が２つに減ることの説明。

ここで提案した方法は，クラスターの最大固有値の和を用いることによ

り，逆転項目について特に事前に配慮することなく分析を進めることができる。

すべてのクラスターには最小１つの項目が含まれる　このことは，次のようにして示される。まず，次のまったく自明な命題を述べておく。

命題２　１つの項目からなる相関行列はスカラーの１であり，その固有値は１である。

これに対して，次の命題は，PCCA の性質を考える上で重要である。

命題３　すでに p_l 個の項目が含まれているクラスターに１つの項目を追加した場合，それらの項目間相関係数の最大固有値の増分 $\Delta\lambda$ は，

$$0 \le \Delta\lambda < 1 \tag{17}$$

の範囲にある。

証明　相関行列 $\mathbf{R}$ から得られる負荷ベクトルを $\mathbf{a}$ とする。$\mathbf{R}$ の最大固有値を λ とすると，$\mathbf{a}$ は $\mathbf{a'a} = \lambda$ と基準化された $\mathbf{R}$ の固有ベクトルである。次に，項目を１つ追加して得られる相関行列を $\mathbf{R}_+$ とし，そこから算出される負荷ベクトルを $\mathbf{b}$ とする。$\mathbf{R}_+$ の最大固有値を λ_+ とすると，$\mathbf{b'b} = \lambda_+$ であり，$\Delta\lambda = \lambda_+ - \lambda$ である。

ここで，あきらかに，$\mathbf{a'a} \le \mathbf{b'b}$ である。なぜなら，相関行列の次数を p とすると，$\mathbf{a}$ は p 次元空間内で p 個の変数の１次結合として最も長いベクトルであり，$\mathbf{b}$ は $p+1$ 次元空間内で $p+1$ 個の変数の１次結合として最も長いベクトルである。ここで，$p+1$ 番目の変数が，他のすべての変数と直交していたとき，$\mathbf{b}=\mathbf{a}$，したがって，$\mathbf{b'b} = \mathbf{a'a} = \lambda$ であり，これが λ_+ の

下限である。すなわち，$\Delta\lambda \geq 0$ である。

次に，$\mathbf{b}' = \begin{bmatrix} \mathbf{b}'_- & b_{p_l+1} \end{bmatrix}$ と記したとき，$\mathbf{b}'_-\mathbf{b}_- \leq \mathbf{a}'\mathbf{a}$ である。なぜなら，変数を１つ増やしたとき，もとの変数がより多く説明されるとすれば，それは命題３に反するからである。他方，この方法では負荷量は相関係数だから，$b_{p+1} \leq 1$ である。したがって，$\mathbf{b}'\mathbf{b} \leq \mathbf{a}'\mathbf{a} + 1$，すなわち，$\lambda_+ \leq \lambda + 1$ であり，したがって，$\Delta\lambda \leq 1$ である。

ここで等号が成立するのは，$\mathbf{b}'_-\mathbf{b}_- = \mathbf{a}'\mathbf{a}$ かつ $b_{p+1} = 1$ の場合，すなわち，**R** の第１主成分によって項目 p+1 が完全に説明される場合であるが，われわれは，項目は相互に１次独立であると仮定したから，このようなことはありえない。したがって，$\Delta\lambda < 1$ である。（証明終わり）

命題２と命題３を合せて考えれば，クラスタリングの手続きで，１つの項目の所属を決定する際，もし空のクラスターが存在すれば，その項目と他の項目との相関関係は無視しても空のクラスターに所属させるのが，基準の最大化につながることは明らかである。したがって空のクラスターは存在しえないことになる。

すべての項目はどれか１つのクラスターに属する　このことは，(17) 式の前半から明らかであろう。もっとも，最適化基準の増加にまったく寄与しない項目が原理的には存在しうるのは事実である。それでも，その項目をどこかのクラスターに所属させることが,（その後の尺度構成にあたっては削除するとしても），分析手続き上，特に有害であるわけではない。この方法では，残余項目のためにゴミ箱を用意する必要がない[iii]。

クラスターの数を増やすと説明力は増加する　これについても，命題２と３から明らかである。クラスター数を１つ増加させれば，空のクラスターが１つできる。どの項目をそこに移動しても，基準の値は増加するからである。

さらに，$q=p$ であれば，すべてのクラスターに１つずつの項目が属し，説明力は100％に到達する。これは，負荷行列が単位行列で，主成分間相関行列がもとの項目間相関行列に等しいという，つまらない（trivial）解であるが，$q<p$ の場合の各クラスターの最大固有値の和を p で除して解の説明力をパーセントで表わすことを正当化するものである。

望ましい単純構造は達成できるか　単純構造とは何だろうか。因子分析の多くの方法には，変数のグルーピング，またはクラスタリングという意味があるとすれば（Harman, 1967），単純構造には分類としての適切性が含意されているはずである。それは，次の芝（1979；p.96）の定義によく表されている。

(1) 各変量は１つの因子にだけ高い負荷を示すこと。

(2) 各因子において高い負荷を示す変量の数はほぼ等しくなること。

まず，(1)は，いわばあいまいさの回避，各変数が一義的に特定のグループに属することが判定できるということである。(2) は，多くの変数が少数のグループに集中してしまわないことであって，これも分類としての良さであろう。

芝の定義の（1）は，ここで提案している方法では，定義により自動的に満たされる。他方，(2) についてはデータ次第と言わざるを得ない。すなわち，この方法には，変数を多くのグループに分散させるという性質は備わっていない。そのことを示唆するのが，次の命題４である。

命題４　項目を $q(<p)$ 個の群にクラスター化するとする。すべての項目間相関係数が等しいとき，すべての分類における最適化規準は等しい。すなわち，いかなるクラスタリングも同等である。

証明　それぞれのクラスターに含まれる変数の数を $p_1, p_2, \ldots, p_q$ とする。$p_1+p_2+\cdots+p_q=p$ である。また，すべての項目間の相関係数を r とする。このとき，クラスター l の変数群間相関行列の最大固有値は，$1+(p_l-1)r$ である（たとえば，Schott, 1997；p.100）。その全クラスターにわたる総和

は，$\varphi = q + (p - q)r$ であり，これは $p_1, p_2, ..., p_q$ にかかわらない。（証明終わり）

結局，最悪の分類とは，1つのクラスターに $p-q+1$ 個の項目が入り，残りの $q-1$ 個の項目はそれぞれ1つのずつ $q-1$ 個のクラスターに分属するというケース，最善はすべてのクラスターに p/q に近い整数の項目が所属するケースである。この点については，さらに検討が必要であるが，ここでは，最悪と最善の典型例を1つずつ上げておこう。

最初は，Spearman の2因子説（共通因子1つと特殊因子1つで2因子と考えているから，普通の解釈では1因子である）にもとづく。$\mathbf{a}' = [0.9 \quad 0.8 \quad 0.7 \quad 0.6 \quad 0.5 \quad 0.4]'$ という負荷ベクトルから生成された相関行列 $\mathbf{R}_s$ から得られる3クラスターは，成員数が4, 1, 1となる最悪のケースである。

$$\mathbf{R}_s = \begin{bmatrix} 1.00 & 0.72 & 0.63 & 0.54 & 0.45 & 0.36 \\ 0.72 & 1.00 & 0.56 & 0.48 & 0.40 & 0.32 \\ 0.63 & 0.56 & 1.00 & 0.42 & 0.35 & 0.28 \\ 0.54 & 0.48 & 0.42 & 1.00 & 0.30 & 0.24 \\ 0.45 & 0.40 & 0.35 & 0.30 & 1.00 & 0.20 \\ 0.36 & 0.32 & 0.28 & 0.24 & 0.20 & 1.00 \end{bmatrix}, \mathbf{A}_s = \begin{bmatrix} 0.89 & 0 & 0 \\ 0.85 & 0 & 0 \\ 0.80 & 0 & 0 \\ 0.73 & 0 & 0 \\ 0 & 1.00 & 0 \\ 0 & 0 & 1.00 \end{bmatrix}$$

他方，次の相関行列 $\mathbf{R}_g$ は，Guttman（1954）の simplex 行列と呼ばれるもので $r_{jk} = r^{|j-k|}$ というルールで作られている。これは，下に見られるように3つのクラスターの成員数が均等な最善の例となる。

$$\mathbf{R}_g = \begin{bmatrix} 1.00 & 0.80 & 0.64 & 0.51 & 0.41 & 0.33 \\ 0.80 & 1.00 & 0.80 & 0.64 & 0.51 & 0.41 \\ 0.64 & 0.80 & 1.00 & 0.80 & 0.64 & 0.51 \\ 0.51 & 0.64 & 0.80 & 1.00 & 0.80 & 0.64 \\ 0.41 & 0.51 & 0.64 & 0.80 & 1.00 & 0.80 \\ 0.33 & 0.41 & 0.51 & 0.64 & 0.80 & 1.00 \end{bmatrix}, \mathbf{A}_g = \begin{bmatrix} 0.95 & 0 & 0 \\ 0.95 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.95 \\ 0 & 0 & 0.95 \end{bmatrix}$$

この2つは，どちらも（意味は違うが）階層性をもつ1因子の相関行列である（Guttman, 1954）にもかかわらず，対照的な結果となるのが面白い。

5. 数値例

この方法の数値例のために，どのような実データを選択するかは重要である。多くの刊行されている尺度は，すでに（少なくとも）探索的因子分析の手続きを経て構造は確定に近い場合が多い。この方法の性質からして，分析例としては，ある程度未完成な尺度が望まれる。

ここで選んだのは，すでに表1に負荷行列等を例示した Mehrabian & Epstein（1972）による感情的共感性尺度（emotional empathy scale）の日本語版（中村，2000）である。この尺度は社会心理学の実験のための個体差要因を作り出すために開発されており，わずか100名弱の大学生のデータしか（少なくとも論文公刊時点では）とられておらず，33個の項目は1因子として扱われている。中村（2000）では，前述のような4因子として分析されているが，その段階でも。特に項目の加除等は行われていないから，ほとんど予備調査段階の項目がそのまま公表されている状態と言える。このことは，研究の最も初期の状況で使われることを想定しているここで提案した方法の適用例として適切である。項目の反応段階は9であり。データは，1990年から1991年にかけて，前述のように，1645名の大学生に実施して得られたものである。

まず，全項目間相関行列の固有値を，降順に序数に対してプロットしたスクリー・プロットを図5に示した。固有値の値は降順に，5.51，3.04，1.79，1.50，1.21，1.07，1.05，1.02，0.99となっており，1を超える固有値は8個存在するものの，その減衰状況をスクリー・プロットで読み取る限り，7以上の因子（クラスター）を検討することは無意味であろう。そこで，3節のアルゴリズムにもとづき，1因子から6因子までの解をもとめ，最適化基準 φ を求めた。その結果は，5.51，8.13，9.64，11.01，12.08，13.07

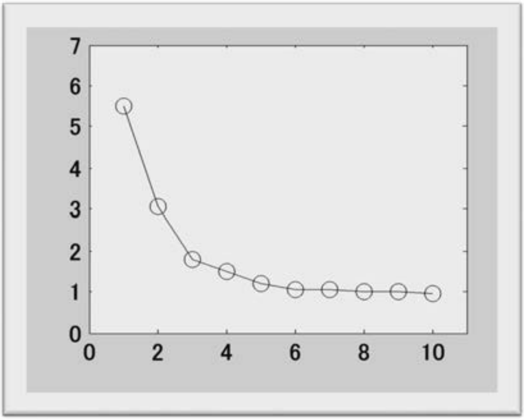

図５　スクリー・プロット

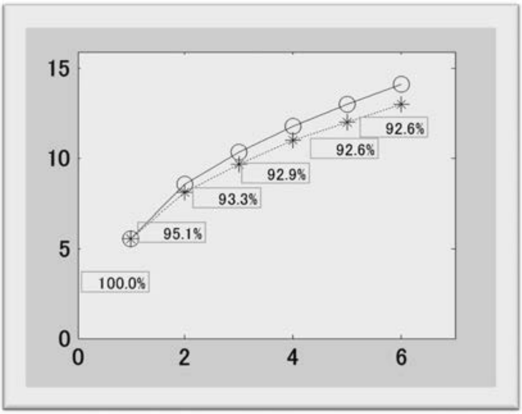

図６　主成分分析○と PCCA＊の説明力の比較。
%の数値は主成分分析に対する PCCA の効率

となった（1 因子は全項目による主成分分析の 1 因子解になるから，これは相関行列の第 1 固有値と当然一致する）。他方，累積固有値によって主成分分析の各因子数における説明力を求めると，5.51，8.55，10.34，11.84，13.06，14.12 となるが，これらを図 6 にプロットした。項目の因子への単

回帰という強い制約条件を課した PCCA の解の説明力が常に対応する主成分解のそれを下回るのは当然である。図中には，主成分解の説明力を分母とし，対応する PCCA の σ^2 を分子として計算された効率（efficiency）をパーセントの形で示している。3 因子以上において，その値は約 93% でほぼ一定となっている。この値を十分に大きいと見るか，小さいと見るかには議論の余地がありうるが，それについては 7 節で再度論じることにしよう。

次に，5 因子解の負荷行列と因子間相関行列を表 2 に示した（表 1 と比較可能になる 4 因子解は，6 節の表 6 に示す）。因子負荷行列の空白部分はすべて 0 である。完全単純構造の効果は大きい。さらに，R^2 は各行の非ゼロの負荷量の 2 乗である。加えて，4 節で述べたように PCCA の場合，列ごとの 2 乗和は対応する因子の説明力と一致する（「因子寄与」と表現している）。この総和は，R^2 の総和と一致し，その値はこの 5 因子解の φ の値である 12.08 である。なお，負荷量の符号は，ここではプログラムのデフォルトとして，因子ごとに符号が正のものが増える方向（具体的には，負荷量の 3 乗の和が正になる向き）に決められており，必ずしも共感性が高くなる方向にはなっていない。実際，因子 3 と因子 5 は，得点が低いほど共感性が高くなる。このことは，因子間相関の符号から読み取ることができる。

因子間相関は，（直接比較はできないにせよ，）表 1 の主成分分析・プロマックス回転の解よりも，全体として高い値が多くなっている。どちらも，（標準化された）項目反応の結果を示しているから，その違いの理由ははっきりしない。ただ，一般に（使用頻度の高い）斜交回転の方法には，パターン行列の要素が 1 を超えるのを防ぐために，因子間相関を低くする機能が備わっているように思える。

次に，異なる因子数の解がどのように相互に違っているかをおおよそ見るために，非ゼロ要素の位置と符号だけを示す表を作成した（表 3）。これで見ると，因子数が増加するにつれて直前の因子数の因子のどれかが 2

表２　感情的共感性尺度の PCCA による分析（q=5）

質問項目	因子負荷量 1	2	3	4	5	R^2
13 友人が動揺しているからといって，自分も動揺するようなことはない。	**-0.72**					0.52
25 まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	**0.71**					0.51
10 まわの人の気分によって自分の気分も大きく左右される。	**0.66**					0.43
9 人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	**0.66**					0.43
7 友人の悩みに感情的に巻き込まれてしまう傾向がある。	**0.64**					0.41
32 周囲が興奮していても，自分は冷静でいられることが多い。	**-0.64**					0.41
16 泣いている人を見ると自分も動揺してしまう。	**0.63**					0.40
20 まわりの人たちがいらいらしていても，自分は平静でいられる。	**-0.58**					0.34
24 人の気持ちに左右されずに判断を下すことができる。	**-0.47**					0.23
5 まわりに神経質な人がいると，自分も神経質になる。					**0.59**	0.34
30 他人の涙を見ると，同情するよりいらいらする。			**0.72**			0.51
6 幸せのあまり泣くというのは，馬鹿げている。			**0.63**			0.40
21 友人が悩み事を話し始めると，別の事柄に話を変えようと努める。			**0.62**			0.39
23 映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。			**0.64**			0.40
4 不幸な人がすっかりしょげているのを見ると，いらだたしくなる。					**0.62**	0.39
26 人がどうしてそんなに動揺しているのかをなかなか理解できない。			**0.50**			0.25
22 他の人の笑いに引きずられて笑ったりはしない。			**0.50**			0.25
15 孤独な人は多分友情を求めていないのであろう。			**0.41**			0.17
3 感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。					**0.67**	0.45
2 人は動物の感情や感受性について思い入れすぎである。					**0.44**	0.19
11 自分がであった外国人の多くは，クールで，感情的ではないように見えた。					**0.42**	0.18
33 小さい子どもはときどき，はっきりとした理由もなく泣くものである。					0.19	0.03
31 映画を観ると強く引き込まれてしまう。		**0.74**				0.54
18 小説の登場人物の気持ちになりきってしまう。		**0.69**				0.48
17 歌によって幸せな気分になることがある。		**0.67**				0.44
28 本や映画の世界に浸るのは，少し馬鹿げている。		**-0.64**				0.41
8 ときどきラブソングの歌詞に深く感動することがある。		**0.64**				0.41
14 人がプレゼントを開けるところを見るのが好きである。		**0.41**				0.17
1 グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。				**0.67**		0.45
27 動物が苦しんでいるのを見るとひどく動揺してしまう。				**0.67**		0.44

29 無力な老人を見ると動揺してしまう。				**0.65**		0.42
19 ひどい目にあっている人を見ると，強い憤りを感じる。				**0.63**		0.40
12 職業訓練所で働くよりは社会福祉関係の仕事をしたい。				**0.52**		0.27
因子寄与	3.68	2.46	2.37	1.99	1.59	12.08
因子間相関						
1	1.00	0.29	-0.26	0.30	0.04	
2	0.29	1.00	-0.47	0.35	-0.15	
3	-0.26	-0.47	1.00	-0.33	0.40	
4	0.30	0.35	-0.33	1.00	-0.08	
5	0.04	-0.15	0.40	-0.08	1.00	

表３　感情的共感性尺度の解の要約

	クラスター数	1	2		3			4				5					6						Promax								
	質問項目	1	1	2	1	2	3	1	2	3	4	1	2	3	4	5	1	2	3	4	5	6	1	2	3	4	1	2	3	4	5
13	友人が動揺しているからといって，自分も動揺するようなことはない。	+		−	−			−				−					−						−				−				
25	まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	−		+	+			+				+					+						+				+				
10	まわの人の気分によって自分の気分も大きく左右される。	−		+	+			+				+					+						+				+				
9	人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	−		+	+			+				+					+						+								
7	友人の悩みに感情的に巻き込まれてしまう傾向がある。	−		+	+			+				+					+						+				+				
32	周囲が興奮していても，自分は冷静でいられることが多い。	+		−	−			−				−									+		−				−				
16	泣いている人を見ると自分も動揺してしまう。	−		+	+			+				+					+						+				+				
20	まわりの人たちがいらいらしていても，自分は平静でいられる。	+		−	−			−				−									+		−				−				
24	人の気持ちに左右されずに判断を下すことができる。	+		−	−			−				−									+		−				−				
5	まわりに神経質な人がいると，自分も神経質になる。	−		+	+			+								+						+		+							+
30	他人の涙を見ると，同情するよりいらいらする。	+	+			+			+					+					+					+							+
6	幸せのあまり泣くというのは，馬鹿げている。	+	+			+			+					+					+					+							+
21	友人が悩み事を話し始めると，別の事柄に話を変えようと努める。	+	+			+			+					+					+					+							
23	映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。	+	+			+			+					+					+					+							
4	不幸な人がすっかりしょげているのを見ると，いらだたしくなる。	+	+			+			+							+						+		+							+
26	人がどうしてそんなに動揺しているのかをなかなか理解できない。	+	+			+			+					+					+					+							+
22	他の人の笑いに引きずられて笑ったりはしない。	+	+			+			+					+					+										−		
15	孤独な人は多分友情を求めていないのであろう。	+	+			+			+					+					+												
3	感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。	+	+			+			+							+						+		+							+
2	人は動物の感情や感受性について思い入れすぎである。	+	+			+			+							+						+									
11	自分がであった外国人の多くは，クールで，感情的ではないように見えた。	+	+			+			+							+						+									
33	小さい子どもはときどき，はっきりとした理由もなく泣くものである。	+	+			+			+							+						+									
31	映画を観ると強く引き込まれてしまう。	−	−				+			+			+					+							+			+			
18	小説の登場人物の気持ちになりきってしまう。	−	−				+			+			+					+							+			+			
17	歌によって幸せな気分になることがある。	−	−				+			+			+					+							+			+			
28	本や映画の世界に浸るのは，少し馬鹿げている。	+	+			+				−			−					−							−			−			
8	ときどきラブソングの歌詞に深く感動することがある。	−	−				+			+			+					+							+			+			
14	人がプレゼントを開けるところを見るのが好きである。	−	−				+			+			+					+													
1	グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。	−	−				+				+				+					+						+				+	
27	動物が苦しんでいるのを見るとひどく動揺してしまう。	−	−				+				+				+					+						+				+	
29	無力な老人を見ると動揺してしまう。	−		+			+				+				+					+						+					
19	ひどい目にあっている人を見ると，強い憤りを感じる。	−	−				+				+				+					+						+				+	
12	職業訓練所で働くよりは社会福祉関係の仕事をしたい。	−	−				+				+				+					+						+				+	

つに別れるという，階層的な構造がだいたい成立していることがわかる。同じ表には，0.40 未満の値を 0 として，同様の記法でプロマックス回転の結果も示している。こちらには，必ずしも階層的関係はないように見える。

続いて，アルゴリズムの性質を検討するために，極大（local maxima；真の最大値でない極値）の発生状況を見ておこう。図 7 は，$q=2\sim6$ の各条件について，1000 回ずつアルゴリズムを実行し，到達した極値を昇順に並べ替えたものである。反復ごとに異なるのは，3 節で記述したアルゴリズムの初期化の部分で，項目を他のクラスターに（仮に）移動する順序がランダム化されるところだけである。

まず，q=2 ではすべてのケースで真の最大値に到達している。因子数が増加するにつれて，真の最大値に達する確率は減少していく。それでも，q=4 までは，過半数のケースにおいて，真の最大値に到達するが，それ以上の場合には，単なる極大に終わる確率は高くなる。なお，q=4 の場合について，真の最大値とそれに続く単なる極大との差が大きいこと（図の＊印の下に大きな落差があること）に注目しよう。

表 4 と表 5 は，q=4 と q=5 の場合における真の最大値とそれに続く 4

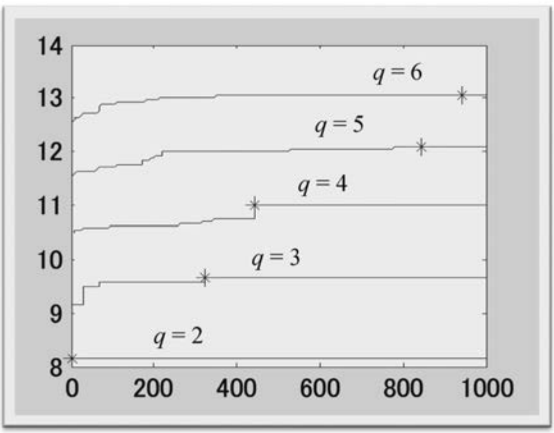

図 7　並べ替えられた極大値。＊より右は最大値

表４　４クラスターの極大解

質問項目	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
13 友人が動揺しているからといって，自分も動揺するようなことはない。	−				−				−					−			−			
25 まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	+				+				+					+			+			
10 まわの人の気分によって自分の気分も大きく左右される。	+				+				+					+			+			
9 人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	+				+				+					+			+			
7 友人の悩みに感情的に巻き込まれてしまう傾向がある。	+				+				+					+			+			
32 周囲が興奮していても，自分は冷静でいられることが多い。	−				−				−							+	−			
16 泣いている人を見ると自分も動揺してしまう。	+				+				+					+			+			
20 まわりの人たちがいらいらしていても，自分は平静でいられる。	−				−				−							+	−			
24 人の気持ちに左右されずに判断を下すことができる。	−				−				−							+	−			
5 まわりに神経質な人がいると，自分も神経質になる。	+							+				+				−	+			
30 他人の涙を見ると，同情するよりいらいらする。		+					+					+	+					+		
6 幸せのあまり泣くというのは，馬鹿げている。		+					+				+		+					+		
21 友人が悩み事を話し始めると，別の事柄に話を変えようと努める。		+					+				+		+					+		
23 映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。		+					+				+		+					+		
4 不幸な人がすっかりしょげているのを見ると，いらだたしくなる。		+						+				+	+					+		
26 人がどうしてそんなに動揺しているのかをなかなか理解できない。		+					+					+	+					+		
22 他の人の笑いに引きずられて笑ったりはしない。		+					+				+		+					+		
15 孤独な人は多分友情を求めていないのであろう。		+					+				+		+						−	
3 感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。		+						+				+	+					+		
2 人は動物の感情や感受性について思い入れすぎである。		+						+				+	+					+		
11 自分がであった外国人の多くは，クールで，感情的ではないように見えた。		+					+				+		+					+		
33 小さい子どもはときどき，はっきりとした理由もなく泣くものである。		+						+				+	+						−	
31 映画を観ると強く引き込まれてしまう。			+			+				+					+			−		
18 小説の登場人物の気持ちになりきってしまう。			+			+				+					+			−		
17 歌によって幸せな気分になることがある。			+			+				+					+			−		
28 本や映画の世界に浸るのは，少し馬鹿げている。			−				+				+		+					+		
8 ときどきラブソングの歌詞に深く感動することがある。			+			+				+					+			−		
14 人がプレゼントを開けるところを見るのが好きである。			+			+				+					+			−		
1 グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。				+		+				+					+				+	
27 動物が苦しんでいるのを見るとひどく動揺してしまう。				+		+				+					+					+
29 無力な老人を見ると動揺してしまう。				+		+				+					+					+
19 ひどい目にあっている人を見ると，強い憤りを感じる。				+		+				+					+				+	
12 職業訓練所で働くよりは社会福祉関係の仕事をしたい。				+		+				+					+				+	
因子寄与	11.006				10.755				10.716				10.645				10.450			

表５　５クラスターの極大解

質問項目	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
13 友人が動揺しているからといって，自分も動揺するようなことはない。	−					−					−					−					−				
25 まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	+					+					+					+					+				
10 まわの人の気分によって自分の気分も大きく左右される。	+					+					+					+					+				
9 人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	+					+					+					+					+				
7 友人の悩みに感情的に巻き込まれてしまう傾向がある。	+					+					+					+					+				
32 周囲が興奮していても，自分は冷静でいられることが多い。	−					−					−					−									+
16 泣いている人を見ると自分も動揺してしまう。	+					+					+					+					+				
20 まわりの人たちがいらいらしていても，自分は平静でいられる。	−					−					−					−									+
24 人の気持ちに左右されずに判断を下すことができる。	−					−					−					−									+
5 まわりに神経質な人がいると，自分も神経質になる。					+					+					+		+					+			
30 他人の涙を見ると，同情するよりいらいらする。			+				+								+		+					+			
6 幸せのあまり泣くというのは，馬鹿げている。			+				+						+				+					+			
21 友人が悩み事を話し始めると，別の事柄に話を変えようと努める。			+				+						+							+		+			
23 映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。			+				+						+				+					+			
4 不幸な人がすっかりしょげているのを見ると，いらだたしくなる。					+		+								+		+					+			
26 人がどうしてそんなに動揺しているのかをなかなか理解できない。			+				+								+		+					+			
22 他の人の笑いに引きずられて笑ったりはしない。			+				+						+							+					+
15 孤独な人は多分友情を求めていないのであろう。			+				+						+							+		+			
3 感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。					+					+					+		+					+			
2 人は動物の感情や感受性について思い入れすぎである。					+		+								+		+					+			
11 自分がであった外国人の多くは，クールで，感情的ではないように見えた。					+					+			+							+		+			
33 小さい子どもはときどき，はっきりとした理由もなく泣くものである。					+					+					+		+					+			
31 映画を観ると強く引き込まれてしまう。		+						+				+						+					+		
18 小説の登場人物の気持ちになりきってしまう。		+						+				+						+					+		
17 歌によって幸せな気分になることがある。		+						+				+						+					+		
28 本や映画の世界に浸るのは，少し馬鹿げている。		−						−				−						−					−		
8 ときどきラブソングの歌詞に深く感動することがある。		+						+				+						+					+		
14 人がプレゼントを開けるところを見るのが好きである。		+						+					−						−				+		
1 グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。				+					+					+					+					+	
27 動物が苦しんでいるのを見るとひどく動揺してしまう。				+					+					+					+					+	
29 無力な老人を見ると動揺してしまう。				+					+					+					+					+	
19 ひどい目にあっている人を見ると，強い憤りを感じる。				+					+					+					+					+	
12 職業訓練所で働くよりは社会福祉関係の仕事をしたい。				+					+					+					+					+	
因子寄与	12.084					12.080					12.062					12.040					12.040				

つの極大値に対応する負荷行列を表３と同様の方法で示したものである。表４（q=4）については，極大の解は真の最適解から著しく異なっていき，本来の解とは似ても似つかぬものとなる。最大化基準そのものの値も，真の最大値より 0.5 以上小さいものが含まれる。それに対して，表５（q=5）については，真の解とあまり違わない解が並んでおり，最適化基準の値も，真の最大値と 0.04 程度の差があるにすぎない。

このように見ると，q=4 の（真の）最適解の基準の値は，いわば周辺の極大値から孤立している一方，q=5 の解はその近傍（Givens & Hoeting, 2013）に類似した解が「まとわって」いることがわかる。その意味で，ユーザーとしては q=4 の解を（あくまでも暫定的に）採用する方向に傾くであろう。この判断は，次の resampling の手続きによってさらに進められる。

6. Resampling による確認的分析

単純構造への回転をともなう因子分析は，前述のように，探索的（exploratory）方法と呼ばれ，解全体に対しても，個々のパラメータの推定値に対しても，推測統計学的な推定，検定の処置はなされないのが普通である。しばしば，得られた因子負荷量から（しばしば完全単純構造の）仮説を導き出し，同じデータにその仮説を当てはめるという形で確認的分析が行われることがあるが，これは，推測統計学的手法の誤用として非難されるであろう。この問題の最も単純で明快な説明は吉村（1969）参照[iv]。

しかしながら，探索的方法といえども，サンプリング誤差の存在に対して，なんらの対策もとらないならば，その結果にもとづいて，さらに進んだ分析，たとえば潜在方程式モデルの利用等に進むことはためらわれるであろう。また，もし尺度を個人のアセスメントに用いる等，個人の処遇にかかわる目的で使用するならば，その尺度について一定の推測統計学的裏づけは必要とされよう。実際，回転を伴う因子分析においても，負荷量の標準誤差を算出する方法は知られている（たとえば，市川，2010）。しかしながら，こうした方法は必ずしも活用されているとは言えない。

本研究では，bootstrap 法として知られている resampling の方法を用いることにする。Bootstrap 法（たとえば，Efron & Tibshirani, 1993）とは，実際に得られている大きさ n の標本から，同じ大きさの標本を多数抽出（resampling）し，それらの標本に同じアルゴリズムを適用して，変動を検討する方法である。存在する標本と同じ大きさの標本を，その標本自体から抽出するということは一見不可能とも思われるが，復元抽出を行うことでそうした「マジック」が可能になる[v]。

本研究では，$n = 1645$ の標本から，resampling による標本を 200 個取り出して，2~6 の各主成分数について，PCCA のアルゴリズムを適用し，それぞれ 200 のクラスターを得た。表 6 は，q=4 の場合について，全サンプルによる最適解と，200 回の resampling による結果を示したものである。さらに，変数ごとに正しいクラスターに配属された比率（的中率）も示した。

ただし，resampling の結果については，必ずしも全サンプルによる結果と同じ順序でクラスターが出現するとは限らず，さらには，独自の内容のクラスターの出現や，（その結果としての）2 つのクラスターの融合といったことも起こりうる。したがって，各 resampling の解析結果ごとに，個々のクラスターを全サンプルの（正）解のクラスターと対応づける必要がある。ここでは，それを次のような手順で行った。

まず，ある resampling であらわれたクラスターの番号を i，全サンプルによるクラスターの番号を j とする。個々の項目は，それぞれ特定の i と j に対応付けられるから，これらをクロス集計した頻度を f_{ij} とする。これらの頻度の j ごとの総和を $f_{.j} = \sum_{i=1}^{q} f_{ij}$ とし，クラスターごとに，$f_{ij}/f_{.j}$ を求め，この値が最大になるクラスター j にクラスター i を対応付ける。この結果，同一のクラスター j に複数のクラスター i が関係づけられることが起こりうる一方，resampling では出現しなかったクラスターもありうることになる。表 6 の ’resampling’ の部分は，このようにして得られたクラスターがどの（全サンプルによる）クラスターと関連付けられ

表６　負荷行列とリサンプリングの結果（q=4）

質問項目	因子負荷量 1	2	3	4	R^2	Resampling 1	2	3	4	的中率
13 友人が動揺しているからといって，自分も動揺するようなことはない。	−0.72				0.51	200	0	0	0	1.000
25 まわりに落ち込んでいる人がいると，自分も平静のままではいられない。	0.71				0.50	200	0	0	0	1.000
10 まわの人の気分によって自分の気分も大きく左右される。	0.66				0.44	200	0	0	0	1.000
9 人に悪い知らせを伝えるときには，自分も冷静でいられなくなる傾向がある。	0.65				0.43	200	0	0	0	1.000
7 友人の悩みに感情的に巻き込まれてしまう傾向がある。	0.64				0.40	200	0	0	0	1.000
32 周囲が興奮していても，自分は冷静でいられることが多い。	−0.63				0.40	200	0	0	0	1.000
16 泣いている人を見ると自分も動揺してしまう。	0.63				0.39	200	0	0	0	1.000
20 まわりの人たちがいらいらしていても，自分は平静でいられる。	−0.59				0.35	200	0	0	0	1.000
24 人の気持ちに左右されずに判断を下すことができる。	−0.47				0.22	200	0	0	0	1.000
5 まわりに神経質な人がいると，自分も神経質になる。	0.30				0.09	100	100	0	0	0.500
30 他人の涙を見ると，同情するよりいらいらする。		0.72			0.52	0	200	0	0	1.000
6 幸せのあまり泣くというのは，馬鹿げている。		0.62			0.38	0	190	10	0	0.950
21 友人が悩み事を話し始めると，別の事柄に話を変えようと努める。		0.58			0.33	0	197	3	0	0.985
23 映画館で周りの人が泣いたり鼻をすすったりすればするほど，おもしろがってしまう。		0.57			0.33	0	183	17	0	0.915
4 不幸な人がすっかりしょげているのを見ると，いらだたしくなる。		0.53			0.28	0	200	0	0	1.000
26 人がどうしてそんなに動揺しているのかをなかなか理解できない。		0.49			0.24	0	200	0	0	1.000
22 他の人の笑いに引きずられて笑ったりはしない。		0.42			0.17	28	110	62	0	0.550
15 孤独な人は多分友情を求めていないのであろう。		0.40			0.16	0	167	7	26	0.835
3 感情をそのまま表に出しているのをみると，いらだたしく思うことが多い。		0.40			0.16	0	200	0	0	1.000
2 人は動物の感情や感受性について思い入れすぎである。		0.34			0.11	0	200	0	0	1.000
11 自分がであった外国人の多くは，クールで，感情的ではないように見えた。		0.33			0.11	0	185	15	0	0.925
33 小さい子どもはときどき，はっきりとした理由もなく泣くものである。		0.14			0.02	8	163	5	24	0.815
31 映画を観ると強く引き込まれてしまう。			0.74		0.54	0	4	196	0	0.980
18 小説の登場人物の気持ちになりきってしまう。			0.69		0.48	0	4	196	0	0.980
17 歌によって幸せな気分になることがある。			0.67		0.44	0	4	196	0	0.980
28 本や映画の世界に浸るのは，少し馬鹿げている。			−0.64		0.41	0	14	186	0	0.930
8 ときどきラブソングの歌詞に深く感動することがある。			0.64		0.41	0	4	196	0	0.980
14 人がプレゼントを開けるところを見るのが好きである。			0.41		0.17	0	12	164	24	0.820
1 グループの中に見知らぬ人が一人ぼっちでいるのを見ると，自分も悲しくなる。				0.67	0.45	0	0	0	200	1.000
27 動物が苦しんでいるのを見るとひどく動揺してしまう。				0.67	0.44	0	0	0	200	1.000
29 無力な老人を見ると動揺してしまう。				0.65	0.42	0	0	200	0	1.000
19 ひどい目にあっている人を見ると，強い憤りを感じる。				0.63	0.40	0	0	200	0	1.000
12 職業訓練所で働くよりは社会福祉関係の仕事をしたい。				0.52	0.27	0	0	200	0	1.000
因子寄与	3.747	2.815	2.458	1.99	11.006					
因子間相関	1.00	−0.20	0.28	0.30						
	−0.20	1.00	−0.44	−0.32						
	0.28	−0.44	1.00	0.35						
	0.30	−0.32	0.35	1.00						

たかを項目ごとに集計したものである。

なお，resampling の分析でも当然極大の問題は発生するから，1つの resampling ごとに 100 回の反復を行い，最大値に対応する解を選択している。図 7 を見る限り，q=6 では真の最大値に到達できない危険性もなくはないが，ほぼ安全な反復数であると考えられる。

こうして得られた，各 200 の resampling の結果は，次の 3 つの基準で評価することができる。

(1) 全サンプルによるクラスターが，resampling において出現する比率。

(2) Resampling によるクラスターに属する項目数の安定度。

(3) 項目ごとの全サンプルによるクラスターへの的中率

このうち，(1) はクラスター自体の存否の判断に，(2) はクラスターの安定性に，(3) は個別項目の安定性にそれぞれ対応すると考えられる。このように resampling によるクラスタリングの評価は，通常の定められたモデルのパラメータ推定の精度以上の意味をもつとも考えられる。

表 7 は，q= 2 ~ 6 の解の各因子について，上記の (1) と (2) を評価するために resampling の結果を要約して示したものである。解の中にそれぞれの因子（クラスター）が存在した割合を，存在率として，また，それぞれの因子に負荷する項目数については，平均値，標準偏差，最大値，最小値を，さらに，当該の因子に全サンプルの解で負荷した項目について，resampling の解でも負荷した（当該のクラスターに所属した）的中率を最後の列に示した。なお，的中率については，解全体について算出したものと，因子ごとに算出したものの両方を示している。

これによると，まず，2 因子解では，2 つの因子は常に存在しており，負荷する項目数の揺らぎも小さく，的中率も 0.97 を超えており，sampling の揺さぶりに対して非常に安定した解が得られていることがわかる。それに対し，5 因子以上の解では，的中率の低い因子が現れ，さらにそうした因子では，負荷する項目数の揺らぎも大きいことから，安定した解は得ら

表7　Resampling の結果の要約

			負荷する項目数				
因子	存在率	平均値	標準偏差	最大値	最小値	的中率
2因子解						0.978
1	1.000	21.91	0.71	23	20	0.982
2	1.000	11.10	0.71	13	10	0.972
3因子解						0.941
1	1.000	9.85	0.62	12	9	0.960
2	1.000	12.97	1.81	18	7	0.926
3	1.000	10.19	1.92	17	5	0.942
4因子解						0.944
1	1.000	9.68	0.60	11	9	0.950
2	1.000	11.69	1.49	19	6	0.915
3	0.980	6.27	1.44	11	0	0.945
4	1.000	5.37	0.57	7	5	1.000
5因子解						0.897
1	1.000	9.25	0.54	12	9	1.000
2	1.000	5.55	0.77	8	4	0.899
3	0.995	8.53	1.77	14	0	0.891
4	1.000	5.21	0.43	7	5	1.000
5	0.880	4.46	2.08	11	0	0.660
6因子解						0.852
1	1.000	7.63	1.52	10	6	1.000
2	1.000	5.73	0.96	8	4	0.897
3	1.000	8.28	1.56	13	5	0.884
4	1.000	5.18	0.39	7	5	1.000
5	0.475	1.67	1.80	5	0	0.473
6	0.965	4.54	1.51	9	0	0.690

れていないように見える。また，3因子解と4因子解を比較すると4因子解には1つだけ存在率が1に達しない因子（「感情移入」）が存在するものの，負荷する項目数の安定度も高く，的中率も全体としては3因子解をわずかに上回っている。

表2に示した5因子解では，表6の4因子解の第2因子が第3と第5の2つの因子に別れた結果となっているが，表7ではその2つの因子の安定性が低いことがわかる。こうしてみると，きわめて安定した結果は得られるものの，内容の多様性という点では劣る2因子解と比較的安定度の高い4因子解が尺度構成の基礎となりうると考えられる。他の変数との関係にもとづいて研究を発展させることを考えるに際して，まず4因子解を採用

するのが合理的であろう[vi]。加えて，表4で最適解の近傍にそれに近い極大解がほとんど存在しないことも（この33項目を所与とする限り）この解を採用したくなる理由である。

そこで改めて表7を検討してみよう。4因子解の第1因子は存在率と安定性には問題ないものの的中率が0.95とやや低い。この原因は，表6からわかるように，resampling においてちょうど100回ずつ第1因子と第2因子に負荷する，項目5にある。この項目は，一読すると他者の感情への感染しやすさを示しているようにも思えるが，他者の感情を迷惑なものと感じる「他者の感情の無理解」を反映しているともとれる。つまりあいまいな項目である。これと，主に第2因子に負荷するものの，その負荷量が著しく低い項目33の2つの項目を削除し，改めて4因子解を求めた結果の要約が表8である。おおよそ満足すべき存在率と安定性を示している。表8には，全サンプルによる解にもとづく，単純和による尺度のα係数も示した。安定した存在である第4因子から作られた尺度のα係数が，項目数の少なさもあってやや低いが，個人のアセスメントでなく変数間の関連にもとづく研究のための最低限の基準はクリアーしていると見なせよう。とりあえず，研究の第1段階としては，これら4つの尺度を構成しえたことで，PCCA の効用を確認できたとしてよいであろう。

表８　２項目を削除した場合の resampling の結果と尺度の信頼性

		負荷する項目数					尺度		
因子	存在率	平均値	標準偏差	最大値	最小値	的中率	平均相関	項目数	α係数
						0.971			
1	1.000	9.54	0.34	10	9	1.000	0.331	9	0.816
2	1.000	10.54	1.00	17	8	0.944	0.169	11	0.692
3	0.995	6.07	1.00	9	0	0.951	0.283	6	0.703
4	1.000	5.27	0.47	7	5	1.000	0.244	5	0.618

7. 討論と今後の問題

本研究では，1次合成変量への回帰を重回帰から単回帰に切り替えるという単純なアイディアによって，完全単純構造を実現するための新しい方

法を提案した。この方法の幾つかの性質を証明するとともに，実データへの適用を通じて一定の実用性を示した。さらに，resampling による確認的分析によって，解を多面的に評価する方法を導入し，これについても実データで効果を確認した。最終的に，若干の項目の削除を行った上で，単純和による尺度構成とその信頼性の評価に至るプロセスの一例を示した。

このように，ここで提案した方法は，ある程度の実用性を示したが，多くの批判がありうることも事実であり，ここではまず，その幾つかについて検討しておこう。

まず，この方法が単純であり，近年数多くの研究がある sparse PCA（たとえば，Jollife, 2002）のアルゴリズム開発の成果を考慮していないことが指摘される。ただ，本研究が目的としているのは，項目反応の単純和としての心理学的尺度構成の方法であり，比較的大きな分散を説明するとともに，ほぼ等質的な重み（この方法では負荷量に比例する）をもつ解が求められるから，本研究の方法は，その目的には十分であるとも考えられる。標本に対して変数の数が著しく多いなど，より厳しい条件のデータや，非常に説明力の小さい主成分に着目する必要があるといったケースでは，本研究の方法の効果には限界があって当然である。

次に，本研究で採用したアルゴリズムが，単純な hill climbing 法であり，極大の問題をうまく回避できていないのではないかという問題がある。これについても，焼きなまし法や遺伝的アルゴリズムのような進んだ方法（たとえば，Givens & Hoeting, 2013）よりも，3 節で示したアルゴリズムはコーディングが簡単で頑健であり，十分な反復数を確保する限り，本研究が想定しているような適用場面では十分であるように思われる。

さらに，質問項目への反応には多様な影響要因が考えられる（たとえば，村上，2008）心理測定データに対して，完全単純構造という制約条件は厳しすぎ，実態を捉えていないのではないかという批判がありうる。しかし，これも最終的には単純和としての比較的少数の尺度が目標であり，そこに

一定の信頼性が求められるとともに，複数の尺度間で共通項目が存在することによって人工的に作り出される尺度間相関のことを考慮すれば，この単純性は適用領域に対して適切なレベルにあると見ることもできよう。

また，主成分分析に比して説明力の低下が著しいことも問題となろう。本研究の適用例で採用した4クラスター解の主成分分析に対する効率，約93％（図6）を高いとみるかどうかには議論があるであろう。しかしながら，負荷行列から因子ごとに目立った項目を選択して単純和によって尺度構成を行うという手続きでは，さらに多くの説明力を切り捨てていることも考えるべきであろう。

確認的分析に採用した resampling の方法についても検討の余地があろう。典型的な resampling 法である bootstrap 法は，通常，定められたモデルにおいてパラメータ推定値の標準誤差を求める方法として用いられるのに対し，本研究ではクラスタリングの安定性という観点で使用している。こうした使用法は筆者の知る限りあまり例がないようであり，理論上の問題がある可能性もある。また，resampling の試行数が200というのも，十分であるかどうか検討の余地があろう。ただ，定められた因子解の標準誤差よりも，因子の成立自体とその安定性，さらに個々の項目の的中率という3つの観点で評価できる個々での用法は魅力的であり，何らかの改善を行ってでも維持することを望みたい。

最後に今後の発展の方向について触れておこう。まず，2節で触れた部分的仮説（partial hypothesis）の導入の問題がある。この方法では，幾つかの項目を特定のクラスターに固定して動かさないという簡単な方法で部分的仮説を導入することができる。この方法の実データでの検討は期待のもてる方向である。

次に，残余項目をアルゴリズムの中で排除していく方法を考えることである。これには，4節で触れたα係数の和の最大化を基準とした方法（Ten Berge & Hofstee, 1999）が考えられる。ただし，この方法の実用化には多くの問題があり，現行の方法での運用の工夫とあわせて慎重な検討が必

要である。

もう1つの適用領域として2値データの分析がある。2値データについては，いわゆる馬蹄現象等，さまざまな余剰次元の問題が指摘される（たとえば，村上，2011；Murakami, 2012）。項目クラスター内では1次元の分析に限定される本研究で示した方法では，こうした余剰次元が原理的には回避できる。これも多くの適用例を積み重ねることによって，その性質と問題点は明らかになっていくであろう。

文　献

足立浩平・村上隆（2011）．非計量多変量解析法　主成分分析から多重対応分析へ　朝倉書店

Bernaard, C.A., & Jennrich, R.I.（2003）. Orthomax rotation and perfect simple structure. *Psychometrika*, 68, 585–588.

Braverman, E. M.（1970）. Methods for the extremal grouping of parameters and the problem of determining essential factors. *Automatíon and Remote Control*, Ⅰ, 108–116.

Carmine, E.G. & Zeller, R.A.（1979）. *Reliability and validity assessment*. London: Sage.

Cronbach, L.J.（1951）. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

DeVellis, R.F.（2003）. *Scale development*: *Theory and applications*. 2nd ed. Thousand Oaks: Sage.

Digman, J.M.（1996）. The curious history of five factor model. In J.S. Wiggins（Ed.）*The five–factor model of personality*, 1–20. New York: Guilford.

Efron, B., & Tibshirani, R.J.（1993）. *An introduction to the bootstrap*. New York: Chapman & Hall.

Escoufier, Y.（1988）. Beyond correspondence analysis. In H.H. Bock（Ed.）, *Classification and related methods of data analysis*. Amsterdam: Elsevier Science Publishers, 505–514.

Givens, G.H. & Hoeting, J.A.（2013）. *Computational Statistics*. 2nd ed.

Guttman, L. (1954). *A new approach to factor analysis : the radix*. In P.F. Lazarsfeld (ed.) Mathematical thinking in the social sciences. New York : Russell & Russell. 258-348.

Harman, H.H. (1967). *Modern factor analysis*. 2nd ed. Chicago : University of Chicago Press.

Harshman, R.A., & Lundy, M.E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder, Jr., Hattie, J.A., & McDonald, R.P. *Reserch methods for multimode data analysis*. New York : Praeger.

市川雅教（2010）．因子分析　朝倉書店

Izelman, A.J. (2008). *Modern multivariate statistical techniques ; Regression, classification, and manifold learning*. New York : Springer.

Jollife, I.T. (2002). *Principal component analysis*. 2nd ed. New York : Springer.

Kiers, H.A.L. (1993). A comparison of techniques for finding components with simple structure. In C.M. Cuadras and C.R. Rao, (eds.) *Multivariate analysis : future directions* 2, Amsterdam : Elsevier Science Publishers, 67-86.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, No.140, 3-55.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40, 525-543

村上　隆（2008）．個人差測定尺度のプロトタイプ理論と主成分のプロトタイプ変換　中京大学現代社会学部紀要，1, 141-177

村上　隆（2011）特別な構造をもつ２値データの相関行列の性質について　中京大学現代社会学部紀要，5, No.1, 107-124.

Murakami, T. (2012). A geometrical interpretation of the horseshoe effect in multiple correspondence analysis of binary data. In W. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, and J. Kunze (eds.) *Challenges at the interface of data analysis, computer science, and optimization*. Berlin : Springer, 101-108.

村上隆（2013）．セミメトリック主成分分析と３値データへの応用　日本行動計量学会第41回大会抄録集，196-199.

中村陽吉（編著）（2000）．対人場面における心理的個人差　測定対象についての分類を中心にして　ブレーン出版

Nunnally, J.C.（1978）. *Psychometric theory. 2nd ed*. New York : McGraw Hill.

Robinson, J.P., Shaver, P.R., &, Wrightsman, L.S.（eds.）（1991）. *Measures of personality and social psychological attitudes*. San Diego : Academic Press.

Schott, J.R.（1997）. *Matrix analysis for statistics*. New York : Wiley.

芝祐順（1979）．因子分析法　東京大学出版会

Streiner, D.L. and Norman, G.R.（2003）. *Health measurement scales : A practical guide to their development and use. 3rd ed*. Oxford : Oxford University Press.

Ten Berge, J.M.F., & Hofstee, W.K.B.（1999）. Coefficients alpha and reliabilities of unrotated and rotated components. *Psychometrika*, 64, 83-90.

吉村功（1969）．数理統計学　培風館

---

i　この値は，素点をそのまま加算して得られる合計点ではなく，標準得点を加算して得られる得点のα係数である。統計ソフトウェア SPSS では，standardized alpha coefficient として出力されているものである。

ii　中村（2000）では，これらの因子は「対他感情移入的動機」,「対人冷淡感情」,「対芸術共感性」,「対弱者共感性」と呼ばれている。

iii　最適化基準をα係数の和に置き換えることによって，不要（有害）項目を PCCA の段階で削除する基準が生まれてくる。ただし，そのような変更によって，各クラスターには最小限２個の項目を所属させなければならなくなるなど，方法の使い勝手が悪くなることも否定できない。

iv　ある週の６日間の通勤時間を記録し，月，火，木曜日がやや長め，残る３日が短めであることを見た後で，同じデータで月，火，木と水，金，土の平均値の差の検定を行い，有意であったことをもって，月，火，木は通勤時間が長くなると結論するという例を,「統計学とは無縁」として退けている。探索的因子分析の結果を仮説として同一のデータで確認的因子分析を行うのは，これと同じ誤りを犯しているように思われる（後半は村上の意見である）。

v　Bootstrap は靴紐である。この名称の由来は日本では「ほら吹き男爵」と呼ばれているドイツの小説にある。湖底に落ちたミュンヒハウゼン男爵は，自分の靴紐を持ち上げることによって浮かび上がる（Efron＆Tibshirani,1993）。

vi　他の変数との関係の分析の結果，余分な因子があるとか，何かが欠けているという可能性が出てくることはもちろんありうる。