正方分割表における点対称性からの隔たりを測る尺度の推定について

東京理科大学大学院 理工学研究科　中島 亮
東京理科大学大学院 理工学研究科　星野 瞭太
東京理科大学 理工学部　藤澤 健吾
東京理科大学 理工学部　田畑 耕治

1. はじめに

行変数と列変数が同じ分類からなる分割表は正方分割表と呼ばれ, 正方分割表解析においては, 変数間に関連が見られ, 統計的独立性は成立しない. そのため変数間の統計的独立性に代わり, 対称性や点対称性に関心がある. 対称性に関するモデルとして, Bowker (1948) の対称 (S) モデル, Read (1977) のグローバル対称 (GS) モデル, McCullagh (1978) の条件付き対称 (CS) モデル等が提案され, 点対称性に関するモデルとして, Wall and Lienert (1976) の点対称 (PS) モデル, Tomizawa (1986) の条件付き点対称 (CPS) モデル, Kurakami *et al.* (2017) のアナザー点対称 (APS) モデル, 逆グローバル対称 (RGS) モデル等が提案されている.

分割表解析において, モデルの当てはまりが悪い場合に, 隔たりの程度を測る尺度に関心がある. 対称性に関する尺度については Tomizawa (1994), Tomizawa (1995), Tomizawa and Saitoh (1999) 等で提案され, 点対称性に関する尺度については Iki and Tomizawa (2019) によって提案されている.

尺度の推定量は漸近不偏推定量であることが知られている. Tomizawa *et al.* (2007) ではテイラー展開の 2 次の項を用いて, S モデルからの隔たりを測る尺度について推定量の改善を行っている. 本報告では, APS, RGS, CPS モデルからの隔たりを測る尺度について推定量の改善を行う. 提案する推定量は, 従来の推定量よりも少ないサンプル数で真値に近づくことをシミュレーションによって確認する.

2. APS モデルからの隔たりを測る尺度の推定について

行変数 X と列変数 Y が順序のある同じ分類からなる $R \times R$ 正方分割表において, (i, j) セル確率を p_{ij} $(i = 1, \ldots, R;\ j = 1, \ldots, R)$ とする. S モデルは次のように定義される (Bowker, 1948):

$$p_{ij} = p_{ji} \quad (i \neq j).$$

S モデルはセル確率の対称構造を示す.

PS モデルは次のように定義される (Wall and Lienert, 1976):

$$p_{ij} = p_{i^* j^*} \quad (1 \leq i, j \leq R),$$

ただし, $i^* = R + 1 - i$, $j^* = R + 1 - j$ である. このモデルは正方分割表において, R が奇数の場合は中心のセル, R が偶数の場合は中心点に関するセル確率の点対称的な確率構造を示す.

APS モデルは次のように定義される (Kurakami *et al.*, 2017):

$$p_{ij} = p_{i^*j^*} \quad (i+j \neq R+1).$$

APS モデルは PS モデルと比べ, 逆対角セルに制約が課されないため, PS モデルよりも制約の弱いモデルである.

モデルが与えられたデータに適合しない場合, モデルからの隔たりを測る尺度に関心がある. Tomizawa (1994) は $p_{ij} + p_{ji} > 0$ $(i = 1, \ldots, R;\ j = 1, \ldots, R)$ を仮定し, S モデルからの隔たりを測る尺度を次のように提案した:

$$\Phi_S = \frac{1}{\delta \log 2} \sum\sum_{i \neq j} p_{ij} \log \frac{2p_{ij}}{p_{ij} + p_{ji}},$$

ただし, $\delta = \sum\sum_{i \neq j} p_{ij}$ である.

Iki and Tomizawa (2019) は $p_{ij} + p_{i^*j^*} > 0$ $(i = 1, \ldots, R;\ j = 1, \ldots, R)$ を仮定し, APS モデルからの隔たりを測る尺度を次のように提案した:

$$\Phi_{APS} = \frac{1}{\Delta \log 2} \sum\sum_{i+j \neq R+1} p_{ij} \log \frac{2p_{ij}}{p_{ij} + p_{i^*j^*}},$$

ただし, $\Delta = \sum\sum_{i+j \neq R+1} p_{ij}$ である.

観測度数 n_{ij} がサンプル数 n $(n = \sum_i \sum_j n_{ij})$ の多項分布に従うと仮定し, p を $R^2 \times 1$ 多項確率ベクトルとする. すなわち,

$$p = (p_{11}, p_{12}, \ldots, p_{1R}, p_{21}, p_{22}, \ldots, p_{2R}, \ldots, p_{R1}, p_{R2}, \ldots, p_{RR})^t,$$

ここで t は転置を表す. $\hat{p}_{ij}$ を標本比率 $(\hat{p}_{ij} = n_{ij}/n)$ とし, p_{ij} を $\hat{p}_{ij}$ で置き換えたベクトルを $\hat{p}$ とする. Tomizawa *et al.* (2007) は $\hat{\Phi}_S$ の漸近バイアスを次のように与えた:

$$E(\hat{\Phi}_S - \Phi_S) = \frac{1}{2n} tr \left(\left[\frac{\partial^2 \Phi_S}{\partial p \partial p^t} \right] (D(p) - pp^t) \right),$$

ここで $D(p)$ は p の i 番目の要素を i 番目の対角要素とする対角行列, tr は行列のトレースを表す. また, $\hat{\Phi}_S$ は, Φ_S の p_{ij} を $\hat{p}_{ij}$ に置き換えた推定量である.

Tomizawa *et al.* (2007) と同様にして, $\hat{\Phi}_{APS}$ の漸近バイアスは次のように与えられる:

$$E(\hat{\Phi}_{APS} - \Phi_{APS}) = \frac{1}{2n} tr \left(\left[\frac{\partial^2 \Phi_{APS}}{\partial p \partial p^t} \right] (D(p) - pp^t) \right).$$

漸近バイアスを取り除くため, 次の推定量を提案する:

$$\tilde{\Phi}_{APS} = \hat{\Phi}_{APS} - \frac{1}{2n} tr \left(\left[\frac{\partial^2 \hat{\Phi}_{APS}}{\partial \hat{p} \partial \hat{p}^t} \right] (D(\hat{p}) - \hat{p}\hat{p}^t) \right),$$

ここで, $[\partial^2 \hat{\Phi}_{APS} / \partial \hat{p} \partial \hat{p}^t]$ は $[\partial^2 \Phi_{APS} / \partial p \partial p^t]$ の p_{ij} を $\hat{p}_{ij}$ で置き換えたものである. 推定量の右辺の二項目を簡略化すると, $\tilde{\Phi}_{APS}$ は次のように与えられる:

$$\tilde{\Phi}_{APS} = \hat{\Phi}_{APS} - \frac{R(R-1)}{4n\hat{\Delta} \log 2}.$$

3. RGS, CPS モデルからの隔たりを測る尺度の推定について

Read (1977), Kurakami *et al.* (2017) では S モデルや APS モデルについて, モデルが成立するための必要十分条件を与えた. 一方で, Tomizawa and Saitoh (1999), Iki and Tomizawa (2019) では尺度の分解定理が Read (1977), Kurakami *et al.* (2017) の結果と対応するように与えられている. 本報告では, RGS モデルと CPS モデルからの隔たりを測る尺度の推定量の漸近バイアスについての分解定理を与える.

GS モデルは次のように定義される (Read, 1977):

$$\delta_U = \delta_L,$$

ただし, $\delta_U = \sum\sum_{i<j} p_{ij}$, $\delta_L = \sum\sum_{i>j} p_{ij}$ である. GS モデルは対角線から上側の確率の和と下側の確率の和が等しい確率構造を示す.

CS モデルは次のように定義される (Read, 1977, McCullagh, 1978):

$$p_{ij} = \gamma p_{ji} \quad (i < j),$$

ここで, γ は非対称性の程度を示すパラメータであり, $\gamma = 1$ のとき S モデルと一致する. また Read (1977) は S モデルが成立するための必要十分条件は GS モデルと CS モデルの両方が成立することであることを示した.

RGS モデルは次のように定義される (Kurakami *et al.*, 2017):

$$\Delta_U = \Delta_L,$$

ただし, $\Delta_U = \sum\sum_{i+j<R+1} p_{ij}$, $\Delta_L = \sum\sum_{i+j>R+1} p_{ij}$ である. RGS モデルは逆対角線から上側の確率の和と下側の確率の和が等しい確率構造を示す.

CPS モデルは次のように定義される (Tomizawa, 1986):

$$p_{ij} = \tau p_{i^*j^*} \quad (i + j < R + 1),$$

ここで, τ は非点対称性の程度を示すパラメータであり, $\tau = 1$ のとき APS モデルと一致する. また Kurakami *et al.* (2017) は APS モデルが成立するための必要十分条件は RGS モデルと CPS モデルの両方が成立することであることを示した.

S モデルの場合と同様に GS, CS モデルが成立しない場合, GS, CS モデルからの隔たりを測る尺度に関心がある. Tomizawa (1995) は $\delta_U + \delta_L > 0$ を仮定し, GS モデルからの隔たりを測る尺度を次のように提案した:

$$\Phi_{GS} = \frac{1}{\log 2}\left(\delta_U^c \log \frac{\delta_U^c}{1/2} + \delta_L^c \log \frac{\delta_L^c}{1/2}\right),$$

ただし, $\delta_U^c = \delta_U/\delta$, $\delta_L^c = \delta_L/\delta$ である.

Tomizawa and Saitoh (1999) は $\delta_U > 0$, $\delta_L > 0$, $p_{ij} + p_{ji} > 0$ を仮定し, CS モデルからの隔たりを測る尺度を次のように提案した:

$$\Phi_{CS} = \frac{1}{\delta \log 2}\sum\sum_{i<j}\left(p_{ij} \log \frac{\delta p_{ij}^c}{\delta_U} + p_{ji} \log \frac{\delta p_{ji}^c}{\delta_L}\right),$$

ただし, $p_{ij}^c = p_{ij}/(p_{ij} + p_{ji})$ である. また Tomizawa and Saitoh (1999) は Φ_S, Φ_{GS}, Φ_{CS} に対して次の定理を与えた:

定理 1. Φ_S は Φ_{GS} と Φ_{CS} の和に等しい.

Iki and Tomizawa (2019) は RGS モデルと CPS モデルからの隔たりを測る尺度をそれぞれ次のように提案した:

$$\Phi_{RGS} = \frac{1}{\log 2}\left(\Delta_U^c \log \frac{\Delta_U^c}{1/2} + \Delta_L^c \log \frac{\Delta_L^c}{1/2}\right),$$
$$\Phi_{CPS} = \frac{1}{\Delta \log 2}\sum\sum_{i+j<R+1}\left(p_{ij} \log \frac{\Delta p_{ij}^*}{\Delta_U} + p_{i^*j^*} \log \frac{\Delta p_{i^*j^*}^*}{\Delta_L}\right),$$

ただし, $\Delta_U^c = \Delta_U/\Delta$, $\Delta_L^c = \Delta_L/\Delta$, $p_{ij}^* = p_{ij}/(p_{ij}+p_{i^*j^*})$ である. また, Iki and Tomizawa (2019) は Φ_{APS}, Φ_{RGS}, Φ_{CPS} に対して次の定理を与えた:

定理 2. Φ_{APS} は Φ_{RGS} と Φ_{CPS} の和に等しい.

APS モデルの場合と同様にして, 尺度の推定量をそれぞれ次のように提案する:

$$\tilde{\Phi}_{RGS} = \hat{\Phi}_{RGS} - \frac{1}{2n} tr\left(\left[\frac{\partial^2 \hat{\Phi}_{RGS}}{\partial \hat{p} \partial \hat{p}^t}\right](D(\hat{p}) - \hat{p}\hat{p}^t)\right),$$
$$\tilde{\Phi}_{CPS} = \hat{\Phi}_{CPS} - \frac{1}{2n} tr\left(\left[\frac{\partial^2 \hat{\Phi}_{CPS}}{\partial \hat{p} \partial \hat{p}^t}\right](D(\hat{p}) - \hat{p}\hat{p}^t)\right).$$

推定量の右辺の二項目を簡略化すると, $\tilde{\Phi}_{RGS}, \tilde{\Phi}_{CPS}$ はそれぞれ次のように与えられる:

$$\tilde{\Phi}_{RGS} = \hat{\Phi}_{RGS} - \frac{1}{2n\hat{\Delta}\log 2},$$
$$\tilde{\Phi}_{CPS} = \hat{\Phi}_{CPS} - \frac{(R+1)(R-2)}{4n\hat{\Delta}\log 2}.$$

$\hat{\Phi}_{APS}$, $\hat{\Phi}_{RGS}$, $\hat{\Phi}_{CPS}$ の漸近バイアスを n 倍したものを Λ_{APS}, Λ_{RGS}, Λ_{CPS} とする. このとき, 次の定理及び系を得る:

定理 3. Λ_{APS} は Λ_{RGS} と Λ_{CPS} の和に等しい.

系 1. $\tilde{\Phi}_{APS}$ は $\tilde{\Phi}_{RGS}$ と $\tilde{\Phi}_{CPS}$ の和に等しい.

定理 4. $\hat{\Phi}_{APS}$, $\hat{\Phi}_{RGS}$, $\hat{\Phi}_{CPS}$ の漸近バイアスの比は $R(R-1)/2 : 1 : (R+1)(R-2)/2$ である.

APS モデルと RGS モデルと CPS モデルの適合度検定の自由度はそれぞれ, $R(R-1)/2$, 1, $(R+1)(R-2)/2$ であり, 定理 4 より, APS モデルと RGS モデルと CPS モデルの尺度の推定量の漸近バイアスの比は各モデルの自由度の比と一致している.

4. 数値シミュレーション

提案した推定量の精度を調べるため, 各モデルからの隔たりを測る尺度について数値シミュレーションを行う. 詳細は当日報告する予定である.

参考文献

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

[2] Iki, K. and Tomizawa, S. (2019). Measure of departure from point symmetry and decomposition of measure for square contingency tables. *Journal of Statistical Theory and Applications*, **19**, 526-533.

[3] Kurakami, H., Negishi, K., and Tomizawa, S. (2017). On decomposition of point-symmetry for square contingency tables with ordered categories. *Journal of Statistics: Advances in Theory and Applications*, **17**, 33-42.

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

[5] Read, C. B. (1977). Partitioning chi-square in contingency tables: A teaching approach. *Communications in Statistics-Theory and Methods*, **6**, 553-562.

[6] Tomizawa, S. (1986). Four kinds of symmetry models and their decompositions in a square contingency table with ordered categories. *Biometrical Journal*, **28**, 387-393.

[7] Tomizawa, S. (1994). Two kinds of measures of departure from symmetry in square contingency tables having nominal categories. *Statistica Sinica*, **4**, 325-334.

[8] Tomizawa, S. (1995). Measures of departure from global symmetry for square contingency tables with ordered categories. *Behaviormetrika*, **22**, 91-98.

[9] Tomizawa, S., Miyamoto, N., and Ohba, N. (2007). Improved approximate unbiased estimators of measures of asymmetry for square contingency tables. *Advances and Applications in Statistics*, **7**, 47-63.

[10] Tomizawa, S. and Saitoh, K. (1999). Kullback-Leibler information type measure of departure from conditional symmetry and decomposition of measure from symmetry for contingency tables. *Calcutta Statistical Association Bulletin*, **49**, 31-39.

[11] Wall, K. D. and Lienert, G. A. (1976). A test for point-symmetry in J-dimensional contingency-cubes. *Biometrical Journal*, **18**, 259-264.