多次元分割表の完全独立性検定における変換統計量の性能について

北海道教育大学・札幌　種市　信裕
北海道教育大学・釧路　関谷　祐里
数学利用研究所　外山　淳

1 はじめに:

多項分布モデルを想定した多次元分割表（M 次元分割表）において, 完全独立性帰無仮説の検定を考える. 検定統計量として, 対数尤度比統計量や Pearson X^2 統計量を含む ϕ-ダイバージェンス統計量の族を考える. 完全独立性帰無仮説のもとでの検定統計量の分布の漸近展開に基づく近似式を導出する. さらに, その近似式を用いて標本数が余り大きくない場合であっても, 極限カイ二乗分布への近似が良い変換統計量の構築をおこない, その性能を調べる.

2 多次元分割表と完全独立性帰無仮説の検定統計量:

ある整数 M $(M \geq 3)$ に対して, 標本数 n を固定した M 次元の $J_1 \times \cdots \times J_M$ 分割表多項分布モデルを考える. つまり, $(j_1, \ldots, j_M)$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ セルにおける度数を $X_{j_1 \cdots j_M}$ ただし, $\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} X_{j_1 \cdots j_M} = n$ とし, $X^* = (X_{1 \cdots 11}, \ldots, X_{1 \cdots 1 J_M}, \ldots, X_{J_1 \cdots J_M})^T$ とすると X^* は多項分布 $\mathrm{Mult}_K(n, p^*)$ ただし, $K = \prod_{m=1}^{M} J_m$, $p^* = (p_{1 \cdots 11}, \ldots, p_{1 \cdots 1 J_M}, \ldots, p_{J_1 \cdots J_M})^T$, $0 < p_{j_1 \cdots j_M} < 1 \quad (j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ および $\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} p_{j_1 \cdots j_M} = 1$ に従うとする. 完全独立性の帰無仮説と検定統計量を表すために次のように記号を定義する. $a_{\cdot(m, j_m)} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_{m-1}=1}^{J_{m-1}} \sum_{j_{m+1}=1}^{J_{m+1}} \cdots \sum_{j_M=1}^{J_M} a_{j_1 \cdots j_m \cdots j_M}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, ただし $a_{j_1 \cdots j_m \cdots j_M}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ は M 個の指標を持つ系列とする. 以下で用いられる記号 $p_{\cdot(m, j_m)}$ と $X_{\cdot(m, j_m)}$ は $a_{\cdot(m, j_m)}$ と同じ方法で定義されているものとする. 本報告においては, M 次元分割表の完全独立性帰無仮説 $H_0^M : p_{j_1 \cdots j_M} = Q(j_1, \ldots, j_M)$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, ただし, $Q(j_1, \ldots, j_M) = \prod_{m=1}^{M} p_{\cdot(m, j_m)}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ を考察する. 帰無仮説 H_0^M を検定するための ϕ-ダイバージェンスに基づく検定統計量 C_ϕ^M は以下のように与えられる. $C_\phi^M = 2n \sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} \hat{Q}(j_1, \ldots, j_M) \phi(\hat{p}_{j_1 \cdots j_M} / \hat{Q}(j_1, \ldots, j_M))$, ただし, $\hat{Q}(j_1, \ldots, j_M) = \prod_{m=1}^{M} \hat{p}_{\cdot(m, j_m)}$, $\hat{p}_{j_1 \cdots j_M} = n^{-1} X_{j_1 \cdots j_M}$ $(j_m = 1, \ldots, J_m; m = 1, \ldots, M)$, $\hat{p}_{\cdot(m, j_m)} = n^{-1} X_{\cdot(m, j_m)}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, ϕ は $\phi(1) = \phi'(1) = 0$ および $\phi''(1) = 1$ を満たす $(0, \infty)$ で定義される実凸関数とする.

3 変換統計量の構築と性能の考察:

帰無仮説 H_0^M のもとで, 検定統計量 C_ϕ^M は自由度 $L = K - \sum_{m=1}^{M} J_m + M - 1$ のカイ二乗分布を極限分布として持つ. Yarnold [1] は多項分布の適合度検定統計量としての Pearson X^2 統計量の単純帰無仮説 H_0 のもとでの漸近展開に基づく分布の近似として $P\{X^2 \leq z | H_0\} \approx K_1(z) + K_2(z)$ を提案した. ここで, $K_1(z)$ は多変量エッジワース展開による項, $K_2(z)$ は不連続性を考慮した離散項である. 本報告では, $P\{C_\phi^M \leq x | H_0^M\}$ の多変量エッジワース展開に基づく近似を考え, それに基づき, 極限カイ二乗分布への収束の速さを改良する変換統計量を構築する. さらに, 数値実験を用いてその性能を統計量の分布の近似の良さおよび検出力において考察する.

参考文献

[1] Yarnold, J. K.: Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Ann. Math. Statist.*, **43**, (1972), 1566–1580.