数理解析研究所講究録
第 1804 巻 2012 年 65-69

二段階法による正規分布の分散の最小リスク問題

新潟大学・自然科学　磯貝　英一 (Eiichi Isogai)
Graduate School of Science and Technology,
Niigata University

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

1. 問題

$X_1, X_2, X_3, \ldots$ は独立で次の同一の正規分布に従う確率変数列とする。

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (-\infty < x < \infty) \tag{1.1}$$

ただし，母平均 $\mu \in (-\infty, \infty)$ と母分散 $\sigma^2 \in (0, \infty)$ はともに未知である。ここでは，母分散 σ^2 の点推定問題を考える。大きさ $n\,(\geq 2)$ の無作為標本 $X_1, X_2, \ldots X_n$ に対して，

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{および} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

とおき，標本分散 S_n^2 により母分散 σ^2 を推定する。このとき，損失関数として

$$L_n(S_n^2; \sigma^2) = (S_n^2 - \sigma^2)^2 + cn, \quad c > 0$$

を考える。ここで，c は既知の正の定数で，単位標本当たりの観測費用を表わす。リスクは

$$R_n(c) = E\{L_n(S_n^2; \sigma^2)\} = 2\sigma^4/(n-1) + cn$$

で与える。リスクをできるだけ小さくするように母分散 σ^2 を推定する問題が最小リスク問題である。上のリスクを最小にする標本の大きさ n の値は

$$n = n^* = n_0 + 1, \quad \text{ここで，} n_0 = \sqrt{2}c^{-1/2}\sigma^2 \tag{1.2}$$

となる。簡単のために n^* を整数とすると，最小リスクは $R_{n^*}(c) = 2cn_0 + c$ となる。σ^2 は未知であるため，(1.2) で与えられる最適標本数 n^* は未知であり，実際には利用できない。そこで，逐次手法を用いてこの問題を考える。

本研究では，過去の経験等から

$$\sigma^2 > \sigma_L^2 \tag{1.3}$$

となる母分散 σ^2 の下界 $\sigma_L^2\,(>0)$ が既知であると仮定して，二段階法を用いる。Mukhopadhyay and Duggan (1997) は正規分布の母平均 μ の逐次推定問題を考え，(1.3) の仮定のもとで観測費用 c が十分小さいとき，二段階法の 2 次漸近有効性を示した。Aoshima and Takada (2000) はこの 2 次漸近有効性を精密化した。Uno and Isogai (2002) はべき

σ^r $(r \neq 0)$ の最小リスク問題を考えた。Mukhopadhyay and Duggan (2001) は指数分布の尺度母数 λ の最小リスク問題に対して，$\lambda > \lambda_L\,(>0)$ となる λ の下界 $\lambda_L\,(>0)$ が既知であることを仮定して二段階法を扱った。本研究では，正規分布 (1.1) の母分散の最小リスク問題に対する二段階法を提案し，観測費用 c が十分小さいとき，リスクの漸近展開を与える。

2. 二段階法

(1.2) の n^* に基づいて，次のような二段階法を定義する。まず，初期標本数として次を与える。

$$m \equiv m(c) = \max\left\{m_0,\ \left[\sqrt{2}c^{-1/2}\sigma_L^2\right]^* + 1\right\} \tag{2.1}$$

ここで，$m_0\,(\geq 2)$ はあらかじめ与えた整数，$[x]^*$ は x より小さい最大整数である。初期標本 $X_1, \ldots, X_m$ に基づいて標本分散 S_m^2 を計算し，

$$N \equiv N(c) = \max\left\{m,\ \left[\sqrt{2}c^{-1/2}S_m^2\right]^* + 1\right\} \quad \text{および} \quad N = M + 1 \tag{2.2}$$

を求める。次に，二段階目の標本抽出を行い，$X_{m+1}, \cdots, X_N$ を得る。この手法では必ず二段階目に少なくとも 1 つの観測を行うのが特徴である。すべての標本 $X_1, \cdots, X_N$ を用いて S_N^2 によって母分散 σ^2 を推定する。このとき，リスクは

$$E\{L_N(S_N^2;\sigma^2)\} = E\{(S_N^2 - \sigma^2)^2 + cN\}$$

で与えられる。最小リスク問題における逐次手法の良さの基準として，次のリグレットを用いる。

$$\omega(c) = E\{L_N(S_N^2;\sigma^2)\} - R_{n^*}(c)$$

リグレットの値が小さいほど，逐次手法は良いと見なすことができる。

3. 主結果

(1.3) を仮定する。

$$T = \sqrt{2}c^{-1/2}S_m^2, \qquad U_c = [T]^* + 1 - T$$

とおき，

$$\Delta_c = -n_0^{-1}E\left[(U_c - \tfrac{1}{2})\{W_{m-1} - (m-1)\}^2\right]$$

とする。ここで，$W_{m-1} = (m-1)S_m^2/\sigma^2$ は自由度 $m-1$ のカイ二乗分布に従う。

Uno and Isogai (2011) により，二段階法 (2.1)，(2.2) は 2 次漸近有効性をもつことが分かる。

定理 3.1. （2 次漸近有効性）$c \to 0$ のとき，

$$E(N - n^*) = \frac{1}{2} + O(c^{1/2}).$$

さて，リグレットの２次漸近展開は次の定理で与えられる。

定理 3.2. $c \to 0$ のとき，

$$\omega(c)/c = 4 + 2(\sigma^2/\sigma_L^2) + \Delta_c + o(1),$$

ここで，

$$|\Delta_c| \le (\sigma_L^2/\sigma^2) + O(c^{1/2})$$

である。

注. (i) 定理 1 より，

$$\omega(c)/c > 5 + o(1) \quad (c \to 0).$$

(ii) Starr and Woodroofe (1972) and Woodroofe (1977) は次の純逐次法を考えた。ただし，仮定 (1.3) は必要としない。

$$N^* = \inf\left\{n \ge m:\ n \ge \sqrt{2}c^{-1/2}S_n^2 + 1\right\}$$

このとき，$c \to 0$ として，

$$[E\{L_{N^*}(S_{N^*}^2;\sigma^2)\} - R_{n^*}(c)]/c = 6 + o(1)$$

が成り立つことを示した。定理 3.2 から，もし $2(\sigma^2/\sigma_L^2) + \Delta_c > 2$ ならば，リスクを減少させるという観点から，純逐次推定方式 $S_{N^*}^2$ は二段階推定方式 S_N^2 より漸近的に優れていると言える。

S_N^2 のバイアスは次の定理で与えられる。

定理 3.3. $c \to 0$ のとき，

$$E(S_N^2) = \sigma^2 - \sqrt{2}c^{1/2} + O(c^{3/4}).$$

定理 3.3 を考慮して，次の偏り補正方式を提案する。

$$\delta_N = S_N^2 + \sqrt{2}c^{1/2} \tag{3.1}$$

偏り補正方式に対するリスクは $E\{L_N(\delta_N;\sigma^2)\} = E\{(\delta_N - \sigma^2)^2 + cN\}$ で与えられる。二段階法 (2.1)，(2.2) と偏り補正方式（3.1）に対するリスクの比較を行う。次の定理から，偏り補正方式 δ_N に対するリスクは二段階推定方式 S_N^2 に対するリスクより２つ分の単位標本当たりの観測費用が少ないことが分かる。従って，偏り補正はリスクを減少させるという意味で効果がある。

定理 3.4. $c \to 0$ のとき，

$$E\{L_N(S_N^2;\sigma^2)\} - E\{L_N(\delta_N;\sigma^2)\} = 2c + O(c^{5/4}).$$

4. シミュレーション結果

$N(0,1)$ を考える。従って，$\sigma^2 = 1$ を推定する。(2.1) で $m_0 = 2$ とし，$n_0 = 50$, すなわち，$c = 0.0008$，$n^* = 51$ に取って，$\sigma_L^2 = 0.2$, 0.4 に対して (2.1)，(2.2) の二段階法を１００万回の繰り返しを行い，その平均を取った。表 1，表 2 において，$\bar{N}$, $\bar{S}_N^2$, $\bar{\delta}_N$, $\bar{\Delta}_c$, $\bar{\omega}_1/c$, $\bar{\omega}_2/c$ はそれぞれ，N, S_N^2, δ_N, $-n_0^{-1}(U_c - 2^{-1})\{W_{m-1} - (m-1)\}^2$, $L_1(c)/c$, $L_2(c)/c$ の平均である。ただし，

$$L_1(c) = L_N(S_N^2; \sigma^2) - R_{n^*}(c) \quad \text{and} \quad L_2(c) = L_N(\delta_N; \sigma^2) - R_{n^*}(c).$$

$s(\cdot)$ は推定量 $(\cdot)$ の標準誤差を表わす。

これらの表から，下界 σ_L^2 が増加するとき，単位標本当たりの観測費用に対するリグレット $\omega(c)/c$ は減少することが分かる。また，リスクを減少させるには偏り補正が有効であることも分かる。従って，シミュレーション結果により定理の正当性が確認できる。

表 1. $\sigma_L^2 = 0.2$, $\sigma^2 = 1$, $n^* = 51$

$\bar{N}$	$\bar{S}_N^2$	$\bar{\delta}_N$	$\bar{\Delta}_c$	$\bar{\omega}_1/c$	$\bar{\omega}_2/c$
51.524682	0.951123	0.991123	-0.000394	18.930961	16.043268
	bias of $\bar{S}_N^2$	bias of $\bar{\delta}_N$			
	-0.048877	-0.008877			
$s(\bar{N})$	$s(\bar{S}_N^2)$		$s(\bar{\Delta}_c)$	$s(\bar{\omega}_1/c)$	$s(\bar{\omega}_2/c)$
0.023586	0.000229		0.000217	0.109879	0.104002

表 2. $\sigma_L^2 = 0.4$, $\sigma^2 = 1$, $n^* = 51$

$\bar{N}$	$\bar{S}_N^2$	$\bar{\delta}_N$	$\bar{\Delta}_c$	$\bar{\omega}_1/c$	$\bar{\omega}_2/c$
51.515002	0.957132	0.997132	-0.000247	9.127811	6.841029
	bias of $\bar{S}_N^2$	bias of $\bar{\delta}_N$			
	-0.042868	-0.002868			
$s(\bar{N})$	$s(\bar{S}_N^2)$		$s(\bar{\Delta}_c)$	$s(\bar{\omega}_1/c)$	$s(\bar{\omega}_2/c)$
0.016155	0.000212		0.000417	0.084408	0.081057

5. 謝辞

本研究は科学研究費の補助金，磯貝英一（23540128），宇野力（20540102）により行われた。

参考文献

AOSHIMA, M. and 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.

MUKHOPADHYAY, N. and DUGGAN, W. T. (1997). Can a two-stage procedure enjoy second-order properties?, *Sankhyā, Series A*, 59, 435–448.

MUKHOPADHYAY, N. and DUGGAN, W. T. (2001). A two-stage point estimation procedure for the mean of an exponential distribution and second-order results, *Statist. Decisions*, 19, 155–171.

STARR, N. and WOODROOFE, M. (1972) Further remarks on sequential estimation: The exponential case, *Ann. Math. Statist.*, 43, 1147–1154.

UNO, C. and ISOGAI, E. (2002) Sequential point estimation of the powers of a normal scale parameter, *Metrika*, 55, 215–232.

UNO, C. and ISOGAI, E. (2011) Asymptotic higher-order properties for a two-stage procedure and their applications, *submitted.*

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