正方分割表における
拡張パリンドロミック対称モデルと対称性の分解

三枝　祐輔　(東京理科大学大学院・理工学研究科)
田畑　耕治　(東京理科大学・理工学部)
富澤　貞男　(東京理科大学・理工学部)

1. はじめに

表1は，1943年から1946年までに英国の王立軍需工場で働いていた年齢30歳から39歳までの女性7477名の左右裸眼視力データである．

表1. 英国人女性の左右裸眼視力データ (Stuart [8])

右眼の視力	左眼の視力				計
	良い (1)	やや良い (2)	やや悪い (3)	悪い (4)	
良い (1)	1520	266	124	66	1976
やや良い (2)	234	1512	432	78	2256
やや悪い (3)	117	362	1772	205	2456
悪い (4)	36	82	179	492	789
計	1907	2222	2507	841	7477

表1のような行と列が同じ分類からなる2元分割表は正方分割表とよばれている．2種類の分類間の対称性や非対称性を調べる統計モデルはこれまでに数多く導入されている．たとえば，対称性に関するモデルとして，対称モデル (Bowker [2])，周辺同等モデル (Stuart [8])，準対称モデル (Caussinus [3]) などがある．非対称性に関するモデルとして，条件付き対称モデル (McCullagh [5])，対角パラメータ対称モデル (Goodman [4])，拡張周辺同等モデル (Tomizawa [11])，m比一般化周辺同等モデル (Tahata and Tomizawa [9]) などがある．また，Caussinus [3] は‘対称モデルが成り立つための必要十分条件は，準対称モデルと周辺同等モデルの両方が成り立つことである’という定理を与えた．

本講演では，累積確率の構造を示すパリンドロミック対称モデル (McCullagh [5]) を一般化したモデルを提案する．また，そのモデルを用いた対称モデルの分解を与える．

2. モデル

行変数X_1と列変数X_2が同じ分類からなる$r \times r$正方分割表を考える．(i, j)セルの観測値の出現確率をp_{ij} $(i = 1, \ldots, r; j = 1, \ldots, r)$ とおく．このとき，対称 (S) モデルは次のように定義される (Bowker [2]; Bishop, Fienberg and Holland [1])：

$$p_{ij} = \psi_{ij} \quad (i = 1, \ldots, r; j = 1, \ldots, r), \quad \text{ただし}\ \psi_{ij} = \psi_{ji}.$$

表1において，Sモデルは，ある女性の右眼視力がiかつ左眼視力がjである確率と右眼視力がjかつ左眼視力がiである確率が等しいという構造を示す．Sモデルは制約の強いモデルであるため，適合度が悪い場合が多い．そのようなときには，より制約を弱めたモデルの適用に関心が移る．Sモデルより弱い制約をもつ対称性に関するモデル

としては，たとえば周辺同等 (MH) モデル，準対称 (QS) モデルがある．MH モデルは次のように与えられる (Stuart [8])：

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \ldots, r-1),$$

ただし，$p_{i\cdot} = \sum_{t=1}^{r} p_{it}$，$p_{\cdot i} = \sum_{s=1}^{r} p_{si}$ である．MH モデルは X_1 と X_2 の周辺分布の対称構造を示す．表 1 においては，ある女性の右眼視力が i である確率と左眼視力が i である確率が等しいことを表す．次に，Caussinus [3] によって提案された QS モデルは次のように定義される：

$$p_{ij} = \alpha_i \beta_j \psi_{ij} \quad (i = 1, \ldots, r; j = 1, \ldots, r), \quad \text{ただし} \ \ \psi_{ij} = \psi_{ji}.$$

ここで，$\{\alpha_i = \beta_i\}$ とおいた QS モデルは S モデルであることに注意する．さらに、X_1 が i であるよりも j $(> i)$ であるオッズと X_2 が s であるよりも t $(> s)$ であるオッズの比率を $\theta_{(i<j;s<t)} = (p_{is}p_{jt})/(p_{js}p_{it})$ とすると，QS モデルはオッズ比を用いて，

$$\theta_{(i<j;s<t)} = \theta_{(s<t;i<j)} \quad (i < j; s < t),$$

とも表すことができる．このことから，QS モデルは分割表の主対角線に関するオッズ比の対称構造を示すといえる．

行分類と列分類に順序があるとする．S モデルより弱い制約をもつ非対称性に関するモデルもいくつか提案されている．たとえば，条件付き対称 (CS) モデルは次のように定義される (McCullagh [5])：

$$p_{ij} = \begin{cases} \delta \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases} \quad \text{ただし} \ \ \psi_{ij} = \psi_{ji}.$$

表 1 において，CS モデルは，ある女性の右眼視力が i かつ左眼視力が j $(> i)$ である確率が，右眼視力が j かつ左眼視力が i である確率の δ 倍であることを示す．すなわち，$\delta > 1$ ならば右眼視力の方が左眼視力に比べて良いといえる．また，$\delta = 1$ とした CS モデルは S モデルである．CS モデルを拡張したモデルとして，対角パラメータ対称 (DPS) モデルが

$$p_{ij} = \begin{cases} \delta_{j-i} \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases} \quad \text{ただし} \ \ \psi_{ij} = \psi_{ji},$$

で与えられている (Goodman [4])．表 1 においては，右眼視力が i かつ左眼視力が j $(> i)$ である確率と右眼視力が j かつ左眼視力が i である確率の比が左右視力の差 $j - i$ にのみ依存することを示す.

累積確率を次のように導入する：

$$G_{ij} = \sum_{s=1}^{i} \sum_{t=j}^{r} p_{st} = \mathrm{P}(X_1 \leq i, X_2 \geq j) \quad (i < j),$$

$$G_{ji} = \sum_{s=j}^{r} \sum_{t=1}^{i} p_{st} = \mathrm{P}(X_1 \geq j, X_2 \leq i) \quad (i < j).$$

このとき，S モデルは次のようにも表すことができる：

$$G_{ij} = \Psi_{ij} \quad (i \neq j), \quad p_{ii} = \Psi_{ii}, \quad \text{ただし}\ \Psi_{ij} = \Psi_{ji}.$$

よって，S モデルは累積確率の対称性を示す．また，MH モデルは次のようにも表現できる：

$$G_{i,i+1} = G_{i+1,i} \quad (i = 1, \ldots, r-1).$$

ここまでセル確率の構造を示すモデルをいくつか挙げたが，累積確率の構造を表すモデルもいくつか提案されている．たとえば，Tomizawa [11] は MH モデルの制約を緩めた拡張周辺同等 (EMH) モデルを次のように導入した：

$$G_{i,i+1} = \Delta G_{i+1,i} \quad (i = 1, \ldots, r-1).$$

このモデルの下では，累積確率の比の対数 $\log(G_{i,i+1}/G_{i+1,i})$ が定数となる．EMH モデルをさらに一般化したモデルとして，$\log(G_{i,i+1}/G_{i+1,i})$ がカテゴリ値 i の 1 次式で表される一般化周辺同等 (GMH) モデル (Tomizawa [13]) と，固定した m $(m = 1, \ldots, r-1)$ に対して，$m-1$ 次式で表される m 比一般化周辺同等 (MH(m)) モデル (Tahata and Tomizawa [9]) が提案されている．GMH モデルは次のように定義される：

$$G_{i,i+1} = \Delta_0 \Delta_1^i G_{i+1,i} \quad (i = 1, \ldots, r-1).$$

MH(m) モデルは次のように定義される：固定した m $(m = 1, \ldots, r-1)$ に対して，

$$G_{i,i+1} = \Delta_i^{(m)} G_{i+1,i} \quad (i = 1, \ldots, r-1),$$

ここに，

$$\Delta_i^{(m)} = \prod_{k=0}^{m-1} \Delta_k^{i^k}.$$

MH(1) モデル，MH(2) モデルはそれぞれ EMH モデル，GMH モデルである．また，McCullagh [5] によって提案されたパリンドロミック対称 (PS) モデルは，次のように与えられる：

$$G_{ij} = \begin{cases} \Delta \dfrac{\alpha_i}{\alpha_{j-1}} \Psi_{ij} & (i < j), \\ \Psi_{ij} & (i > j), \end{cases} \quad p_{ii} = \Psi_{ii}, \quad \text{ただし}\ \Psi_{ij} = \Psi_{ji}.$$

ただし，一般性を失うことなく $\alpha_1 = 1$ とする．特に $\alpha_1 = \cdots = \alpha_{r-1}$ とした PS モデルは CS モデルである．さらに，$\Delta = 1$ かつ $\alpha_1 = \cdots = \alpha_{r-1}$ とした PS モデルは S モデルである．また，PS モデルの Δ を Δ_i と置き換えたモデルは一般化パリンドロミック対称 (GPS) モデルである (McCullagh [5])．Tomizawa [12] は，‘PS モデルが成り立つための必要十分条件は，EMH モデルと GPS モデルの両方が成り立つことである’ という定理を与えた．

3. 拡張パリンドロミック対称モデル

3.1. モデル提案

本節では，McCullagh [5] によって導入された PS モデルの一般化を与える．まず，次のモデルを考える：

$$G_{ij} = \begin{cases} \Delta_0 \Delta_1^i \dfrac{\alpha_i}{\alpha_{j-1}} \Psi_{ij} & (i < j), \\ \Psi_{ij} & (i > j), \end{cases} \quad p_{ii} = \Psi_{ii}, \quad \text{ただし} \ \Psi_{ij} = \Psi_{ji}.$$

このモデルを拡張パリンドロミック対称 (EPS) モデルと呼ぶこととする．特に $\Delta_1 = 1$ のとき EPS モデルは PS モデルである．また，このモデルの $j-i=1$ での構造は，GMH モデルに一致する．このことから，EPS モデルは，累積確率の比 $G_{i,i+1}/G_{i,i+1}$ がカテゴリ値 i に応じて指数関数的に変化するときに，適合度が良いと考えられる．このモデルの下で，$\Delta_0\Delta_1^i > 1$ のとき行変数は列変数より確率的に小さく，$\Delta_0\Delta_1^i < 1$ のとき行変数は列変数より確率的に大きい．よって，このモデルのもつパラメータ Δ_0，Δ_1 は行変数と列変数の周辺分布関数の大小関係を推測するのに有用である．

さらに，m 比一般化パリンドロミック対称 (PS(m)) モデルを次のように考える (Saigusa, Tahata and Tomizawa, [7])：固定した m ($m = 1, \ldots, r-1$) に対して，

$$G_{ij} = \begin{cases} \Delta_i^{(m)} \dfrac{\alpha_i}{\alpha_{j-1}} \Psi_{ij} & (i < j), \\ \Psi_{ij} & (i > j), \end{cases} \quad p_{ii} = \Psi_{ii}, \quad \text{ただし} \ \Psi_{ij} = \Psi_{ji},$$

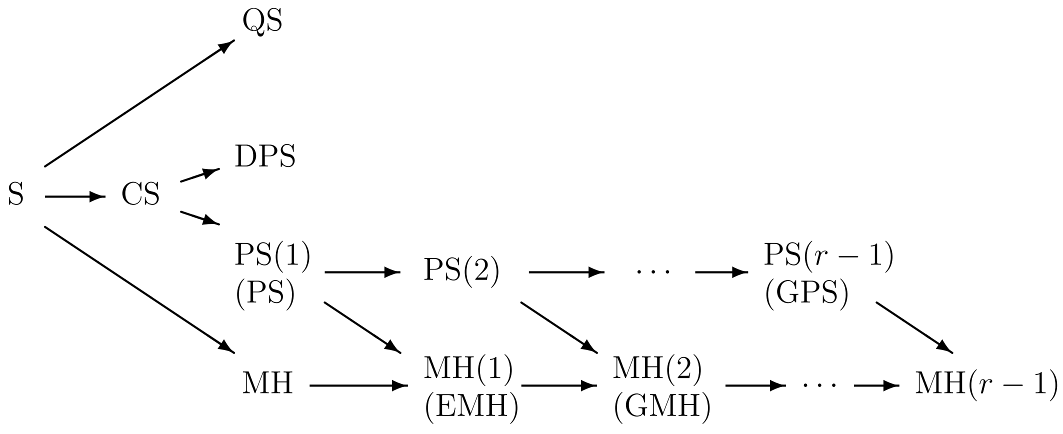
ここに，

$$\Delta_i^{(m)} = \prod_{k=0}^{m-1} \Delta_k^{i^k}.$$

PS(1) モデル，PS(2) モデル，PS($r-1$) モデルはそれぞれ PS モデル，EPS モデル，GPS モデルである．

各モデルの関係を図 1 に示す.

図 1 ： 各モデルの関係



ただし，“$M_1 \to M_2$” はモデル M_1 がモデル M_2 を示すとする.

3.2. 応用例

表 2 は，著名なアメリカ黒人の母親と父親 390 組の学歴データで，(1) 8th grade or less，(2) Part high school，(3) High school，(4) College の順序をもつカテゴリに分類した 4×4 分割表である．表 3 は，表 2 のデータに各モデルを適用した結果を示す.

表2. 著名なアメリカ黒人の母親と父親の学歴データ (Mullins and Sites [6])

	父親の学歴				
母親の学歴	(1)	(2)	(3)	(4)	計
(1)	81	3	9	11	104
(2)	14	8	9	6	37
(3)	43	7	43	18	111
(4)	21	6	24	87	138
計	159	24	85	122	390

表3. 表2のデータに対して適用した各モデルに対する尤度比カイ二乗統計量 G^2 値

モデル	自由度	G^2 値
S	6	36.18*
QS	3	4.17
CS	5	15.40*
DPS	3	10.96*
MH	3	32.15*
MH(1)	2	12.94*
MH(2)	1	0.01
PS(1)	3	14.07*
PS(2)	2	1.40
PS(3)	1	1.39

*印は 5% 有意を示す

表3の結果に加えて，MH(2) モデルが成り立つと仮定した下でのPS(2) モデルの条件付き適合度検定が有意水準5% で採択されることから，PS(2) モデルの方がMH(2) モデルよりよく当てはまっているといえる．さらに，PS(2) モデルの下で，パラメータの最尤推定値は $\hat{\Delta}_0 = 0.191$，$\hat{\Delta}_1 = 1.534$ である（すなわち，$\hat{\Delta}_0\hat{\Delta}_1 = 0.293$，$\hat{\Delta}_0\hat{\Delta}_1^2 = 0.449$，$\hat{\Delta}_0\hat{\Delta}_1^3 = 0.690$）．したがって，パラメータと周辺分布関数の関係から，母親の学歴は父親の学歴に比べて高い傾向があると推測できる．

4. モデル分解

4.1. 対称モデルの分解

Caussinus [3] は，‘S モデルが成り立つための必要十分条件は，QS モデルと MH モデルの両方が成り立つことである’ という定理を与えた．この定理はS モデルの適合度が悪いとき，その原因を調べる上で有用である．本節では，前節で与えたPS(m) モデルを用いたS モデルの分解を与える．

行変数と列変数の原点まわりの k 次モーメント一致 (MO(k)) 構造は次のように表される：固定した k (≥ 1) に対して，

$$E(X_1^k) = E(X_2^k).$$

特に $k = 1$ のとき，ME (平均一致) と記す．

累積部分対称 (CSS) モデルは次のように定義される (Tomizawa, Miyamoto and Ouchi [14])：

$$G_{i,i+2} = G_{i+2,i} \quad (i = 1, \ldots, r-2).$$

Tahata, Yamamoto and Tomizawa [10] はこれらのモデルを用いたSモデルの分解を与えた：

定理 1. Sモデルが成り立つための必要十分条件は，PSモデル，MEモデルおよびCSSモデルのすべてが成り立つことである．

平均値まわりのl次モーメント一致 (MA(l)) 構造は次のように表される：固定したl (≥ 2) に対して，

$$\mu_l^{X_1} = \mu_l^{X_2}, \quad \text{ただし}\ \mu_l^{X_t} = E((X_t - E(X_t))^l) \quad (t = 1, 2).$$

特に$l = 2$のとき，VE (分散一致) と記す．また，歪度一致 (SE) 構造は次のようになる：

$$\frac{\mu_3^{X_1}}{(\mu_2^{X_1})^{3/2}} = \frac{\mu_3^{X_2}}{(\mu_2^{X_2})^{3/2}}.$$

尖度一致 (KE) 構造は次のようになる：

$$\frac{\mu_4^{X_1}}{(\mu_2^{X_1})^2} = \frac{\mu_4^{X_2}}{(\mu_2^{X_2})^2}.$$

このとき，次の定理および系を得る．

定理 2. 固定したm ($m = 1, \ldots, r-1$) に対して, Sモデルが成り立つための必要十分条件は，PS(m)モデル，MO(k)モデル ($k = 1, \ldots, m$) およびCSSモデルのすべてが成り立つことである．

系 1. (i)，(ii)，(iii) は同値である．
- (i) Sモデルが成り立つ．
- (ii) 固定したm ($m = 2, \ldots, r-1$) に対して，PS(m)モデル，MEモデル，MA(l)モデル ($l = 2, \ldots, m$) およびCSSモデルのすべてが成り立つ．
- (iii) $r \geq 5$のとき，PS(4)モデル，MEモデル，VEモデル，SEモデル，KEモデルおよびCSSモデルのすべてが成り立つ．

特に$m = 1$とおいた定理2は定理1である．

4.2. 応用例

表4は，1982年に東京理科大学において得られたデータで，18歳から約25歳の4746名の左右視力に関する4×4分割表である．表5は，表4のデータに各モデルを適用した結果を示す．

表4. 東京理科大学学生の左右視力データ (Tomizawa [11])

右眼の視力	左眼の視力				計
	良い (1)	やや良い (2)	やや悪い (3)	悪い (4)	
良い (1)	1291	130	40	22	1483
やや良い (2)	149	221	114	23	507
やや悪い (3)	64	124	660	185	1033
悪い (4)	20	25	249	1429	1723
計	1524	500	1063	1659	4746

表5. 表4のデータに対して適用した各モデルに対する尤度比カイ二乗統計量G^2値

モデル	自由度	G^2値
S	6	16.95*
QS	3	5.71
MH	3	11.18*
MH(1)	2	0.56
MH(2)	1	0.41
PS(1)	3	1.98
PS(2)	2	1.81
PS(3)	1	1.47
MO(1)	1	9.94*
MO(2)	1	10.38*
MO(3)	1	10.06*
MA(2)	1	0.05
MA(3)	1	7.27*
SE	1	6.74*
KE	1	0.85
CSS	2	3.86

*印は 5% 有意を示す

表5より，Sモデルの当てはまりが悪いことが分かるので，定理2を用いてその原因を考える．たとえば$m=2$とおいた定理2を用いると，PS(2) モデルとCSSモデルよりもむしろMO(1) モデルとMO(2) モデルの示す確率分布の構造が崩れていることが（すなわち行変数と列変数の原点まわりの１次モーメントと２次モーメントが一致していないことが），Sモデルの当てはまりが悪い原因であると推測できる．

参考文献

[1] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge.

[2] Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association* **43**, 572-574.

[3] Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse* **29**, 77-182.

[4] Goodman, L. A. (1979). Multiplicative models for square contingency tables with ordered categories. *Biometrika* **66**, 413-418.

[5] McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* **65**, 413-418.

[6] Mullins, E. l. and Sites, P. (1984). The origins of contemporary eminent black Americans: A three-generation analysis of social origin. *American Sociological Review* **49**, 672-685.

[7] Saigusa, Y., Tahata, K. and Tomizawa, S. (2014). An extended asymmetry model for square contingency tables with ordered categories. *Model Assisted Statistics and Applications*, to appear.

[8] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412-416.

[9] Tahata, K. and Tomizawa, S. (2008). Generalized marginal homogeneity model and its relation to marginal equimoments for square contingency tables with ordered categories. *Advances in Data Analysis and Classification* **2**, 295-311.

[10] Tahata, K., Yamamoto, K. and Tomizawa, S. (2012). Decomposition of symmetry using palindromic symmetry model in a two-way classification. *Journal of Statistics Applications and Probability* **1**, 201-204.

[11] Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *Journal of the Japan Statistical Society* **14**, 35-42.

[12] Tomizawa, S. (1989). Decompositions for conditional symmetry model into palindromic symmetry and modified marginal homogeneity models. *The Australian Journal of Statistics* **31**, 287-296.

[13] Tomizawa, S. (1995). A generalization of the marginal homogeneity model for square contingency tables with ordered categories. *Journal of Educational and Behavioral Statistics* **20**, 349-360.

[14] Tomizawa, S., Miyamoto, N. and Ouchi, M. (2006). Decompositions of symmetry model into marginal homogeneity and distance subsymmetry in square contingency tables with ordered categories. *Revstat: Statistical Journal* **4**, 153-161.