複数の系列カテゴリー変数の区間値による尺度化について

○宮埜寿夫¹・橋本貴充¹

(¹大学入試センター研究開発部)

キーワード：最適尺度，対応分析，多重対応分析

Interval-valued scaling of successive categorical variables

Hisao MIYANO¹ and Takamitsu HASHIMOTO¹

(¹National Center for University Entrance Examinations)

Key words: Optimal scaling, Correspondence analysis, Multiple correspondence analysis

目 的

系列カテゴリー変数の区間値による尺度化は，最適尺度法あるいは対応分析法による方法の単純な変形により与えられる(宮埜, 橋本;2009)。ここでは，その方法の系列カテゴリー変数が複数である場合への拡張を試みる。まず，カテゴリー変数がひとつの場合について，区間値による尺度化の方法の概要を与える。そして次に，その方法を拡張する形で複数の変数がある場合について尺度化の方法を導く。最後に考察として，導かれた方法とひとつの変数に対する尺度化の方法の関係について議論する。

尺度化の方法—1変数の場合—

いま， I 個の対象に対して，連続する c 個のカテゴリーを値としてとる変数 X について測定を行ったところ，頻度データ n_{ij} , $i = 1, 2, \dots, I$, $j = 1, 2, \dots, c$ が得られたとする。また， j 番目のカテゴリーの区間尺度値を $[t_j, t_{j+1}]$ ，その区間の中心値を $\bar{t}_j = (t_j + t_{j+1})/2$ とする。このとき，最適尺度の場合と同様に，対象の尺度値について全体の平均 0 (すなわち， $\sum_j n_{.j} \bar{t}_j = 0$)，全分散一定の制約条件の下に，対象間の分散を最大にする区間の境界 t_j , $j = 1, 2, \dots, c+1$ を求めると，境界は $\bar{F}t = \lambda Gt$ を満たすことが導かれる。ただし， $t = (t_1, t_2, \dots, t_{p+1})'$ ， $\bar{F} = (\bar{f}_{jk})$ ， $\bar{f}_{jk} = \sum_{i=1}^I \frac{1}{n_{i.}} \bar{n}_{ij} \bar{n}_{ik}$ ， $\bar{n}_{ij} = (n_{i(j-1)} + n_{ij})/2$ ， $n_{i0} = n_{i(c+1)} = 0$ ，

$$G = \frac{2}{3} \begin{pmatrix} \bar{n}_{.1} & n_{.1}/4 & 0 & 0 & 0 & \dots \\ n_{.1}/4 & \bar{n}_{.2} & n_{.2}/4 & 0 & 0 & \dots \\ 0 & n_{.2}/4 & \bar{n}_{.3} & n_{.3}/4 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

とする。

境界の満たす方程式は， $\bar{N} = (\bar{n}_{ij})$ ， $R = \text{diag}(n_{i.})$ とすると， $\bar{N}'R^{-1}\bar{N}t = \lambda Gt$ のように表される。行列 G は，任意の j について $\bar{n}_{.j} > (n_{.j-1} + n_{.j})/4$ のとき正定である。したがって， G は正定とすると方程式の解は $R^{-1/2}\bar{N}G^{-1/2}$ の特異値分解 (SVD) により求められる。すなわち，右特異値ベクトルを e とするとき，解は $t = G^{-1/2}e$ により与えられる。なお， $c+1$ 次元の 1 ベクトルを $\mathbf{1}$ ， $\bar{n}_{.} = (\bar{n}_{.1}, \bar{n}_{.2}, \dots, \bar{n}_{.(c+1)})'$ とすると， $\bar{F}\mathbf{1} = G\mathbf{1} = \bar{n}_{.}$ であるから， $\lambda = 1$ に対応する不適解を持つ。この解を取り除いた解を得るには， $\bar{N}$ の代わりに $\tilde{N} = (\bar{n}_{ij} - \bar{n}_{i.}\bar{n}_{.j}/n_{.})$ を使えばよい。

尺度化の方法—多変数の場合—

複数のカテゴリー変数がある場合についても 1 変数の場合と同様な議論を行うことができる。いま， p 個の変数 X_j , $j = 1, 2, \dots, p$ について測定を行い，頻度データ n_{ijk} , $i = 1, 2, \dots, I$, $j = 1, 2, \dots, p$, $k = 1, 2, \dots, c_j$ が得られたとする。ここで， c_j は変数 X_j のカテゴリー数とする。また， X_j の k 番目のカテゴリーの区間尺度値を $[t_{jk}, t_{j(k+1)}]$ ，その区間の中心値を $\bar{t}_{jk} = (t_{jk} + t_{j(k+1)})/2$ とする。このとき，1 変数の場合と同様に各変数について $\sum_k n_{.j} \bar{t}_{jk} = 0$ ，全体の分散一定の制約条件の下に，対象間の分散を最大にする区間の境

界 t_{jk} , $j = 1, 2, \dots, p$, $k = 1, 2, \dots, c_j + 1$ を対象内における変数間の独立性を仮定して求めると，境界は $Ft = \lambda F_G t$ を満たすことが導かれる。ただし， $t = (t'_1, t'_2, \dots, t'_p)'$ ， $t_j = (t_{jk})$ ， $F = (\bar{F}_{gh})$ ， $\bar{F}_{gh}$ は変数 X_g , X_h に対応する $(c_g + 1) \times (c_h + 1)$ の行列 $\bar{F}_{gh} = (\sum_i (p/n_{i.}) \bar{n}_{ig} \bar{n}_{ih})$ ， F_G は行列 F の対角ブロック行列 $\bar{F}_{gg}$ を行列 G_g ，

$$G_g = \frac{2}{3} \begin{pmatrix} \bar{n}_{.g1} & n_{.g1}/4 & 0 & 0 & 0 & \dots \\ n_{.g1}/4 & \bar{n}_{.g2} & n_{.g2}/4 & 0 & 0 & \dots \\ 0 & n_{.g2}/4 & \bar{n}_{.g3} & n_{.g3}/4 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

により置き換えた行列とする。

行列 F_{gh} , G_g について，すべての要素が 1 である長さ g の列ベクトルを $\mathbf{1}_g$ ， $\bar{n}_{.g} = (\bar{n}_{.g1}, \bar{n}_{.g2}, \dots, \bar{n}_{.g(c_g+1)})'$ とすると，

$$1) F_{gh} \mathbf{1}_h = \bar{n}_{.g}.$$

$$2) G_g \mathbf{1}_g = \bar{n}_{.g}.$$

が成り立つ。このことは，複数のカテゴリー変数がある場合には行列 F_G の正値性が仮定できないことを意味する。したがって，ここでは F_G は非負であることを仮定する。

方程式 $Ft = \lambda F_G t$ は， $\bar{N}_g = (\bar{n}_{ig})$ ， $N = (\bar{N}_1, \bar{N}_2, \dots, \bar{N}_p)$ ， $R = \text{diag}(n_{i.})$ とすると， $N'R^{-1}Nt = \lambda F_G t$ のように表される。したがって， F_G が非負であるならば， F_G の正の固有値からなる固有値行列を Λ_G ，対応する固有ベクトル行列を U_G とすると， F_G は $U_G \Lambda_G U'_G$ により表され， $F_G^{-1/2}$ も一意に表される。したがって，境界の満たす方程式の解は， $R^{-1/2} N F_G^{-1/2}$ の特異値分解により求められる。

なお，1 変数の場合と同様に， $Ft = \lambda F_G t$ は $F_G t = (\bar{n}'_1, \bar{n}'_2, \dots, \bar{n}'_p)'$ となるような t ，例えば $t = (\mathbf{1}'_1, \mathbf{1}'_2, \dots, \mathbf{1}'_p)'$ を不適解として持つ。この不適解を解からあらかじめ取り除くには，行列 $F_0 = p/n_{.}(\bar{n}_{.g} \bar{n}'_{.g})$ を F より差し引いた行列 $\tilde{F} = F - F_0$ を F の代わりに使えばよい。

考 察

系列カテゴリー変数がひとつの場合に対する区間値による尺度構成の方法を多変数の場合について拡張し，その特徴を明らかにした。多変数の場合についても，両側固有値問題を解く問題に帰着されるが，解が存在するための条件である F_G の非負性についてはもう少し議論する必要があると思われる。より一般的にはカテゴリー変数に系列的な変数とそうでない変数とが混在している場合への拡張も考えられる。しかし，この場合については本論文での議論から明らかなように，比較的容易に拡張できると思われる。また，この方法では変数ごとに独立した尺度化を考えていることになるが，一般的には変数に共通した尺度化について議論する必要があると思われる。

引用文献

宮埜・橋本 (2009). 連続するカテゴリーの区間値による尺度化について，日本心理学会第 73 回大会予稿集，486.