最大確率推定量の２次の有効性

高　橋　倫　也

最大確率推定量の2次の有効性

0. 序

この論文では，実確率変数の列 X_1, X_2, $\cdots$, X_n, ……が互いに独立に未知の母数 θ を含む同一の分布に従い，その密度関数が $f(x|\theta)$ で与えられる場合の推定問題について考える。

正則条件の成り立つ場合には，最尤推定量 (m.l.e.) が漸近的に有効推定量になることが Cramér〔1〕等によって示されている。

しかし，実際には正則条件の成り立たない場合（例えば，指数分布）もしばしばある。そのようなときには m.l.e. は必ずしも有効推定量にはなりえない。

そこで，Weiss と Wolfowitz〔2〕は，有効性というものから検討しなおし，損失関数に依る最大確率推定量 maximum probability estimator (m.p.e.) を提示した。そうして，彼らは，正則，非正則を問わず m.p.e. はある推定量のクラスの中で漸近的にリスクを最小にする（W－有効）ことと，正則な場合 m.l.e. を特殊な場合として含むことを示した（2節を見よ)。ところがW－有効な推定量がいろいろ存在する（例 3.2 を見よ）ことがある。

一方，正則条件の成り立つ場合には，有効推定量（BAN estimator とよぶ）のクラスに対して Rao〔3〕は2次の有効性の定義を与えている。

そこで我々は，非正則な場合の2次の有効性の1つの定義与えてW－有効である推定量の優劣の比較ができることを議論しよう（3節)。

1. 漸近分布を求める方法

非正則の場合，m.p.e. は $\min_{1\leq i\leq n} X_i$ や $\max_{1\leq i\leq n} X_i$ の関数として表わされる（このことは3節で見る）ことが多い。そこで，$\min X_i$ の分布の計算について調べてみよう。

次のような密度関数 $f(x)$ を考える。

$$
(1.1)\quad
\begin{aligned}
&f(x)=0 && x<0\\
&f(0)=c<0\\
&f(x)\geqq 0 && x>0\\
&f_+'(0)\ \text{が存在する。}
\end{aligned}
$$

X_i が (1.1) を満たす密度関数 $f(x)$ を持つならば，直接計算により $n\min_{1\leqq i\leqq n} X_i$ は漸近的に指数分布に従うことがわかる。

ここで，関数の正則変動性が基本的に分布を決定することに着目すれば，議論は簡明になる。従って，上述のことを De Haan〔4〕にそって示そう。

補題　$f(x)$ は (1.1) を満たしている密度関数とする。このとき

$$F(x)=\int_0^{x^{-1}} f(y)dy \qquad x>0$$

は指数 -1 の正則変動関数である。

証明
$$\lim_{t\to+0}\frac{\int_0^t f(y)dy-f(0)t}{t^2}=\lim_{t\to+0}\frac{f(t)-f(0)}{2t}=\frac{1}{2}f_+'(0)$$

であるから，

$$\int_0^t f(y)dy=f(0)t+\frac{1}{2}f_+'(0)t^2+o(1)t^2$$

となる。よって

$$F(x)=\int_0^{x^{-1}} f(y)dy=f(0)x^{-1}+\frac{1}{2}f_+'(0)x^{-2}+o(1)x^{-2}$$

となり

$$\lim_{s\to\infty}\frac{F(xs)}{F(s)}=\lim_{s\to\infty}\frac{f(0)(xs)^{-1}+\frac{1}{2}f_+'(0)(xs)^{-2}+o(1)(xs)^{-2}}{f(0)s^{-1}+\frac{1}{2}f_+'(0)s^{-2}+o(1)s^{-2}}=x^{-1}$$

を得る。すなわち，$F(x)$ は指数 -1 の正則変動関数である。

ここで，Y_1, Y_2, …, Y_n, ……も (1.1) を満たす密度関数 $g(x)$ を持つ確率変数とすると，

(1.2)　$$\lim_{n\to\infty}P\{n\min X_i<x\}=\lim_{n\to\infty}P\{n\min Y_i<x\}=1-e^{-cx} \qquad x>0$$

となる。ところが次の定理が成り立つ。

定理　もし $f_+'(0)>g_+'(0)$ ならば，$\min X_i$ の方が $\min Y_i$ よりも次の意味で収速が早い。

任意の $x>0$ に対して，n が十分大きければ，

(1.3)　$$P\{\min X_i<x/n\}>P\{\min Y_i<x/n\}$$

すなわち

(1.4) $P\{n \min X_i \in (0, x)\} > P\{n \min Y_i \in (0, x)\}$

が成り立つ。

このことは，補題と同様の計算によって簡単に示される。ところで，式 (1.4) は有限段階でW－有効性の比較が可能であることを示していることに注意しよう。

2. Weiss と Wolfowitz の結果

Weiss と Wolfowitz〔2〕の結果について簡単に述べる。$n \in N$ とし，$(P_{\theta,n})$ $\theta \in \Theta$ を標本空間 $\mathfrak{X}_n$ 上の確率分布族とする。ここに Θ（母数空間）と可能な決定の空間 $\overline{\Theta}$ とは R^m の連結閉部分集合とする。$f_n(\cdot|\theta)$ で，σ－有限な測度 μ_n に関する $P_{\theta,n}$ の密度を，$L_n(a, \theta)$ で $\theta \in \Theta$, $a \in \overline{\Theta}$ における非負な損失関数を表わすことにする。さらに，与えられた零列 $h_2(n) > 0$ に対して

$$s(n) = \sup_{\theta, a}\{L_n(a, \theta) : \|\theta - a\| \leqq h_2(n)\}$$

と定義する，ここで x, $y \in R^m$ に対して

$$\|x-y\| = \max_{1 \leqq i \leqq m}|x_i - y_i| \quad \text{または} \quad \|x-y\| = \{\sum_{i=1}^{m}(x_i - y_i)^2\}^{1/2}$$

であり，$s(n)$ は任意の n に対して有限とする。

次のような積分を考える。

$$I(d) = \int_{\{\|d-\theta\| \leqq h_2(n)\}} (s(n) - L_n(d, \theta)) f_n(\xi_n | \theta) d\theta \qquad \xi_n \in \mathfrak{X}_n$$

このとき，$Y_n(\xi_n)$ が m.p.e. であるとは，任意の $\xi_n \in \mathfrak{X}_n$ に対して，

$$I(Y_n) / \sup_d I(d) \geqq 1 - \ell_n \qquad \ell_n > 0$$

（ここに，ℓ_n は零列である）を満たす推定量であることを意味する。

$\theta_0 \in \Theta$ を任意に固定し，

$$H_n = \{\theta \in \overline{\Theta} : \|\theta - \theta_0\| \leqq h_1(n)\}$$

とおく。ここに $h_1(n) > 0$ は零列で，

$$\lim_{n \to \infty} h_2(n) / h_1(n) = 0$$

を満たすものである。θ の関数 $F_n(\theta)$ が H_n で一様に $F(\theta)$ に収来するとは，

$$\lim_{n \to \infty} [\sup_{H_n} |F_n(\theta) - F(\theta)|] = 0$$

を意味するものとする。

このとき，次の三つの定理が示される。

定理2.1　Y_n は次の条件を満たす推定量とする。

H_n で一様に

(2.1)　$\lim_{n\to\infty} E(L_n(Y_n,\theta)\mid\theta)=\beta(\theta_0)$

(2.2)　$\lim_{n\to\infty} s(n)P_n(\|Y_n-\theta\|>h_2(n)\mid\theta)=0$

(2.3)　$\lim_{n\to\infty}\int_{B_n(\theta)} L_n(Y_n(\xi_n),\theta)f_n(\xi_n\mid\theta)d\mu_n(\xi_n)=0$

となる。ここに

(2.4)　$B_n(\theta)=\{\xi_n\in\Upsilon_n:\|Y_n(\xi_n)-\theta\|>h_2(n)\}$

である。もし T_n が次の2つの条件 H_n で一様に

(2.5)　$\lim_{n\to\infty}[E\{L_n(T_n,\theta)\mid\theta\}-E\{L_n(T_n,\theta_0)\mid\theta_0\}]=0$

(2.6)　$\lim_{n\to\infty} s(n)P_n(\|T_n-\theta\|>h_2(n)\mid\theta)=0$

を満たす推定量ならば

(2.7)　$\beta(\theta_0)\leqq\underline{\lim}_{n\to\infty} E(L_n(T_n,\theta_0)\mid\theta_0)$

が成り立つ。

さらに，損失関数に制限をつけると，次の2つの定理が成り立つ。

定理2.2　すべての $n\geqq n_0$ について，$L_n(a,\theta)$ が $\|a-\theta\|$ の単純増加関数ならば，定理2.1は条件 (2.6) がなくても成り立つ。

定理2.3　すべての $n\geqq n_0$ について，$\|a-\theta\|\geqq h_2(n)$ ならば，$L_n(a,\theta)=s(n)$ のとき定理 2.1 は条件 (2.6) がなくても成り立つ。

本稿で考える損失関数は $L_n(a,\theta)=\mathrm{X}\{\|a-\theta\|>r/h_1(n)\}$ に限定する。すなわち定理2.3が成り立つ場合である。

3.　最大確率推定量の2次の有効性

定理 2.1 より m.p.e. ならば漸近的に良い推定量になることがわかる。しかし，W－有効であるような推定量は必ずしも1つではない。このとき，どのような規準によって m.p.e.に優劣をつけるべきか，という問題が生じる。以下，このことを例をあげて考えていこう。

次のような密度関数を考える。

$$f(x|\theta)=f(x-\theta)$$
$$f(x)>0 \qquad a\leqq x\leqq b$$
$$\quad =0 \qquad x<a,\ x>b$$
$f_+'(a)$ と $f_-'(b)$ が存在する

ここで $f(x|\theta)$ は〔2〕の5（4）（p.47）と同様の仮定を満たしているとする。このとき

$$E_{\theta_0}\left\{\frac{\partial}{\partial\theta}\log f(X|\theta_0)\right\}=C(\theta_0)$$

とおくと，

$$1=\int_{a+\theta}^{b+\theta} f(x|\theta)dx$$

であるから，両辺を微分して

$$C(\theta_0)=f(a)-f(b)$$

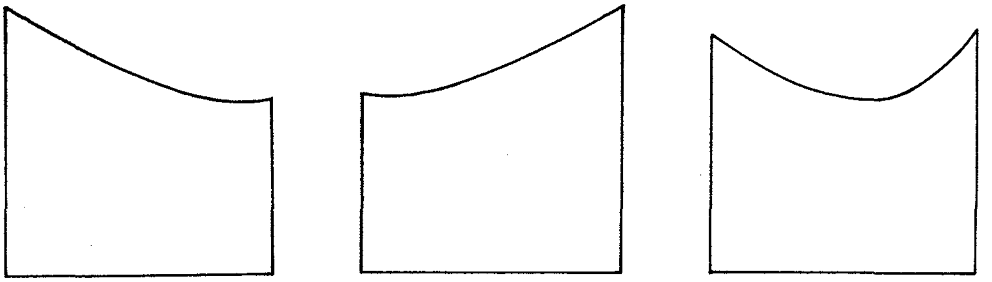
が得られる。このとき，$R=(-r,\ r)$ $(r>0)$ に関する m.p.e. を Z_n とすると〔2〕と同様にして，

(1) $f(a)>f(b)$ ならば $Z_n=W_n-r/n$

(2) $f(a)<f(b)$ ならば $Z_n=V_n+r/n$

(3) $f(a)=f(b)$ ならば Z_nは〔m^*, m^{**}〕の任意の点である。ただし，Z_n は可測で $m^*=\min\{W_n-r/n,\ V_n+r/n\}$,　$m^{**}=\max\{W_n-r/n,\ V_n+r/n\}$ とする。ここに，$W_n=\min X_i-a$, $V_n=\max X_i-b$ である。

図1



(1) $f(a)>f(b)$　(2) $f(a)<f(b)$　(3) $f(a)=f(b)$ $(f'_+(a)+f'_-(b)>0)$

このとき問題となるのは(3)である。ここで reasonable な m.p.e. は，$Z_n^{(1)}=W_n-r/n$ と $Z_n^{(2)}=V_n+r/n$ の2つである。一般性を失なわずに，$\theta_0=0$ とできる。そこで，

$$\beta_n^{(i)}=P\{n\,Z_n^{(i)}\notin(-r,\ r)\} \qquad i=1,2$$

とおく。これは $L_n(a,\ \theta)=\chi\{\|a-\theta\|>r/n\}$ なる損失関数によるリスクである。補題と同様

の計算により，

$$\beta_n^{(1)}=1-P\{nZ_n^{(1)}\in(-r,\ r)\}$$

$$=\left\{1-\frac{2rf(a)+(2r^2f_+'(a)+o(1))\dfrac{1}{n}}{n}\right\}^n$$

$$=\left\{1-\frac{2rf(a)}{n}\right\}^n\left\{1-\frac{2r^2f_+'(a)+o(1)}{n^2(1-2rf(a)/n)}\right\}^n\longrightarrow\exp\{-2rf(a)\}$$

となる，同様にして

$$\beta_n^{(2)}=\left\{1-\frac{2rf(b)}{n}\right\}^n\left\{1+\frac{2r^2f_-'(b)+o(1)}{n^2(1-2rf(b)/n)}\right\}^n\longrightarrow\exp\{-2rf(b)\}$$

を得る。従って，次のような量が考えられる。

$$\left(\frac{\beta_n^{(2)}}{\beta_n^{(1)}}\right)^n=\frac{\left\{1+\dfrac{2r^2f_-'(b)+o(1)}{n^2(1-2rf(b)/n)}\right\}^{n^2}}{\left\{1-\dfrac{2r^2f_+'(a)+o(1)}{n^2(1-2rf(a)/n)}\right\}^{n^2}}\qquad(\because\ f(a)=f(b))$$

$$\longrightarrow\frac{\exp\{2r^2f_-'(b)\}}{\exp\{-2r^2f_+'(a)\}}=\exp\{2r^2(f_-'(b)+f_+'(a))\}$$

$$<1\qquad f_-'(b)+f_+'(a)<0\ \text{のとき}$$

$$>1\qquad f_-'(b)+f_+'(a)>0\ \text{のとき}$$

このことは n が十分大きいとき

$$\lim_{n\to\infty}(\beta_n^{(2)}/\beta_n^{(1)})<1\ \ \text{ならば}\ \ \beta_n^{(1)}>\beta_n^{(2)}$$

$$\lim_{n\to\infty}(\beta_n^{(2)}/\beta_n^{(1)})>1\ \ \text{ならば}\ \ \beta_n^{(1)}<\beta_n^{(2)}$$

を示している。

よって次の定義が考えられる。

定義　2つの m.p.e. の列 $\{Z_n^{(1)}\}$，$\{Z_n^{(2)}\}$ とそれらのリスクの列 $\{\beta_n^{(1)}(\theta)\}$，$\{\beta_n^{(2)}(\theta)\}$ とがあるとき，次のような2次の相対有効率

$$SE(1,\ 2\ ;\theta)=\lim_{n\to\infty}SE_n(1,\ 2\ ;\theta)\ =\lim_{n\to\infty}\left(\frac{\beta_n^{(2)}(\theta)}{\beta_n^{(1)}(\theta)}\right)^n$$

が存在し，任意の $\theta\in\Theta$ に対して

$$SE(1,\ 2\ ;\theta)\geqq1$$

が成立し，少くとも1つの $\theta'\in\Theta$ が存在して

$$SE(1,\ 2\ ;\theta')>1$$

ならば，$\{Z_n^{(1)}\}$ は $\{Z_n^{(2)}\}$ より2次の意味で有効である。

よって，上の例では

$f_-'(b)+f_+'(a)>0$ ならば $\{Z_n^{(1)}\}$ が2次の有効

$$f_-'(b)+f_+'(a)<0 \quad \text{ならば} \quad \{Z_n^{(2)}\} \text{ が2次の有効}$$

となる。

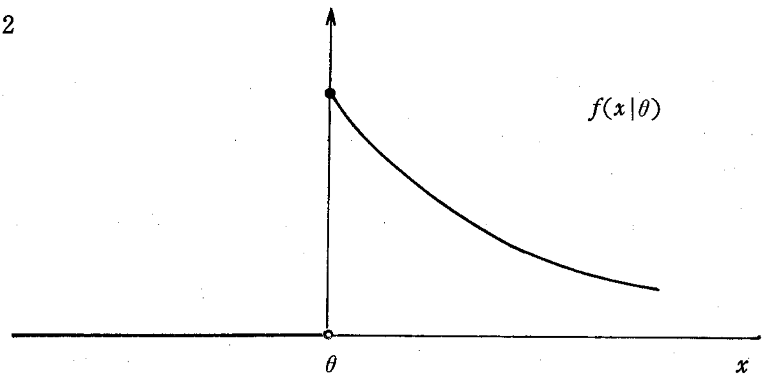
以下〔2〕の提示した例について考察していく。

例3.1〔2〕の5(2)(p.41)の例について，2次の有効性を考えるために次の条件を加える。

$f(\theta|\theta)=h(\theta)>0$

$f_+'(\theta|\theta)$ が存在する

図2



このとき，m.p.e. は $Z_n=\min X_i-n/r$ である。この場合，reasonable な m.p.e. で Z_n に対抗できるようなものは考えられない。しかし，密度関数によってリスクの収速がどうなるかみてみよう，密度関数が $f(x|\theta)$ の場合の Z_n のリスクを

$$\beta_n^{(f)}(\theta)=P_\theta\{n(Z_n-\theta)\notin(-r,\ r)\}$$

とおくと，

$$\beta_n^{(f)}(\theta)=\left\{1-\frac{2rh(\theta)+(2r^2f_+'(\theta|\theta)+o(1))\dfrac{1}{n}}{n}\right\}^n$$

$$=\left\{1-\frac{2rh(\theta)}{n}\right\}^n\left\{1-\frac{2r^2f_+'(\theta|\theta)+o(1)}{n^2(1-2rh(\theta)/n)}\right\}^n\longrightarrow\exp\{-2rh(\theta)\}$$

を得る。一方

$$f_e(x|\theta)=h(\theta)e^{-h(\theta)(x-\theta)} \qquad x\geqq\theta$$

$$=0 \qquad x<\theta$$

なる密度関数に対しては

$$\beta_n^{(e)}(\theta)=\beta_n^{(f_e)}(\theta)=\exp\{-2rh(\theta)\}=\left\{1-\frac{2rh(\theta)-(2r^2h^2(\theta)+o(1))\dfrac{1}{n}}{n}\right\}^n$$

$$=\left\{1-\frac{2rh(\theta)}{n}\right\}^n\left\{1+\frac{2r^2h^2(\theta)+o(1)}{n^2(1-2rh(\theta)/n)}\right\}^n$$

従って

$$SE(e,\ f\,;\theta)=\lim_{n\to\infty}\left(\frac{\beta_n^{(f)}(\theta)}{\beta_n^{(e)}(\theta)}\right)^n=\exp\{-2r^2(h^2(\theta)+f_+'(\theta|\theta))\}$$

$<1\qquad h^2(\theta)+f_+'(\theta|\theta)>0$ のとき

$>1\qquad h^2(\theta)+f_+'(\theta|\theta)<0$ のとき

すなわち，n が十分大きいとき

$SE(e,\ f\,;\theta)<1$ ならば $\beta_n^{(f)}(\theta)\nearrow\exp\{-2rh(\theta)\}$

$SE(e,\ f\,;\theta)>1$ ならば $\beta_n^{(f)}(\theta)\searrow\exp\{-2rh(\theta)\}$

であることがわかる。

例3.2　〔2〕の5(4)(p. 47) の例について考えてみる。この場合2次の有効性について考えられるのは，$C(\theta_0)=f(\theta_0|\theta_0)-f(A(\theta_0)|\theta_0)A'(\theta_0)=0$ の場合である。このとき，m.p.e. を Z_n とすると

Z_n は〔$m^*,\ m^{**}$〕の任意の点でよい。ただし，Z_n は可測，$m^*=\min\{W_n-r/n,\ A^{-1}(V_n)+r/n\}$，$m^{**}=\max\{W_n-r/n,\ A^{-1}(V_n)+r/n\}$，$W_n=\min X_i$，$V_n=\max X_i$ である。

このとき reasonable と考えられる2つの m.p.e. $Z_n^{(1)}=W_n-r/n$ と $Z_n^{(2)}=A^{-1}(V_n)+r/n$ について2次の相対有効率を計算しよう。そのために次の仮定を付加えよう。

$f(\theta_0|\theta_0)=h(\theta_0)>0,\ f(A(\theta_0)|\theta_0)=g(\theta_0)>0$

$f_+'(\theta_0|\theta_0),\ f_-'(A(\theta_0)|\theta_0)$ と $A''(\theta_0)$ が存在する。

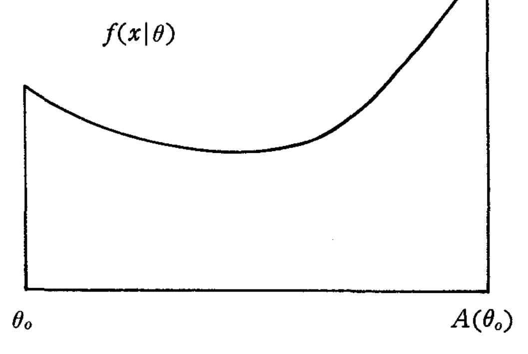
このとき

$$\beta_n^{(i)}(\theta_0)=P_{\theta_0}\{n(Z_n^{(i)}-\theta_0)\notin(-r,\ r)\}\qquad i=1,\ 2$$

であり，

$$\beta_n^{(1)}(\theta_0)=\left\{1-\frac{2rf(\theta_0|\theta_0)}{n}-\frac{2r^2f_+'(\theta_0|\theta_0)+o(1)}{n^2}\right\}^n$$

図3



$$\beta_n^{(2)}(\theta_0)=\left\{1-\frac{2rf(A(\theta_0)|\theta_0)A'(\theta_0)}{n}+\frac{2r^2[f_-'(A(\theta_0)|\theta_0)A'^2(\theta_0)+f(A(\theta_0)|\theta_0)A''(\theta_0)]+o(1)}{n^2}\right\}^n$$

となる。よって $C(\theta_0)=0$ に注意して

$$SE(1,\ 2\ ;\theta_0)=\lim_{n\to\infty}\left(\frac{\beta_n^{(2)}(\theta_0)}{\beta_n^{(1)}(\theta_0)}\right)^n$$

$$=\exp\{2r^2[f_+'(\theta_0|\theta_0)+f_-'(A(\theta_0)|\theta_0)A'^2(\theta_0)+f(A(\theta_0)|\theta_0)A''(\theta_0)]\}$$

を得る。これによって，$Z_n^{(1)}$ と $Z_n^{(2)}$ の2次の有効性がわかる。

謝辞

m.p.e. のセミナーで御指導下さり，本稿を書くにあたり様々な御助言を下さった，基礎工学部の稲垣講師と，本稿を読まれ様々な御意見を下さった丘本教授に深く感謝いたします。

〔1〕 Cramér, H. (1946). Mathematical Methods of Statistics, Princeton University Press.
〔2〕 Weiss, L. and Wolfowitz, J. (1974). Maximum Probability Estimators, Lecture Notes in Mathematics, Vol.424, Springer-Verlag.
〔3〕 Rao, C.R. (1963). Criteria and Estimation in Large Samples, Sankhyā 25, 189-206.
〔4〕 De Haan, L. (1970). On Regular Variation and its Application to the Weak Convergence of Sample Extremes. MC Tract 32, Mathematisch Centrum Amsterdam.
〔5〕 竹内啓 (1974). 漸近理論，教育出版
〔6〕 Kuß, U. (1975). Maximum Probability Estimators in the Case of Exponential Distribution, Metrika, 22, 129-146.

SECOND ORDER EFFICIENCY OF MAXIMUM PROBABILITY ESTIMATORS

RINYA TAKAHASHI

L. Weiss and J. Wolfowitz developed a theory of maximum probability estimators (m.p.e.'s). They showed that m.p.e.'s have the property of minimizing the limiting value of the risk, so-called W-efficiency. Any sequence of estimators which is W-efficient is a sequence of m.p.e.'s.

Then to compare two sequences of m.p.e.'s we define second order efficiency in non-regular case and decide the priority of them, which is explained in the examples shown in the paper.