$r \times s$ 分割表の独立性検定統計量の分布の近似について

種市　信裕（鹿児島大学 理学部）

関谷　祐里（北海道教育大学　教育学部　釧路校）

$r \times s$ 分割表において，フルマルチノミアルのモデルを考える。すなわち，$\boldsymbol{X} = (X_{11}, \ldots, X_{1r}, \ldots, X_{s1}, \ldots, X_{sr})' \sim \mathrm{Mult}_{rs}\ (n; p_{11}, \ldots, p_{1r}, \ldots, p_{s1}, \ldots, p_{sr})$ とする。ただし，$\sum_{i=1}^{s}\sum_{j=1}^{r} X_{ij} = n$, $\sum_{i=1}^{s}\sum_{j=1}^{r} p_{ij} = 1$, $0 < p_{ij} < 1$, $(i = 1, \ldots, s, j = 1, \ldots, r)$ である。このとき，独立性の帰無仮説

$$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}, \qquad (i = 1, \ldots, s,\ j = 1, \ldots, r)$$

を検定するための統計量として，パワーダイバージェンス [1] に基づく独立性検定統計量

$$R^a = 2n \sum_{i=1}^{s}\sum_{j=1}^{r} I^a(\hat{p}_{ij}, \hat{p}_{i\cdot}\hat{p}_{\cdot j}),$$

ただし，

$$I^a(e,f) = \begin{cases} \{a(a+1)\}^{-1} e\Big\{\Big(\dfrac{e}{f}\Big)^a - 1\Big\} & (a \neq 0, -1) \\ e \log\Big(\dfrac{e}{f}\Big) & (a = 0) \\ f \log\Big(\dfrac{f}{e}\Big) & (a = -1), \end{cases}$$

をさらに一般化した ϕ-ダイバージェンス [2] に基づく独立性検定統計量は,

$$C_\phi = 2n \sum_{i=1}^{s}\sum_{j=1}^{r} \hat{p}_{i\cdot}\hat{p}_{\cdot j}\phi\left(\frac{\hat{p}_{ij}}{\hat{p}_{i\cdot}\hat{p}_{\cdot j}}\right)$$

として定義される。ここで，$\phi(t)$ は $t > 0$ において凸で $\phi(1) = \phi'(1) = 0$ と $\phi''(1) = 1$ を満たす関数，$\hat{p}_{ij} = X_{ij}/n, (i = 1, \ldots, s, j = 1, \ldots, r)$, $\hat{p}_{i\cdot} = X_{i\cdot}/n, (i = 1, \ldots, s)$, $\hat{p}_{\cdot j} = X_{\cdot j}/n, (j = 1, \ldots, r)$, $X_{i\cdot} = \sum_{j=1}^{r} X_{ij}$, $X_{\cdot j} = \sum_{i=1}^{s} X_{ij}$ である。

本報告では，独立性の帰無仮説のもとでの局所エッジワース展開の導出をおこない，それに基づき，独立性検定統計量 C_ϕ における下側確率 $P(C_\phi < x|H_0)$ の漸近近似について，Yarnold [3] による多項分布における Pearson カイ二乗検定統計量の議論を参考に，統計量の分布の連続項と離散項の２つの部分に分けた考察をおこなった。その結果，確率変数の従う分布が連続分布であることを仮定した場合における漸近展開式を導出した。また，離散項の近似に関しても，その評価式や，評価式が有効である場合の条件の導出を

おこなった。一方，この統計量の１次と２次のモーメントの近似式の導出をおこなった。これらの漸近展開式や近似式に関する具体的なパワーダイバージェンス統計量に対する表現も与えた。

独立性帰無仮説のもとでの確率変数の従う分布が連続分布であることを仮定した場合における ϕ-ダイバージェンスに基づく検定統計量の下側確率の漸近展開式に基づいて，対数尤度比統計量におけるバートレット変換の一般化である改良変換統計量の導出をおこなった。一方，モーメントの近似式に基づく変換統計量の構築もおこなった。

このようにして得られた，改良変換統計量およびモーメント修正タイプの変換統計量を，具体的なパワーダイバージェンス統計量に適用し性能の評価をおこなった。その結果，モーメント修正タイプの近似式および改良変換統計量は共に，もとの統計量に比べ，カイ二乗分布への収束の速さが非常に改良されていることが示され，両者を比較すると改良変換統計量のほうが優る場合が多いことがわかった。またそれらの変換統計量は検出力においても良好な性能を示すことが示された。

さらに，分割表のフルマルチノミアルのモデルにおいて，局所対立仮説

$$H_{1n} : p_{ij} = p_{i\cdot}p_{\cdot j} + \frac{c_{ij}}{\sqrt{n}}, \quad \left(\sum_{i=1}^{s}\sum_{j=1}^{r} c_{ij} = 0\right),\ (i = 1, \ldots, s, j = 1, \ldots, r)$$

のもとでの局所エッジワース展開の導出をおこない，確率変数の従う分布が連続分布であることを仮定した場合における ϕ-ダイバージェンスに基づく検定統計量 C_ϕ の下側確率 $P(C_\phi < x|H_{1n})$ の漸近展開式の導出をおこなった。また，具体的なパワーダイバージェンス統計量 R^a に適用し，その性能の評価をおこなった。

参考文献

[1] N. Cressie and T. R. C. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. B, 46 (1984), 440–464.

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

[3] K. Zografos, Asymptotic properties of Φ-divergence statistic and its applications in contingency tables, Int. J. Math. Stat. Sci. 2 (1993), 5–21.