

最小次元解析の課題

青山学院大学文学部 丸山久美子

はじめに:

最小次元解析 (Minimum Dimension Analysis) ^(3*) とは多数事象から成る諸現象の中に含まれている不必要と思われる部分を切り捨て、現象の本質となるべく少ない変数でおさえながら、その本質をみきわめようとする目的のもとに考案された多次元尺度構成法の1つである。つまり、現象を集約し、それを簡潔に表現しようとするオツカムの剃刀-parsimony-を徹底させたものといえよう。元来、Minimum Dimension という概念は多くの多変量統計解析が考案されてきた時点で、この分析技法の性質上、つねに使用され、議論の的になるべき筈であるが、かならずしも、この概念をのっけから議論しようとするものではなく、多くは分析技法の中で次元の減少 (reducing dimensionality)、最小次元性 (minimum dimensionality)、さらには測定値結少 (data reduction) といっ

た説明概念にとどめる程度であつたように思われる。例えば因子分析法などは複雑なデータをすっきりさせ、わかり易いものにすむための方法で、データを2次元平面空間、及至は3次元立体空間で表現しようとするものである。この種の空間的表現はもっぱら2次元平面の布置図を描くにとどめることが多い。とはいえ、それだからといって、このデータは2次元で全てが説明されるとするわけではない。所謂、次元(軸)の数は変数の数だけ存在するわけで、ここで言っている次元というのは数学的次元ではなく、経験的に言うならばクラス分け程度の意味をもつものであり、多くの変数をもなるべく効率的に分類することを目的としている。その場合の分類の基準は相互に相関の高いもの同士、及至は似ている程度の高いもの同士が集まるように工夫され、その基準にてらして次元の解釈をするのである。というのは、多くの変数の中には相互に関係の強いものが含まれており、その現象を説明に要する変数はすつと少なくなる筈だからである。変数間の相互関連性を検討し、なるべく少変数で説明するためにクラス分けするという分析技法の開発される所以はここにある。事前に集団分割が出来ていれば、このような手間を省くことが出来るかもしれないが、複雑な現象一般においては事前にこのような処理の出来ないものが多いために、因子分析等に頼つ

て現象解析の一步を踏み出すことになる。そこでは次元の数は少ない方がよいが、かならずしも最小でなくてもよい。どこまでの次元がその現象にとって有意な情報を提供するかも模索するための模索法なども考案されているが、あくまでも便宜である。そこで、どこまでの次元がその現象の本質をとらえているかを明らかにするために模索法とは全く異なった視点から分析することと考えてみよう。

Torgerson, W. S (1952) の多次元尺度構成法(MDS)が考案されて以来、その一般解ともいうべき方法として、Shepard, R. N (1962) はなるべくゆるい条件の下で最小の次元をみつけ出してゆくための分析法を非計量的(Non-metric) MDS と称して MDS 一般に 1 つの道を開いた。因子分析法、Torgerson 流の MDS (もつぱら、これを metric MDS と言っている) よりも、もっと parsimonious なものである。すなわち、ある 1 つの最小次元を保持しようとする目的関数(一般に単調関数)を定め、それを 1 次元の場合、2 次元の場合といった具合に低次元から始めて、次元毎にあてはまりの状況を考え、なるべく 3 次元位で収めてしまおうという方法を採用している。つまり、所与のデータは $r_{ij} = f(d_{ij})$ という形式を採用。(この場合、 f は単調関数で、 d_{ij} は距離を表わす) どうしても、3 次元で収まらない場合ははじめからやりなおして、所謂、変数選択を

したり、変数間の相互作用過程などを考えるためにクロス集計からやりなおしをするという試行錯誤的立場を採る方法である。Shepard の方法のアルゴリズムを整えた Kruskal, J.B. (1964) の場合ははじめから高次元(ほとんど3次元とあらかじめ決めてある)より始めて、順次、次元を下げてあてはまりの状況を検討するというプロセスを進める。このような Shepard-Kruskal タイプの非計量的 MDS では一見錯綜(confusion)しているようなデータであっても、その性質がすでに明らかなるものであるという予想が成立しているものでなければ見通しが暗い。つまり、あらかじめその現象について何らかの見通しがついているような場合にごくあたりまえの結果しか得られないことが多い。というのは、初期値の与え方がデータそのものを全く無視した機械的な与え方(Shepard は真間の距離が相互に全て等しくなるような正則単体(regular simplex)の頂点の座標, Kruskal は全てランダムな値)をするので、かならずしもそのデータに適合するようなものではないからである。この意味において、最小次元を得るまで強引に iteration をくり返し、何とか結論を得ようとするわけだから、結果についても妥当なものであるという保証がなされにくい。そこで、測定値の再現性問題が真実となってくる。さらに又、3次元の空間で全てを説明しようとすでに次元の

数と定めていなければならないわけだからそのデータが3次元位で収まるという前提が必要であり、複雑な社会現象などには適合性が低いものと言わざるを得ない。その点以下にのべる Guttman, L (1968) の最小空間分析 (Smallest Space Analysis, 一般に SSA と称している), 林知正夫 (1974) の最小次元解析 (一般に MDA と称している) では初期値の与え方がそのデータの性質を忠実に反映するという意味でユニークである。

SSA と MDA :

SSA はその分析の当初からあらかじめデータの性質をよく吟味するためにフェイス・デザイン (facet design) という要因を事前に定めておくようなデータの配置構造を考慮しておくことが必要である。これには、Guttman 独自のランク・イメージの原理というものによってデータをランクに変換するといった操作が入る。フェイスで組まれたデータは距離順位数 D_{ij} に変換される。これは、ある距離 d_{ij} があり、これを近いものから遠いものに順位付けたものを距離順位数 D_{ij} とするのである。そのデータの最大固有値の固有ベクトルが初期値として与えられるわけである。単調性の条件も弱単調性, 準単調性, 強単調性などデータの性質によって様々な単調性が考えられている。MDA では弱単調性の条件から出発してデータが順位のついた4個のグループに分類されている。そ

のように分類したものを分類するもの。これは、タイが多く、変数が多い時に適用すれば効果的である。初期値も SSA と同様最大固有値の固有ベクトルである。従って、SSA と MDA の求める結果は初期値の与え方が同じである限り、アルゴリズムにそのちがいがあつたとしても、そういった差異は認められない。各々の方法は、このようにして、1次元の場合から始めて、適合度係数 (SSA の場合は $\mu = U/\sqrt{VW}$, 但し、 $U = \sum \sum e_{ij} \hat{D}_{ij} d_{ij}$, $V = \sum \sum e_{ij} \hat{D}_{ij}^2$, $W = \sum \sum e_{ij} d_{ij}^2$, $e_{ij} = \begin{cases} 1 & D_{ij} \text{が与えられている時,} \\ 0 & \text{与えられていない時,} \end{cases}$ $\hat{D}_{ij} = f(D_{ij})$, D_{ij} は d_{ij} のランク・イメージによって得た値を各々示す。), MDA では G 値にクラス分けしたものの間の相関比 η^2) を用いる。ちなみに Shepard-Kruskal タイプでは $\text{Stress} = \sqrt{S^*/T^*}$, $S^* = \sum \sum \{\hat{d}_{ij} - f(d_{ij})\}^2$, $T^* = \sum \sum \hat{d}_{ij}^2$ でこの Stress を最小にするまで反復計算をくり返す。その意味において、Stress-Minimum Procedure ともいわれている。)の様子を検討し、あてはまりの具合が悪ければ2次元にするというくり返し反復法を用いる。その結果によって、そのデータの性質が1次元で説明がつくか、及至は2次元、3次元と次元を伸ばして行かねばならないものかも μ や η^2 の値から判断するのである。これらの方法は、ともに評価出来ないものであり、あるデータには SSA が良く、あるデータには MDA が良いといった具

合にデータの性質に依存した結果からしか議論することは出来ない。最小次元解析およびMDが一般の特徴はこのようになっているので、各々の方法の妥当性を論ずることは意味のないことであるという結論を生む。従って、今後は広範囲のデータを用いて各々の方法を計算してみ、その後、データの分類を行うことにより、どのようなデータにはこの方法が良くあてはまるというようなところに問題点をしぼるこの方がより効果的であるように思われる。

あわりに：

最小次元解析の主要目的である最適最小次元性といかに決まるかという課題について今少し考察してみることしよう。前述のように、これらの方法は Torgerson 以来、相当数の分析手法が開発され、その内容もしだいに現象解明と密接に結びついたかたちで工夫されるようになり、従って、なるべくゆかい条件の下で融通性のあるものが多く出現してきた。空間の名称もかならずしも Euclid 空間ではなく、Non-Euclid 空間で分析出来るほどに柔軟性を持たせている。例えばある距離 $d_{ij} = \left(\sum (x_i - x_j)^p \right)^{1/p}$ がある時 $p=2$ の場合が Euclid であり、この p の値によって空間の名称が変わってくる。 p の値もデータの性質に依存する。このように考えれば、いかなる分析手法が開発されたとしてもデータの性質に依存するわけだ

から、事前にデータに対するより一層の認識がせられてい
る。従って、今後は何れかものごとたにした錯綜データをやめ
くもに分析するというよりも、あらかじめそのセット的次元操作
によるデータの配置、個人差など十分に考慮したり、集団
分割をしたり、あるいは個人差を完全に処理しうるような分
析法の工夫を可なりと方向に沿って発展してゆくべきよう
に思われる。なお、現在のところ、もっぱら、計量的アプロ
ーチによる個人差を考慮したMDSが開発されているが、最
小次元解析においても、もし事前にこのようなかたちで個人
個人の重みづけを考慮しうるような方法が開発されることが
期待されるであろう。

又、分析技法が多種あるということ、どれを用いるのが
妥当であるかも方法論の側から論ずることが必要である。す
なわち、Shepard-Kruskalタイプもさることながら、SSA, MD
Aの場合もデータの再現性問題について深く議論するべきこ
とが要請されている。

(注) Minimum Dimension AnalysisというのはGuttmanのSSAが開発された
当時、その手法の名称にちなんでつけられたもので、当時は併知2つの一連の数量化
理論の中で、むしろ不変性をせまうれいたK-1型数量化法とその逐次近似的
手法、最小次元を見つね出すという目的が明らかであったところ、SSAと対比せ
ながら、検証した際には、はじめて使用された名称である。(Maruyama, K. 1969, 1970)
但し、K-1型数量化法はその分析法の性質上、制約条件が厳しく、非計量的と
いうよりも、計量的(metric)、乃至は準計量的(semi-metric)な分析法とい
えるので、データの性質が明らかに定まっているような、しかも、少変数のものに

しか適用可能ではない。その意味において MDA は K-L 型数量化への一般化したものといえよう。

References

Guttman, L. 1968. A general nonmetric technique for finding the Smallest Space Analysis for a configuration of points. *Psychometrika*, 33, 469-506.

Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematical-statistical point of view. *Ann. Inst. Statist. Math.*, 2, 69-98. 1974. Minimum Dimension Analysis MDA—one of the methods of multidimensional quantification (MDQ), *Behaviormetrika*, No 1, 1-24.

Kruskal, J.B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, 29, 1-29. Nonmetric multidimensional scaling: A numerical method, *Psychometrika*, 29, 115-129.

Maruyama, K. 1969. On the minimum dimension analysis(I)—New approach by quantification theory—, *Jap. Psychol. Res.*, 11, 134-146. Comparison between SSA and K-L type technique concerning metric and nonmetric problem. Unpublished manuscript for the Symposium on Minimum Dimension Theory, *Inst. Statist. Math.*, 1970

Shepard, R.N. 1962. The analysis of proximities: Multidimensional scaling with unknown distance function. Part I, II. *Psychometrika*, 27, 125-140, 219-246.

Torgerson, W.S. 1952. Multidimensional scaling: I Theory and method. *Psychometrika*, 17, 401-419.