数理解析研究所講究録
第231巻 1975年 104-112

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（法））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すて，現象の本質を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ッカムの剃刀－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縮少（reducing dimensionality），最小次元性（minimum dimensionality），さらには測定値縮少（data reduction）といっ

た説明概念にとどめる程度であったように思われる。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るための方法で，データを2次元平面空間，乃至は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志，乃至は似ている程度の高いもの同志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ずっ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くあるために，因子分析等に頼っ

て現象解析の一歩を踏み出することになる。そこでは次元の数は少ない方が良いがかならずしも最小でなくても良い。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を検定するための検定法なども考案されているが，あくまでも便宜的である。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検定法とは全く異なった視点から分析することを考えてみよう。

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(もっぱら，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うるような目的関数(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はじめて，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っている。つまり，所与のデータは $r_{ij}=f(d_{ij})$ という形式を採る。(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次元とあらかじめ定めてある)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辿る。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は点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剣となってくる。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

数を定めている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えない。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），林知己夫(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ァセット・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え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従ってデータをランクに変換するという操作が入る。ファセット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づ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の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から分析をすすめる。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ったとしても、そうたいした差異は認められない。各々の方法は、このようにして、1次元の場合からはじめて、適合度係数（SSAの場合は $\mu = U/\sqrt{VW}$,

但し、$U=\sum\sum e_{ij}\hat{D}_{ij} d_{ij}$, $V=\sum\sum e_{ij}\hat{D}_{ij}^2$, $W=\sum\sum e_{ij} d_{ij}^2$,

$e_{ij}=\begin{cases}1 & D_{ij}\text{が与えられている時}\\ 0 & \text{しからざる時}\end{cases}$ $\hat{D}_{ij}=f(D_{ij})$, 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個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 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次元にすすむ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、乃至は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を μ や η^2 の値から判定するのである。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MDS一般の特徴はこのようなところにあるので、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て、その後、データの分類を行うことにより、こ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との方がより効果的であるように思われる。

おわりに：

最小次元解析の主要目的である最適最小次元性をいかに決定するかという課題について今少し考察してみることにしよう。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、なるべくゆるい条件の下で融通性のあるものが多く台頭してきた。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 $d_{ij}=\{(\sum(x_i-x_j)^p)\}^{1/p}$ がある時、$P=2$ の場合がEuclidでこれら P の値によって空間の名称が変ってくる。P の値もデータの性質に依存する。このように考えれば、いかなる分析技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まられている。従って、今後は何もかもごったにした錯綜データをやみくもに分析するというよりも、あらかじめファセット的次元操作によるデータの配置、個人差などを十分に考慮したり、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するなどの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出来るであろう。

又、分析技法が多種あるということで、どれを用いるのが妥当であるかを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、MDA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、その方法の名称にちなんでつけられたもので、当所は林知己夫の一連の数量化理論の中でもっとも検証をせまられていたK-L型数量化法をその逐次近似的手法、最小次元を見つけ出すという目的が明らかであったことからSSAと対比させながら検証した時に、はじめて使用された名称である。(Maruyama, K. 1969, 1970) 但し、K-L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型数量化法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 Sapce 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 Teory, 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. Psyhcometrika, 17, 401-419.