因子分析におけるある種の順位カテゴリーの最適尺度化について

Some Comments on the Optimal Scaling of a Certain Type of Ordered Categories in Factor Analysis.

(1991年4月3日受理)

福 森 護
Mamoru Fukumori

Key words: Factor analysis, ordered categories, optimal scaling

Abstract

This paper proposes an algorithm which helps apply a certain type of ordered categories, such as k-point scale, to factor analysis. For this purpose, a comparison is made of the data which exhibit small variations in factor analysis, and the validity of the proposed idea is discussed.

1．はじめに

因子分析法は，心理学をはじめとする人文科学や社会科学などの分野においては，次に示されるような多段階評定尺度（k-point scale）によるスコアで実施されることが多い。

大賛成	賛成	少し賛成	少し反対	反対	大反対
1	2	3	4	5	6

しかし，このような多段階評定尺度は順位カテゴリーであり，各スコア間の順位関係は保証されているものの，スコア間の間隔については等間隔であるとは断定できない。したがって，順位カテゴリーの最適な尺度化を考える必要があると思われる。そこで，本研究では，スコアを微小に変化させることによる順位カテゴリーの最適尺度化のアルゴリズムの提案を行い，また，簡単な数値例によって，順位カテゴリーによる因子分析の妥当性について検討を行う。

2．最適尺度化の手順

2.1 データの構造

データは，多段階尺度評定法（k-point scale）により測定されていることを前提とする。また，k段

階のそれぞれにt_1，t_2，…，t_kのようなスコアが与えられているものとする。ここで，例えばk＝5としたとき，尺度の任意性から，$t_1=1$，…，$t_5=5$というスコアを与えることも，対称性を仮定して，$t_1=-t_5=-1$，$t_2=-t_4$，$t_3=0$というスコアを与えることも可能である。

2.2 分散共分散のスコアによる表現

全個体数をn，第i項目のk番目のカテゴリーと第j項目のℓ番目のカテゴリーの両方に対応する個体数をf (ik, jℓ) としたとき，分散共分散は次のように表現できる。

$$f(ik, j\cdot)=\sum_{\ell'} f(ik, j\ell'),\quad f(i\cdot, j\ell)=\sum_{k'} f(ik', j\ell)$$

$$S_{ij}=(n-1)^{-1}\sum_{k'}\sum_{\ell'} f(ik', j\ell')\, t_{k'}t_{\ell'}$$
$$-\{n(n-1)\}^{-1}\left(\sum_{k'} f(ik', j\cdot)\, t_{k'}\right)\left(\sum_{\ell'} f(i\cdot j\ell')\, t_{\ell'}\right) \cdots\cdots ①$$

2.3 スコアの微小変化の影響

ある特定のt_kでの微分を考える。

$$\frac{\partial}{\partial t_k}\sum_{k'}\sum_{\ell'} f(ik', j\ell')t_{k'}t_{\ell'}=\sum_{\ell'} f(ik, j\ell')t_{\ell'}+\sum_{k'} f(ik', jk)t_{k'}$$

$$\frac{\partial}{\partial t}\left[\left(\sum_{k'} f(ik', j\cdot)t_{k'}\left(\sum_{\ell'} f(i, j\ell')t_{e'}\right)\right]$$
$$=f(ik, j\cdot)\sum_{\ell'} f(i\cdot, j\ell')t_{\ell'}+f(i, jk)\sum_{k'} f(ik', j\cdot)t_{k'}$$

故に

$$\frac{\partial_{sij}}{\partial t_k}=(n-1)^{-1}\sum_{\ell'}\{f(ik, j\ell')+f(i\ell'+jk)\}t_{\ell'}$$
$$-\{n(n-1)\}^{-1}\left[f(ik, j\cdot)\sum_{\ell'} f(i\cdot j\ell')t_{\ell'}+f(i\cdot jk)\sum_{\ell'} f(i\ell', j\cdot)t_{\ell'}\right] \cdots\cdots ②$$

即ち，あるt_kだけを$t_k+\varepsilon_k$に微小に変化させると，Sは，

$$S\mapsto S+\varepsilon_k S_k^{(1)},\quad S_k^{(1)}\equiv\frac{\partial S}{\partial t_k} \cdots\cdots ③$$

（ここで，$\partial S/\partial t_k$は，$\partial_{sij}/\partial t_k$を（i，j）要素とする行列）

のように変化する。

分散共分散Sの微小変化が，△とTにどのように影響するかについては，Tanaka and Odaka (1989) [2),3)]で議論されている。

③のような変化に対応する△とTの変化を

$$\hat{\Delta}\mapsto\hat{\Delta}+\varepsilon_k\hat{\Delta}_k^{(1)},\quad \hat{T}\mapsto\hat{T}+\varepsilon_k\hat{T}_k^{(1)} \cdots\cdots ④$$

と表すと，$\hat{\Delta}_k^{(1)}$，$\hat{T}_k^{(1)}$は主因子法，最尤推定法のいずれの場合についても容易に求めることができる。④のε_kの係数は，

$$\hat{\Delta}_k^{(1)}=\frac{\partial\hat{\Delta}}{\partial t_k},\quad \hat{T}_k^{(1)}=\frac{\partial\hat{T}}{\partial t_k} \cdots\cdots ⑤$$

を意味する。次に，

$$\mathrm{vecd}(\hat{\Delta}) \underline{\underline{d}} (\hat{\Delta}_{11}\hat{\Delta}_{22}\cdots\cdots\hat{\Delta}_{pp})^T$$
$$\mathrm{vech}(\hat{T}) \underline{\underline{d}} (T_{11},\ T_{12},\ \cdots\cdots,\ T_{1p};\ T_{22},\ T_{23}\cdots\cdots,\ T_{2p};\ \cdots\cdots;\ T_{pp})^T$$

とすると，

$$\frac{\partial Q}{\partial \varepsilon_k}=\frac{\partial Q}{\partial_{\mathrm{vecd}}(\hat{\Delta})^T}\mathrm{vecd}(\hat{\Delta}_k{}^{(1)})+\frac{\partial Q}{\partial_{\mathrm{vecd}}(\hat{T})^T}\mathrm{vecd}(\hat{T}_k{}^{(1)}) \cdots\cdots ⑥$$

となり，④式の$\partial Q/\partial \varepsilon_k$が評価できると，Qに対するsteepest directionは次のように定まる。

$$\frac{\Delta t_k}{[\partial Q/\partial t_k]_{\hat{\Delta}\hat{T}}}=\mathrm{CONST},\ \ k=1,\ 2,\ \cdots\cdots,\ K \cdots\cdots ⑦$$

kのうち，free parametersに対して，⑦で定まる方向にεだけ変化させる。
即ち，

$$t_k \mapsto t_k+\varepsilon t_k{}^{(1)},\ \ t_k{}^{(1)}=c\left[\frac{\partial Q}{\partial t_k}\right]_{\hat{\Delta}\hat{T}},\ \ \sum_k (t_k{}^{(1)})^2=1 \cdots\cdots ⑧$$

①の右辺のt_kを$t_k(\varepsilon)$で書き換えて，

$$\begin{aligned}S_{ij}(\varepsilon)&=(n-1)^{-1}\sum_{k'}\sum_{\ell'}f(ik',\ j\ell')t_{k'}(\varepsilon)t_{\ell'}(\varepsilon)\\&\quad-\{n(n-1)\}^{-1}\Big(\sum_{k'}f(ik',\ j\bullet)t_{k'}(\varepsilon)\Big)\Big(\sum_{\ell'}f(i\bullet j\ell')t_{\ell'}(\varepsilon)\Big)\\&=S_{ij}+\varepsilon S_{ij}{}^{(1)}+(\varepsilon^2/2)S_{ij}{}^{(2)} \cdots\cdots ⑨\end{aligned}$$

ただし，

$$\begin{aligned}S_{ij}{}^{(1)}&=(n-1)^{-1}\left(\sum_{k'}\sum_{\ell'}f(ik',\ j\ell')t_{k'}{}^{(1)}t_{\ell'}+\sum_{k'}\sum_{\ell'}f(ik',\ 1\ell')t_{k'}t_{\ell'}{}^{(1)}\right)\\&\quad-\{n(n-1)\}^{-1}\Big([\sum_{k'}f(ik',\ j\bullet)t_{k'}\}\{\sum_{\ell'}f(i\bullet j\ell')t_{\ell'}{}^{(1)}\}\\&\quad+\{\sum_{k'}f(ik',\ j\bullet)t_{k'}{}^{(1)}\}\{\sum_{\ell'}f(i\bullet,\ j\ell')t_{\ell'}{}^{(1)}\}] \cdots\cdots ⑩\end{aligned}$$

$$\begin{aligned}S_{ij}{}^{(2)}&=2(n-1)^{-1}\sum_{k'}\sum_{\ell'}f(ik',\ j\ell')t_{k'}{}^{(1)}t_{\ell'}{}^{(1)}\\&\quad-2\{n(n-1)\}^{-1}[\{\sum_{k'}f(ik',\ j\bullet)t_{k'}{}^{(1)}\}\{\sum_{\ell'}f(i\bullet\ j\ell')t_{\ell'}{}^{(1)}\}] \cdots\cdots ⑪\end{aligned}$$

なお，$S \mapsto S+\varepsilon S^{(1)}+(\varepsilon^2/2)S^{(2)}$のように展開されるときの$\hat{\Delta}$と$\hat{T}$の展開式

$$\hat{\Delta}(\varepsilon)=\hat{\Delta}+\varepsilon\hat{\Delta}^{(1)}+(\varepsilon^2/2)\hat{\Delta}^{(2)}$$
$$\hat{T}(\varepsilon)=\hat{T}+\varepsilon\hat{T}^{(1)}+(\varepsilon^2/2)\hat{T}^{(2)}$$

については，Tanaka and Castãno-Tostado (1990) において議論済みである。

以上の結果から，最適尺度化の手順をまとめると，次のようになる。

1）特定のt_kに対して
　$S_k(1) \equiv \partial S/\partial t_k$
　を②式により計算する。

2）$S \to S+\varepsilon_k S_k{}^{(1)}$を微小変化させたときの$\hat{\Delta}$と$\hat{T}$の微小変化

$\hat{\Delta} \mapsto \hat{\Delta} + \varepsilon_k \hat{\Delta}_k^{(1)}$, $T \mapsto T + \varepsilon_k \hat{T}_k^{(1)}$

の係数$\hat{\Delta}_k^{(1)}$, $\hat{T}_k^{(1)}$を求める。(Tanaka and Odaka (1989))

ここで，１）と２）をパラメータの数だけ繰り返す。

３）steepest direction ($t_1^{(1)}$, $t_2^{(1)}$, …) を⑦，⑧式により求める。

４）その方向にεだけ変化させたときの$S=(s_{ij})$ の展開式の係数$s_{ij}^{(1)}$, $s_{ij}^{(2)}$を⑨～⑪式により求める。

５）このようなSの変化に対応する△とTの２次の展開式の係数$\hat{\Delta}^{(1)}$, $\hat{\Delta}^{(2)}$；$\hat{T}^{(1)}$, $\hat{T}^{(2)}$を求める。(Tanaka and Castãno-Tostado (1990)[1])

６）その方向への最適なεを，

$Q(\hat{\Delta} + \varepsilon \hat{\Delta}^{(1)} + (\varepsilon^2/2)\hat{\Delta}^{(2)},\ \hat{T} + \varepsilon \hat{T}^{(1)} + (\varepsilon^2/2)\hat{T}^{(2)})$

をεの関数として調べることにより定める。

７）その点において因子分析を再実行し，$\hat{\Delta}$と$\hat{T}$を更新する。

８）収束判定を行い，未収束なら１）へ戻る。

３．スコアの微小変化による因子分析の結果の評価

ここに，６段階評定(大反対，反対，少し反対，少し賛成，賛成，大賛成)で測定された100サンプル13変量のデータがある。ここでは，対称性を仮定して，この６段階に対して，大反対$=-t_3$，反対$=-t_2$，少し反対$=-t_1$，少し賛成$=t_1$，賛成$=t_2$，大賛成$=t_3$とし，スコアの微小変化による因子分析の結果の評価を行う。$t_3=1$とすると，未知パラメータはt_1とt_2の２つとなるため，t_1とt_2の２つのパラメータだけを微小変化させることで評価を行うことができる。

このデータをもとに，制約式：$0<t_1<t_2<1$において，t_1とt_2を微小に変化させることを考える。各尺度間の距離を等間隔にした場合には，$t1=0.2$，$t2=0.6$となるので，これを初期値として，初期値を中心に0.1ごとのメッシュに切り，種々のスコアのもとで因子分析を行うことによりその結果を評価する。評価の基準についてはいくつかの方法が考えられるが，今回は第３因子までの累積寄与率を基準にして，主因子法（反復解法，$\varepsilon=0.00001$）の場合について評価を行う。

表１は，t_1とt_2を$0<t_1<t_2<1$のもとで，0.1間隔で変化させ，主因子法(反復解法，$\varepsilon=0.00001$)を適用したときの４因子の累積寄与率（%）を示したものである。また，図１は表１の累積寄与率について，横軸にt_1を，縦軸にt_2をとり，等間隔の場合（$t_1=0.2$，$t_2=0.6$）よりも累積寄与率の高かったばあいのスコアに☆印をつけたものである。

表１　0.1間隔でスコアを変化させたときの累積寄与率（%）

t_2t_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	55.792							
0.3	56.634	55.876						
0.4	56.470	56.258	55.645					
0.5	55.951	56.772	56.187	55.630				
0.6	55.071	55.915	55.362	54.639	54.204			
0.7	54.855	54.712	54.191	53.182	53.526	53.191		
0.8	54.344	54.206	53.714	53.184	52.755	52.380	51.812	
0.9	53.585	54.445	53.578	52.895	52.127	51.819	50.318	49.717

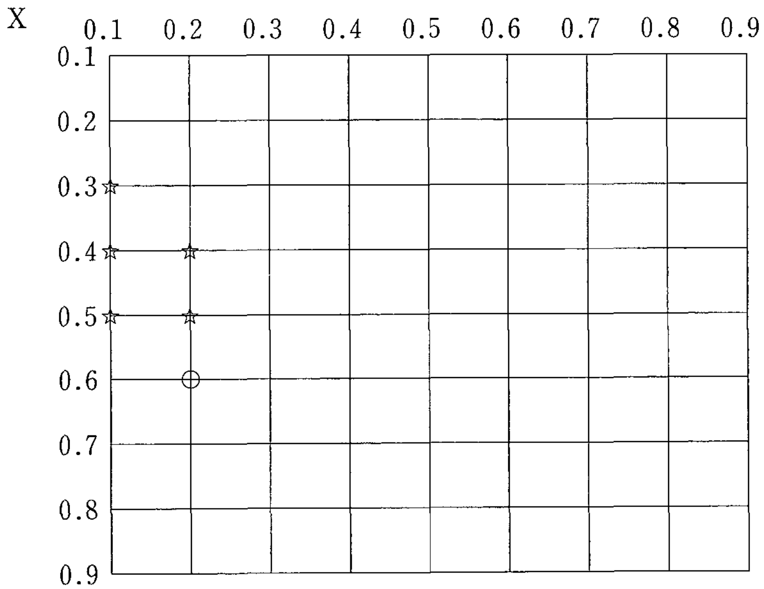

図１　等間隔スコアと微小変化スコアの比較図

これをみると，いくつかのスコアにおいては，等間隔の場合よりも累積寄与率が高くなっていることがわかる。実際の因子構造の変化については，今後検討が必要であろうが，累積寄与率を高めるという方向で，最適なスコアを与えることが可能であることを示唆する結果といえる。また，今後，より種々な順位データに適用してみることにより，等間隔地点をスタート点として，山登り的に最適なスコアを探索するアルゴリズムを一般化できるものと考えられる。

今回の報告では，順位カテゴリーによる因子分析の結果の妥当性に関する問題提起とそのアルゴリズムに関するアイデアの提案を行ったが，今後に残された問題点として，次のようなものが考えられる。

1）今回は，対称性を仮定することにより，より少ないパラメータによる因子分析の評価を行ったが，非対称性を仮定して，全てのパラメータを微小変化させた場合の結果の評価を行う必要がある。

2）パラメータの制約式をつけて解くかどうかで制約無しの最適化，制約有りの最適化の問題となるた

め，制約式の有無について検討を行う必要がある。

3）結果の評価方法としては，今回は累積寄与率を用いたが，他の方法も考えられるため，評価基準について検討を行う必要がある。

4）今回は主因子法のみで評価を行ったが，最尤推定法の場合なども別に行う必要がある。

参 考 文 献

1）Tanaka, Y. and Castãno-Tostado, E. (1990). Sensitivity analysis in multivariate methods: Decomposition of an arbitrary influence into a finite number of components. Comm. Statist., A19, No. 4.

2）Tanaka, Y. and Odaka, Y. (1989a) Influential observations in principal factor analysis. Psychometrika, 54, 475-485.

3）Tanaka, Y. and Odaka, Y. (1989b) Sensitibity analysis in maximum likelihood factor analysis. Comm, Statist., A18, 4067-84.