数理解析研究所講究録
1506 巻 2006 年 137-142

非正則な位置尺度母数分布族における位置母数の逐次点推定について

筑波大・数理物質科学研究科　小池 健一　(Ken-ichi Koike)
Institute of Mathematics, University of Tsukuba

1　はじめに

$X_1, X_2, \ldots$ を，互いにいずれも平均 $E(X_1)=\mu$，分散 $V(X_1)=\sigma^2(>0)$ の分布に従う確率変数列とする (μ, σ^2 は未知)．損失関数として二乗損失，一標本観測当たりの費用を $d(>0)$ として，μ に関する 点推測問題を考える．μ を標本平均 $\bar{X}_n=\sum_{i=1}^n X_i/n$ で推測すると，リスク関数は

$$r'_n := E(\bar{X}_n-\mu)^2+dn=\sigma^2/n+dn$$

となる．σ が既知であれば，リスク関数は $n=n'_d:=\sigma/\sqrt{d}$ に一番近い整数で最小値をとる．簡単のため，n'_d を整数とすると，リスク関数の最小値は $r'_{n'_d}=2dn'_d=2\sqrt{d}\sigma$ となる．しかし，σ は未知なので，非逐次の推定方式でこのリスク関数の最小値を達成するようなものは存在しない．そこで，母集団分布を正規分布として，Robbins (1959) は次の停止則を考えた:

$$T'_d := \{n \geq m'_d \mid n^2 \geq v_n/d\} \qquad \left(v_n := \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X}_n)^2\right).$$

ただし，m'_d は初期標本数とする．Chow and Yu (1981) は，正規性の条件を外して，ある条件の下で，逐次推定方式 $(T'_d, \bar{X}_{T'_d})$ が漸近リスク有効，すなわち，$r'_{T'_d}/r'_{n'_d} \to 1$ $(d \to 0)$ となることを示した．さらに，Chow and Martinsek (1982) は，$(T'_d, \bar{X}_{T'_d})$ が有界リグレット，すなわち，$(r'_{T'_d}-r'_{n'^{(1)}_d})/d=O(1)$ であることを示した．

非正則な確率分布の典型例として，一様分布に対する逐次推定問題を扱ったものには，例えば，Akahira and Koike (2005), Akahira and Takeuchi (2003), Chaturvedi et al. (2001), Govindarajulu (1997), Wald (1950) など多くがある．

本論文では，有界な台をもつ位置尺度母数分布族に対して，位置母数の逐次点推定を考え，漸近有効性，Robbins の逐次推定方式との比較を行う．同様の結果として，逐次区間推定に関して，Koike (2006) があるが，これは本論文の結果とある意味で一致していることが分かる．

2 台の両端で密度関数が正になる場合

この節では，密度関数が有界な台を持ち，台の両端で正となる場合を考える．まず，極値の漸近分布を，Akahira (1975), Akahira and Takeuchi (1995), Koike (2006) と同様にして求める．

$Z_1, Z_2, \ldots$ を，互いに独立にいずれも (Lebesgue 測度に関する) 密度関数 $f_0(x-\theta)$ $(\theta \in \mathbb{R}^1)$ をもつ確率変数列とする．次を仮定する:

(A1) $f_0(x)$ は有界な台 $(-a, a)$[1]$(a > 0)$ をもつ，すなわち，$f_0(x) > 0$ $(-a < x < a)$, $f_0(x) = 0$ (その他) とする．また，$f_0(x)$ は $(-a, a)$ で 2 回連続微分可能とする．

(A2) $f_0(x)$ は次を満たす:

$$
\begin{cases}
\displaystyle\lim_{x\to -a+0} f_0(x) = c(>0), & \displaystyle\lim_{x\to a-0} f_0(x) = c'(>0), \\
\displaystyle\lim_{-a\to -a+0} f_0'(x) = h, & \displaystyle\lim_{x\to a-0} f_0'(x) = h'.
\end{cases}
$$

ただし，$c(>0)$, $c'(>0)$, h, h' は定数.

$Z_{(1)} := \min_{1\le i\le n} Z_i$, $Z_{(n)} := \max_{1\le i\le n} Z_i$, $U := n(Z_{(1)}+a-\theta)$, $V := n(Z_{(n)}-a-\theta)$ とすると次の補題を得る (Koike (2006)).

補題 1. 条件 (A1), (A2) の下で, (U, V) の同時密度関数 $f_{U,V}^{(n)}(u, v)$ は

$$
\begin{aligned}
&f_{U,V}^{(n)}(u,v) \\
&= \begin{cases}
\exp\{-(uc-vc')\}\Bigg[cc' + \dfrac{1}{n}\Bigg\{-cc' + cc'\Bigg(2(uc-vc') - \left(\dfrac{hu^2}{2} - \dfrac{h'v^2}{2}\right) \\
\quad -\dfrac{1}{2}(uc-vc')^2\Bigg) + huc' + h'vc\Bigg\}\Bigg] + o\left(\dfrac{1}{n}\right) & (v < 0 < u), \\
0 & (\text{その他})
\end{cases}
\end{aligned}
$$

となる.

次に，有界な台 $(\theta-\xi a, \theta+\xi a)$ をもつ位置尺度母数を考える．$X_1, X_2, \ldots, X_n, \ldots$ を，互いに独立にいずれも密度関数 $(1/\xi)f_0((x-\theta)/\xi)$ に従う確率変数列とする．ただし，$\theta \in \mathbb{R}$, $\xi > 0$ とする．$i = 1, 2, \ldots$ に対して $Y_i := (X_i - \theta)/\xi$ とおき，$Y_{(1)} := \min_{1\le i\le n} Y_i$, $Y_{(n)} := \max_{1\le i\le n} Y_i$ とする．$S := n(Y_{(1)} + Y_{(n)})/2$, $T := n(Y_{(1)} - Y_{(n)} + 2a)/2$ とすると，(S, T) の漸近同時密度関数は

$$
f_{S,T}(s,t) = \begin{cases}
2cc'\exp\{-(c-c')s - (c+c')t\} & (t > |s|), \\
0 & (\text{その他})
\end{cases}
$$

[1] f_0 の台が $(-a, b)$ $(a \neq b)$ のとき，基準化したミッドレンジは $n \to \infty$ としたとき θ に確率収束しない．

となる．S の漸近周辺密度関数は

$$f_S(s) = \begin{cases} Ke^{-2cs} & (s \geq 0), \\ Ke^{2c's} & (s < 0) \end{cases}$$

となる．ただし，$K = 2cc'/(c+c')$. 従って，S と S^2 の漸近期待値は

$$E(S) \approx K\left\{\int_0^\infty se^{-2cs}ds + \int_{-\infty}^0 se^{2c's}ds\right\} = \frac{c'-c}{2cc'},$$
$$E(S^2) \approx K\left\{\int_0^\infty s^2e^{-2cs}ds + \int_{-\infty}^0 s^2e^{2c's}ds\right\} = \frac{c'^2-cc'+c^2}{2(cc')^2}$$

となる．よって，次を仮定する:

(A3) $$E(S^2) \to A, E(S^4) = O(1) \qquad (n \to \infty).$$

母集団分布が一様分布 $U(-1,1)$ であるとき, (A3) は満たされる．実際,

$$E(S^2) = \frac{2n^2}{(n+1)(n+2)} \to 2,$$
$$E(S^4) = \frac{24n^4}{(n+1)(n+2)(n+3)(n+4)} = O(1)$$

となることが分かる.

θ をミッドレンジ $M_n = (X_{(1)} + X_{(n)})/2$ で推測すると, リスク関数は

$$r_n := E(M_n - \theta)^2 + dn$$

となる．ただし，$d(>0)$ は一標本観測当たりの費用とする．$S = n(M_n - \theta)/\xi$ であり，(A3) から, r_n は $(A\xi^2/n^2) + dn$ で近似される．これは，$n = n_d^{(1)} := (2A\xi^2/d)^{1/3}$ のときに最小値 $r_{n_d^{(1)}} = 3(A\xi^2d^2)^{1/3}/2^{2/3}$ をとる．しかし，ξ は未知なので，非逐次推定方式でこのリスクを達成することは不可能である．レンジ $R_n := X_{(n)} - X_{(1)}$ は，$n \to \infty$ とすると $2a\xi$ に概収束するので，次の停止則を考える:

$$T_d^{(1)} := \left\{n \geq m_d^{(1)} \mid n^3 \geq AR_n^2/(2a^2d)\right\}.$$

ただし，$m_d^{(1)}$ は初期標本数で $d^{-l} \leq m_d^{(1)} = o(d^{-1/3})$ $(0 < l < 1/3)$ を満たすとする．このとき次の定理を得る.

定理 1. 条件 (A1), (A2), (A3) の下で, $d \to 0$ とすると,

$$\text{(i)}\quad T_d^{(1)}/n_d^{(1)} \xrightarrow{\text{a.s.}} 1, \quad \text{(ii)}\quad E\left(T_d^{(1)}\right)/n_d^{(1)} \to 1, \quad \text{(iii)}\quad r_{T_d^{(1)}}/r_{n_d^{(1)}} \to 1$$

となる.

定理1と Chow and Yu (1981) から, $d \to 0$ としたとき,

$$\frac{r_{T_d^{(1)}}}{r'_{T'_d}} \approx \frac{3(A\xi^2 d^2)^{1/3}/2^{2/3}}{2\sqrt{d}\sigma} \to 0$$

となる．ただし，$\sigma^2 = V(X_1)$ とする．従って，逐次推定方式 $(T_d^{(1)}, M_{T_d^{(1)}})$ は標本数の意味で $(T'_d, \bar{X}_{T'_d})$ より漸近的に優れている．同様の現象が，逐次区間推定の場合に Koike (2006) により示されている．

3　台の両端で密度関数が0になる場合

この節では，密度関数が有界な台を持ち，台の両端で0となる場合を考える．

$Z_1, Z_2, \ldots$ を，互いに独立にいずれも (Lebesgue 測度に関する) 密度関数 $f_0(x-\theta)$ $(\theta \in \mathbb{R}^1)$ をもつ確率変数列とする．次を仮定する:

(A4) $f_0(x)$ は次を満たす:

$$f_0(x) \approx g(x+a)^\gamma \quad (x \to -a+0), \quad f_0(x) \approx g'|x-a|^\gamma \quad (x \to a-0).$$

ただし，γ, g, g' は正の定数とする[2]. $U' := n^{1/(\gamma+1)}(Z_{(1)}+a-\theta)$, $V' := n^{1/(\gamma+1)}(Z_{(n)}-a-\theta)$ とすると, 補題1と同様にして次を得る.

補題 2. 条件 (A1), (A4) の下で, (U', V') の同時密度関数 $f^{(n)}_{U',V'}(u,v)$ は，$n \to \infty$ のとき

$$f^{(n)}_{U',V'}(u,v) \to \begin{cases} gg'(-uv)^\gamma \exp\{-\frac{g'}{\gamma+1}(-v)^{\gamma+1} - \frac{g}{\gamma+1}u^{\gamma+1}\} & (v < 0 < u), \\ 0 & (\text{その他}) \end{cases}$$

となる.

第2節と同様にして，$S' := n^{1/(\gamma+1)}(Y_{(1)}+Y_{(n)})/2$, $T' := n^{1/(\gamma+1)}(Y_{(1)}-Y_{(n)}+2a)/2$ とおくと, (S', T') の漸近同時密度関数を求めることが出来る．さらに，S'^2 の漸近期待値 is $E(S'^2) \to B(>0)$ が計算できる．よって，次を仮定する:

$$\text{(A5)} \qquad E(S'^2) \to B, E(S'^4) = O(1) \qquad (n \to \infty).$$

条件 (A5) の下で, $n \to \infty$ のとき,

$$E\left(n^{2/(\gamma+1)} M_n^2\right) \to B\xi^2 \tag{3.1}$$

---

[2]収束の次数 γ が両端で異なる場合，基準化したミッドレンジは $n \to \infty$ としたとき θ に確率収束しない.

となる．θ をミッドレンジ $M_n = (X_{(1)} + X_{(n)})/2$ で推定すると，リスク関数は

$$r_n = E(M_n - \theta)^2 + dn$$

となる．ただし，$d(>0)$ は一標本観測当たりの費用とする．$S' = n^{1/(\gamma+1)}(M_n - \theta)/\xi$ であり，(3.1) から，r_n は $K\xi^2 n^{-2/(\gamma+1)} + dn$ で近似される．これは，$n = n_d^{(2)} := (2B\xi^2/d)^{1/3}$ で最小値

$$r_{n_d^{(2)}} = \left(\frac{d(\gamma+1)}{2B\xi^2}\right)^{2/(\gamma+3)} \left(1 + \frac{2B\xi^2}{\gamma+1}\right)$$

をとる．しかし，ξ は未知なので，非逐次推定方式でこのリスクを達成することは不可能である．レンジ $R_n := X_{(n)} - X_{(1)}$ は，$n \to \infty$ とすると $2a\xi$ に概収束するので，次の停止則を考える:

$$T_d^{(2)} := \left\{ n \geq m_d^{(2)} \mid n^{(\gamma+3)/(\gamma+1)} \geq BR_n^2/(2a^2 d(\gamma+1)) \right\}.$$

ただし，$m_d^{(2)}$ は初期標本数で，$d^{-l} \leq m_d^{(2)} = o(d^{-(\gamma+1)/(\gamma+3)})$ $(0 < l < (\gamma+1)/(\gamma+3))$ を満たすとする．このとき次を得る．

定理 2. 条件 (A1), (A4), (A5) の下で，$d \to 0$ とすると，

$$\text{(i)} \quad T_d^{(2)}/n_d^{(2)} \overset{\text{a.s.}}{\to} 1, \quad \text{(ii)} \quad E\left(T_d^{(2)}\right)/n_d^{(2)} \to 1, \quad \text{(iii)} \quad r_{T_d^{(2)}}/r_{n_d^{(2)}} \to 1$$

となる．

定理 2 と Chow and Yu (1981) から，$d \to 0$ のとき，

$$\frac{r_{T_d^{(2)}}}{r'_{T'_d}} \approx \frac{\left(\frac{d(\gamma+1)}{2B\xi^2}\right)^{2/(\gamma+3)} \left(1 + \frac{2B\xi^2}{\gamma+1}\right)}{2\sqrt{d}\sigma} \to \begin{cases} 0 & (0 < \gamma < 1), \\ \text{定数} & (\gamma = 1), \\ \infty & (\gamma > 1) \end{cases}$$

となる．ただし，$\sigma^2 = V(X_1)$ とする．よって，逐次推定方式 $(T_d^{(2)}, M_{T_d^{(2)}})$ は標本数の意味で $(T'_d, \bar{X}_{T'_d})$ より，$0 < \gamma < 1$ のときは漸近的に良く，$\gamma > 1$ のときは悪い．同様の現象が，逐次区間推定の場合に Koike (2006) により示されている．また，上記の結果は，非逐次で位置母数を扱った場合の竹内 (1974), Akahira (1975), Akahira and Takeuchi (1995) の結果と，ある意味で一致している．

参考文献

Akahira, M. (1975). Asymptotic theory for estimation of location in non-regular cases, I: order of convergence of consistent estimators. *Rep. Stat. Appl. Res., JUSE*, **22**, 8–26.

Akahira, M. and Koike, K. (2005). Sequential interval estimation of a location parameter with the fixed width in the uniform distribution with an unknown scale parameter. *Sequential Analysis*, **24**, 63–75.

Akahira, M. and Takeuchi, K. (1995). *Non-Regular Statistical Estimation.* Lecture Notes in Statistics 107, Springer, New York.

Akahira, M. and Takeuchi, K. (2003). The information inequality in sequential estimation for the uniform case. *Sequential Analysis*, **22**, 223–232.

Chaturvedi, A., Surinder, K. and Sanjeev, K. (2001). Multi-stage estimation procedures for the "range" of two-parameter uniform distribution. *Metron*, **59**, 179–186.

Chow, Y.S. and Martinsek, A.T. (1982). Bounded regret of a sequential procedure for estimation of the mean. *Ann. Statist.*, **10**, 909–914.

Chow, Y.S. and Yu, K.F. (1981). On the performance of sequential procedure for the estimation of the mean. *Ann. Statist.*, **9**, 184–189.

Govindarajulu, Z. (1997). A note on two-stage and sequential fixed-width intervals for the parameter in the uniform density. *Statist. Prob. Letters*, **36**, 179–188. (Erratum: *Statist. Prob. Letters*, (1999), **42**, 213–215).

Koike, K. (2006). Sequential interval estimation of a location parameter with the fixed width in the non-regular case. To appear in *Sequential Analysis*, **25**.

Robbins, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics (Harold Cramér Volume)*, Almquist and Wiksell, Stockholm, 235–245.

Serfling, R.J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist.*, **41**, 1227–1234.

竹内啓. (1974). 統計的推定の漸近理論. 教育出版.

Wald, A. (1950). *Statistical Decision Function.* Wiley, New York.