正方分割表における潜在分布に基づく対称性のモデル

生亀　清貴　(日大経済)*

1. はじめに

分割表解析において，多くの人の関心は，行分類と列分類に関連性が存在するかどうか，すなわち，行分類と列分類が独立かどうかである．独立性が成り立たない場合には，たとえばPearsonの相関係数を用いて変数間の連関の強さを推定したり，種々の連関モデルを用いて解析を行う．連関に関するモデルとしては，Goodman (1979b)の一様連関モデル，Agresti (1983a)の線形線形連関モデルなどが提案されている（Goodman, 1981a, 1981bやLiu and Agresti, 2005なども参照されたい）．

一方，行と列が同じ分類からなる正方分割表においては，観測値の多くが主対角セル付近に集中する傾向にあるため，統計的独立性が成り立たないことが多い．そこで，正方分割表の解析においては，分類間の独立性よりも，分類間にどのような関連性が存在するか，すなわち対称性に関心が移り，対称性に関する統計モデルが用いられる．対称性に関するモデルとしては，Bowker (1948)の対称モデル，Caussinus (1965)の準対称モデル，Stuart (1955)の周辺同等モデルなどが提案されている．また，対称性に関するモデルが成り立たない場合には，分類間の非対称性の構造に関心がある．非対称性に関するモデルとしては， McCullagh (1978)の条件付対称モデル，Goodman (1979a)の対角パラメータ対称モデル，Agresti (1983b)の線形対角パラメータ対称モデルなどが導入されている．さらに，Caussinus (1965)は定理「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」を与えた．このような定理をここでは分解定理と呼ぶ．2節では，種々の対称性・非対称性のモデルを紹介し，それらのモデルを用いた分解定理を述べる（さらなる対称性・非対称性に関するモデルや分解定理については，Tomizawa and Tahata (2007)やIki, Tahata and Tomizawa (2009), Iki, Yamamoto and Tomizawa (2014)などを参照されたい）．

さて，分割表の中には，本来データが連続型の変量でいくつかの切断点を設けて離散型の変量に変換し，分割表としたものが多く存在する．ここで，元々の連続型の変量が従う分布を潜在分布と呼ぶことにする．Agresti (1983b)は，潜在分布として周辺分散の等しい2変量正規分布が想定されるとき，線形対角パラメータ対称モデルが分割表データによく適合することを述べた．また，Tomizawa (1991)は等分散性を仮定しない2変量正規分布を潜在分布とするとき，分割表データによく適合するとされるモデルを提案した．本講演では，潜在分布として種々の分布を想定したときに，良く適合するモデルの導入について，近年得られた結果を述べる．具体的な潜在分布としては，3節では正規分布，4節ではt分布，5節では対数正規分布をそれぞれ想定している．実データ解析や，モデルの当てはまりの良さをみるシミュレーションについては，当日報告を行う．

2010 Mathematics Subject Classification: 62H17
キーワード：正方分割表, 対称性, 順序カテゴリ, 潜在分布, 正規分布, t分布, 対数正規分布
* 〒101-8360 東京都千代田区神田三崎町1-3-2
e-mail: iki.kiyotaka@nihon-u.ac.jp

2. 対称性・非対称性に関するモデルと分解

行と列が同じ分類からなる $R \times R$ 正方分割表において，(i, j) セル確率を p_{ij} とする $(i = 1, \ldots, R;\ j = 1, \ldots, R)$. 対称 (S) モデルは次のように定義される (Bowker, 1948):

$$p_{ij} = \psi_{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R),$$

ただし $\psi_{ij} = \psi_{ji}$ (Agresti, 2013, p. 426; Bishop, Fienberg and Holland, 1977, p. 282 なども参照されたい). このモデルは分割表の主対角線に関するセル確率の対称性を示している. S モデルは制約が非常に強いため，実データ解析においては成り立たないことが多い．そこで S モデルの制約を緩めたモデルとして，周辺同等 (MH) モデルや準対称 (QS) モデルが提案されている．MH モデルは次のように定義される (Stuart, 1955):

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

ただし $p_{i\cdot} = \sum_{t=1}^{R} p_{it}$, $p_{\cdot i} = \sum_{s=1}^{R} p_{si}$. このモデルは周辺確率の対称性を表している．QS モデルは次のように定義される (Caussinus, 1965):

$$p_{ij} = \mu\alpha_i\beta_j\psi_{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R),$$

ただし $\psi_{ij} = \psi_{ji}$. 特に $\{\alpha_i = \beta_i\}$ とおいた QS モデルは S モデルである．また，i 行と j 行 $(i < j)$, s 列と t 列 $(s < t)$ のオッズ比を $\theta_{(ij;st)}$ $(= p_{is}p_{jt}/p_{it}p_{js})$ とおくと，QS モデルは次のようにも表される:

$$\theta_{(ij;st)} = \theta_{(st;ij)} \quad (i < j; s < t).$$

すなわち，QS モデルはオッズ比の対称性を表している．

Caussinus (1965) は S モデルに関する分解定理を次のように与えた.

定理 1. S モデルが成り立つための必要十分条件は，QS モデルと MH モデルの両方が成り立つことである．

この分解定理は実データ解析において，S モデルが成り立たないとき，その原因を探るのに有効である．

さらに，S モデルが成り立たないときには，分類間にどのような非対称の構造が存在するかにも関心がある．順序カテゴリ $R \times R$ 正方分割表に対して，McCullagh (1978) は次の条件付対称 (CS) モデルを導入した:

$$p_{ij} = \begin{cases} \delta\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

ただし $\psi_{ij} = \psi_{ji}$. 特に $\delta = 1$ とおいた CS モデルは S モデルである．Goodman (1979a) は対角パラメータ対称 (DPS) モデルを次のように提案した:

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

ただし $\psi_{ij} = \psi_{ji}$. 特に $\delta_1 = \cdots = \delta_{r-1}$ $(= \delta)$ とおいた DPS モデルは CS モデルである．Agresti (1983b) は線形対角パラメータ対称 (LDPS) モデルを次のように定義した:

$$p_{ij} = \begin{cases} \theta^{j-i}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

ただし $\psi_{ij} = \psi_{ji}$. 特に $\theta = 1$ とおいた LDPS モデルは S モデルであり, また LDPS モデルは QS モデルおよび DPS モデルの特別な場合である．LDPS モデルは別表現として次のようにも表すことができる:

$$\frac{p_{ij}}{p_{ji}} = \theta^{j-i} \quad (i < j).$$

よってこのモデルは対称的なセル確率の比が主対角線からの距離 $j-i$ に依存して指数的に変化するという構造を示している．Tomizawa (1991) は拡張線形対角パラメータ対称 (ELDPS) モデルを次のように提案した:

$$p_{ij} = \begin{cases} \theta^{j-i}\gamma^{j^2-i^2}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

ただし $\psi_{ij} = \psi_{ji}$. 特に $\gamma = 1$ のときは LDPS モデルである．ELDPS モデルは別表現として次のようにも表すことができる:

$$\frac{p_{ij}}{p_{ji}} = \theta^{j-i}\gamma^{j^2-i^2} \quad (i < j).$$

よってこのモデルは (i, j) セル確率が対称的な (j, i) セル確率の $\theta^{j-i}\gamma^{j^2-i^2}$ 倍であることを示している．

行変数を X，列変数を Y とし，平均一致 (ME) モデルを $\mathrm{E}(X) = \mathrm{E}(Y)$, 分散一致 (VE) モデルを $\mathrm{Var}(X) = \mathrm{Var}(Y)$ としてそれぞれ定義する．さらに ME モデルと VE モデルの両方の構造を満たすモデルを平均分散一致 (MVE) モデルと呼ぶ．このとき Yamamoto, Iwashita and Tomizawa (2007) は次の定理を与えた．

定理 2. S モデルが成り立つための必要十分条件は，LDPS モデルと ME モデルの両方が成り立つことである．
定理 3. S モデルが成り立つための必要十分条件は，ELDPS モデルと MVE モデルの両方が成り立つことである．

ここで，2 変量正規分布に従う確率変数 (U, V) を考える, ただし $\mathrm{E}(U) = \mu_1$, $\mathrm{E}(V) = \mu_2$, $\mathrm{Var}(U) = \sigma_1^2, \mathrm{Var}(V) = \sigma_2^2$, $\mathrm{Corr}(U, V) = \rho$. このとき同時確率密度関数 $f(u, v)$ は次を満たす:

$$\frac{f(u, v)}{f(v, u)} = a^{v-u}b^{v^2-u^2}, \tag{1}$$

ただし,

$$a = \exp\left[\frac{1}{1-\rho^2}\left\{\left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\right) + \frac{\rho(\mu_2 - \mu_1)}{\sigma_1\sigma_2}\right\}\right],$$

$$b = \exp\left[\frac{1}{2(1-\rho^2)}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\right].$$

特に $\mathrm{Var}(U)=\mathrm{Var}(V)$, すなわち $\sigma_1^2=\sigma_2^2\ (=\sigma^2)$ のとき，確率密度関数 $f(u,v)$ は次を満たす:

$$\frac{f(u,v)}{f(v,u)}=a^{v-u}, \tag{2}$$

ただし，

$$a=\exp\left[\frac{1}{1-\rho}\left(\frac{\mu_2-\mu_1}{\sigma^2}\right)\right].$$

したがって，LDPS モデルの別表現と式 (2) の対応から，分割表データの潜在分布として周辺分散が等しい 2 変量正規分布が想定される場合には LDPS モデルがよく適合すると考えられる (Agresti, 1983b)．同様に，ELDPS モデルの別表現と式 (1) の対応から，周辺分散の異なる 2 変量正規分布が想定される場合には ELDPS モデルがよく適合すると考えられる．

3. 潜在分布に正規分布を想定した場合の対称性に関するモデル

2 変量正規分布に従う確率変数 (U,V) を考える, ただし $\mathrm{E}(U)=\mu_1$, $\mathrm{E}(V)=\mu_2$, $\mathrm{Var}(U)=\mathrm{Var}(V)=\sigma^2$, $\mathrm{Corr}(U,V)=\rho$. このとき同時確率密度関数 $f(u,v)$ は式 (2) のように表せるが，次のようにも表現できる:

$$f(u,v)=ca_1^{(u-v)^2}a_2^{u-v}b_1^{(u+v)^2}b_2^{u+v}, \tag{3}$$

ただし，

$$\begin{aligned}
c&=\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\exp\left(-\frac{(\mu_1-\mu_2)^2}{4\sigma^2(1-\rho)}-\frac{(\mu_1+\mu_2)^2}{4\sigma^2(1+\rho)}\right),\\
a_1&=\exp\left(-\frac{1}{4\sigma^2(1-\rho)}\right),\quad a_2=\exp\left(\frac{\mu_1-\mu_2}{2\sigma^2(1-\rho)}\right),\\
b_1&=\exp\left(-\frac{1}{4\sigma^2(1+\rho)}\right),\quad b_2=\exp\left(\frac{\mu_1+\mu_2}{2\sigma^2(1+\rho)}\right).
\end{aligned}$$

順序カテゴリ $R\times R$ 正方分割表に対して，Tahata, Yamamoto and Tomizawa (2009) は次の正規分布型対称 (NDS) モデルを提案した:

$$p_{ij}=\xi\alpha_1^{(i-j)^2}\alpha_2^{i-j}\beta_1^{(i+j)^2}\beta_2^{i+j}\quad (i=1,\ldots,R;j=1,\ldots,R).$$

式 (3) との対応より，NDS モデルは潜在分布として周辺分散が等しい正規分布が想定される場合に当てはまりが良いとされるモデルである．NDS モデルの下では次が成り立つ:

$$\frac{p_{ij}}{p_{ji}}=(\alpha_2^2)^{i-j}\quad (i>j).$$

したがって，NDS モデルは LDPS モデルの特別な場合である．

また式 (3) に対して $a_2=1$ とおくと（すなわち，$\mathrm{E}(U)=\mathrm{E}(V)=\mu$），次のように表現することができる:

$$f(u,v)=ca^{u^2+v^2}b^{u+v}d^{uv}, \tag{4}$$

ただし，

$$
\begin{aligned}
c &= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{\mu^2}{\sigma^2(1+\rho)}\right), \\
a &= \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)}\right), \quad b = \exp\left(\frac{\mu}{\sigma^2(1+\rho)}\right), \quad d = \exp\left(\frac{\rho}{\sigma^2(1-\rho^2)}\right).
\end{aligned}
$$

順序カテゴリ $R \times R$ 正方分割表に対して，Yamamoto, Nakane and Tomizawa (2016) は制限正規分布型対称 (RNDS) モデルを次のように提案した:

$$
p_{ij} = \xi \alpha^{i^2+j^2} \beta^{i+j} \gamma^{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R).
$$

式 (4) との対応より，RNDS モデルは潜在分布として周辺平均と周辺分散がともに等しい正規分布が仮定される場合に当てはまりが良いとされるモデルである．$\alpha_2 = 1$ のときの NDS モデルは RNDS モデルである．

さらに，2 変量正規分布に従う確率変数 (U, V) に対して，$\mathrm{E}(U) = \mu_1$, $\mathrm{E}(V) = \mu_2$, $\mathrm{Var}(U) = \sigma_1^2, \mathrm{Var}(V) = \sigma_2^2$, $\mathrm{Corr}(U, V) = \rho$ とする．このとき同時確率密度関数 $f(u, v)$ は式 (1) のように表せるが，次のようにも表現できる:

$$
f(u, v) = c a_1^{u^2} a_2^{v^2} b_1^u b_2^v d^{uv}, \tag{5}
$$

ただし，

$$
\begin{aligned}
c &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{\mu_1^2}{2\sigma_1^2(1-\rho^2)} - \frac{\mu_2^2}{2\sigma_2^2(1-\rho^2)} + \frac{\mu_1\mu_2\rho}{\sigma_1\sigma_2(1-\rho^2)}\right), \\
a_1 &= \exp\left(-\frac{1}{2\sigma_1^2(1-\rho^2)}\right), \quad a_2 = \exp\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\right), \\
b_1 &= \exp\left(\frac{\mu_1\sigma_2 - \mu_2\sigma_1\rho}{\sigma_1^2\sigma_2(1-\rho^2)}\right), \quad b_2 = \exp\left(\frac{\mu_2\sigma_1 - \mu_1\sigma_2\rho}{\sigma_1\sigma_2^2(1-\rho^2)}\right), \\
d &= \exp\left(\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\right).
\end{aligned}
$$

順序カテゴリ $R \times R$ 正方分割表に対して，Saigusa, Goda, Yamamoto and Tomizawa (2018) は非制限正規分布型対称 (UNDS) モデルを次のように提案した:

$$
p_{ij} = \xi \alpha_1^{i^2} \alpha_2^{j^2} \beta_1^i \beta_2^j \gamma^{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R).
$$

式 (5) との対応より，UNDS モデルは潜在分布として正規分布が仮定される場合に当てはまりが良いとされるモデルである．特に $\alpha_1 = \alpha_2$ かつ $\beta_1 = \beta_2$ のとき，UNDS モデルは RNDS モデルである．また UNDS モデルは ELDPS モデルの特別な場合であることに注意する．

Saigusa et al. (2018) は次の分解定理を与えた．

定理 4. RNDS モデルが成り立つための必要十分条件は，UNDS モデルと MVE モデルの両方が成り立つことである．

4. 潜在分布に t 分布を想定した場合の対称性に関するモデル

自由度 m $(m > 2)$ の 2 変量 t 分布に従う確率変数 (U, V) を考える, ただし $\mathrm{E}(U) = \mu_1$, $\mathrm{E}(V) = \mu_2$, $\mathrm{Var}(U) = m\sigma_1^2/(m-2), \mathrm{Var}(V) = m\sigma_2^2/(m-2)$, $\mathrm{Corr}(U, V) = \rho$. このとき確率密度関数 $g(u, v)$ は,

$$g(u,v) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\left(1 + \frac{Q(u,v)}{m}\right)^{-\frac{m+2}{2}},$$

ただし,

$$Q(u,v) = \frac{1}{1-\rho^2}\left[\left(\frac{u-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho}{\sigma_1\sigma_2}(u-\mu_1)(v-\mu_2) + \left(\frac{u-\mu_2}{\sigma_2}\right)^2\right];$$

Muirhead (2005, p. 48). 確率密度関数 $g(u, v)$ は次のようにも表すことができる:

$$g(u,v) = c\left[1 + \frac{1}{m}(a_1u + b_1v + a_2u^2 + b_2v^2 + d(u,v))\right]^{-\frac{m+2}{2}}, \tag{6}$$

ただし

$$\begin{aligned}
c &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},\\
a_1 &= \frac{2}{\sigma_1(1-\rho^2)}\left(\frac{\rho\mu_2}{\sigma_2} - \frac{\mu_1}{\sigma_1}\right), \quad b_1 = \frac{2}{\sigma_2(1-\rho^2)}\left(\frac{\rho\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}\right),\\
a_2 &= \frac{1}{\sigma_1^2(1-\rho^2)}, \quad b_2 = \frac{1}{\sigma_2^2(1-\rho^2)},\\
d(u,v) &= \frac{1}{1-\rho^2}\left(-\frac{2\rho}{\sigma_1\sigma_2}uv + \frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2}\right),
\end{aligned}$$

であり，$d(u, v) = d(v, u)$. また, 確率密度関数 $g(u, v)$ は次の関係を満たす:

$$\Bigl(g(u,v)\Bigr)^{-\frac{2}{m+2}} - \Bigl(g(v,u)\Bigr)^{-\frac{2}{m+2}} = k_m(v-u) + l_m(v^2-u^2) \quad (u < v), \tag{7}$$

ただし,

$$\begin{aligned}
k_m &= \frac{2\bigl\{\sigma_2^2\mu_1 - \sigma_1^2\mu_2 + \rho\sigma_1\sigma_2(\mu_1-\mu_2)\bigr\}}{m\sigma_1^2\sigma_2^2(1-\rho^2)}\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{\frac{2}{m+2}},\\
l_m &= \frac{\sigma_1^2 - \sigma_2^2}{m\sigma_1^2\sigma_2^2(1-\rho^2)}\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{\frac{2}{m+2}}.
\end{aligned}$$

特に $\mathrm{Var}(U) = \mathrm{Var}(V)$, すなわち $\sigma_1^2 = \sigma_2^2$ $(= \sigma^2)$ のとき，確率密度関数 $g(u, v)$ は次のように表すことができる:

$$g(u,v) = c\left[1 + \frac{1}{m}(a_1u + b_1v + d(u,v))\right]^{-\frac{m+2}{2}}, \tag{8}$$

ただし

$$
\begin{aligned}
c &= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}},\\
a_1 &= \frac{2}{\sigma^2(1-\rho^2)}\left(\rho\mu_2-\mu_1\right), \quad b_1=\frac{2}{\sigma^2(1-\rho^2)}\left(\rho\mu_1-\mu_2\right),\\
d(u,v) &= \frac{1}{1-\rho^2}\left(-\frac{2\rho}{\sigma^2}uv+\frac{1}{\sigma^2}(u^2+v^2)+\frac{\mu_1^2+\mu_2^2-2\rho\mu_1\mu_2}{\sigma^2}\right),
\end{aligned}
$$

であり，$d(u,v)=d(v,u)$. さらに $g(u,v)$ は次の関係をみたす:

$$
\Big(g(u,v)\Big)^{-\frac{2}{m+2}}-\Big(g(v,u)\Big)^{-\frac{2}{m+2}}=k_m(v-u) \quad (u<v), \tag{9}
$$

ただし,

$$
k_m=\frac{2(\mu_1-\mu_2)}{m\sigma^2(1-\rho)}\Big(2\pi\sigma^2\sqrt{1-\rho^2}\Big)^{\frac{2}{m+2}}.
$$

$R\times R$ 順序カテゴリ正方分割表に対して Iki, Ishihara and Tomizawa (2013) は次の t 分布型対称（TS(m)）モデルを提案した. 固定した m ($m>2$) に対して,

$$
p_{ij}=\gamma\left[1+\frac{1}{m}\Big(\alpha i+\beta j+\phi(i,j)\Big)\right]^{-\frac{m+2}{2}},
$$

ただし $\phi(i,j)=\phi(j,i)$. 式 (8) との対応より，TS(m) モデルは潜在分布として周辺分散が等しい自由度 m の t 分布が想定されるとき, よく適合すると考えられる. TS(m) モデルの下で,

$$
\lim_{m\to\infty}\frac{p_{ij}}{p_{ji}}=\theta^{j-i},
$$

ただし

$$
\theta=\exp\left[\frac{1}{2}(\alpha-\beta)\right].
$$

すなわち，m を大きくすると，TS(m) モデルは LDPS モデルに近づく．TS(m) モデルは次のようにも表すことが可能である:

$$
p_{ij}^{-\frac{2}{m+2}}-p_{ji}^{-\frac{2}{m+2}}=\eta_m(j-i) \quad (i<j),
$$

ただし η_m は未知パラメータとする. この表現は式 (9) と対応していることに注意する．特に $\eta_m=0$ とおいたこのモデルは S モデルである. このモデルは p_{ij} の $-2/(m+2)$ 乗と p_{ji} の $-2/(m+2)$ 乗の差が主対角線からの距離 $j-i$ に比例することを示している. 行変数 X, 列変数 Y の分布関数をそれぞれ $F_i^X=\sum_{k=1}^{i}p_{k\cdot}$, $F_i^Y=\sum_{l=1}^{i}p_{\cdot l}$ とする ($i=1,\ldots,R-1$). 任意の m に対する TS(m) モデルの下で, $\eta_m>0$ と $p_{ij}<p_{ji}$ ($i<j$) は同値である. すなわち $\eta_m>0$ と $F_i^X<F_i^Y$ ($i=1,\ldots,R-1$) が同値である. よって, このモデルの下でのパラメータ η_m は, X と Y のどちらが確率的に大きい (小さい) かを推測するのに有効である.

さらに，Iki, Okada and Tomizawa (2018) は次の拡張 t 分布型対称（ETS(m)）モデルを提案した:

$$
p_{ij}=\gamma\left[1+\frac{1}{m}(\alpha_1 i+\beta_1 j+\alpha_2 i^2+\beta_2 j^2+\psi(i,j))\right]^{-\frac{m+2}{2}},
$$

ただし $\psi(i,j)=\psi(j,i)$. 式 (6) との対応より，このモデルは潜在分布として周辺分散が異なる自由度 m の 2 変量 t 分布が想定されるとき, よく適合すると考えられる. ETS(m) モデルの下で，

$$\lim_{m\to\infty}\frac{p_{ij}}{p_{ji}}=\theta_1^{j-i}\theta_2^{j^2-i^2},$$

ただし

$$\theta_1=\exp\left[\frac{1}{2}(\alpha_1-\beta_1)\right],\ \theta_2=\exp\left[\frac{1}{2}(\alpha_2-\beta_2)\right].$$

すなわち，m を大きくすると，ETS(m) モデルは ELDPS モデルに近づく．ETS(m) モデルは次のようにも表すことができる:

$$p_{ij}^{-\frac{2}{m+2}}-p_{ji}^{-\frac{2}{m+2}}=\eta_m(j-i)+\gamma_m(j^2-i^2)\quad(i<j).$$

ただし η_m,γ_m は未知パラメータとする. この表現は式 (7) と対応していることに注意する．特に $\gamma_m=0$ のときは ETS(m) モデルである. このモデルは, p_{ij} の $-2/(m+2)$ 乗と p_{ji} の $-2/(m+2)$ 乗の差が, i と j の二乗差 j^2-i^2 と, 主対角線からの距離 $j-i$ の線形結合によって表せる.

5. 潜在分布に対数正規分布を想定した場合の対称性に関するモデル

2 変量対数正規分布に従う確率変数 (U,V) を考える，ただし $U>0$ かつ $V>0$. $Z_1=\log U$, $Z_2=\log V$ とおくと, Z_1, Z_2 は 2 変量正規分布に従う, ただし $\mathrm{E}(Z_1)=\mu_1$, $\mathrm{E}(Z_2)=\mu_2$, $\mathrm{Var}(Z_1)=\sigma_1^2$, $\mathrm{Var}(Z_2)=\sigma_2^2$, $\mathrm{Corr}(Z_1,Z_2)=\rho$. このとき, U と V の同時確率密度関数 $h(u,v)$ は,

$$h(u,v)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}uv}\exp\left[-\frac{1}{2(1-\rho^2)}R(u,v)\right],$$

ただし

$$R(u,v)=\left(\frac{\log u-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{\log u-\mu_1}{\sigma_1}\right)\left(\frac{\log v-\mu_2}{\sigma_2}\right)+\left(\frac{\log v-\mu_2}{\sigma_2}\right)^2.$$

同時確率密度関数 $h(u,v)$ は次を満たす:

$$\frac{h(u,v)}{h(v,u)}=a^{\log v-\log u}b^{(\log v)^2-(\log u)^2},\tag{10}$$

ただし

$$a=\exp\left[\frac{1}{1-\rho^2}\left\{-\left(\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}\right)-\frac{\rho(\mu_1-\mu_2)}{\sigma_1\sigma_2}\right\}\right],$$
$$b=\exp\left[\frac{1}{2(1-\rho^2)}\left\{\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right\}\right].$$

$\mathrm{Var}(Z_1)=\mathrm{Var}(Z_2)$, すなわち $\sigma_1^2=\sigma_2^2\ (=\sigma^2)$ のとき，確率密度関数 $h(u,v)$ は次のように表すことができる:

$$\frac{h(u,v)}{h(v,u)}=a^{\log v-\log u},\tag{11}$$

ただし

$$a=\exp\left[\frac{\mu_2-\mu_1}{\sigma^2(1-\rho)}\right].$$

$R\times R$ 順序カテゴリ正方分割表に対して，Iki and Tomizawa (2018) は次の対数正規分布型対称 (LNS) モデルを提案した:

$$p_{ij}=\alpha^{\log i}\beta^{\log j}\psi_{ij}\quad(i=1,\ldots,R;j=1,\ldots,R),$$

ただし $\psi_{ij}=\psi_{ji}$. このモデルは次のようにも表される:

$$\frac{p_{ij}}{p_{ji}}=\theta^{\log j-\log i}\quad(i<j),$$

ただし $\theta=\beta/\alpha$. 式 (11) との対応より，LNS モデルは潜在分布として対数をとった後の周辺分散が等しい 2 変量対数正規分布が想定されるとき, よく適合すると考えられる. 特に $\theta=1$ とおいたこのモデルは S モデルである.

さらに，Iki and Tomizawa (2018) は次の拡張対数正規分布型対称 (ELNS) モデルを提案した:

$$p_{ij}=\alpha_1^{\log i}\alpha_2^{(\log i)^2}\beta_1^{\log j}\beta_2^{(\log j)^2}\psi_{ij}\quad(i=1,\ldots,R;j=1,\ldots,R),$$

ただし $\psi_{ij}=\psi_{ji}$. このモデルは次のようにも表される:

$$\frac{p_{ij}}{p_{ji}}=\theta_1^{\log j-\log i}\theta_2^{(\log j)^2-(\log i)^2}\quad(i<j),$$

ただし $\theta_1=\beta_1/\alpha_1$ かつ $\theta_2=\beta_2/\alpha_2$. 式 (10) との対応より，ELNS モデルは潜在分布として対数をとった後の周辺分散が異なる 2 変量対数正規分布が想定されるとき, よく適合すると考えられる. 特に $\theta_2=1$ とおいたこのモデルは LNS モデルである.

行変数 X と列変数 Y に対して対数平均一致 (LME) モデルを $\mathrm{E}(\log X)=\mathrm{E}(\log Y)$, 対数分散一致 (LVE) モデルを $\mathrm{V}(\log X)=\mathrm{V}(\log Y)$ としてそれぞれ定義する．さらに LME モデルと LVE モデルの両方の構造を満たすモデルを対数平均分散一致 (LMVE) モデルと呼ぶ．このとき Iki and Tomizawa (2018) は次の定理を与えた.

定理 5. S モデルが成り立つための必要十分条件は，LNS モデルと LME モデルの両方が成り立つことである.

定理 6. S モデルが成り立つための必要十分条件は，ELNS モデルと LMVE モデルの両方が成り立つことである.

ただし，定理 5 は Kurakami, Yamamoto and Tomizawa (2011) が与えた定理の特別な場合であることに注意する.

6. 適合度検定

$R\times R$ 正方分割表において，(i,j) セル観測度数を n_{ij}, 対応する期待度数を m_{ij} とする $(i=1,\ldots,R;j=1,\ldots,R)$, ただし $n=\sum_{i=1}^R\sum_{j=1}^R n_{ij}$. $\{n_{ij}\}$ が多項分布に従うとし，モデルの下での m_{ij} の最尤推定量を $\hat{m}_{ij}$ とする．S モデルの下での $\{\hat{m}_{ij}\}$ の値は，直接

求めることができる．その他のモデルの下での $\{\hat{m}_{ij}\}$ の値は，対数尤度方程式に対して Newton-Raphson 法を適用して求めるか，または対数線型モデルに対する繰り返し計算法（たとえば，Darroch and Ratcliff, 1972）などを用いて求める．モデル M の適合度検定は，たとえば尤度比カイ二乗統計量 $G^2(M)$ を用いて行う:

$$G^2(M) = 2\sum_{i=1}^{R}\sum_{j=1}^{R} n_{ij}\log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right),$$

ただし，$\hat{m}_{ij}$ はモデル M の下での m_{ij} の最尤推定量である．各モデルの自由度を表 1 に示す．

表 1: 各モデルの自由度.

モデル	自由度
S	$R(R-1)/2$
LDPS	$(R+1)(R-2)/2$
ELDPS	$(R^2-R-4)/2$
MVE	2
ME	1
NDS	R^2-5
RNDS	R^2-4
UNDS	R^2-6
TS(m)	$(R+1)(R-2)/2$
ETS(m)	$(R^2-R-4)/2$
LNS	$(R+1)(R-2)/2$
ELNS	$(R^2-R-4)/2$
LMVE	2
LME	1

参考文献

[1] Agresti, A. (1983a). A survey of strategies for modeling cross-classifications having ordinal variables. *Journal of the American Statistical Association*, **78**, 184-198.

[2] Agresti, A. (1983b). A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters*, **1**, 313-316.

[3] Agresti, A. (2013). *Categorical Data Analysis.* 3rd ed. New Jersey: Wiley.

[4] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: The MIT Press.

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

[6] 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.

[7] Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, **43**, 1470-1480.

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

[9] Goodman, L. A. (1979b). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537-552.

[10] Goodman, L. A. (1981a). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, **76**, 320-334.

[11] Goodman, L. A. (1981b). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347-355.

[12] Iki, K., Ishihara, T. and Tomizawa, S. (2013). Bivariate t-distribution type symmetry model for square contingency tables with ordered categories. *Model Assisted Statistics and Applications: An International Journal*, **8**, 315-319.

[13] Iki, K., Okada, M. and Tomizawa, S. (2018). An extended bivariate t-distribution type symmetry model for square contingency tables. *Open Journal of Statistics*, **8**, 249-257.

[14] Iki, K., Tahata, K. and Tomizawa, S. (2009). Ridit score type quasi-symmetry and decomposition of symmetry for square contingency tables with ordered categories. *Austrian Journal of Statistics*, **38**, 183-192.

[15] Iki, K. and Tomizawa, S. (2018). Log-normal distribution type symmetry model for square contingency tables with ordered categories. *Austrian Journal of Statistics*, **47**, 39-48.

[16] Iki, K., Yamamoto, K. and Tomizawa, S. (2014). Quasi-diagonal exponent symmetry model for square contingency tables with ordered categories. *Statistics and Probability Letters*, **92**, 33-38.

[17] Kurakami, H., Yamamoto, K. and Tomizawa, S. (2011). Generalized exponential symmetry model and orthogonal decomposition of symmetry for square tables. *The Open Statistics and Probability Journal*, **3**, 1-6.

[18] Liu, I., and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Sociedad Española de Estadística e Investigación Operativa Test*, **14**, 1-73.

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

[20] Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory.* New Jersey: John Wiley & Sons.

[21] Saigusa, Y., Goda, S., Yamamoto, K. and Tomizawa, S. (2018). Unrestricted normal distribution type symmetry model for square contingency tables with ordered categories. *Journal of Biometrics and Biostatistics*, **9**, 1-6.

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

[23] Tahata, K., Yamamoto, K. and Tomizawa, S. (2009). Normal distribution type symmetry model for square contingency tables with ordered categories. *The Open Statistics and Probability Journal*, **1**, 32-37.

[24] Tomizawa, S. (1991). An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *Metron: International Journal of Statistics*, **49**, 401-409.

[25] Tomizawa, S. and Tahata, K. (2007). The analysis of symmetry and asymmetry: orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. *Journal de la Sociétè Française de Statistique*, **148**, 3-36.

[26] Yamamoto, H., Iwashita, T. and Tomizawa, S. (2007). Decomposition of symmetry into ordinal quasi-symmetry and marginal equimoment for multi-way tables. *Austrian Journal of Statistics*, **36**, 291-306.

[27] Yamamoto, K., Nakane, H. and Tomizawa, S. (2016). Symmetry model based on bivariate normal distribution for square contingency tables with ordered categories. *Journal of Statistics: Advances in Theory and Applications*, **15**, 71-84.