

カテゴリー評定データの乗法型非対称 MDS による ベイズ分析の試み

○岡本安晴

日本女子大学人間社会学部

1 はじめに

非対称 MDS の 1 つとして Okada and Imaizumi (1987) があるが、Okada et al. (1987) では非対称性が加減算で表されている。Okada らのモデルはブランドスイッチングデータなど社会学的データに適用されている。心理量としての距離を対象とする場合、岡本(2010)は精神物理学的方法を考慮して非対称性を乗法的に表すモデルを提案した。しかし岡本(2010)では、nonmetric モデルに拡張した場合、退化解が得られたと報告されている。これに対して岡本(2011)は、nonmetric モデルではなく、物理量と感覚量の関係を表す精神物理学の尺度の 1 つであるべき法則を用いて感覚量としての距離を表し、データとしての非類似度は距離の対数値を中心とする対数正規分布に従うとする確率モデルを提案した。Stevens (1975) は、prothetic 感覚は近似的に対数正規分布に従うとしている。しかし、心理学におけるデータ収集法としては、(非) 類似度のマグネチュード推定の他に、カテゴリー評定もよく用いられる。系列カテゴリー判断のモデル (Guilford, 1954) を距離の対数値に適用したモデルを試みる。

2 モデル

岡本 (2010) の乗法型非対称 MDS モデルを設定する。

$$(1) \quad d_{ij} = \left\{ \sum_{t=1}^p (x_{it} - x_{jt})^2 \right\}^{1/2} \times \frac{r_j}{r_i}$$

ここで、 $\{x_{it}\}$ は対象 i の座標値、 r_i は対象 i の重みを表す。

対象 i から対象 j への距離 (非類似度) の評定カテゴリーがカテゴリー h であることを $R_{ij,h}$ で表す。
このとき、確率モデルを

$$(2) \quad P(R_{ij,h} | d_{ij}) = \Phi\left(\frac{C_h - \log(d_{ij})}{\sigma}\right) - \Phi\left(\frac{C_{h-1} - \log(d_{ij})}{\sigma}\right)$$

と設定する。 C_h はカテゴリー境界で次式

$$-\infty = C_0 < C_1 < \dots < C_{K-1} < C_K = +\infty$$

を満たすものとする。

3 解法（アルゴリズム）

ベイズ的方法によって事後分布を求める。対象*i*から対象*j*への距離の評定カテゴリーがカテゴリー*h*であった頻度を $f_{ij,h}$ で表し、尤度関数を

$$L(x_{it}, r_i, C_h, \sigma) = \prod_{i \neq j} \prod_h P(R_{ij,h} | d_{ij})^{f_{ij,h}}$$

とおく。事前分布は、

$$p_0(x_{it}, r_i, C_h, \sigma) = \begin{cases} \exp\left\{-\lambda_1 \sum_{i,t} (x_{it} - x_{it,0})^2\right\} \times \exp\left\{-\lambda_2 \sum_i (r_i - r_{i,0})^2\right\} \\ 0 \end{cases} \quad \begin{matrix} r_i, \sigma > 0, C_1 < \dots < C_{K-1} & \text{のとき} \\ \text{上記以外} & \text{のとき} \end{matrix}$$

と設定する。モデル（1）における x_{it} 、 r_i の不定性の問題を回避するために、岡本(2011)と同様、初期値 $x_{it,0}$ および $r_{i,0}$ を中心とする分布を事前分布として設定している。初期値は、岡本(2010)の方法による解である。 σ と C_h の制約条件が満たされないときは、事前分布の値は0とする。さらに σ と C_h は、有界ではあるが十分に広い範囲において一様分布に従うものとする。

以上の設定の下で事後分布は

$$(3) \quad P(x_{it}, r_i, C_h, \sigma) \propto p_0(x_{it}, r_i, C_h, \sigma) \times L(x_{it}, r_i, C_h, \sigma)$$

により与えられる。式（3）で与えられる事後分布に対して Metropolis within Gibbs 法により事後分布の推定を行う。詳しくは、以下の手順である。

評定者*k*の対象*i*から対象*j*への距離の評定カテゴリー値を $\delta_{ij}^{(k)}$ とおく。 $\delta_{ij}^{(k)}$ の評定者にわたる平均値を $\bar{\delta}_{ij}$ とおく。すなわち、

$$\bar{\delta}_{ij} = \frac{1}{n} \sum_{k=1}^n \delta_{ij}^{(k)}$$

である。ここで、 n は評定者の総数である。 $\bar{\delta}_{ij}$ に対して岡本（2010）の方法による解を求め、得られた布置の座標値を $x_{it,0}$ 、重みを $r_{i,0}$ とおく。これらの値により事前分布 $p_0(x_{it}, r_i, C_h, \sigma)$ を設定して、MCMC 法を以下の Metropolis within Gibbs 法によって行う。

第*s*試行における各パラメータの値を $x_{it}^{(s)}$ 、 $r_i^{(s)}$ 、 $C_h^{(s)}$ 、 $\sigma^{(s)}$ とおいたとき、 $x_{it}^{(s)} \rightarrow r_i^{(s)} \rightarrow C_h^{(s)} \rightarrow \sigma^{(s)}$ の順に1つずつ $x_{it}^{(s+1)}$ 、 $r_i^{(s+1)}$ 、 $C_h^{(s+1)}$ 、 $\sigma^{(s+1)}$ に更新する。更新のときの提案分布は、第*s*試行における値を平均値とする正規分布を用いる。パラメータ $r_i^{(s)}$ および $\sigma^{(s)}$ は正值である、 $C_h^{(s)}$ の間には順序関係があるという制約条件があるが、これは事前分布に設定されているので正規分布を提案分布として用いることができる。初期値 $x_{it}^{(0)}$ 、 $r_i^{(0)}$ 、 $C_h^{(0)}$ 、 $\sigma^{(0)}$ および提案分布の分散は、メインのMCMCの前に準備としてのMCMCを行い、この準備としてのMCMCにおける学習アルゴリズムによって初期値および提案分布の分散の値を設定する。

4 分析例

9個の対象の間の6カテゴリーによる非類似度評定を50人に行わせたデータをコンピュータ・シミュレーションにより作成した。分析結果を、平均評定値とモデルによる距離との散布図として図1、図2に示す。推定値は事後分布の平均値である。横軸がデータとしての平均評定値

を表し、縦軸が式（1）によるモデル値である。初期値として与えられる岡本（2010）の解による散布図（図 1）に対して、式（2）に基づいてベイズ的に与えられる解（事後分布の平均値）による散布図（図 2）の方がデータとモデル値の間の単調性がよい。

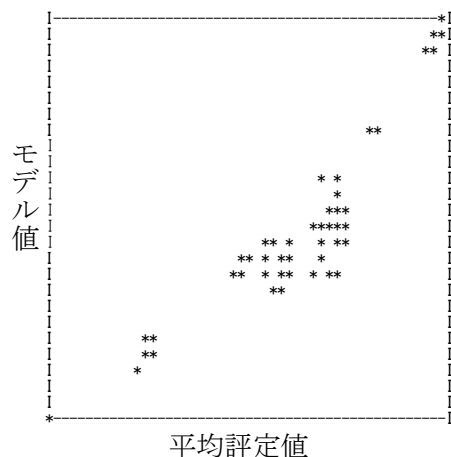


図 1 初期値による散布図。

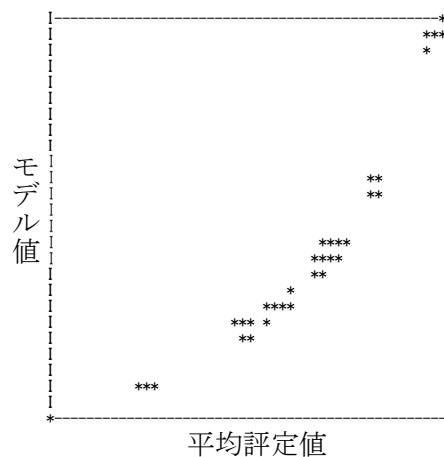


図 2 ベイズ解による散布図。

初期値による布置、およびベイズ的に与えられる解（事後分布の平均値）による布置を図 3 および図 4 に示す。シミュレーションで作成したデータでは、布置は 3×3 の等間隔の格子点とし

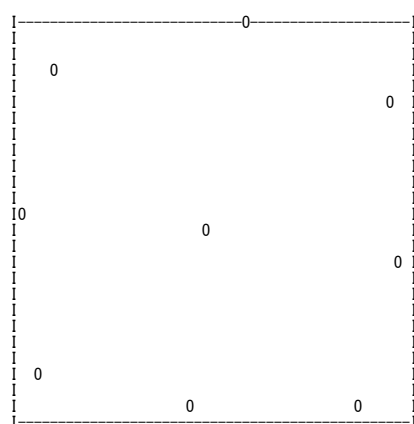


図 3 初期値による布置。

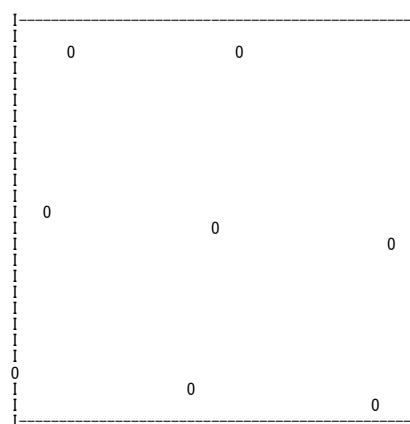


図 4 ベイズ解による布置

て与えられており、この布置に基づいて 6 カテゴリーの評定データが 50 人分作成されている。初期解の布置（図 3）よりベイズ解の布置（事後分布の平均値）の方が等間隔の格子点の布置に近づいている。

重み r_i は、シミュレーションによるデータ生成においては 1 : 2 の比率に設定されている（表 1）。初期解での値は、生成時に 1 であった対象（対象 1, 3, 5, 7, 9）に対する値の平均値は 0.976 であり、生成時に 0.5 であった対象（対象 2, 4, 6, 8）に対する値の平均値は 0.63 である。その比は、 $0.976/0.63 \approx 1.5$ となる。ベイズ解の場合は、生成時に 1 であった対象に対する値の平均値は 1.904 であり、0.5 であった対象に対する値の平均値は 0.95 であるので、その比は

1.904/0.95 $\approx$ 2.0となる。初期値よりベイズ解の方が重みの比率がデータ生成時のものに近くなっている。

表 1 重み r_i の値。

	データ生成時	初期値	ベイズ解
対象 1	1	0.96	1.86
対象 2	0.5	0.63	0.95
対象 3	1	0.98	1.94
対象 4	0.5	0.63	0.94
対象 5	1	1.00	1.90
対象 6	0.5	0.63	0.95
対象 7	1	0.98	1.94
対象 8	0.5	0.63	0.96
対象 9	1	0.96	1.88

以上より、評定値の平均値と解による距離（平均値によるモデル値）との単調関係、対象の布置、重みの比率において、ベイズ解は初期値より改善されていることが確認できた。

5 まとめ

Okada ら（1987）の社会学的データに対する非対称 MDS を、心理学的距離に対応するものとして精神物理学的知見を踏まえて乗法型に変更したものを提案した（岡本、2010）。さらに、心理学的データを対象にするものとして心理尺度を取り入れたモデルを提案し、心理尺度としてべき法則を取り上げたものを報告している（岡本、2011）。これは、マグネチュード推定法によるデータを想定するものであるが、心理学においてはカテゴリー評定もよく用いられる方法であることから、本報告では系列カテゴリー判断のモデルに非対称 MDS を組み込んだものを提案した。提案されたモデルによるシミュレーション・データの分析結果は満足のいくものであった。心理学においてカテゴリー評定によるデータが多いことから、本提案モデルの有用性が期待される。

引用文献

- Guilford, J. P. (1954). *Psychometric methods*, 2nd ed. McGraw-Hill Book Company, Inc.
- Okada, A. and Imaizumi, T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, 21, 81-96.
- 岡本安晴（2010）. 乗法型非対称多次元尺度法. 行動計量学会第 38 回大会抄録集、122–123.
- 岡本安晴（2011）. パラメトリック乗法型非対称 MDS のベイズ分析. 行動計量学会第 39 回大会抄録集、435–438.
- Stevens, S. S. (1975). *Psychophysics*. John Wiley & Sons, Inc.