セミパラメトリック検定統計量の Hoeffding 分解

長崎大学　経済

永井　圭二

1　序

Woodroofe(1983) は、Savage-Sethuraman(1966) による二標本問題でレーマン対立仮説を検定する順位による逐次確率比検定に対し、非線形更新定理が用いられるかどうかという問題を提起した。この報告では今までの研究者とは異なる方法で順位の対数尤度比を Chernoff-Savage 的に高次に展開し、その問題を解決する。

2　Lehmann 対立仮説の順位による尤度比検定

独立な確率変数列 $(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ が観測されるものとする。 ここで X_i は分布 F に従い、Y_i は分布 G に従うものとする。ここでレーマン対立仮説の検定 $H_0: G = F$ vs $H_1: G = F^{\Delta}, \Delta > 0$ を考える。Savage (1956) によれば、順位の対数尤度比はつぎのように書ける。

$$l_n = n \log \Delta + n \int_{-\infty}^{\infty} \log\left(\frac{F_n + G_n}{F_n + \Delta G_n}\right) d(F_n + G_n). \tag{2.1}$$

Lai(1975) は l_n を Chernoff-Savage 統計量と見なして、ランダムウォーク（独立同一な確率変数の和）と残余項の和に書いた。すなわち、

$$\begin{aligned} l_n &= S_n + \xi_n, \\ S_n &= nS(\Delta, F, G) + n(1-\Delta)\int_{-\infty}^{\infty} \frac{G_n - G}{F + \Delta G} dF + n(\Delta - 1)\int_{-\infty}^{\infty} \frac{F_n - F}{F + \Delta G} dG, \end{aligned} \tag{2.2}$$

ここで $S(\Delta, F, G)$ は Kulback-Leibler 情報量で

$$S(\Delta, F, G) = \log(\Delta) + \int_{-\infty}^{\infty} \log\left(\frac{F + G}{F + \Delta G}\right) d(F + G).$$

残余項に対して Lai はどのような小さな $\mu > 0$ に対しても $n^{-\mu}\xi_n \to 0$ という漸近的な結果を得た。これに対し、U 統計量の理論を用いることにより、次の結果が得られる。

レーマン対立仮説 $G = F^{\Delta}$ を検定する順位の対数尤度比 l_n はランダムウォークと残余項の和に書ける；$l_n = S_n + \xi_n$. ここで S_n は (2.2) で定義されたものと同じである。残余項 ξ_n は次の関係を満足する。

$$E\left[\max_{n \le k \le 2n} \left(\xi_k - c^*(k)\right)^2\right] = O(\log n), \tag{2.3}$$

$c^*(n)$ は次のような数列である。$H = \frac{F+G}{2}$ として、

$$\begin{aligned} c^*(n) \;=\; & \int_{H > n^{-1}} \left(\frac{1}{F+G} - \frac{1}{F + \Delta G}\right)(1-F)dF + \int_{H > n^{-1}} \left(\frac{1}{F+G} - \frac{\Delta}{F+\Delta G}\right)(1-G)dG \\ & + \int_{H > n^{-1}} \left\{\left(\frac{-1}{(F+G)^2} + \frac{1}{(F+\Delta G)^2}\right)F(1-F) + \left(\frac{-1}{(F+G)^2} + \frac{\Delta^2}{(F+\Delta G)^2}\right)G(1-G)\right\} dH. \end{aligned}$$

さらに G が実際にレーマン対立仮説であるとすると、すなわち $A \neq 1$ に対して $G = F^A$ とすると、

$$\{\max_{0 \le j \le n} |\xi_{n+j}|, n \ge 1\} \quad \text{は一様可積分、、} \quad \xi_n \to V \quad \text{弱収束、（} n \to \infty \text{）}. \tag{2.4}$$

ここで V はある可積分な確率変数である。

3 順位による逐次確率比検定の期待標本数の漸近展開

Savage and Sethuraman (1966) は順位による逐次確率比検定 $N = \inf\{n \ge 1 : l_n < a \text{ or } l_n > b\}$ $(a < 0 < b)$ を定義した。Berk (1973) によれば、その期待標本数の 1 次の漸近近似は

$$\begin{aligned} EN &= \frac{b}{S(\Delta, F, G)}(1 + o(1)) \quad \text{if } S(\Delta, F, G) > 0\,, \\ EN &= \frac{|a|}{|S(\Delta, F, G)|}(1 + o(1)) \quad \text{if } S(\Delta, F, G) < 0, \end{aligned}$$

によって与えられる。$(\min(|a|, b) \to \infty)$ Woodroofe (1983) は、この順位による逐次確率比検定で第一種と第二種の誤りを犯す確率の 2 次の近似を求めたが、同時に非線形更新定理を期待標本数の 2 次の漸近展開に使えるかどうかという問題を提起した。その解答として次の結果を得る。

もし $S(\Delta, F, G) > 0$ ならば、

$$E(N) = \frac{b - c^*(b)}{S(\Delta, F, G)} + o(\log b) \qquad (\,|a|, b \to \infty). \tag{3.1}$$

$G = F$ の場合、

$$EN = \frac{|a|}{|S(\Delta, F, F)|} + \frac{(c^{**} + o(1))}{|S(\Delta, F, F)|} \log |a| \quad (|a|, b \to \infty). \tag{3.2}$$

ここで $c^{**} = \frac{1}{2}(1 - \Delta)^2/(1 + \Delta)^2$。さらに $G = F^A$、$A \neq 1$、かつ $S(\Delta, F, F^A) > 0$ とすると、

$$E(N) = \frac{b + r - EV}{S(\Delta, F, F^A)} + o(1) \quad (\,|a|, b \to \infty). \tag{3.3}$$

ここで、$r = ES^2_{\tau_+}/(2ES_{\tau_+})$、 $\tau_+ = \inf\{n;\, S_n > 0\}$ である。

参考文献

[1] Keiji Nagai. Nonparametric change-point detection for two populations. *Sequantial Analysis, Issues III-IV* Vol.17. 1998.