数理解析研究所講究録
第1560巻 2007年 155-161

位置尺度母数分布族における位置母数の逐次区間推定について

筑波大・数理物質　小池 健一　(Ken-ichi Koike)
Institute of Mathematics, University of Tsukuba
筑波大・数理物質　赤平 昌文　(Masafumi Akahira)
Institute of Mathematics, University of Tsukuba

1 はじめに

一般に，位置尺度母数をもつ分布の平均に対して，非逐次で固定幅の区間推定をしようとしても，任意の母数に対して被覆確率を信頼係数を満たすように構成するのは不可能であることが，Lehmann (1951) により示されている．

平均と分散が未知の分布族において，Chow and Robbins (1965) は，平均に対する固定幅の逐次信頼区間を構成している．この信頼区間は，区間幅 $2d$ を 0 に近づけたとき漸近一致性を持ち，しかも，分散が既知のときの最小の標本数（標本の大きさ）と逐次推定に必要とされる標本数との比が 1 に収束するなどよい性質をもつことが分かる．この推定方式は母集団の分布型を仮定しなくてもよい長所をもつため，その2次の漸近展開を求めるなど多くの研究がある (例えば，Woodroofe (1977))．

非正則な分布に対する逐次区間推定としては，区間 $(0,\theta)$ 上の一様分布に関して, Graybill and Connell (1964), Akahira (1993), Bose (2001), Mukhopadhyay and Cicconetti (2002) など多くの研究がある．この場合には，標本の最大値が完備十分統計量になり，未知母数も一変数であるので，比較的取り扱いやすい．また，区間 $(\theta-(1/2),\theta+(1/2))$ 上の一様分布に関して， Wald (1950), Akahira and Takeuchi (2003) は，補助統計量を用いたある推定方式が，非逐次推定方式よりも優れた逐次推定方式となることを示した．

一方，位置尺度母数をもつ一様分布，すなわち区間 $(\theta-(\xi/2),\theta+(\xi/2))$ 上の一様分布に関しては，ほとんど研究がなされていなかったが，最近，Akahira and

Koike (2005) は，位置尺度母数をもつ一様分布族において，平均に対する固定幅の信頼区間を構成し，さらに，Koike (2007a) は，密度の台の端点での 0 への収束の速さが両端で一致する切断分布の位置尺度母数分布族を考え，その位置母数に対する固定幅の信頼区間を構成した．その結果，その台の端点で密度関数が急激に変化する場合には，新しい推定方式が Chow and Robbins (1965) の方式より漸近 (的に) 標本数の意味で優れていることが分かった．このことは，密度の台の端点に関する情報を取り入れて推定を行えば，標本数を十分に節約できることを示している．

本論では，Koike(2007a) の結果をさらに一般化し，台の端点での 0 への収束の速さが両端で一致しない切断分布の位置尺度母数分布族を考え，その位置母数に対する固定幅の信頼区間を構成し，それが漸近一致性，漸近有効性などをもつことを示す．また，その信頼区間の被覆確率を数値的に求めて，その良さを確認する．

2 切断分布における極値統計量の分布の漸近展開

まず，$X_1, X_2, \ldots, X_n, \ldots$ を，互いに独立にいずれも（ルベーグ測度に関する）密度関数 $f\left((x-\theta)/\xi\right)/\xi$ ($\theta \in \mathbb{R}^1$, $\xi > 0$) をもつ確率分布に従う確率変数列とする．いま，密度 $f(x)$ は有界な台 $(-a, a)$ ($a > 0$) をもつ, すなわち

$$f(x) \begin{cases} > 0 & (-a < x < a), \\ = 0 & (\text{その他}) \end{cases}$$

とし，区間 $(-a, a)$ において 2 回連続微分可能とする．さらに

$$A := \lim_{x \to -a+} (x+a)^{-\gamma} f(x) > 0, \quad B := \lim_{x \to a-} (a-x)^{-\eta} f(x) > 0, \qquad (1)$$

を満たすと仮定する．ただし，γ と η は非負の定数とする．

密度 f の台が開区間 $(-a, b)$ $(a \neq b)$ であるとき，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．また，(1) を満たす $f(x)$ は，$x \to -a+0$ としたとき $(x+a)^{\gamma}$ のオーダーで 0 に収束し，$x \to a-0$ のときは $(a-x)^{\eta}$ のオーダーで 0 に収束する．従って，$0 \leq \gamma, \eta < 1$ のときには台の両端で密度関数の値は急激に変化し，$\gamma, \eta > 1$ のときには滑らかに変化することになる．なお，こ

れらの条件は, Akahira (1975), Akahira and Takeuchi (1995, pp. 81, 148); Koike (2007a, b) で与えられたものと同様である.

本論では, (1) において, $\gamma \neq \eta$ の場合を考える ($\gamma = \eta$ の場合は, Koike (2007a) を参照). まず, 一般性を失うことなく $\gamma > \eta$ と仮定して良い. いま $U := n^{1/(\gamma+1)}[\{(X_{(1)} - \theta)/\xi\} + a]$, $V := n^{1/(\eta+1)}[\{(X_{(n)} - \theta)/\xi\} - a]$ とおくと, Koike (2007a, b) と同様にして, $n \to \infty$ のときの (U, V) の漸近同時確率密度関数は

$$f_{U,V}(u, v) = \begin{cases} ABu^{\gamma}(-v)^{\eta} \exp\left\{-\frac{A}{\gamma+1}u^{\gamma+1} - \frac{B}{\eta+1}(-v)^{\eta+1}\right\} & (v < 0 < u), \\ 0 & (\text{その他}) \end{cases}$$

となることが分かる. これを用いると, ミッドレンジ $M_n := (X_{(1)} + X_{(n)})/2$ は θ に対して漸近的に偏りをもち, M_n を用いて θ を推定するのは適当でないことが分かる. そこで, θ の推定量として,

$$\delta_n = X_{(1)}\left(\frac{1}{2} + \frac{\kappa}{4a}n^{-1/(\gamma+1)}\right) + X_{(n)}\left(\frac{1}{2} - \frac{\kappa}{4a}n^{-1/(\gamma+1)}\right) \quad (n \geq 1)$$

をとる. ただし, $\kappa := A^{-1}\{(\gamma+1)/A\}^{1/(\gamma+1)}\Gamma((\gamma+2)/(\gamma+1))$ とする. ここで, $Z := (U/2) - \kappa$ とおくと, $n \to \infty$ のとき, Z の漸近密度関数は

$$f_Z(z) = \begin{cases} A2^{\gamma+1}(z+\kappa)^{\gamma} \exp\left\{-\frac{A}{\gamma+1}2^{\gamma+1}(z+\kappa)^{\gamma+1}\right\} & (z > -\kappa), \\ 0 & (z \leq -\kappa) \end{cases}$$

となる. いま, $0 < \alpha < 1$ に対して,

$$\int_{-l_0}^{l_0} f_Z(z)dz = 1 - \alpha \qquad (2)$$

となるように l_0 をとる. ξ が既知とすると, $0 \leq d \leq a\xi$ に対して,

$$P\{|\delta_n - \theta| \leq d\} = P\left\{n^{1/(\gamma+1)}|\delta_n - \theta|/\xi \leq n^{1/(\gamma+1)}d/\xi\right\} \approx \int_{-n^{1/(\gamma+1)}d/\xi}^{n^{1/(\gamma+1)}d/\xi} f_Z(z)dz \qquad (3)$$

となる. ただし, "$\approx$" は δ_n の $n \to \infty$ のときの漸近分布による近似を表す. 従って,

$$n \geq (l_0\xi/d)^{\gamma+1} =: n_d^{(1)},$$

となる n をとれば, 上記の確率は $1 - \alpha$ 以上になる. 言い換えれば, $n_d^{(1)}$ は漸近的な必要最小標本数となる. いま, 停止則として

$$T_d^{(1)} := \inf\left\{n \geq n_0 \;\middle|\; \frac{R_n}{n^{1/(\gamma+1)}} \leq \frac{2ad}{l_0}\right\}$$

を考える．ただし，$n_0(\geq 2)$ は初期標本数とする．この停止則は閉であるだけでなく，有界であることが分かる．

定理　条件 (1) の下で, $d \to 0+$ のとき，次が成り立つ.

$$
\begin{aligned}
&\text{(i) } P\left\{\left|\delta_{T_d^{(1)}} - \theta\right| \leq d\right\} \to 1 - \alpha \quad (\text{漸近一致性}),\\
&\text{(ii) } T_d^{(1)}/n_d^{(1)} \xrightarrow{\text{a.s.}} 1,\\
&\text{(iii) } E(T_d^{(1)})/n_d^{(1)} \to 1 \qquad\qquad (\text{漸近有効性}).
\end{aligned}
$$

かつて，平均 $\mu := E(X_1)$ の推定問題において, Chow and Robbins (1965) は停止則

$$
T_d^{CR} := \left\{n \geq n_0 \mid n \geq u_{\alpha/2} d^{-2} s_n^2\right\}
$$

を提案した．ただし，$u_{\alpha/2}$ を標準正規分布の上側 $100(\alpha/2)\%$ 点，$n_0(\geq 2)$ を初期標本数とする．上の定理と Chow and Robbins (1965) から, $d \to 0+$ のとき

$$
\frac{T_d^{(1)}}{T_d^{CR}} \approx \frac{(l_0\xi/d)^{\gamma+1}}{u_{\alpha/2}^2\sigma^2/d^2} = \frac{(l_0\xi)^{\gamma+1}}{u_{\alpha/2}^2\sigma^2} d^{-\gamma+1} \to \begin{cases} 0 & (0 \leq \gamma < 1), \\ \text{定数} & (\gamma = 1), \\ \infty & (\gamma > 1) \end{cases}
$$

となることが分かる．ただし，$\sigma^2 = V(X_1)$ とする．従って，逐次推定方式 $(T_d^{(1)}, \delta_{T_d^{(1)}})$ は，漸近標本数の意味で，$0 \leq \gamma < 1$ のときには $(T_d^{CR}, \bar{X}_{T_d^{CR}})$ よりも良く，$\gamma > 1$ のときには良くないことが分かる．言い換えると，密度関数がその台の端点で急激に変化する場合には $(T_d^{(1)}, \delta_{T_d^{(1)}})$ が $(T_d^{CR}, \bar{X}_{T_d^{CR}})$ に比べて漸近的に優れていることになる．この結果は，$\eta \to \gamma$ とすれば，Koike (2007a, b) のものと一致している．なお，非逐次の場合について，位置尺度母数を持つ非正則な分布族に対して，Akahira (1975), Akahira and Takeuchi (1995) などに同様の結果がある．

3 数値例

前節の定理は，区間幅 $2d$ について $d \to 0+$ とした極限を考えたものであったが，$d \to 0+$ としない場合まで与えられた逐次推定方式の挙動について保証するものではない．そこで，この節では，数値例を用いてその効率について調べる．まず，前節の $f(x)$ を

$$f(x) = \frac{\gamma+1}{2^{\gamma+1}}(x+1)^{\gamma} \quad (-1 \le x \le 1, \gamma > 0)$$

とする．推定量 δ_n は位置共変なので，$\theta = 0$ として一般性を失わない．ここでは特に，$\gamma = 1$ の場合を考え，これらの分布に従う疑似乱数を 10000 回発生させて，$d = 0.01(0.01)0.05$, $\xi = 1(1)5$, $\alpha = 0.10$ として，定理の逐次推定方式 $(T_d^{(1)}, \delta_{T_d^{(1)}})$ の被覆確率を数値的に求めた結果が下表である．この場合には，逐次推定方式の被覆確率がほぼ信頼係数 0.90 に等しいことが分かる．

表　$(T_d^{(1)}, \delta_{T_d^{(1)}})$ の被覆確率

$\xi \setminus d$	0.01	0.02	0.03	0.04	0.05
1	0.8971	0.8893	0.8879	0.8813	0.8738
2	0.8999	0.8970	0.8975	0.8874	0.8841
3	0.8933	0.8979	0.8994	0.8941	0.8894
4	0.8993	0.8912	0.9000	0.8974	0.9004
5	0.8961	0.8926	0.8993	0.8927	0.8983

4 おわりに

本論において，非正則な位置尺度母数分布族において，漸近一致性，漸近有効性などをもつ逐次推定方式を提案して，数値的観点からもその逐次推定方式の良さを確認した．今後，もっと様々な非正則分布族の母数の逐次推定についても考察する必要がある．

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. (1993). Two-stage sequential estimation procedures for the uniform distribution. (In Japanese). *Proc. Sympos., Res. Inst. Math. Sci., Kyoto University*, **842**, 151–156.

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.

Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.*, **48**, 600–607.

Bose, A. (2001). A boundary crossing problem with application to sequential estimation. *Sequential Analysis*, **20**, 65–76.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean, *Ann. Math. Statist.*, **36**, 457–462.

Ghosh, M. and Mukhopadhyay, N. (1979). Sequential point estimation of the mean when the distribution is unspecified. *Commun. Statist.-Theory & Methods*, **8**, 637–652.

Ghosh, M. and Mukhopadhyay, N., and Sen, P. K. (1997). *Sequential Estimation.* Wiley, New York.

Graybill, F. A. and Connell, T. L. (1964). Sample size required to estimate the parameter in the uniform density within d units of the true value. *J. Amer. Statist. Assoc.*, **59**, 550–556.

Koike, K. (2007a). Sequential interval estimation of a location parameter with the fixed width in the non-regular case. *Sequential Analysis*, **26**, 63–70.

Koike, K. (2007b). Sequential point estimation of location parameter in location-scale family of non-regular distributions. To appear in *Sequential Analysis.*

Lehmann, E. L. (1951). *Notes on the Theory of Estimation.* Berkeley: Univ. of Calif. Press.

Mukhopadhyay, N. and Cicconetti, G. (2002). Second-order properties of a two-stage point estimation procedure for the range in a power family distribution. *Calcutta Statist. Assoc. Bull.*, **52**, 219–234.

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

Woodroofe, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.*, **5**, 984–995.