

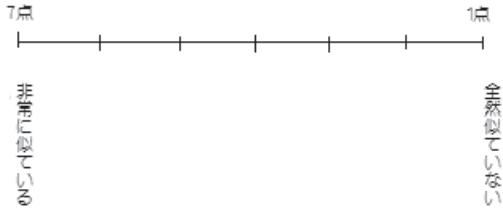
# 多次元尺度法概論とそのアルゴリズム

奥 喜 正

## 1. 多次元尺度法概説

多次元尺度法(MultiDimensional Scaling ; MDS)は人間の感覚器官, 視覚や聴覚器官が, ある刺激に対してどのように反応するか, そのしくみを心理的距離によって捉えようとする試みから発展したデータ解析メソッドである. その応用分野は, 当初は計量心理学の分野であった. その後, マーケティング戦略における分析メソッドとして, 認知マップとしての製品マップの作成に貢献した. 最近では, MDSの研究者である Everitt が Institute of Psychiatry(精神病理学)に所属しているように, 精神病理学や臨床心理学の分野で, 欧米では応用が始まっている[1][2]. 人間の内臓の状況を把握するための画像診断装置がCT, MRであるように, 人の内的世界の状態を把握するための画像的診断ツールの一つとして, MDSが有効なツールになる日が到来することを期待したい. そこで, 本稿では多次元尺度法の概説と, その解を得るためのアルゴリズムについて言及する.

さて, 多次元尺度法とは, 対象間の「非類似性データ」量を距離データに変換して, 潜在する少数の次元や背後にある認知構造を明らかにしたいときに利用する多変量データ解析手法の一つである. 非類似性データ(dissimilarity data ; proximity data)とは, 例えば, 英語圏の国際観光都市(対象)に対する観光客が持つイメージを7点満点で評定してもらおう場合を想定する. 「ロンドン-エジンバラ」という2都市の対象どうしのイメージ相違度を7点満点で評定してもらい非類似性データ  $\delta_{12}$  を得る. おそらく, 「ロンドン-ラスベガス」という2都市間のイメージ相違度という非類似性  $\delta_{13}$  のほうが  $\delta_{12}$  よりも大きくなるであろうことは容易に想像できよう. MDSは, 直感的には非類似性に基づいて対象群を多次元空間に位置づける手法で, 非類似性の大きい対象どうしは遠くに, 小さい対象どうしは近くに位置づけるデータ解析メソッドとも言える.



さて，対象  $i, j, k, l$  間の非類似性データ  $\delta_{ij}, \delta_{kl}$  が得られたとする．それらを視覚的に認識できる距離量  $d_{ij}, d_{kl}$  に，できるだけデータの大小関係(順序関係)のみを保存されるように単調変換する「対応」が，非計量的多次元尺度法である(注)．ここで，距離とは三角不等式

$$d_{ik} \leq d_{ij} + d_{jk} \quad \text{for } \forall i, j, k, l \in P \quad (1)$$

を満足する非負の量のことである．諸対象間の違い(非類似性)を，対象点の位置の相違として認識するためには，非類似性データ量が距離に変換されている必要がある．なお，本稿ではユークリッド空間における距離に限定する．ここで，集合  $P$  は解析対象の全体集合を示す．そして

$$\delta_{ij} < \delta_{kl} \Rightarrow d_{ij} \leq d_{kl} \quad \text{for } \forall i, j, k, l \in P \quad (2)$$

となるように，非類似性データ量を，その大小関係(順序関係)のみを出来るだけ保って距離に変換することが非計量的MDSというメソッドにとって，本質的な事柄である．各々の対象を点として表現する空間，すなわち最終解をMDSでは布置(Configuration)と呼ぶ．布置の次元を  $T$  とすると，対象  $i$  と対象  $j$  の次元  $t$  における座標がそれぞれ  $x_{it}, x_{jt}$  のように得られたとすると多次元尺度法とは

$$\delta_{ij} \stackrel{m}{=} d^*_{ij} \approx d_{ij} = \sqrt{\sum_{k=1}^T (x_{it} - x_{jt})^2}$$

のように定式化できる．ここで，記号「 $\approx$ 」は可能な限り値が近いこと，記号「 $\stackrel{m}{=}$ 」は広義の単調増加関係を意味する． $d^*_{ij}$  は条件式(2)を完全に満足する擬似距離量であるが，必ずしも所与の次元  $T$  で実在する距離量ではなく，得られた布置の距離  $d_{ij}$  から「クルスカルの単調回帰」というアルゴリズムを使用して計算される疑似距離量が  $d^*_{ij}$  であって，デイスパリティとも呼ばれる[3]．すなわち，

$$\delta_{ij} < \delta_{kl} \Rightarrow d^*_{ij} \leq d^*_{kl} \quad \text{for } \forall i, j, k, l \in P \quad (3)$$

という関係が必ず成立するものが擬似距離量，デイスパリティである．よって，デイスパリティ  $d^*_{ij}$  は非類似性データを  $\delta_{ij}$  とすると，広義の単調増加関数を  $f$  として， $d^*_{ij} = f(\delta_{ij})$  と明示的に書ける．

本稿では，SPSSのBaseで実行可能な「通常のユークリッド距離モデル」に限定して説明する．すなわち，対象  $i$  と対象  $j$  を表す列ベクトルの位置ベクトルをそれぞれ

$\mathbf{x}_i, \mathbf{x}_j$ とすると

$$d_{ij}^{(s)2} = (\mathbf{x}_i - \mathbf{x}_j)^T W_S (\mathbf{x}_i - \mathbf{x}_j) \quad (4)$$

のように、被験者Sにおける対象iと対象jとの距離 $d_{ij}^{(s)}$ が定義される[1]. 通常のユークリッドモデルは式(4)で、 $W_S = I$  (Iは単位行列)の場合に相当する. 重みつきモデル (Weighted Euclidian Model) は、Wは対角行列となり、その対角成分には被験者sが軸kを重要視する程度、重み $w_k^{(s)}$ が並ぶ.

ところで、ある次元T (通常、 $T=2\sim4$ ) で完全に式(2)を満足するような布置を得ることは一般に困難である. そこで、与えられた非類似性データ $\delta_{ij}$ の順序関係と、求められた布置から計算される $d_{ij}$ との順序関係がどの程度、一致しているかを評価する適合度基準に、ストレス(正確には、クルスカルのストレス1式)と呼ばれる、モデル適合度の評価規準Sを次のように導入する[3].

$$S = \sqrt{\sum_{i < j} \sum_{i < j} (d_{ij} - d_{ij}^*)^2} / \sqrt{\sum_{i < j} \sum_{i < j} d_{ij}^2}$$

一般的には、最小二乗規準を想定したストレス1式が、布置の良否、モデル適合度の評価規準に利用される. 現在の解、布置で順序関係が完全に保存されていれば、すなわち式(2)が成立すれば、 $S = 0$ となる. 大小関係の維持が悪くなるにつれて、つまり、適合度が不良になるにつれてSの値は大きくなるが、ストレス1式の値Sは最悪でも0.2未満であることが要求される. そうでない場合は、布置の次元Tを1次元上げて、改めて布置を求めなおす.

つぎに、クルスカルによるプログラム、MDSCAL (MultDimensional SCALing)に基づいてMDSの解を求めるアルゴリズムを簡単に説明する. これは、つぎのステップから構成される.

- ①非類似性データ群  $\{\delta_{ij}\}$  が与えられているとする
- ②布置 (configuration) の次元を設定する
- ③初期布置Xを定める
- ④布置Xを基準化した布置をZとする
- ⑤現在の布置から距離量  $\{d_{ij}\}$  を計算する
- ⑥非類似性 $\delta_{ij}$ と単調関係にあるディスパリティ $d_{ij}^*$ を $d_{ij}$ から計算する
- ⑦適合度指標、ストレス値Sを $d_{ij}$ と $d_{ij}^*$ とから計算する
- ⑧Sが大きければ、最急降下法などの逐次近似法を用いて、適合度指標Sの値が小さくなる方向に布置Zを微小変化させて、 $Z_{NEW} \leftarrow Z_{OLD} + \Delta Z$ のように更新して、ストレスSを改善して、ステップ④に戻る
- ⑨Sが一定以下に小さくなれば、最終布置 $Z_{final}$ とする

SPSSで使用されているALSCAL(Alternating Least squares SCALing)やPROXSCAL(PROXimity SCALing)のアルゴリズムの基本的枠組みは、MDSCALを変化させたものである[4]. 特に、本稿で利用するPROXSCALは⑦、⑧を変更したものと見做せば理解しやすいが、そのアルゴリズムについては後述、説明する[5].

MDSによって分析するデータ形式について説明する[6]. 図1で被験者が一人の場合が、Two-way, つまりデータ行列が1個の場合である. このケースでは通常の「重みなしユークリッドモデル」で分析すればよい. それに対して、被験者が複数の場合、例えば、年度ごとに得られるプロ野球成績データや、経時的に行われる患者群の健康診断データなどはThree-way dataとなる. 対応する解析法には、被験者間の個人差や測定時点差など、いわゆる、INDSCALと呼ばれる個人差を考慮する「重みつきユークリッドモデル」が代表的なモデルである[7].

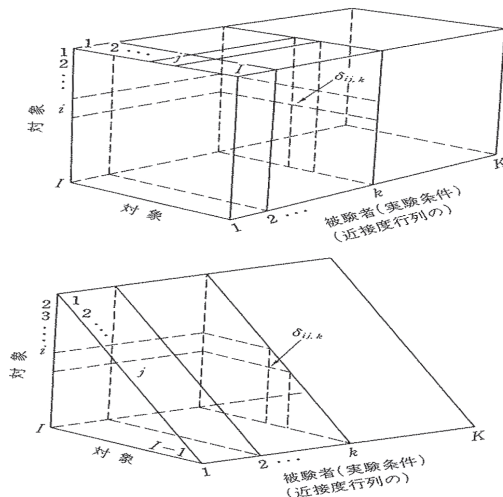


図1. 3元データの図解 文献[6]

## 2. 通常のユークリッドモデルでの分析例

行が属性で列が対象のTwo-Way data を通常のユークリッドモデルにより分析する具体例によって、MDS解析プロセスの概略を説明する. この調査では、被験者、具体的には筆者の4年ゼミに所属するアルコール類をよく飲む学生7人が、各種アルコール飲料に対してもつイメージを、居酒屋とホテルバーという異なる2か所で実際に飲酒してもらってから、10属性に関してそれぞれのアルコール飲料について5点満点で評価してもらった. 調査用紙への記入による属性型アンケートデータがこの分析例の入力データとなる. それぞれの成分は評定合計値を解答者数で除したものである. よって、小数点がデータにでるのである. これは、5種類のアルコール飲料を全て飲むことが出来ない

学生も当然のことながら居たから、解答者数は対象アルコールごとに異なるからである。

表1. バーと居酒屋でのアルコールイメージデータ (入力データ)

銘柄	ドライ_Bar	赤ワイン_B	日本酒_B	ブランデー_B	ウイスキー_B	ドライビール	赤ワイン	日本酒	ブランデー	ウイスキー
スマートな	3.00	3.80	3.00	4.00	4.00	2.14	4.00	2.75	3.00	3.33
男性的な	3.80	2.40	4.00	5.00	3.50	4.43	2.33	4.50	4.00	4.33
都会的な	3.00	4.60	2.00	4.00	4.00	2.14	5.00	2.75	4.00	4.33
飲みやすい	4.40	4.40	3.00	3.00	4.00	3.86	3.00	2.50	2.00	2.00
若者向きな	3.20	3.00	2.00	3.00	3.00	3.29	3.33	2.50	2.00	1.67
高級な	2.60	4.80	4.00	5.00	4.00	1.29	4.67	2.00	5.00	4.33
ムードがある	3.20	4.80	4.00	4.00	4.50	2.14	5.00	1.75	4.00	3.67
さわやかな	4.00	3.40	4.00	3.00	3.00	3.43	3.00	2.00	2.00	2.00
家庭的な	3.40	2.60	3.00	3.00	2.50	4.29	1.67	3.75	2.00	3.33
きつい	1.40	2.80	3.00	5.00	3.00	1.71	3.00	4.25	5.00	5.00

アルコール飲料間の非類似性  $\delta$  はつぎのように計算する. 例えば, Barでドライビールを嗜んだ場合と, バーで赤ワインを飲んだ場合の両アルコール飲料間の非類似性は, 添え字を'1=ドライ\_Bar', '2=赤ワイン\_B' とすれば,  $\delta_{12}$  と表現できて, 表1より

$$\begin{aligned} \delta_{12} &= \sqrt{(3.00-3.80)^2 + (3.80-2.40)^2 + \dots + (3.40-2.60)^2 + (1.40-2.80)^2} \\ &= 3.945 \end{aligned}$$

のように, ユークリッド距離的に非類似性データ  $\delta_{ij}$  は求められる. 同様に10個の全対象間について,  $(10 \cdot 9)/2 = 45$ 個の非類似性データ群  $\{\delta_{ij}\}$  を計算して, 非類似性行列  $\Delta$  を表2のように作成する.

一般的に, 対象  $i, j$  の  $p$  個の属性に関する属性データがそれぞれ  $\mathbf{i} = (x_{i1}, x_{i2}, \dots, x_{ip})$ ,  $\mathbf{j} = (x_{j1}, x_{j2}, \dots, x_{jp})$  であるときに, 対象  $i$  と対象  $j$  との非類似性データ  $\delta_{ij}$  を計算するには

$$\delta_{ij} = \left[ \sum_{T=1}^p |x_{iT} - x_{jT}|^r \right]^{1/r}$$

あるいは, 属性ごとにデータを標準化した場合は

$$\delta_{ij} = \left[ \sum_{T=1}^p |z_{iT} - z_{jT}|^r \right]^{1/r}$$

という式を利用すればよい. この分析例は  $r = 2$  を採用してユークリッド距離的な非類似性データを使用した.  $r = 1$  のときは「市街地距離」という. どのような  $r$  の値を採用するかは, 刺激に対して被験者がどのように反応するかによる.

表 2. 非類似性データ行列

近接行列										
	ドライ_Bar	赤ワイン_Bar	日本酒_Bar	ブランデー_Bar	ウイスキー_Bar	ドライビール	赤ワイン	日本酒	ブランデー	ウイスキー
ドライ_Bar	.									
赤ワイン_Bar	3.945	.								
日本酒_Bar	3.124	3.842	.							
ブランデー_Bar	5.095	3.868	3.606	.						
ウイスキー_Bar	3.211	1.646	2.958	2.958	.					
ドライビール	2.494	6.059	4.216	6.158	5.080	.				
赤ワイン	4.817	1.845	4.333	3.887	2.205	6.710	.			
日本酒	4.407	5.943	4.070	4.479	4.724	3.553	6.070	.		
ブランデー	5.793	4.281	3.67	2.45	3.57	6.691	3.712	4.479	.	
ウイスキー	5.520	4.517	3.89	2.33	3.88	6.136	4.282	3.698	1.667	.

変換プロット

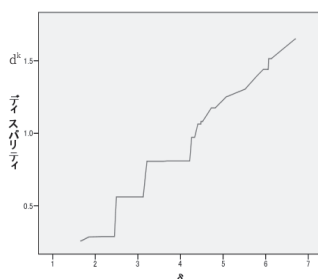


図 2.  $\delta$  と  $d^*$  の関係

この非類似性データ  $\delta_{ij}$  から対応する距離データ  $d_{ij}$  を計算し、続いて、距離データ  $d_{ij}$  と非類似性データ  $\delta_{ij}$  から「クルスカルの単調回帰」というアルゴリズムを利用して擬似距離量、ディスペリティ  $d^*_{ij}$  を計算する。 $\delta_{ij}$  と  $d^*_{ij}$  との関係を散布図で示すと図 2 のようになっている、この散布図から  $\delta$  と  $d^*$  との関係は、広義の単調増加関係が成立していること、つまり、 $\delta$  と  $d^*$  との間では、データ間の大小関係が完全に保存されていることが了解できる。

表 3. 距離行列

距離										
	ドライ_Bar	赤ワイン_Bar	日本酒_Bar	ブランデー_Bar	ウイスキー_Bar	ドライビール	赤ワイン	日本酒	ブランデー	ウイスキー
ドライ_Bar	.000									
赤ワイン_Bar	.887	.000								
日本酒_Bar	.522	.771	.000							
ブランデー_Bar	1.224	.811	.748	.000						
ウイスキー_Bar	.808	.266	.552	.590	.000					
ドライビール	.578	1.420	.786	1.518	1.277	.000				
赤ワイン	1.143	.284	.926	.704	.376	1.645	.000			
日本酒	1.077	1.454	.717	1.045	1.202	.893	1.542	.000		
ブランデー	1.392	.987	.0.9	0.18	.773	1.649	.853	1.085	.000	
ウイスキー	1.303	1.072	.0.78	0.3	.825	1.481	.998	.846	.249	.000

ところが、非類似性データと距離との間では、必ずしも大小関係が保存されていないことが表3から発見できる。例えば、第3列の対象（具体的には、日本酒\_bar）と第9列の対象9（9：ブランデー）、第10列の対象10（10：ウイスキー）の非類似性データの大小関係は、表2より  $\delta_{39}$  (3.87) <  $\delta_{3,10}$  (3.89) であるが、対応する距離量は  $d_{39}$  (0.9) >  $d_{3,10}$  (0.78) のように変換されている。つまり、必ずしも大小関係を保持した変換が、限定された2次元布置では実現できていないことがわかる。ここで、カッコ内の数値は非類似性データ量、距離をそれぞれ示す。

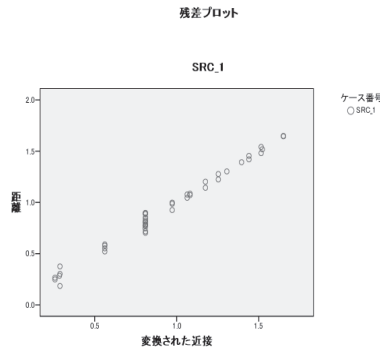


図3.  $d_{ij}^*$  の  $d_{ij}$  の関係

そこで、 $\delta_{ij}$ と単調増加関係にある  $d_{ij}^*$  と、実際に得られた布置における距離  $d_{ij}$  との剥離の二乗  $(d_{ij} - d_{ij}^*)^2$  の和を、距離の総和  $\sum_{i < j} d_{ij}^2$  で除した「ストレス1式」によって非類似性データ量の順序関係がどの程度、保存されて距離量に単調変換されているかを評価することになる。「ストレス1式」による解の良し悪し程度の評価判断には表4が参考になる[3]。適合度指標には、「ストレス1式」が代表的であるが、その他にはALSCALの最適化の目的関数として使用される「S-ストレス」や[4]、PROXSCALで利用している「正規化した原ストレス」「rawストレス」がある。このデータ分析における各種ストレス値は以下のとおりである。当該データ解析ではPROXSCALをアルゴリズムとして利用して、そのストレス1式の値は2次元解で0.0429となり0.05未満であるので、得られた2次元布置は良好な (Good) の最終布置といえる。

表4. ストレス1式による適合度の評価規準 文献[3]

Stress (%)	Goodness-of-fit
20	Poor
10	Fair
5	Good
2.5	Excellent
0	Perfect

表5. 2元アルコールデータ解析の適合度

ストレスと適合の測定	
正規化された原ストレス	.00184
ストレス-I	.04289 <sup>f</sup>
ストレス-II	.11073 <sup>e</sup>
S-ストレス	.00287 <sup>b</sup>
説明された散らばり (D.A.F.)	.99816
Tuckerの適合係数	.99908

PROXSICAL は正規化された原ストレスを最小化します。

a. 最適尺度因子 = 1.002。

b. 最適尺度因子 = 1.000。

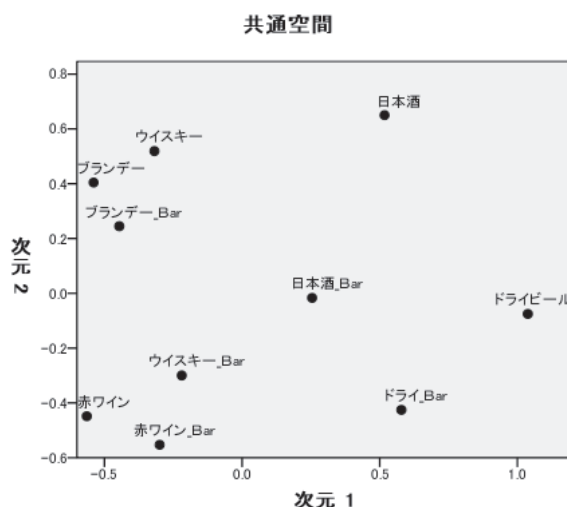


図4. アルコール飲料の製品マップ (添え字 Barはバーで嗜んだ場合を示す)

そして、アルコール飲料認知空間を確定することと、アルコール飲料を嗜む場所の違いで消費者のイメージ認知がどの程度、変動するかを把握するという2個の分析目的のために、認知マップ(製品マップ)を作成した。認知マップとは、被験者の心の中での対象やブランドの位置づけを可視化したものである。被験者が消費者で対象が製品の場合は、認知マップは製品マップと呼ぶことになる。

この分析例では、アルコール飲料の製品空間は2次元空間で捉えられて、製品マップの横軸は「飲みやすい-アルコール度が強い」、縦軸は「中年男性的な-スマートな」と解釈でき、ゼミ学生達は各種アルコール飲料について、「飲みやすさ」と「スマートさ」の2軸による認知空間で飲酒行動をしている様子が推察できた(図4)。つぎに、アルコール飲料を楽しむ場所の違いの影響を調べる。バーで嗜んだ場合のドライビール、日本酒のマップでの対象点が、居酒屋での対象点よりも左下にかなり移動していること



から、バーでビールなどの大衆酒を飲むと「スマートで多少アルコール度を強く感じて、程良く酔えるのであろう。他方、ウイスキーや赤ワインなどの高級酒では、バーでの対象点が居酒屋のそれよりも右下に移動しているの、バーでウイスキーなどを飲んだ際、スマートで気持ちよく、アルコール度数があまり気にならないので、お酒がすすんで飲み過ぎるかもしれない。また、バーで楽しむ場合、赤ワインとウイスキーが製品マップにおいて、それらの位置が隣接していることから、それらが競合関係、代替財になっていることは興味深い。ベテランバーテンダーの話によると、実際にワインを楽しんでから、ウイスキーを飲む常連客が2割程度、バーでは居るようである。

### 3. MDSアルゴリズムとPROXSCAL

多次元尺度法は統計的データ解析ツールのなかで、情報工学的色彩が強いメソッドとも見做せて、PCの性能が現在ほど優れていない時代には最終解を得るのにかなりの時間を要した。そこで、多次元尺度法のアルゴリズム的側面について説明する。

多次元尺度法のアルゴリズムには、MDSCALの他にALSCAL (Alternating Least squares SCALing) やPROXSCAL (PROXimity SCALing) などがあるが、基本的な枠組みはMDSCALのところで書いたように以下のとおりで、⑦、⑧がそれぞれ異なると見做せばよい。

- ①非類似性データ  $\{\delta_{ij}\}$  が与えられているとする
- ②布置 (configuration) の次元を設定する
- ③初期布置  $X$  を定める
- ④布置  $X$  を基準化して布置  $Z$  を得る
- ⑤現在の布置から距離  $\{d_{ij}\}$  を計算する
- ⑥非類似性データ  $\delta_{ij}$  と単調関係にあるディスパリティ  $d_{ij}^*$  を  $d_{ij}$  から計算する
- ⑦適合度指標の各種ストレス値  $S$  を  $d_{ij}$  と  $d_{ij}^*$  とから計算する
- ⑧適合度指標の  $S$  が大きければ、最急降下法などの逐次近似法で  $S$  が小さくなるように、現在の布置  $Z$  を微小変化させて、 $Z' \leftarrow Z + \Delta Z$  のように更新して、ステップ④に戻る
- ⑨  $S$  が小さければ最終布置  $Z_{\text{final}}$  を得る

ステップ③について補足説明をする。計量的MDSにより得られる布置を初期布置に採用して、出来るだけ最終布置  $X_{\text{Final}}$  に近いものを利用することが、局所最小問題を防ぐための合理的な方法である。

ここで、局所最小について説明する。いま、複数の極小点をもつ関数  $H$  について考

えてみる。この場合、複数の極小点のうち、関数Hの最小値を与える最小点をGlobal Minimaといい、そうでない極小点をLOCAL Minima (局所最小)という。局所最小点に不幸にも関数Hが収束してしまうことを「局所最小」問題といい、最急降下法でも、あるいはMajorization Algorithmを使用しても防ぐことが出来ない厄介な問題である。この問題には、複数の初期値群から逐次近似法をスタートさせるしか対策法はない。

さて、計量的MDSのプロセスは以下のとおりである。非類似性データ  $\delta$  に適切な定数を加算して距離量  $d$  に変換する。布置座標行列 (布置行列) を求めるために対象  $i$  と対象  $j$  の距離  $d_{ij}$  を、ダブルセンタリング法で原点からの位置ベクトル  $i, j$  による内積  $b_{ij}$  に変換する。対象  $i$  の次元  $t$  の座標を  $x_{it}$  とすると

$$b_{ij} = (i, j) = \sum_{t=1}^p x_{it} x_{jt}$$

と書いて、 $x_{it}$  を成分とする行列  $X$  は布置座標行列に他ならない。実対称行列の内積行列  $B = [b_{ij}]$  は固有値分解が可能である。すなわち

$$B = T D T^T \text{ (固有値分解)} \Leftrightarrow T^T B T = D \text{ (実対称行列の対角化問題)}$$

が成立して、また、 $B = X X^T$  が成立するので

$$B = X X^T = T D^{1/2} D^{1/2} T \text{ より } X = T D^{1/2}$$

とできるから、 $T D^{1/2}$  が初期布置行列の候補として得られる。ここで、 $D$  は対角成分に行列  $B$  の固有値が並ぶ対角行列、 $T$  は対応する固有ベクトルが並ぶ直交行列である。しかるに、 $p$  次元初期布置を求める場合は、行列  $X$  の最初の  $p$  個の列によって形成される行列を初期布置行列  $X_{\text{start}}$  にすればよい [1]。その他に、入力データについて主成分分析が可能である場合は、その解を初期布置に利用することもある。

つぎに、本稿のデータ解析で利用したPROXSCALについて説明する [5] [11]。このプログラムはSPSSのCategoryオプションに搭載されている。PROXSCALでは、上記のステップ⑦でストレス1式の代わりに、rawストレス、すなわち、

$$\sigma_r(X) = \sum_{i < j} w_{ij} (\delta_{ij} - d_{ij}(X))^2$$

を使用する。かつ、ステップ⑧でMDSALでは最急降下法を利用するが、PROXSCALではMajorization Algorithm を採用した反復式

$$X_{k+1} = V^+ B (X_k) X_k$$

を使用する。

PROXSCALは良好な布置を求めるために、最適化手法に Majorization Algorithm (M.A.) を利用して、局所最適解への収束が保証される [10] [12]。M.A.を利用するので、最急降下法を利用するプログラムよりも数値計算の観点からは局所最終解への収束が良好となるとも言われている [14]。

具体的には、PROXSCAL は rawストレス、 $\sigma_r(X)$  を最小にするように最終解を求

める[8].  $X$ を布置行列とすると $\sigma_r(X)$ は以下のように定義されて

$$\begin{aligned}\sigma_r(X) &= \sum_{i < j} w_{ij} (\delta_{ij} - d_{ij}(X))^2 \\ &= \eta_\delta^2 + \text{tr } X^T V X - 2 \text{tr } X^T B(X) X \\ &\leq \eta_\delta^2 + \text{tr } X^T V X - 2 \text{tr } X^T B(Z) Z \equiv \tau(X, Z)\end{aligned}$$

となる[5].  $\eta_\delta^2$ は定数項である. そして,  $\sigma_r(X)$ を最小にする代わりに $\tau(X, Z)$ を最小にするというMajorization algorithm がPROXSCALでは採用される. ここで, 行列 $B(Z)$ は

$$\begin{aligned}B &= [-w_{ij} \delta_{ij}/d_{ij}] \quad \text{あるいは} \quad [-w_{ij} d_{ij}^*/d_{ij}] \quad (i \neq j \text{ のとき}) \\ B &= [-\sum_{j \neq i} b_{ij}] \quad (i = j \text{ のとき})\end{aligned}$$

という対称行列で,  $V$ は重み行列である.

$\tau(X, Z)$ を行列 $X$ で微分すると

$$\begin{aligned}\nabla \tau(X, Z) &= \nabla (\text{tr } X^T V X) - \nabla (2 \text{tr } X^T B(Z) Z) \\ &= 2V X - 2B(Z) Z\end{aligned}$$

となる. ここで,  $\nabla \tau(X, Z) = 0$ とおくと

$$V X = B(Z) Z$$

となる. 行列 $V$ は一般にランク落ちをするので, 行列 $V$ には通常の逆行列は存在しない. そこで,  $V$ のムーアペンローズ (一般) 逆行列を $V^+$ とすると $V^+$ は一意に定まって (注), 上式から行列 $X$ の推定値 $\hat{X}$ は

$$\hat{X} = V^+ B(Z) Z$$

のように, 布置 $X$ の推定行列は求まる. ここで, ムーアペンローズ逆行列 $V^+$ は具体的には

$$V^+ = (V + 11^T)^{-1} + N^{-2} 11^T$$

という形式になる[11]. よって, 反復式は

$$X_{k+1} = V^+ B(X_k) X_k \quad (k=1, 2, 3, \dots)$$

となって, この反復式を利用して布置 $X$ を更新する.

つぎに, INDSCALを分析モデルとして使用した場合に, 同一データに対するPROXSCALによる解とALSCALの解のモデル適合度の相違状況をストレス1式で評価して比較検討する.

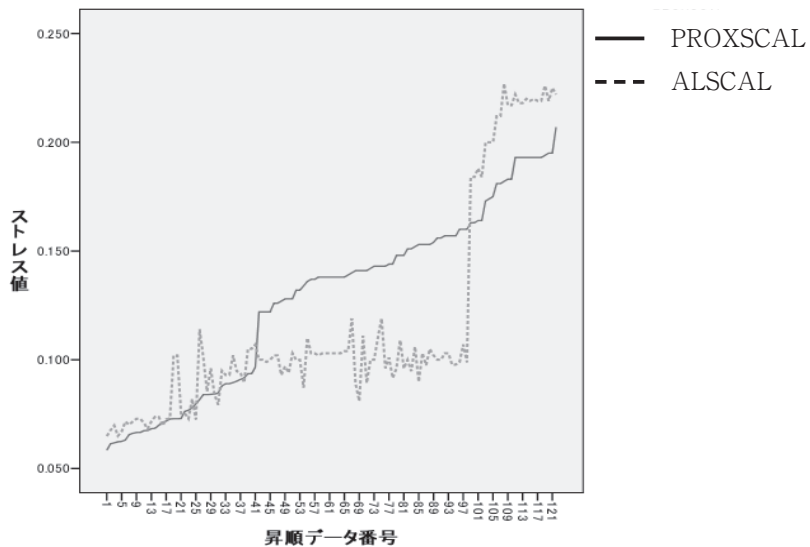


図5. ALSICALとPROXSCALのストレス1式値

粗死亡率データと12年間プロ野球成績データから乱数を発生させて $24 \times 3 + 50 = 122$ 個のサブデータを作成した。それぞれのサブデータセットに対してINDSCALを分析モデルとしたデータ解析をALSICALおよびPROXSCALによって実行した。PROXSCAL解のストレス1式の値とALSICAL解の当該のストレス値を記録した。対の2列のデータ群をPROXSCAL解のストレス値を基軸にして昇順に並べ替えて作成される $122 \times 2$ 型のストレス1式値データ行列 $Z_s$ を折れ線グラフで図解したものが図5である。

図5から読み取れるように、PROXSCALは当該ストレス1式の値が0.16以上の、モデルに対して適合度が不良なサブデータに対してはALSICALよりも適合度の良い解を与えるようである。かつ、PROXSCALのストレス値が0.1未満の適合度良好なデータの場合でもALSICALよりも、やや適合度が良い最終解を与える「データ依存性」の特性が、PROXSCALでは見受けられた [13]。一方、ALSICALは中程度の適合度データに対して、良好な布置を与えることが認められるが、ストレス値が0.15以上の適合度不良のデータに対して良好な布置を得ることは弱いようだ[15]。

このように使用するアルゴリズムの相違によって、同一データに対して、かなり適合度が異なる布置(最終解)をもたらされる事実はMDSが情報工学的色彩の強いツールであることを示す良い証でもある。

#### 4. MDSの数値計算法—最急降下法とM.A.

非計量的多次元尺度法は反復計算で最終解を求めていくので計算機、CPUの性能が現在ほど優れていない時代では、最終解を求めるために相当の実行時間を費やした。製

品マップを作成するために利用する因子分析に比べて、計算時間がMDSでは、異常に長かったことが、その普及を日本で遅らせた一因であったのかもしれない。逆に言うと、MDSはコンピュータの性能にかなり依存する「Computer intensiveなメソッド」であることが、大きなメソッドの特徴でもあり、MDSを統計学の範疇で扱うよりも情報工学分野で扱う傾向が見受けられる時代もあった。そこで、MDSの数値計算で古典的に利用した最急降下法について説明する。つぎに、Majorization Algorithmを説明するが、この数値解析法は最急降下法に比べても複雑なプロセスから成り立っているので計算時間をさらに費やす。この算法を利用したPROXSCALはPCの性能が向上した1990年代後半までは、その実用化は目立たなかった。このような数値計算法を観ることで、MDSがコンピュータサイエンスの性格が強いメソッドであることが実感されたら幸いである。主成分分析や因子分析などで必ず登場する「固有値」という言葉すら、非計量的MDSでは登場しないことから、他の多変量データ解析手法とは色彩が異なるメソッドと感ずる。

#### 4-1. 最急降下法

多次元尺度法の解（布置）を逐次改良していく反復計算には、最急降下法が伝統的に採用されてきた [1]。いま、最適化すべきストレス式を  $n$  変数関数  $S(\mathbf{x})$  とおくとランダウの記号を用いると

$$S(\mathbf{x} + \Delta \mathbf{x}) - S(\mathbf{x}) = \sum_j \Delta x_j \frac{\partial S}{\partial x_j} + o(|\Delta \mathbf{x}|) \quad (\Delta \mathbf{x} \rightarrow 0)$$

となる。ここで、 $\mathbf{x} = (x_1, x_2, \dots, x_n)$  である。この式は難しそうに視えるが、対応する1変数関数  $S(x_1)$  の場合について示してみると、関数  $S(x_1)$  の点  $x_1$  での微分係数の定義式を拡張したものに他ならないことがわかる。つまり

$$\begin{aligned} S(x_1 + \Delta x_1) - S(x_1) &= S'(x_1) \Delta x_1 + o(|\Delta x_1|) && (\Delta x_1 \rightarrow 0) \\ &= \Delta x_1 \frac{ds}{dx_1} + o(|\Delta x_1|) && (\Delta x_1 \rightarrow 0) \end{aligned}$$

であるので、 $n$  変数関数の場合は1変数の場合の微分係数の定義式を拡張しただけである。ここで勾配、 $-\text{gradient} = (\dots, -\frac{\partial S}{\partial x_j}, \dots)$  が点  $\mathbf{x}$  における「局所的な」最急降下方向 (gradient, grad) である。最急降下法は現在の布置を表すベクトル  $\mathbf{x}$  から  $-\text{grad} \cdot \Delta t$  ( $\Delta t$  はステップサイズ) だけ、 $\mathbf{x} - \text{grad} \cdot \Delta t$  へ点をずらして解  $\mathbf{x}$  を逐次改良していく方法である。

具体的には、対象が10個で2次元布置の場合、10行2列の布置行列  $X = [x_{ij}]$  は、行列  $X$  の成分数は20になるから、 $x_{11} = y_1, x_{21} = y_2, \dots, x_{10,1} = y_{10}, x_{21} = y_{11}, \dots, x_{10,2} = y_{20}$  とおくと、この場合のストレス  $S$  の式は、 $S = S(\mathbf{y}) = S(y_1, y_2, \dots, y_{20})$  とする20変数関数になる。20変数関数の非線形最適化問題を扱う訳であるから、CPUの性能が現在ほど優れていない時代には、MDSの実行時間は相当長い時間を費やした

のである。

最急降下法について、つぎの簡単な関数の2変数関数Fで説明する。

$$F(x_1, x_2) = x_1^2 + 4x_2^2 \text{ とする.}$$

点  $(x_1, x_2)$  における最も勾配が急な方向は、ベクトル  $\text{gradient} = \left( \frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2} \right) = (2x_1, 8x_2)$  によって与えられて、ステップサイズを  $\Delta t = 0.2$  とする。

よって、反復式は  $\mathbf{x}_{(k+1)} = \mathbf{x}_{(k)} - \text{grad} \cdot \Delta t$  であるから、初期値  $\mathbf{x}_{(0)} = (2.0, 2.0)$  とすると

$$\mathbf{x}_{(1)} = \mathbf{x}_{(0)} - \text{grad} \cdot \Delta t = (2.0, 2.0) - 0.2(4.0, 16.0) = (2.0 - 0.8, 2.0 - 3.2) = (1.2, -1.2)$$

すなわち、 $F(\mathbf{x}_{(1)}) = 7.2$  となる。

$$\begin{aligned} \mathbf{x}_{(2)} &= \mathbf{x}_{(1)} - \text{grad} \cdot \Delta t = (1.2, -1.2) - 0.2(2.4, -9.6) = (1.2 - 0.48, -1.2 + 1.92) = \\ &= (0.72, 0.72) \quad \text{よって } F(\mathbf{x}_{(2)}) = 2.59 \text{ と求まる.} \end{aligned}$$

$$\mathbf{x}_{(3)} = \mathbf{x}_{(2)} - \text{grad} \cdot \Delta t = (0.72, 0.72) - 0.2(1.44, 5.76) = (0.432, -0.432) \quad F(\mathbf{x}_{(3)}) = 0.932$$

表6. 最急降下法の収束状態

n	X <sub>1</sub>	X <sub>2</sub>	F(X)
0	2.000	2.000	20.000
1	1.200	-1.200	7.200
2	0.720	0.720	2.592
3	0.432	-0.432	0.933
4	0.259	0.259	0.336
5	0.156	-0.156	0.121
6	0.093	0.093	0.044
7	0.056	-0.056	0.016
8	0.034	0.034	0.006
9	0.020	-0.020	0.002
10	0.012	0.012	0.001

逐次、関数  $F(\mathbf{x})$  の値が最小値0に近づいていく様子が表6から理解できる。なお、最急降下法が局所最小点に到達する保証はない。

#### 4-2. Majorization Algorithm

最急降下法は局所最小点への収束も保証されていない。そこで、局所最小点への収束が優れている majorization algorithm のMDSへの応用が考案されPROXSCALの開発につながった。そこで、PROXSCAL が利用する Majorization Algorithm (MA) について、下記の整関数  $F(x)$  の場合で説明する[11][12][14]。

閉区間  $[-1.5, 2.0]$  を定義域とする4次関数

$$F(x) = 6 + 3x + 10x^2 - 2x^4 \tag{4-1}$$

の最小値を与える点  $x$  を求めたいとする。このとき、関数

$$G(x, y) = 6 + 3x + 10x^2 - 8xy^3 + 6y^4 \tag{4-2}$$

が  $F(x)$  の majorizing function で、 $(y$  を定数と見做せば) 解析が容易な関数  $G(x, y)$  を  $F(x)$  の代わりに考察することが MA の発想である。Majorizing function  $G(\cdot)$  は  $G(\cdot) \geq F(\cdot)$  という条件を満たすことが必要である。そこで、

$$G(x, y) - F(x) = 2x^4 - 8xy^3 + 6y^4 \equiv H(x, y)$$

においてこの条件を確認する。ここで、しばらく  $y$  を定数と考えて

$$\begin{aligned} \partial H(x, y) / \partial x &= 8x^3 - 8y^3 = 8(x - y)(x^2 + xy + y^2) \\ &= 8(x - y)[(x + 1/2y)^2 + 3/4y^2] \quad \dots \textcircled{1} \\ x^2 + xy + y^2 &> 0 \quad \text{for } \forall (x, y) \neq (0, 0) \end{aligned}$$

が成立する。また、 $\partial^2 H / \partial x^2 = 24x^2 > 0 \quad \text{for } \forall x \neq 0 \quad \dots \textcircled{2}$

ゆえに、 $\textcircled{1}$ 、 $\textcircled{2}$ より  $\partial H(x, y) / \partial x = 0$  より関数  $H$  は  $x = y = a$  (定数) で最小になり、かつ  $x = y$  のとき最小値  $H(x, x) = 0$  となるので

$$H(x, y) \geq 0 \Leftrightarrow G(x, y) \geq F(x) \quad (\text{等号は } x = y \text{ のとき})$$

すなわち、

$$6 + 3x + 10x^2 - 8xy^3 + 6y^4 \geq 6 + 3x + 10x^2 - 2x^4$$

が成立する。等号は  $x = y$  のときに成立する。

そこで、関数  $G$  で  $y = a$  (定数) において、4次関数  $F(x)$  の定義域における最小点を探す代わりに、解析が容易な2次関数  $G(x, a)$  の最小化問題を扱うことが M.A. の本質的な事柄である。

$$G(x, a) = 6 + 3x + 10x^2 - 8xa^3 + 6a^4 \quad (4-3)$$

$$\partial G / \partial x = 3 + 20x - 8a^3$$

$\partial G / \partial x = 0$  において

$$x = 1/20 (8a^3 - 3) \quad (4-4)$$

を得る。 $x$  を  $X(n+1)$ 、 $a$  を  $X(n)$  と書きかえると式(4-4)は

$$X(n+1) = 2/5 \{X(n)\}^3 - 3/20 \quad (n=1, 2, 3, \dots)$$

という漸化式になる。 $X(1) = 1.4$  と初期値を与えると表9

表9. MAの収束状況

n	x(n)	x(n+1)
1	1.400000	0.947600
2	0.947600	0.190357
3	0.190357	-0.147241
4	-0.147241	-0.151277
5	-0.151277	-0.151385
6	-0.151385	-0.151388
7	-0.151388	-0.151388
8	-0.151388	-0.151388

のように点  $-0.151$  の近傍で関数  $G$ ，すなわち，関数  $F$  は区間  $[-1.5, 2.0]$  において最小になる．この関係は下記の図からも理解できると窺え[12]，本稿をおわることにする．

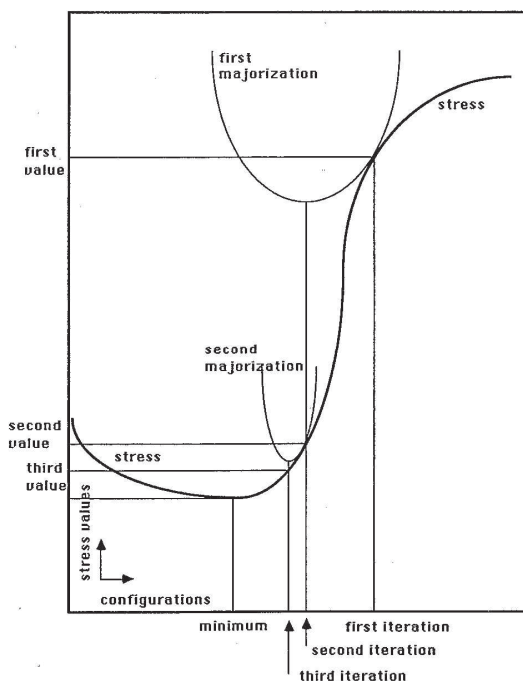


図 6. M.A.の図解 文献[12]

注

### 1. 距離

平面上の距離（距離量）を拡張して，一般的な距離を定義する．任意の点  $a, b, c$  について

$$d(a, b) + d(b, c) \geq d(a, c)$$

という性質をもつ，非負で対称性のものを距離という．例えば  $p$  次元空間における都市ブロック距離  $C(i, j)$  は

$$\begin{aligned} C(i, j) &= \sum_{T=1}^p |x_{iT} - x_{jT}| \\ &= \sum_{T=1}^p |x_{iT} - x_{kT} + x_{kT} - x_{jT}| \\ &\leq \sum_{T=1}^p \{|x_{iT} - x_{kT}| + |x_{kT} - x_{jT}|\} = \sum_{T=1}^p |x_{iT} - x_{kT}| + \sum_{T=1}^p |x_{kT} - x_{jT}| \\ &= C(i, k) + C(k, j) \end{aligned}$$



という性質が成立するので、都市ブロック距離  $C(i, j)$  は距離である。

## 2. 計量的多次元尺度法と非計量的多次元尺度法

非類似性データ量  $\delta_{ij}$  と距離量  $d_{ij}$  との関係に線形関係を仮定する、すなわち

$$d_{ij} = a + \beta \delta_{ij} + \varepsilon_{ij} \quad (a, \beta \text{ は定数})$$

とするものを計量的MDSという。

それに対して、非類似性データ  $\delta_{ij}$  と擬似距離量  $d_{ij}$  との関係に、広義の単調増加関係関数  $f$  のみを仮定するもの、すなわち

$$f(\delta_{ij}) = d_{ij}^*$$

とするものを非計量的MDSという。クルスカルの単調回帰により  $d_{ij}^*$  を求めて、布置を最急降下法などの数値計算法を使用して逐次近似的に求める。

## 3. ムーアペンローズ一般逆行列

線形回帰方程式  $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}$ 、あるいは、必ずしも解を持たない方程式  $\mathbf{y} = \mathbf{A}\mathbf{x}$  において、 $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \rightarrow \min$  のように  $\mathbf{x}$  を推定することを考える。このとき、さらに、ノルム  $\|\mathbf{x}\|$  を最小にする  $\mathbf{A}$  の逆演算子  $\mathbf{A}^+$  をムーアペンローズ一般逆行列という。すなわち、形式的には  $\mathbf{x} = \mathbf{A}^+ \mathbf{y}$  と書けるが通常は  $\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{y}$  と記して、一意に  $\mathbf{A}^+$  は求まる。すなわち、通常の線形回帰モデルにおける最小二乗解はムーアペンローズ一般逆行列を  $\mathbf{G}$  として、 $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$  により得ている。ムーアペンローズ一般逆行列  $\mathbf{A}^+$  は、ノルム最小の最小二乗解を与える一般逆行列とも呼ばれて、通常的一般逆行列とは異なる。

## 参考文献

- [1] Everitt, B.S. and Rabe-Hesketh, S. The Analysis of Proximity Data. Kendall's Library of Statistics 4. Arnold. (1997)
- [2] Treat, A.T et al. Assessing Clinically Relevant Perceptual Organization with Multidimensional Scaling Techniques, *Psychological Assessment*, Vol.14, No 3, 239-252. (2002)
- [3] Kruskal, J.B. Multidimensional Scaling by Optimizing Goodness of fit to a Nonmetric Hypothesis, *Psychometrika*, 29, 1-27. (1964)
- [4] Takane, Y., Young, F.W. and De Leeuw, J. Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, Vol 47, pp.7-67 (1977)
- [5] Commandeur, J.J.F & Heiser, W.J. Mathematical Derivations in the Proximity Scaling (Proxscal) of Symmetric Data Matrices (Tech. Rep. No. RR 93-03) Leiden,

The Netherlands: Department of Data Theory. (1993)

- [ 6 ] Arabie, P., Carroll, J.D. and Desarbo, W.S. Three-way Scaling and Clustering. Newbery Park, CA: Sage (1987) (邦訳：岡太・今泉 (共訳) 3元データの分析 共立出版 (1990))
- [ 7 ] Carroll, J.D. and Chang, J.J. Analysis of Individual Differences in Multi-Dimensional Scaling via an N-way Generalization of Eckart-Young Decomposition. *Psychometrika*, Vol 35, pp.283-319 (1970)
- [ 8 ] 奥喜正・前鶴政和 INDSCALによる重み係数を利用した市場細分化, 日本経営数学会誌, Vol28, No2, pp.61-72. (2007)
- [ 9 ] Meulman, J.J., Heiser, W.J. & SPSS. SpssCategories 10.0. Chicago: SPSS. (1999)
- [10] Kier, H.A.L. Majorization as a tool for optimizing a class of matrix functions. *Psychometrika*, 55, 417-428. (1990)
- [11] Borg, I and Groenen, P. Modern Multidimensional Scaling: Theory and Applications 2<sup>nd</sup> ed. Springer. (2005)
- [12] De Leeuw, J. Convergence of the Majorization Method for Multidimensional Scaling. *Journal of Classification*, 5, 163-180. (1988)
- [13] 奥喜正 ストレス1式による多次元尺度法PROXSCALアルゴリズムの適合度特性, 日本計算機統計学会第24回大会論文集, pp49-53. (2010)
- [14] Bartholomew, D.J, Steele, F, Moustaki, I & Galbrath, J.I. Analysis of Multivariate Social Science Data, 2<sup>nd</sup> ed Chapman & Hall/CRC. (2008)
- [15] 奥喜正 多次元尺度法PROXSCALアルゴリズムによる解の適合度—ALSCALとの比較において, 日本経営数学会誌, Vol32, pp.1-15. (2010)