数理解析研究所講究録
第2047巻 2017年 101-106

二標本指数分布における三段階信頼区間問題

磯貝 英一[1](新潟大学)
Eiichi Isogai
Niigata University

宇野 力 (秋田大学 教育文化学部)
Chikara Uno
Department of Mathematics, Akita University

§1 序

$\{X_{i1}, X_{i2}, \ldots\}$ $(i=1,2)$ は互いに独立な2つの確率変数列とし, $X_{i1}, X_{i2}, \ldots$ は互いに独立にいずれも次の確率密度関数をもつ分布に従うとする:

$$f(t;\mu_i,\sigma_i) = (1/\sigma_i)\exp\left(-(t-\mu_i)/\sigma_i\right) I(t \geq \mu_i).$$

上の式において, $I(\cdot)$ は定義関数を表し, 4つの母数 μ_1, $\mu_2 \in (-\infty,\infty)$, σ_1, $\sigma_2 \in (0,\infty)$ はすべて未知であるとする. この指数分布を $\mathrm{Exp}(\mu_i,\sigma_i)$ とかくことにする. b_1, b_2 $(b_1 b_2 \neq 0)$ を任意に与えられた定数とするとき, 本稿では位置母数の線形結合である $\delta = b_1\mu_1 + b_2\mu_2$ に対する固定幅の区間推定を考える. 任意に与えられた定数 $d \in (0,\infty)$ と $\alpha \in (0,1)$ に対して, 2つの標本 $\{X_{11},\ldots,X_{1n_1}\}$, $\{X_{21},\ldots,X_{2n_2}\}$ にもとづいて, δ に対する幅 $2d$ の信頼区間 J を構成する. このとき, すべての固定された値 μ_1, μ_2, σ_1, σ_2, α, d に対して $P\{\delta \in J\} \geq 1-\alpha$ となるような適切な標本の大きさを求めたい. この問題に対して, 本稿では, 三段階法といわれる逐次手法を与えて, その性質を調べる.

§2 先行研究

$X_1, X_2, \ldots$ は互いに独立に指数分布 $\mathrm{Exp}(\mu,\sigma)$ に従う確率変数列とする. ここで, 位置母数 $\mu \in (-\infty,\infty)$ と尺度母数 $\sigma \in (0,\infty)$ はともに未知である. μ に対する信頼水準 $1-\alpha$ $(0<\alpha<1)$ の固定幅の信頼区間を考える. $X_1,\ldots,X_n$ $(n \geq 2)$ を観測したとき, $X_{n(1)} = \min\{X_1,\ldots,X_n\}$ とし, 固定幅 $d\,(>0)$ の信頼区間 $I_n = [X_{n(1)}-d,\, X_{n(1)}]$ をつくる. このとき, すべての固定された値 μ, σ, α, d に対して $P\{\mu \in I_n\} \geq 1-\alpha$ となるには, 標本の大きさを $n \geq a\sigma/d \equiv n_0$ とすればよい. ここで, $a = \ln(1/\alpha)\,(>0)$ である. しかし, σ は未知なので, n_0 も未知である. これに対して, いくつかの逐次手法が提案されてきた.

[1]本研究は, 科学研究費助成事業（学術研究助成基金助成金）基盤研究 (C) 課題番号 26400193 （研究代表者 磯貝英一）から研究助成を受けています。

(イ) 二段階法: Ghurye (1958) は以下のような Stein (1945) 型の二段階法を与えた. まず, 初期標本 $X_1, \ldots, X_m$ $(m \geq 2)$ により,

$$U_m = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - X_{m(1)}) \quad \text{および} \quad N \equiv N(d) = \max\left\{m,\ \langle \frac{b_m U_m}{d} \rangle + 1\right\}$$

を求める. ここで, b_m は自由度 (2, $2m-2$) の F-分布の上側 100α%点であり, $\langle x \rangle$ は x の整数部分を表す. $N > m$ のとき, 第二段階の標本として $X_{m+1}, \ldots, X_N$ を抽出し, 信頼区間 $I_N = [X_{N(1)} - d,\ X_{N(1)}]$ をつくると, すべての固定された値 μ, σ, d, α に対して, $P\{\mu \in I_N\} \geq 1-\alpha$ (一致性) が成り立つ. しかしながら, $\liminf_{d\to 0} E(N-n_0) = +\infty$ となり, N は 2 次の漸近有効性をもたない.

尺度母数 σ は未知であるが, $\sigma > \sigma_L$ という既知の下限 $\sigma_L\,(> 0)$ があるとき, Mukhopadhyay and Duggan (1999) は上記の二段階法の初期標本の大きさを

$$m \equiv m(d) = \max\{m_0,\ \langle a\sigma_L/d \rangle + 1\} \quad (m_0 \geq 2)$$

と修正し, $N^\dagger = \max\{m,\ \langle b_m U_m/d \rangle + 1\}$ に対して, $d \to 0$ のとき, 次のことを示した:

$$\frac{a\sigma}{2\sigma_L} + o(n_0^{-1/2}) \leq E(N^\dagger - n_0) \leq \frac{a\sigma}{2\sigma_L} + 1 + o(n_0^{-1/2}),$$

$$1 - \alpha + o(n_0^{-1}) \leq P\{\mu \in I_{N^\dagger}\} \leq 1 - \alpha + a\alpha n_0^{-1} + o(n_0^{-1}).$$

第 1 式より, 修正された二段階法 $N^\dagger$ は 2 次の漸近有効性をもつ. Aoshima and Takada (2000) および Aoshima and Aoki (2000) により, この結果は次のように精密化された:

$$E(N^\dagger - n_0) = \frac{a\sigma}{2\sigma_L} + \frac{1}{2} + O(n_0^{-1/2}),$$

$$P\{\mu \in I_{N^\dagger}\} = 1 - \alpha + \frac{a\alpha}{2} n_0^{-1} + o(n_0^{-1}).$$

Isogai, Kobayashi and Uno (2011) はさらに高次の漸近展開式を与えている.

(ロ) 純逐次法: 次の停止規則によって標本の大きさを決める逐次手法を純逐次法といい, これは尺度母数 σ の下限の仮定がなくても 2 次の漸近有効性をもつ:

$$N = \inf\{n \geq m :\ n \geq a\,U_n/d\}.$$

Swanepoel and van Wyk (1982) より, $m \geq 3$ ならば, $d \to 0$ のとき

$$E(N - n_0) = \eta_1 + o(1),$$

$$P\{\mu \in I_N\} = 1 - \alpha + \left(\eta_1 - \frac{a}{2}\right) \alpha a n_0^{-1} + o(n_0^{-1})$$

が成り立つ. ここで, η_1 はある定数で, $\eta_1 \approx -0.253$ であり, $\eta_1 - (a/2) < 0$ となる. しかし, 純逐次法は上述の一致性をもたない.

(ハ) 三段階法: この問題に対して, Mukhopadhyay and Mauromoustakos (1987) は Hall (1981) が提唱した三段階法といわれる逐次手法を与えている. まず, 初期標本として $X_1, \ldots, X_m$ $(m \geq 2)$ をとり, あらかじめ与えた $\rho \in (0,1)$ に対して

$$M_1 = \max\{m,\ \langle \rho a U_m/d \rangle + 1\}$$

を求め, $M_1 > m$ ならば, 第二段階の標本として $X_{m+1}, \ldots, X_{M_1}$ をとる. さらに,

$$M_2 = \max\{M_1,\ \langle a U_{M_1}/d \rangle + 1\}$$

を求め, $M_2 > M_1$ ならば, 第三段階の標本として $X_{M_1+1}, \ldots, X_{M_2}$ をとる. これらの標本をあわせて, 信頼区間 $I_{M_2} = [X_{M_2(1)} - d,\ X_{M_2(1)}]$ をつくる. $d \to 0$ のとき, Mukhopadhyay and Mauromoustakos (1987) は

$$E(M_2 - n_0) = \frac{1}{2} - \rho^{-1} + o(1),$$

$$P\{\mu \in I_{M_2}\} = 1 - \alpha + \eta_2 n_0^{-1} + o(n_0^{-1}) \quad (\eta_2 \text{ はある負の定数})$$

を示しているが, この結果が成り立つには, $d \to 0$ のとき $m \to \infty$ となる増大条件が必要と思われる. Mukhopadhyay and Mauromoustakos (1987) 論文にはこの増大条件に関する記述がないが, Mukhopadhyay(1990) では増大条件を述べている. 上のことから, 三段階法は 2 次の漸近有効性をもつ. しかし, 上の三段階法は一致性をもたない.

Holm (1995) は, 正規分布の母平均の推定について, 一致性をもつ三段階法を提案しているが, それは 2 次の漸近有効性をもたなかった. Takada (2006) はこの Holm の手法を改良して一致性と 2 次の漸近有効性を併せもつ三段階法を与えている.

次に, 第 1 節で述べた二標本指数分布において $\delta = \mu_1 - \mu_2$ $(b_1 = 1,\ b_2 = -1)$ の場合を考える. 各母集団から $\{X_{i1}, \ldots, X_{in_i}\}$ $(n_i \geq 2,\ i = 1, 2)$ を観測したとき,

$$X_{i\,n_i(1)} = \min\{X_{i1}, \ldots, X_{in_i}\},\ \ U_{i\,n_i} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - X_{i\,n_i(1)}) \tag{1}$$

とする. $\underline{n} = (n_1, n_2)$ と表記し, 次の信頼区間を考える.

$$\begin{aligned} J(\underline{n}) &= [\,X_{1\,n_1(1)} - X_{2\,n_2(1)} \pm d\,] \\ &= [\,X_{1\,n_1(1)} - X_{2\,n_2(1)} - d,\ \ X_{1\,n_1(1)} - X_{2\,n_2(1)} + d\,]. \end{aligned}$$

$a = \ln(1/\alpha)$ とするとき, $n_i \geq a\sigma_i/d \equiv C_i$ $(i = 1, 2)$ ならば, すべての固定された値 μ_1, μ_2, σ_1, σ_2, α, d に対して, $P\{\delta \in J(\underline{n})\} \geq 1 - \alpha$ となる. しかし, C_1 と C_2 はともに未知である.

この問題に対して, Mukhopadhyay and Hamdy (1984) は σ_1 と σ_2 が未知であるが等しい場合と, 未知でかつ等しくない場合に分けて, 二段階法と純逐次法を与えた. 二段階法はどちらの場合でも一致性をもつが 2 次の漸近有効性をもたない. 純逐次法はどちらの場合でも 2 次の漸近有効性をもつ. σ_1 と σ_2 が未知であるが等しい場合に, Mukhopadhyay and Hamdy (1984) は純逐次法による信頼区間の被覆確率の 2 次近似式が次のようになることを示した: $d \to 0$ のとき

$$1 - \alpha + \left(\eta_3 + \frac{1}{2} - \frac{a}{4}\right) \alpha\sigma^{-1} d + o(d),$$

ここで, $\sigma = \sigma_1 = \sigma_2$ であり, $\eta_3 \in (0.3735,\ 0.6715)$ はある定数である. Mukhopadhyay and Mauromoustakos (1987) は σ_1 と σ_2 が未知であるが等しい場合に, Mukhopadhyay and Padmanabhan (1993) は σ_1 と σ_2 が未知でかつ等しくない場合にそれぞれ三段階法を与え, 2 次の漸近有効性と信頼区間の被覆確率の 2 次近似式を示した.

Isogai and Futschik (2010) は, $\delta = b_1\mu_1 + b_2\mu_2$ に対する二乗誤差損失のもとでの有界リスク点推定問題において, 純逐次法を提案している.

§3 主結果

第 1 節の二標本指数分布における $\delta = b_1\mu_1 + b_2\mu_2$ の区間推定を考えるとき, (1) 式の $X_{in_i(1)}$ $(i = 1, 2)$ に対して, 信頼区間 $J(\underline{n}) = [\,b_1 X_{1n_1(1)} + b_2 X_{2n_2(1)} \pm d\,]$ をつくる. a_0 を $(1 + a_0)e^{-a_0} = \alpha$ を満たす正数とし,

$$C_i = \frac{a_* |b_i| \sigma_i}{d}, \quad \text{ここで} \quad a_* = \begin{cases} a = \ln(1/\alpha) & (b_1 b_2 < 0 \text{ のとき}) \\ a_0 & (b_1 b_2 > 0 \text{ のとき}) \end{cases}$$

とするとき, $n_i \geq C_i$ $(i = 1, 2)$ ならば, すべての固定された値 μ_1, μ_2, σ_1, σ_2, α, d に対して, $P\{\delta \in J(\underline{n})\} \geq 1 - \alpha$ となることがわかる. しかし, C_1 と C_2 は未知であるので, 本稿では三段階法を提案する.

まず, 大きさ $m\,(\geq 2)$ の初期標本 $X_{i1}, \ldots, X_{im}$ $(i = 1, 2)$ をとる. ここでは, 次の増大条件を仮定する: $d \to 0$ のとき,

$$\text{ある } r > 1 \text{ に対して}, \ m \equiv m(d) = O(d^{-1/r}).$$

任意に $\rho_i \in (0, 1)$ $(i = 1, 2)$ を固定し, (1) で定義される U_{im} に対して

$$T_i \equiv T_i(d) = \max\left\{m,\ \langle \rho_i \frac{a_* |b_i| U_{im}}{d} \rangle + 1\right\}$$

を求め, $T_i > m$ ならば, 第二段階の標本 $X_{im+1}, \ldots, X_{iT_i}$ $(i = 1, 2)$ をとる. さらに,

$$N_i \equiv N_i(d) = \max\left\{T_i,\ \langle \frac{a_* |b_i| U_{iT_i}}{d} \rangle + 1\right\}$$

を計算し, $N_i > T_i$ ならば, 第三段階の標本 $X_{iT_i+1}, \ldots, X_{iN_i}$ $(i = 1, 2)$ をとる. $\underline{N} = (N_1, N_2)$ と表記するとき, すべての標本 $X_{i1}, \ldots, X_{iN_i}$ $(i = 1, 2)$ を用いて, 信頼区間 $J(\underline{N}) = [\,b_1 X_{1\,N_1(1)} + b_2 X_{2\,N_2(1)} \pm d\,]$ をつくる. このとき, Isogai and Uno (2017) は以下の結果を示している.

定理. $d \to 0$ のとき, 次のことがいえる.

(i) $E(N_i) = C_i + \eta_i + o(1) \quad (i = 1, 2)$, ここで $\eta_i = (1/2) - \rho_i^{-1} < 0$.

(ii) $P\{\delta \in J(\underline{N})\} = 1 - \alpha + A_d\, d + o(d)$, ただし

$$A_d = \begin{cases} \dfrac{1}{4}\alpha \displaystyle\sum_{i=1}^{2} \left(2\eta_i - (a+1)\rho_i^{-1}\right) (|b_i|\sigma_i)^{-1} & (b_1 b_2 < 0 \text{ のとき}) \\[2ex] a_0 e^{-a_0} \displaystyle\sum_{i=1}^{2} \left(\frac{1}{2}\eta_i - \frac{1}{6}a_0\rho_i^{-1}\right) (|b_i|\sigma_i)^{-1} & (b_1 b_2 > 0 \text{ のとき}) \end{cases}$$

である.

上の結果より, $A_d < 0$ であることがわかる. $b_1 = 1$ かつ $b_2 = -1$ $(\delta = \mu_1 - \mu_2)$ とする. このとき, 上の三段階法は Mukhopadhyay and Padmanabhan (1993) の手法と同じであり, $A_d = -(\alpha/4)\sum_{i=1}^{2}(a + 3 - \rho_i)/(\rho_i\sigma_i)$ となる. Mukhopadhyay and Padmanabhan (1993) は上の三段階法をある定数 ε_i $(i = 1, 2)$ を用いて

$$N_i^\dagger \equiv N_i^\dagger(d) = \max\left\{T_i,\ \langle \frac{a_*|b_i|U_{iT_i}}{d} + \varepsilon_i \rangle + 1\right\}, \quad \underline{N}^\dagger = (N_1^\dagger, N_2^\dagger)$$

と修正することにより, $d \to 0$ のとき, $E(N_i^\dagger) = C_i + \eta_i + \varepsilon_i + o(1)$ となり, さらに $\varepsilon_i = (a + 3 - \rho_i)/(2\rho_i)$ と選ぶと, $P\{\delta \in J(\underline{N}^\dagger)\} = 1 - \alpha + o(d)$ となることを示した.

また, $b_1 = b_2 = \frac{1}{2}$ $(\delta = (\mu_1 + \mu_2)/2)$ のときは, $A_d = a_0 e^{-a_0} \sum_{i=1}^{2} \left(\eta_i - \frac{1}{3}a_0\rho_i^{-1}\right) \sigma_i^{-1}$ となる. ここで, $\alpha = 0.05$ のとき, $a_0 \fallingdotseq 4.74386$ である.

参考文献

Aoshima, M., Aoki, M. (2000). Two-stage procedure having exact consistency and second-order properties for the s best selection. *Sequential Anal.* **19**, 115-131.

Aoshima, M., Takada, Y. (2000). Second-order properties of a two-stage procedure for comparing several treatments with a control. *J. Japan Statist. Soc.* **30**, 27-41.

Ghurye, S.G. (1958). Note on sufficient statistics and two-stage procedures. *Ann. Math. Statist.* **29**, 155–166.

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* **9**, 1229-1238.

Holm, S. (1995). Confidence sets of fixed size with predetermined confidence level. *Commun. Statist.–Theory Meth.* **24**, 1521–1536.

Isogai, E., Futschik, A. (2010). Sequential estimation oa a linear function of location parameters of two negative exponential distributions. *J. Statist. Plann. Inference* **140**, 2416–2424.

Isogai, E., Kobayashi, K., Uno, C. (2011). Higher order approximations by a two-stage procedures for a negative exponential distribution. *J. Statist. Plann. Inference* **141**, 3304–3312.

Isogai, E., Uno, C. (2017). Three-stage confidence intervals for a linear combination of locations of two negative exponential distributions, submitted.

Mukhopadhyay N (1990). Some properties of a three-stage procedure with applications in sequential analysis. *Sankhyā* **A52**, 218–231.

Mukhopadhyay, N., Duggan, W. (1999). On a two-stage procedure having second-order properties with applications. *Ann. Inst. Statist. Math.* **51**, 621–636.

Mukhopadhyay, N., Hamdy, H.I. (1984). On estimating the difference of location parameters of two negative exponential distributions. *Canadian J. Statist.* **12**, 67–76.

Mukhopadhyay, N., Mauromoustakos, A. (1987). Three-stage estimation procedures for the negative exponential distributions. *Metrika* **34**, 83–93.

Mukhopadhyay, N., Padmanabhan, A.R. (1993). A note on three-stage confidence intervals for the difference of locations: the exponential case. *Metrika* **40**, 121–128.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16**, 243-258.

Swanepoel, J.W.H., van Wyk, J.W.J. (1982). Fixed width confidence intervals for the location parameter of an exponential distribution. *Commun. Statist.* **A11**, 1279–1289.

Takada, Y. (2006). Asymptotic second-order efficiency of three-stage procedure with a warranted confidence level. *Metrika* **63**, 19–31.