

ノンパラメトリックな正規化変換とその応用

九州大学経済学研究院 前園宜彦

1. 序

統計量の n^{-1} の項までの漸近分布に基づく漸近信頼区間を構成し, その信頼限界の比較を行う. 本報告では Fujioka & Maesono (2000) より高次の正規化変換とスチューデント化統計量のエッジワース展開に基づく信頼区間の比較を論ずる. $X_1, X_2, \dots, X_n$ を互いに独立で同じ分布 $F_\theta(x)$ に従う確率変数とし, 母数 θ に関連する統計量 $S = S(X_1, \dots, X_n)$ を考える. この S をスチューデント化したものが漸近 U -統計量であると仮定する. すなわち

$$\begin{aligned} & \frac{\sqrt{n}(S - \theta)}{\hat{\sigma}} \\ = & n^{-1/2}\delta + n^{-1/2} \sum_{i=1}^n g_1(X_i) + n^{-3/2} \sum_{i=1}^n \tilde{g}_1(X_i) + n^{-3/2} \sum_{1 \leq i < j \leq n} g_2(X_i, X_j) \\ & + n^{-5/2} \sum_{1 \leq i < j < k \leq n} g_3(X_i, X_j, X_k) + R_{n,\ell} \end{aligned}$$

ここで δ はバイアス, $\hat{\sigma}^2$ は $\sigma^2 = n\text{Var}(S)$ の一致推定量でさらに

$$(A1) \quad E[g_1(X_1)] = E[\tilde{g}_1(X_1)] = 0, \quad E[g_1^2(X_1)] = 1,$$

$$(A2) \quad E[g_2(X_1, X_2)|X_1] = E[g_3(X_1, X_2, X_3)|X_1, X_2] = 0 \quad a.s.$$

$$(A3) \quad P\{|R_{n,\ell}| \geq n^{-1}(\log n)^{-3/2}\} = o(n^{-1})$$

を満たすと仮定する. この表現は Hoeffding-分解 (Hoeffding (1961)) として知られ, 漸近的な性質を議論するときに有用なものである.

2. 正規化変換

(A1) ~ (A3) の仮定を満たす統計量 S に対して Fujioka & Maesono (2000) はエッジワース展開の $n^{-1/2}$ と n^{-1} の項を同時に消去する単調な正規化変換を提案し, 漸近的な逆関数を求めて, その有効性を U -統計量について議論している. この結果を利用すると, スチューデント化統計量の漸近的な α -点の近似は

$$Q_{Nor}(\alpha) = z_\alpha - n^{-1/2}(\hat{p}z_\alpha^2 + \hat{q}) + n^{-1}\left\{\left(\frac{5}{3}\hat{p}^2 - \hat{u}\right)z_\alpha^3 + (2\hat{p}\hat{q} - \hat{v})z_\alpha\right\}$$

となる. ただし z_α は標準正規分布の α -点, $\hat{p}, \hat{q}$ は

$$\begin{aligned} p &= -\frac{1}{6}E[g_1^3(X_1)] - \frac{1}{2}E[g_1(X_1)g_1(X_2)g_2(X_1, X_2)] \\ q &= \frac{1}{6}E[g_1^3(X_1)] + \frac{1}{2}E[g_1(X_1)g_1(X_2)g_2(X_1, X_2)] - \delta \end{aligned}$$

のジャックナイフ推定量で, $\hat{u}, \hat{v}$ は Fujioka & Maesono (2000) により議論されている 4 次のキウムラントに関連する定数の推定量である. このとき $z_\alpha (= O(1))$ に対して

$$P\left(\frac{\sqrt{n}(S - \theta)}{\hat{\sigma}} \leq Q_{Nor}(\alpha)\right) = \alpha + o(n^{-1})$$

が成り立つ.

この $Q_{Nor}(\cdot)$ を使うと θ の信頼係数 $1 - \alpha$ 左片側漸近信頼区間は

$$\theta \leq s - \frac{\hat{\sigma}}{\sqrt{n}} Q_{Nor}(\alpha)$$

となる. 実際にこの変換に基づく信頼区間を構成するためには, 3 次のキウムラントの推定量 $\hat{p}, \hat{q}$ の漸近的な表現が必要になる. 元の統計量 S が U -統計量のときはこの表現が求まっており, それを利用して漸近信頼区間が構成できる.

3. スチューデント化統計量のエッジワース展開の基づく近似

スチューデント化統計量のエッジワース展開に基づくコーニッシュ・フィッシャー展開は

$$Q_{Stu}(\alpha) = z_\alpha - n^{-1/2}(\hat{p}z_\alpha + \hat{q}) + n^{-1}\left\{(2\hat{p}^2 - \hat{p}\hat{q} - \frac{\hat{\kappa}_{42}^*}{72})z_\alpha^3 + (2\hat{p}\hat{q} - \frac{\hat{q}^2}{2} - \frac{\hat{\kappa}_{43}^*}{72})z_\alpha\right\}$$

で与えられる. ただし $\hat{\kappa}_{42}^*, \hat{\kappa}_{43}^*$ は 4 次のキウムラントに関連した推定量である. この α -点の近似は, n^{-1} の項までのエッジワース展開に基づくものであるが, 一般には

$$P\left(\frac{\sqrt{n}(S - \theta)}{\hat{\sigma}} \leq Q_{Stu}(\alpha)\right) = \alpha + o(n^{-1/2})$$

である.

4. U -統計量への応用

スチューデント化 U -統計量のエッジワース展開は Maesono (1997) により求められ, また 3 次のキウムラントの推定量 $\hat{p}, \hat{q}$ の漸近的な表現が Maesono (1998) によって求められている. したがって S が U -統計量のときは, 正規化変換 $Q_{Nor}(\alpha)$ とコーニッシュ・フィッシャー展開 $Q_{Stu}(\alpha)$ が実際に構成できる. U -統計量の代表的な例である標本平均と標本不偏分散について, 今までの議論を適用し, その性質を吟味したところ, データの従う分布が正規分布 $N(\mu, \sigma^2)$ のときは

$$Q_{Nor}(\alpha) = Q_{Stu}(\alpha) + R_{n,\ell}$$

が成り立つことが分かった. 他の分布のときも $Q_{Nor}(\alpha)$ と $Q_{Stu}(\alpha)$ の差は小さいことがシミュレーションの結果から推察される.

Fujioka, Y. & Y. Maesono (2000). *J. Statist. Plan. Inf.*, **83**, 47-74.

Hoefding, W. (1961). *Univ. of N. C. Instit. of Stat. Mimeo Series*. No.302.

Maesono, Y. (1997). *J. Statist. Plan. Inf.*, **61**, 61-84.

Maesono, Y. (1998). *Bull. Informatics & Cybernetics* **30**, 51-68.