

## 離散データ 2 方向累積和に基づく交互作用順序効果推測

明星大学 広津千尋, 東京都健康長寿医療センター研究所 鶴田陽和

### 1. 概要

2 項分布に従うある事象に対する要因効果( $J$ 水準)を $I$ 層上で比較する問題を考える. 数理的には  $I \times J \times 2$  分割表( $y_{ijk}$ )として定式化されるが, 本稿では特に $I$ 層に年齢層別のように自然な順序があり, かつ, 要因水準にも例えば喫煙習慣の程度のような順序がある場合を想定する. このように 2 方向に順序効果を考える交互作用解析はこれまで主に分散分析モデルで論じているが, 本稿は 2 項分布モデルに対する一般的な定式化を論じる.

### 2. 数理モデル

第 $I$ 層のデータを $y_{ijk}$ , その確率を $p_{ijk}(p_{ij\cdot} = 1)$ で表す. ただし,  $i = 1, \dots, I$ は層を表し,  $j = 1, \dots, J$ は要因水準,  $k = 1, 2$  は成否の反応を表す. この時,  $y_{ijk}$ を並べたベクトルを $\mathbf{y}$ として,

確率分布は $g(\mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^J \{y_{ij\cdot}! \prod_{k=1}^2 (p_{ijk}^{y_{ijk}} / y_{ijk}!)\}$ で与えられる. ここで 対数線形模型:

$\log p_{ijk} = \gamma_i + \varphi_{ij} + \tau_{ik} + \omega_{ijk}$ を想定した上, 周辺和 $y_{ij\cdot}, y_{i\cdot k}$ および $y_{\cdot jk}$ を与えた相似検定を考えると, 確率分布は $g(\mathbf{y}, \boldsymbol{\omega}) = C^{-1}(\boldsymbol{\omega}; y_{ij\cdot}, y_{i\cdot k}, y_{\cdot jk}) \exp(\sum_{i=1}^I \sum_{j=1}^J \omega_{ij1} y_{ij1}) / \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^2 y_{ijk}!$ という簡単な形になる. ただし,  $\mathbf{y}$ は改めて $y_{ij1}$ を並べたベクトルとし, 対数オッズ比 $\omega_{ij1}$ が推論対象である交互作用効果を表す.  $\boldsymbol{\omega}$ はそのベクトル表示である. なお, ここで $\omega_{ij2} = 0$ と定義している.

### 3. 交互作用傾向性仮説

本稿で考える交互作用傾向性仮説は $H: (\mathbf{D}'_I \otimes \mathbf{D}'_J) \boldsymbol{\omega} \geq \mathbf{0}$ , および $K: (\mathbf{D}'_I \otimes \mathbf{L}'_J) \boldsymbol{\omega} \geq \mathbf{0}$ の二通りである. ただし,  $\mathbf{D}'_n$ は $n$ 次差分行列,  $\mathbf{L}'_n$ は $n$ 次 2 階差分行列であり, 具体形は例えば Hirotsu(2017)に与えられている. 仮説 $H$ は交互作用効果が添え字 $i$ と $j$ に関して単調に変化することを意味する.  $K$ はそれが $j$ に関して凸性変化を示し, その程度が $i$ に関して単調に変化することを表す.

### 4. 検定統計量

傾向性仮説検定に関する完全類の理論(Hirotsu 1982)から導かれる基本統計量は, 単調性に対して累積和, 凸性に対して 2 重累積和である. 面白いことにこれらは, 仮説を定義する差分行列の一般逆行列から導かれる. そこで,  $H$ に対して $\mathbf{y}$ に関する 2 方向累積和,  $K$ に対しては $i$ 方向に累積和,  $j$ 方向に 2 重累積和を考えることになる. 本稿ではそれぞれに対して規準化最大統計量を提案し, それを $\max \max \text{acc.}(t_1 \& t_1)$ および $\max \max \text{acc.}(t_1 \& t_2)$ 法と呼称する. これらは 1 元配置における $\max \text{acc.} t$ 法の素直な拡張であり, 累積和のマルコフ性を利用した正確で効率の良いアルゴリズムを提案出来る. 1 元配置の場合より複雑であるが, 通常のサイズで有効に働く.

### 5. 終わりに

この検定問題では通常ロジスティック回帰が用いられるが, 適用には何らかの適合度検定が必要である. 単調性や凸性は, より緩やかな仮定で受け容れ易い. 応用例での比較を当日与える.

### 参考文献

- Hirotsu, C. (2017). *Advanced analysis of variance*. Wiley, New York.
- Hirotsu, C. (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika* 69, 567-577.