数理解析研究所講究録
723巻 1990年 87-97

事前情報に基づく統計的推測方式

鹿児島大・理　稲田浩一（ Koichi Inada ）

1. はじめに

統計的推測における初期の方法は、主として一部の母数を除いて完全に規定された数学模型で表わされた母集団からの無作為観測データに基づき母集団の母数についての情報を得ることであり、そこに科学的・客観的推論として大きな意義を有している。その後、母集団についての既存の事前情報を利用する推測方式も研究され、いわゆる Bayesian 学派の誕生をみた。そこでは、既存の知識を母数の事前分布として導入し、これを標本の提供する情報に組み合わせる推測方式が研究されている。このように標本から得られる情報に加えて母数についての事前情報を組み合わせる推測方式は自然な着想と思われる。

本研究では、データと既存の事前情報を組み合わせる推測方式として Bayes的な立場と異なる統計的推測方式を提唱し、その推測性質を調べ、さらに他の推測方式との比較をする。

すなわち、正規母集団の平均の推測に関し、平均がある区間内にあるという事前情報のもとで、データとこの事前情報を結合した推測方式を提唱し、予備検定推定量・Thompson(1968)の推定量との比較をする。また、正規母集団の分散の推測に関しても、同様の状況のもとでの推測方式を提唱し、予備検定推定量との比較をする。ここでは、推測方式の比較基準として平均2乗誤差を用いる。

2. 事前情報に基づく平均の推定

分散σ^2が既知の正規母集団 $N(\mu, \sigma^2)$からの大きさnの無作為標本 $X_1, X_2, \cdots, X_n$を用いて、平均μの推定問題を考える。通常、推定量として標本平均$\bar{X}$を用いるが、平均μに関してある種の事前情報がある場合は、その事前情報を利用すべきである。例えば、平均μは0に近いという事前情報がある場合は、この事前情報を利用した推定量としては予備検定推定量 PT が考えられる。

<u>予備検定推定量 PT</u>

$$PT = \begin{cases} 0 & \text{if} \quad |\bar{X}| / \sigma_n < C \\ \bar{X} & \text{if} \quad |\bar{X}| / \sigma_n \geq C \end{cases}$$

ここでは、平均μが0を含む区間$[-\mu_0, \mu_0]$にあるという事前情報のもとでの平均μの推定問題について論じる。

Thompson(1968)はこのような状況のもとで平均μの推定量THを与えている。

<u>Thompsonの推定量 TH</u>

$$TH = \bar{X} + \frac{\sigma_n^2}{4\mu_0} \log \left[\frac{(\bar{X} - \mu_0)^2 + \sigma_n^2}{(\bar{X} + \mu_0)^2 + \sigma_n^2} \right]$$

本研究では、3つの推定量 T1, T2, SE を提案し、平均2乗誤差(MSE)を用いて6つの推定量 T1, T2, SE, PT, TH, $\bar{X}$ の間の比較を行う。

<u>推定量 T1, T2, SE</u>

[Type I]

$$T(a) = \begin{cases} \bar{X} & \text{if } |\bar{X}| \leqq \mu_0 \\ \bar{X} + a\sigma_n & \text{if } \bar{X} < -\mu_0 \\ \bar{X} - a\sigma_n & \text{if } \bar{X} > \mu_0 \end{cases} \qquad (a \geqq 0)$$

[Type II]

$$MT(B) = \begin{cases} \bar{X} + b\sigma_n & \text{if } \bar{X} \leqq 0 \\ \bar{X} - b\sigma_n & \text{if } \bar{X} > 0 \end{cases} \qquad (b \geqq 0)$$

[Type III]

$$SE(w) = w\bar{X} \qquad (0 \leqq w \leqq 1)$$

推定量 T(a), MT(b), SE(w) の平均2乗誤差は、次式で与えられる。

$$MSE[T(a)] = \sigma_n^2[a^2\{1-\Phi(\theta_\circ-\theta)+\Phi(-\theta_\circ-\theta)\}$$
$$-2a\{\phi(\theta_\circ-\theta)+\phi(-\theta_\circ-\theta)\}+1]$$

$$MSE[MT(b)] = \sigma_n^2\{b^2-4b\phi(\theta)+1\}$$

$$MSE[SE(w)] = \sigma_n^2\{(w-1)^2\theta^2+w^2\}$$

ただし、$\theta=\mu/\sigma_n, \theta_\circ=\mu_\circ/\sigma_n$ で Φ と ϕ は標準正規分布の c.d.f. と p.d.f. である。

a_1, b_1, w_1の値は、それぞれ次のような Minimax基準によって定め、$T1=T(a_1)$, $T2=MT(b_1)$, $SE=SE(w_1)$ とする。

$$MSE[T(a_1)] = \inf_{a \geqq 0} \sup_{\mu \in [-\mu_\circ, \mu_\circ]} MSE[T(a)]$$

$$MSE[MT(b_1)] = \inf_{b \geqq 0} \sup_{\mu \in [-\mu_\circ, \mu_\circ]} MSE[MT(b)]$$

$$MSE[SE(w_1)] = \inf_{0 \leqq w \leqq 1} \sup_{\mu \in [-\mu_\circ, \mu_\circ]} MSE[SE(w)]$$

b_1 と w_1 は、次式で与えられるが、a_1 については簡単な形で求めることはできないが存在性は示される。

$$b_1 = 2\phi(\theta_\circ)$$

$$w_1 = \frac{\theta_\circ^2}{1+\theta_\circ^2}$$

いくつかの$\theta_{\circ}$に対する a_1, b_1, w_1の値は表1で与えられる。

表1. a_1, b_1, w_1 の値

$\theta_{\circ}$	a_1	b_1	w_1
0.0	.798	.798	.000
0.5	.972	.704	.200
1.0	.866	.484	.500
1.5	.804	.259	.692
2.0	1.281	.108	.800
2.5	1.492	.035	.862
3.0	1.568	.009	.900

$\theta_{\circ}$= 0.5, 1.0, 1.5 について、推定量の MSE/σ_n^2 のグラフを描き、推定量T1, T2, SE, 予備検定推定量PT(α=0.05) と Thompsonの推定量TH の間の比較を行う。

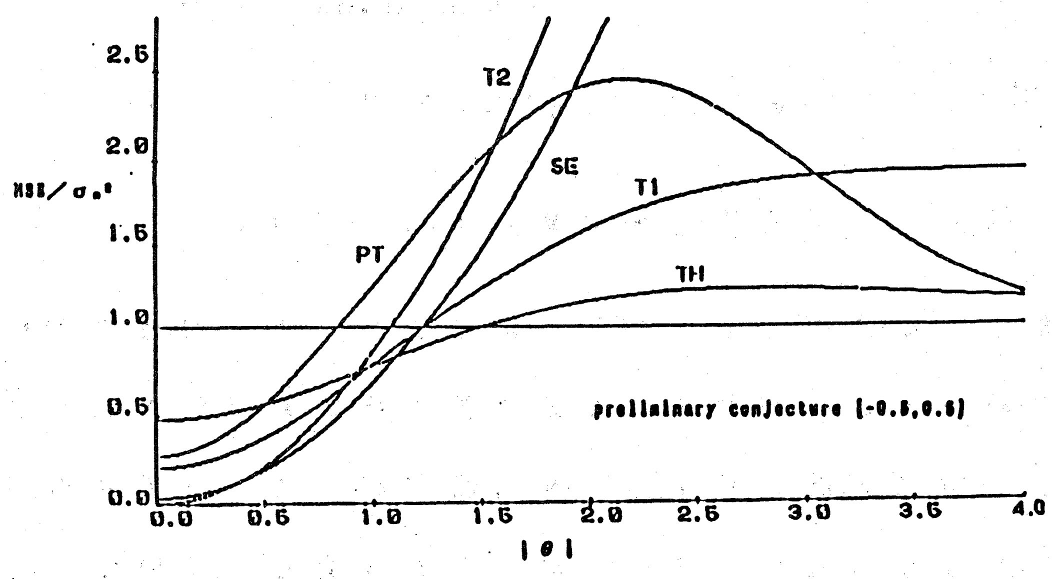

図1. $\bar{X}$ に対する推定量の平均2乗誤差

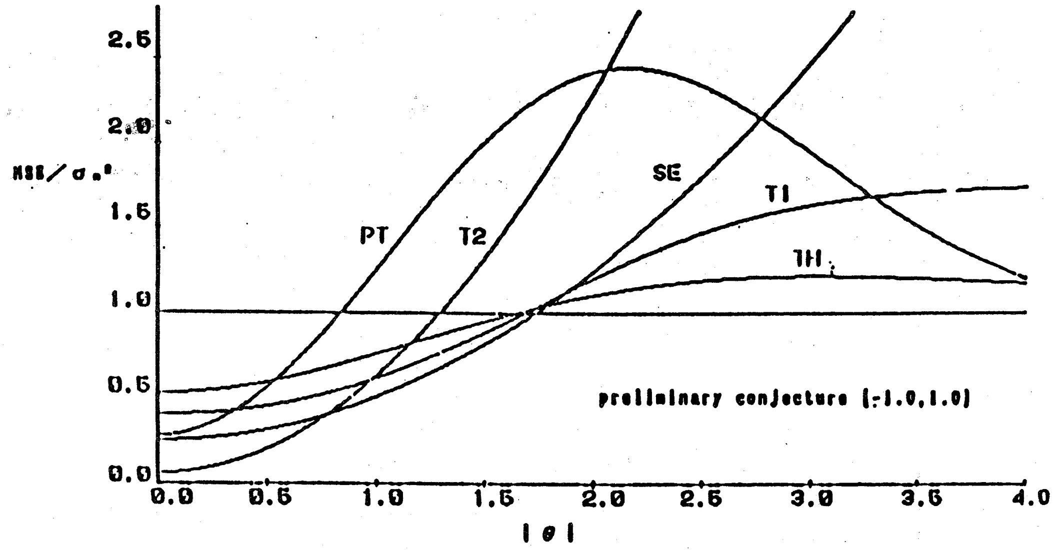

図 2.　$\bar{X}$ に対する推定量の平均２乗誤差

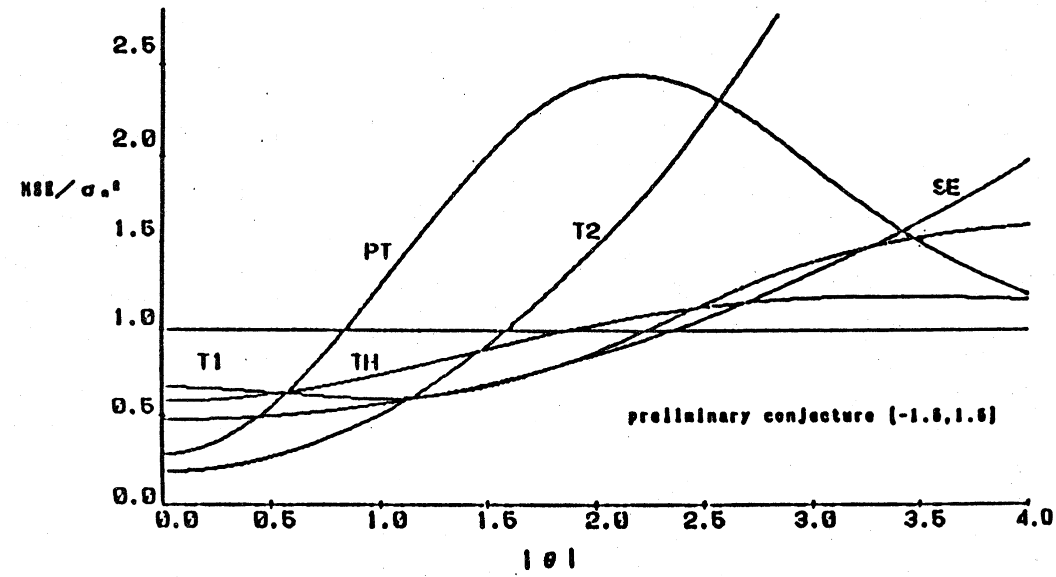

図 3.　$\bar{X}$ に対する推定量の平均２乗誤差

推定量の比較に際して、MSE/σ_n^2をより小さく同時にMSE/$\sigma_n^2<1$を満たすθの範囲をより広く持つ推定量が望ましいと思われる。更に区間に関する事前情報が誤っていた場合も考慮しなければならない。これらのことを考慮して、図１，２，３より著者は推定量T1を推奨する。

3. 事前情報に基づく分散の推定

正規母集団 $N(\mu, \sigma^2)$ からの大きさnの無作為標本 $X_1, X_2, \cdots, X_n$ を用いて、分散 σ^2の推定問題を考える。通常、推定量として不偏分散推定量 $U = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$ を用いるが、分散 σ^2に関してある種の事前情報がある場合は、その事前情報を利用すべきである。例えば、分散 σ^2は既知の値に σ_0^2に近いという事前情報がある場合は、この事前情報を利用した推定量としては予備検定推定量PTが考えられる。

<u>予備検定推定量 PT</u>

$$PT = \begin{cases} \sigma_0^2 & \text{if } \chi^2_{n-1}(1-\alpha/2)/(n-1) < U/\sigma_0^2 < \chi^2_{n-1}(\alpha/2)/(n-1) \\ U & \text{otherwise} \end{cases}$$

ここでは、分散 σ^2は既知の値 σ_0^2を含む区間 $[\sigma_0^2/C_0, \sigma_0^2 C_0]$にあるという事前情報のもとでの分散 σ^2の推定問題について論じる。 2つの推定量 T1, T2 を提案し、平均2乗誤差(MSE)を用いて4つの推定量 T1, T2, PT, U の間の比較を行う。

<u>推定量 T1, T2, SE</u>

[Type I]

$$T(w, C) = \begin{cases} \sigma_0^2 & \text{if } C^{-1} < U/\sigma_0^2 < C \\ wU & \text{if } U/\sigma_0^2 \geqq C \qquad (0 < w \leqq 1, C > 0) \\ w^{-1}U & \text{if } U/\sigma_0^2 \leqq C^{-1} \end{cases}$$

[Type II]

$$T(w) = \begin{cases} wU & \text{if} \quad U/\sigma_o^2 > 1 \\ w^{-1}U & \text{if} \quad U/\sigma_o^2 \leqq 1 \end{cases} \qquad (0 < w \leqq 1)$$

推定量 T(w,C),T(w)の平均2乗誤差は、次式で与えられる。

$$\begin{aligned} MSE[T(w,C)] = {} & \sigma^4[\{w^{-2}I_a(\beta_1)+w^2(1-I_b(\beta_1))\}(n+1)(n-1) \\ & -2\{w^{-1}I_a(\beta_2)+w(1-I_b(\beta_2))\}+1] \\ & +(2\sigma_o^2\sigma^2-\sigma_o^4)\{I_a(\beta_3)-I_b(\beta_3)\} \end{aligned}$$

$$\begin{aligned} MSE[T(w)] = {} & \sigma^4[\{w^{-2}I_d(\beta_1)+w^2(1-I_d(\beta_1))\}(n+1)(n-1) \\ & -2\{w^{-1}I_d(\beta_2)+w(1-I_d(\beta_2))\}+1] \end{aligned}$$

ただし、$I_t(\beta)$は不完全ガンマ関数比で $\beta_1=(n+3)/2$, $\beta_2=(n+1)/2$, $\beta_3=(n-1)/2$, $a=(n-1)C^{-1}\sigma_o^2/(2\sigma^2)$, $b=(n-1)C\sigma_o^2/(2\sigma^2)$, $d=(n-1)\sigma_o^2/(2\sigma^2)$である。

w_1, C_1, w_2の値は、各々次のようなMinimax基準によって定め、$T1=T(w_1,C_1)$, $T2=T(w_2)$とする。

$$MSE[T(w_1,C_1)] = \inf_{0<w\leqq 1,\, C>0} \ \sup_{\sigma^2 \varepsilon [\sigma_o^2/C_o,\ \sigma_o^2 C_o]} MSE[T(w,C)]$$

$$MSE[T(w_2)] = \inf_{0<w\leqq 1} \ \sup_{\sigma^2 \varepsilon [\sigma_o^2/C_o,\ \sigma_o^2 C_o]} MSE[T(w)]$$

いくつかのnとC_oに対する w_1, C_1, w_2の値は、表2で与えられる。

表 2. w_1, C_1, w_2 の値

n	C_0	w_1	C_1	w_2	n	C_0	w_1	C_1	w_2
7	1.5	.422	2.37	.618	30	1.5	.880	1.14	.893
	2.0	.626	1.60	.665		2.0	.929	1.08	.930
	2.5	.686	1.46	.701		2.5	.935	1.07	.935
10	1.5	.591	1.69	.698	50	1.5	.942	1.06	.944
	2.0	.739	1.35	.760		2.0	.960	1.04	.960
	2.5	.783	1.28	.791		2.5	.961	1.04	.961
20	1.5	.801	1.25	.833	100	1.5	.978	1.02	.978
	2.0	.884	1.13	.887		2.0	.980	1.01	.980
	2.5	.900	1.11	.900		2.5	.980	1.01	.980

C_0= 1.5, 2.0, 2.5 について、推定量 T1, T2, 予備検定推定量 PT(α =0.05) の間の比較を、次のようなUに対する効率(efficiency)を用いて行う。

$$e(T1) = MSE(U) / MSE(T1)$$
$$e(T2) = MSE(U) / MSE(T2)$$
$$e(PT) = MSE(U) / MSE(PT)$$

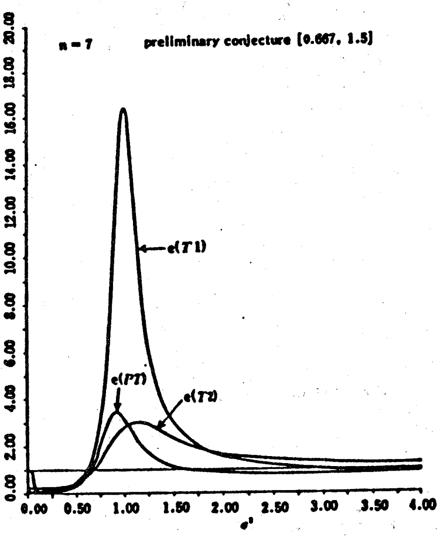

図 4. U に対する推定量の効率

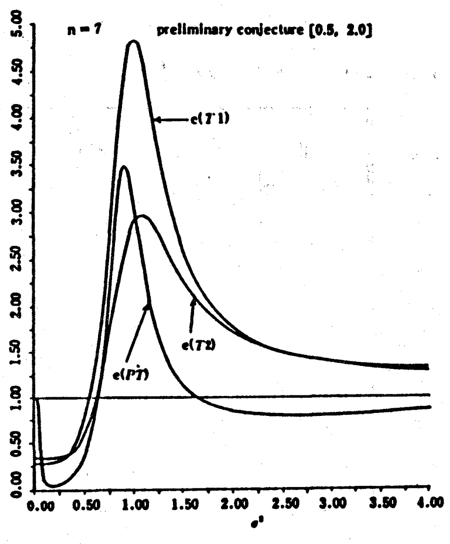

図 5.　U に対する推定量の効率

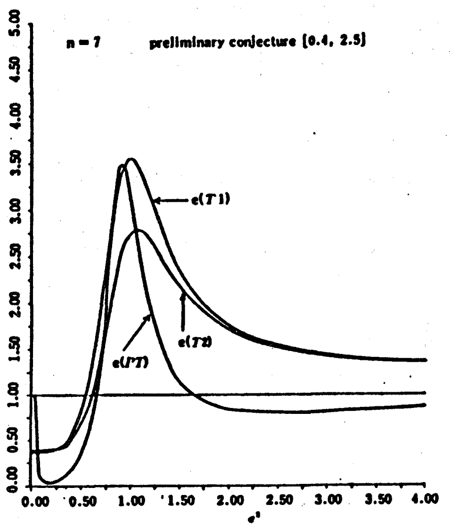

図 6.　U に対する推定量の効率

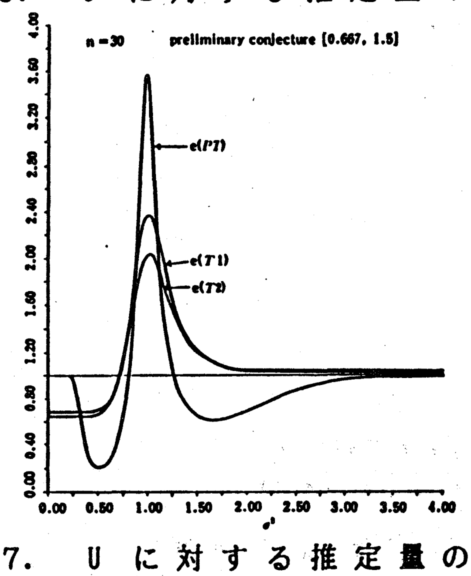

図 7.　U に対する推定量の効率

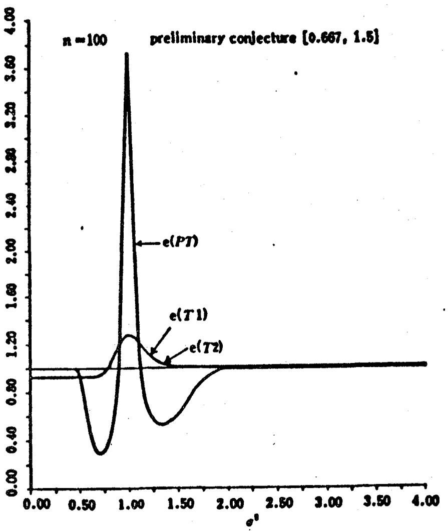

図 8. U に対する推定量の効率

推定量の比較に際して、効率e(T)をより大きく同時にe(T)>1を満たすθの範囲をより広く持つ推定量が望ましいと思われる。更に区間に関する事前情報が誤っていた場合も考慮しなければならない。これらのことを考慮して、図4,5,6,7,8より著者は推定量T1を推奨する。

参考文献

[1] Inada, K.(1989). Estimation of variance after preliminary conjecture. Bulletin of Informatics and Cybernetics,Vol.23, No.3～4,183-198.

[2] 北川敏男(1984). 「推測過程論」. 共立出版.

[3] Thompson, J.R.(1968). Accuracy borrowing in the estimation of the mean by shrinkage to an interval. J. Am. Statist. Assoc. 63, 953-963.