

プールされたグループ標本にもとづく

最小コントラスト推定値の存在について

川崎 匠大 中村 忠 (Tadashi Nakamura)

### 1. 問題設定

$(T', T)$  は  $\mathcal{R} = (-\infty, \infty)$  の開部分区間;  $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$  は  $(T', T)$  上の確率分布の族で, 母数空間  $\Theta$  は空でない任意の集合;  $\{x_{hi}\} (h=1, \dots, l)$  は  $(r_h + 2)$  個の数から成る集合で,  $x_{h0} = T' < x_{h1} < \dots < x_{hr_h} < x_{hr_h+1} = T$  をみたす;  $X_{h1}, \dots, X_{hr_h}$  は独立で, 同一分布  $P_{\theta_0} \in \mathcal{P}$  に従う確率変数とする。

各  $X_{hi}, 1 \leq h \leq l; 1 \leq i \leq r_h$ , については  $\{X_{hi} \in \mathcal{C}_{hi}\}$  という情報しか得られないものとする。ただし,  $\mathcal{C}_{hi} \in \{[x_{h0}, x_{h1}), \dots, [x_{hr_h}, x_{hr_h+1})\}$ . このとき,  $\mathcal{C} = \{\mathcal{C}_{hi}; 1 \leq h \leq l, 1 \leq i \leq r_h\}$  は プールされたグループ標本と呼ばれる。  $\mathcal{C}_{hj} = [x_{hk}, x_{hk+1})$  となる  $\mathcal{C}_{hj}, 1 \leq j \leq r_h$ , の個数を  $n_{hk}$  で表わす。 プールされたグループ標本  $\mathcal{C}$  にもとづく最小コントラスト推定値(MCE)は次の最小化問題の最適解として定義される。

$$(I) \quad \inf \left\{ \sum_{h=1}^l S(q_h) \sum_{k=0}^{r_h} A(n_{hk}/q_h, P_\theta([x_{hk}, x_{hk+1})); \theta \in \Theta \right\} \text{ を見つけよ}$$

ここに  $S(x)$  は  $(0, \infty)$  上で定義された正值関数,  $A(p, x)$  は

$(0, 1] \times (0, 1] \cup \{0\} \times [0, 1]$  上で定義された実数関数で次をみたすものとする。

(1.1) 各  $p \in (0, 1]$  を固定すれば,  $A(p, z)$  は,  $z$  の関数として,  
 $(0, 1]$  上で連続である。

(1.2)  $A(0, z)$  は  $[0, 1]$  上で定数である。

(1.3) 次の条件をみたす定数  $K(A)$  と  $[0, 1]$  上で定義された非負値関数  $d(p)$  が存在する。

(i) 全ての  $p \in (0, 1]$  に対し,  $\lim_{z \rightarrow 0} A(p, z) = K(A)d(p)$  .

(ii)  $p=0$  かつそのときに限り  $d(p) = 0$  .

この問題の目的関数の値は次のルールにもとづいて計算するものとする。

$$A(p, 0) = K(A)d(p) \quad (\forall p \in (0, 1]), \quad t \cdot (\pm\infty) = \pm\infty \quad (\forall t \in (0, \infty)),$$

$$(\pm\infty) + (\pm\infty) = \pm\infty, \quad \text{空集合上の和は } 0 \text{ である。}$$

よく知られた MCE の例としては最小カイ 2 乗推定値, 最尤推定値, 修正最小カイ 2 乗推定値, 最小カルバック・ライプラー分離度推定値, 最小 2 乗推定値, 最小ヘルガー距離推定値などがある(付録参照)。

本論文では“いかなる条件の下で MCE が存在するか? その条件を求めよ。” に対する一つの解法(境界確率解析法)を提案する。以下通じて,  $\Theta$  上で  $P_0(\mathcal{C}_{hk}) \equiv 0$  又は 1 となる  $\mathcal{C}_{hk}$  は存在しないものと仮定する。

## 2. 境界確率解析法

境界確率解析法は次の4つのステップから成る。

ステップ 1. 次の (i) ~ (iii) をみたす正の整数  $m$  と  $(m+2)$  個の点

$x_0, x_1, \dots, x_m, x_{m+1}$  を決定する。

$$(i) \quad x_0 = T' < x_1 < \dots < x_m < x_{m+1} = T.$$

(ii)  $x_i$  と  $x_j$  は  $i \neq j$  なる限り  $\mathcal{P}$ -同値ではない, すなわち

$$P_\theta((T', x_i)) \neq P_\theta((T', x_j)) \text{ となる } \theta \in \Theta \text{ が存在する.}$$

(iii)  $\mathcal{C}_{hk}, 1 \leq h \leq l; 1 \leq k \leq g_h$  の各端点はある  $x_i, 0 \leq i \leq m+1$  と

$\mathcal{P}$ -同値であり, 更に各  $x_i, 1 \leq i \leq m$  はある  $\mathcal{C}_{hk}, 1 \leq h \leq$

$l; 1 \leq k \leq g_h$  の端点と  $\mathcal{P}$ -同値である。

ステップ 2. 非負の整数  $g_{hij}, 1 \leq h \leq l; 0 \leq i < j \leq m+1$ , 写像

$\mathbf{F}: \Theta \rightarrow \mathcal{Z} \equiv \{(z_1, \dots, z_m) \in \mathbb{R}^m; 0 \leq z_1 \leq \dots \leq z_m \leq 1\}$ , 関数  $L:$

$\mathcal{Z} \rightarrow \bar{\mathcal{R}} = [-\infty, \infty]$  を次のように定義する。

$g_{hij} =$  端点  $x_i, x_j$  のいずれかと  $\mathcal{P}$ -同値であるような

$\mathcal{C}_{hk}, 1 \leq k \leq g_h$ , の個数。

$$\mathbf{F}(\theta) = (P_\theta((T', x_1)), \dots, P_\theta((T', x_m))).$$

$$L(\mathbf{z}) = \sum_{h=1}^l S(g_h) \sum_{0 \leq i < j \leq m+1; p_{hij} \neq 0} A(p_{hij}, z_j - z_i),$$

ここに  $z_0 = 0, z_{m+1} = 1, \mathbf{z} = (z_1, \dots, z_m), p_{hij} = g_{hij} / g_h$  である。

ステップ 3. 境界確率集合  $\partial F(\Theta) \equiv \overline{F(\Theta)} - F(\Theta)$  の構造を決定する。

ステップ 4. 次の条件が成立するための十分条件を求める。

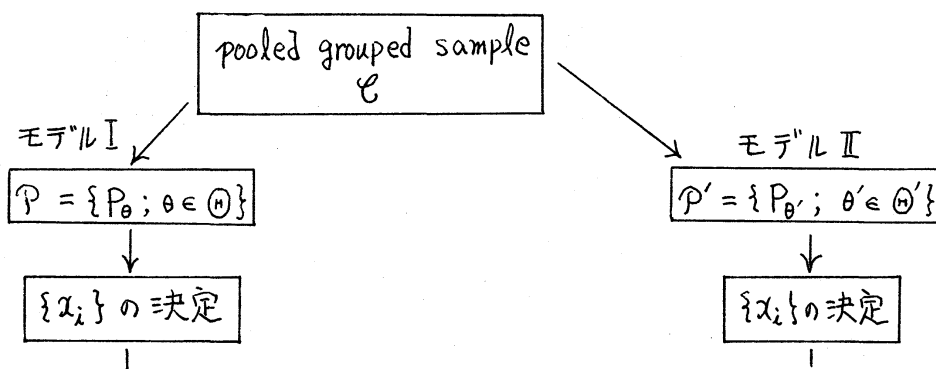
(2.1)  $L(\mathbf{z}') \leq M_b \equiv \inf\{L(\mathbf{z}); \mathbf{z} \in \partial F(\Theta)\}$  となる  $\mathbf{z}' \in \partial F(\Theta)$  が存在する。

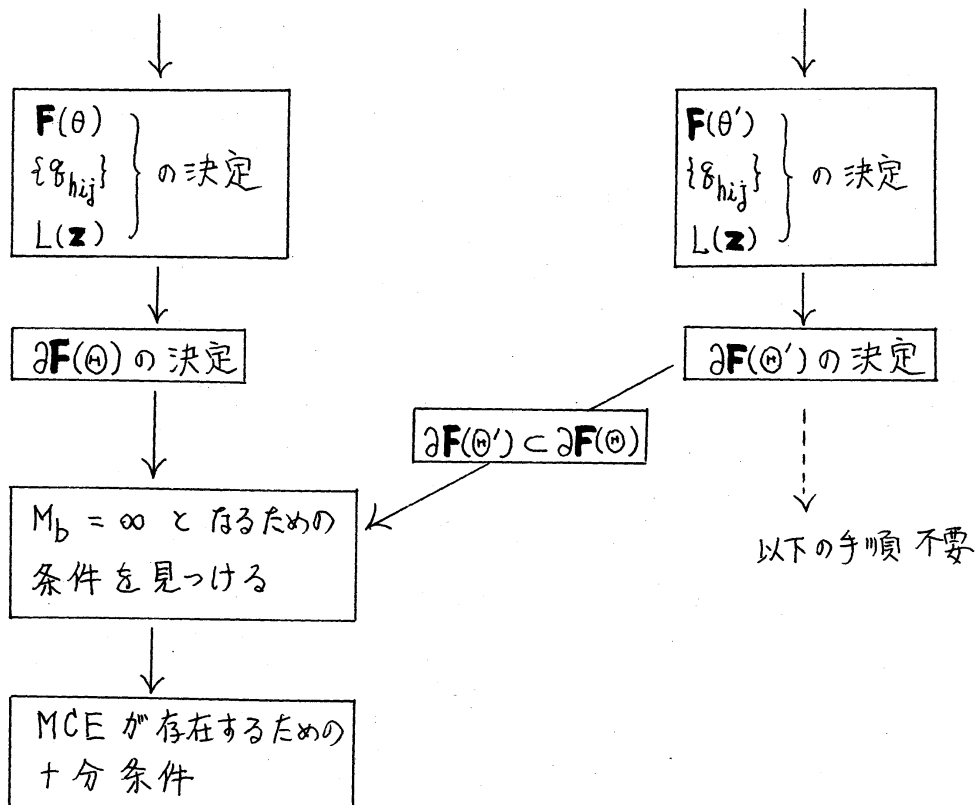
条件 (2.1) は MCE が存在するための必要・十分条件である(注.[3])。従ってステップ 4 で求めた十分条件は MCE が存在するための十分条件になっている。ステップ 1~4 を実行して MCE が存在するための十分条件を求める方法を 境界確率解析法 という。

### 3. 境界確率集合の構造と $M_b$ の評価

前節で述べたように、境界確率解析法においてはステップ 3~4 が重要な役割を果たす。特にステップ 3 における境界確率集合の果たす役割は、次の図で示されるように、顕著である。

境界確率解析法の流れ図 ( $K(A) = \infty$  の場合)





モデル I においてすべての手続きが完了したとしよう。別のモデル II を考える必要が生じた時、ステップ 4 を実行する必要のない場合がしばしばある。例えば、 $\partial F(\theta') < \partial F(\theta)$  ならば、モデル I で求めた MCE が存在するための十分条件はモデル II に対しても有効である。これは、境界確率解析法においては、 $\lambda$  の分布が属すと仮定された分布族ではなく境界確率集合に注目すればよいことを示唆している。更に境界確率集合は分布の形に対して、かなり不変的である (cf. [1], [2])。

組  $(P, \mathcal{C})$  が  $k$ -正則である とは次の条件が成立するときをいう。ただし、 $k$  は  $1 \leq k \leq m$  なる整数である。

(3.1) 各  $\mathbf{z} \in \partial F(\Theta)$  に対し,  $\mathbf{z}$  の成分で 0, 1 ではないものの内, 相異なるものの個数は高々  $k-1$  である。

多くの分布族に対し,  $k$ -正則性は  $m$  に関して不変である。

例えば, 位置母数を持つ正規分布族は任意の  $m (\geq 1)$  に対し, 1-正則である。又位置及び尺度母数を持つ正規分布族は  $m=1$  のとき 1-正則であり,  $m \geq 2$  のとき 2-正則である。

$M_b = \infty$  とする条件を求めるために次の記号を導入する。

$$N''(h; \mathbf{z}) = \{(i, j); 0 \leq i < j \leq m+1, q_{hij} \neq 0 \text{ and } z_j - z_i = 0\}$$

定理 3.1.  $K(A) = \infty$  及び組  $(P, \mathcal{C})$  の  $k$ -正則性を仮定する。

$$\inf \left\{ \sum_{h=1}^l \sum_{(i,j) \in N''(h; \mathbf{z})} g_{hij}; \mathbf{z} \in \partial F(\Theta) \right\} > 0$$

ならば,  $M_b = \infty$  である。

定理 3.1 で示されたように,  $K(A) = \infty$  の場合には MCE の存在のための判定法を得るのは簡単である。しかしながら  $K(A) < \infty$  の場合には,  $K(A) = \infty$  のときほど簡単ではなく, 与えられた関数  $S(\mathbf{z})$ ,  $A(p, \mathbf{z})$  にもとづく固有の工夫が必要となる。次節より応用を述べるのであるが, ここでは  $K(A) = \infty$  についてのみ議論をすることにする。  $K(A) < \infty$  なる場合については文献 [3] を参照されたい。

## 4. 応用 I

以下通じて  $K(A) = \infty$  を仮定する。  $K(A) = \infty$  となる推定法としては次がある。

例 4.1 (最小カイ 2 乗法).  $S(x) = x$ ,  $A(p, x) = (x-p)^2/x$  とする。このとき MCE は最小カイ 2 乗推定値 (MCSE) と呼ばれる。

例 4.2 (最尤法).  $S(x) = 1$ ,  $A(p, x) = p \log(p/x)$  とする。このとき MCE は最尤推定値 (MLE) と呼ばれる。

この節では組  $(\mathcal{P}, \mathcal{C})$  が 1-正則である場合について MCE が存在を保証する実用的な判定法を求める。次は明らか。

$$\text{組 } (\mathcal{P}, \mathcal{C}) \text{ が 1-正則である} \Leftrightarrow \partial F(\Theta) \subset \{\mathbf{a}_0, \dots, \mathbf{a}_m\},$$

$$\text{ここに } \mathbf{a}_i = (0, \dots, 0, 1, \dots, 1).$$

特に  $\mathbf{0} = (0, \dots, 0)$ ,  $\mathbf{1} = (1, \dots, 1)$  とおく。  $g_{hi}$  及び  $g_{i0}$  を次のように定義する。

$$g_{hi} = \sum_{j=i+1}^{m+1} g_{hij}, \quad i=0, \dots, m; \quad g_{i0} = \sum_{h=1}^l g_{hi}, \quad i=0, \dots, m.$$

同様にして  $g_{h0j}$ ,  $g_{00j}$  等も定義される。

定理 3.1 より次を得る。

定理 4.1. 次の条件がみたされれば MCE が存在する。

$$(4.1) \quad \mathbf{a}_k \notin \partial F(\Theta) \text{ 又は } \sum_{j=1}^k g_{00j} + \sum_{i=k+1}^m g_{i0} \neq 0, \quad k=0, \dots, m.$$

適当な制約の下で、条件(4.1)はMCEが存在するための必要条件となる。

定理 4.2.  $F(\Theta) \subset (0, 1)^m$  がすべての  $z \in (0, 1)$  に対し、 $A(1, 1) < A(1, z)$  を仮定する。MCEが存在するための必要・十分条件は条件(4.1)が成立することである。

## 5. 応用 II

この節では組  $(P, \varphi)$  が 2-正則である場合について MCE が存在を保証する実用的な判定法を与える。次は明らか。

$$\begin{aligned} \text{組 } (P, \varphi) \text{ が 2-正則である} &\Leftrightarrow \partial F(\Theta) \subset \{\mathbf{a}_0, \dots, \mathbf{a}_m\} \cup A \cup \bigcup_{i=1}^m A_i, \\ \text{ここに, } \mathbf{a}_i(z) &= (0, \dots, \underbrace{0}_{i-1}, z, 1, \dots, 1), \\ A &= \{z\mathbf{1}; 0 < z < 1\}, \\ A_i &= \{\mathbf{a}_i(z); 0 < z < 1\}. \end{aligned}$$

定理 5.1. 条件(4.1)及び次の条件が満たされれば MCE は存在する。

$$(5.1) \quad \partial F(\Theta) \cap A = \emptyset \quad \text{又は} \quad \sum_{1 \leq i < j \leq m} \varphi_{ij} \neq 0.$$

$$(5.2) \quad \partial F(\Theta) \cap A_k = \emptyset \quad \text{又は} \quad \sum_{j=1}^{k-1} \varphi_{\cdot j} + \sum_{i=k+1}^m \varphi_{i \cdot} \neq 0, \quad k=1, \dots, m.$$



適当な制約の下で，条件 (4.1), (5.1), (5.2) は MCE が存在するための必要条件となる。

定理 5.2.  $l=1$ ,  $\mathcal{F}_i = \{z_i; (z_1, \dots, z_m) \in \mathbf{F}(\Theta)\}$ ,  $1 \leq i \leq m$  とする。

次の条件が満たされているものとする。

(5.3) すべての  $(z_1, \dots, z_m) \in \mathbf{F}(\Theta)$  に対し,  $0 < z_1 \leq z_m < 1$ 。

(5.4)  $\partial \mathbf{F}(\Theta) \cap A \neq \emptyset$  なる限り  $\{z_1; z \in \mathcal{F}_1\} \subset \partial \mathbf{F}(\Theta)$  又は  $\{z_1; z \in \mathcal{F}_m\} \subset \partial \mathbf{F}(\Theta)$ 。

(5.5)  $\partial \mathbf{F}(\Theta) \cap A_i \neq \emptyset$  なる限り,  $\{a_i(z); z \in \mathcal{F}_i\} \subset \partial \mathbf{F}(\Theta)$ ,  $i=1, \dots, m$ 。

(5.6)  $\{z_i^*; a_i(z_i^*)\} \cap \mathcal{F}_i = \emptyset$ ,  $i=1, m$ .  $z_i^* = \sup\{z; z \in \mathcal{F}_i\}$  が  $z_m^* = \inf\{z; z \in \mathcal{F}_m\}$ 。

(5.7) 固定した  $p \in (0, 1]$  に対し,  $A(p, z)$  は  $(0, 1]$  上で  $z$  に関して狭義単調減少である。

このとき MCE が存在するための必要・十分条件は条件 (4.1), (5.1), (5.2) が成立することである。

定理 5.3.  $l \geq 2$ ,  $M < \infty$ ,  $\sum_{1 \leq i < j \leq m} \delta_{ij} \neq 0$  とする。条件 (5.3) 及び (5.5) - (5.7) が満たされているとき, MCE が存在するための必要・十分条件は条件 (4.1) と (5.2) が成立することである。

すぐわかるように定理 5.2 及び 5.3 は最尤法に適用出しても最小カイ 2 乗法には適用出さな。後者の方に適用出さるものとして次がある。

定理 5.4.  $l=1$  とする。条件 (5.3) 及び 次の条件が満たされていゝものとする。

$$(5.8) \quad \partial F(\Theta) \cap A \neq \emptyset \text{ ならば } A \subset \partial F(\Theta).$$

$$(5.9) \quad \partial F(\Theta) \cap A_k \neq \emptyset \text{ ならば } A_k \subset \partial F(\Theta), \quad k=1, \dots, m.$$

$$(5.10) \quad \text{固定した } p \in (0, 1] \text{ に対し, 不等式 } A(p, p) < A(p, z) \text{ が}$$

$$\text{すべての } z \in (0, 1] - \{p\} \text{ に対して成立する.}$$

このとき MCE が存在するための必要・十分条件は条件 (4.1), (5.1), (5.2) が成立することである。

## 6. 応用 III

この節では組  $(\mathcal{P}, \mathcal{C})$  が 2-正則であり,  $\partial F(\Theta) \cap A \neq \emptyset$  が  $\mathcal{P}$ -ルされたグループト標本  $\mathcal{B}$  が 2 値反応標本である場合について MCE の存在を保証する実用的な判定法を与える。  $\mathcal{P}$ -ルされたグループト標本が 2 値反応標本であるとは, すべての  $h, 1 \leq h \leq l$  に対し,  $\gamma_h = 1$  のときに言う。この場合必然的に  $\sum_{1 \leq i < j \leq m} \delta_{ij} = 0$  となる。次の条件を考える。

(P.1)  $z \in \partial F(\Theta) \cap A$  となる各  $z$  に対し, 次の条件をみたすよ

うな正数  $t_0$ ,  $(0, t_0)$  から  $\Theta$  内への写像  $P(t)$  及び  $(0, t_0)$  上の正値関数  $w(t)$  が存在する。

$$(6.1) \quad \lim_{t \rightarrow 0} \mathbf{F}(P(t)) = \mathbf{z}^1.$$

(6.2)  $P_{P(t)}((T', x_i))$ ,  $1 \leq i \leq m$ , は  $(0, t_0)$  上で微分可能であつて  $W(x_i; \mathbf{z}) \equiv \lim_{t \rightarrow 0} w(t) dP_{P(t)}((T', x_i))/dt$  が存在して有限である。

定理 6.1. 条件 (P.1) と次の条件がみたされているものとする。

(6.3)  $L(\mathbf{z}^1) = \inf \{ L(\mathbf{z}^1); \mathbf{z}^1 \in \partial \mathbf{F}(\Theta) \cap \mathcal{A} \}$  とする  $\mathbf{z}^1 \in \partial \mathbf{F}(\Theta) \cap \mathcal{A}$  が存在する。

(6.4) 固定した  $p \in (0, 1]$  に対し,  $A(p, \mathbf{z})$  は  $(0, 1)$  上で連続微分可能である。

このとき 条件 (4.1), (5.2) 及び次の不等式がみたされれば MCE が存在する。

$$(6.5) \quad \sum_{h=1}^l S(q_h) \sum_{1 \leq j \leq m; p_{hoj} \neq 0} A'(p_{hoj}, \mathbf{z}^*) W_j \\ < \sum_{h=1}^l S(q_h) \sum_{1 \leq i \leq m; p_{hi m+1} \neq 0} A'(p_{hi m+1}, 1 - \mathbf{z}^*) W_i$$

$$\text{こゝに } A'(p_{hij}, \mathbf{z}) = dA(p_{hij}, \mathbf{z})/d\mathbf{z}, \quad W_i = W(x_i; \mathbf{z}^*).$$

推定方法が指定されたとき不等式(6.5)がどのように成るかその例を述べる。ここでは最小カイ2乗法について述べるが、最尤法については文献[1]を参照されたい。

例6.1(最小2乗法).  $\partial F(\Theta) \cap A = A$ ,  $g_{\cdot 0 \cdot} > 0$ ,  $g_{\cdot \cdot m+1} > 0$  を仮定する。  $\partial F(\Theta) \cap A = A$  となる分布族の例については文献[1], [2]を参照されたい。このとき定理6.1における  $z^*$  は次の方程式の解である。

$$(6.6) \quad \frac{a}{z^2} - \frac{b}{(1-z)^2} = c,$$

ここには  $a = \sum_{h=1}^l g_h \sum_{j=1}^m p_{hoj}^2$ ,  $b = \sum_{h=1}^l g_h \sum_{i=1}^m p_{hi, m+1}^2$ ,  $c = \sum_{h=1}^l g_h (\sum_{j=1}^m C(p_{hoj}) - \sum_{i=1}^m C(p_{hi, m+1}))$ ,  $C(x)$  は集合  $(-\infty, 0) \cup (0, \infty)$  の特性関数である。  $a > 0$ ,  $b > 0$  だから方程式(6.6)は  $(0, 1)$  内に唯一つの解を持つ。不等式(6.5)は次のようになる。

$$(6.7) \quad \sum_{h=1}^l g_h \sum_{i=1}^m (C(p_{hoi}) - C(p_{hi, m+1})) W_i < \frac{c}{b} + \frac{a}{z^{*2}} \left( \frac{1}{a} \sum_{h=1}^l g_h \sum_{j=1}^m p_{hoj}^2 W_j - \frac{1}{b} \sum_{h=1}^l g_h \sum_{i=1}^m p_{hi, m+1}^2 W_i \right).$$

特に各  $h, 1 \leq h \leq l$ , に対し,  $n_{h0} \neq 0$ ,  $n_{h1} \neq 0$  ならば不等式(6.7)は次のようになる。

$$(6.8) \quad \frac{\sum_{h=1}^l g_h \sum_{i=1}^m p_{hi, m+1}^2 W_i}{\sum_{h=1}^l g_h \sum_{i=1}^m p_{hi, m+1}^2} < \frac{\sum_{h=1}^l g_h \sum_{j=1}^m p_{hoj}^2 W_j}{\sum_{h=1}^l g_h \sum_{j=1}^m p_{hoj}^2}.$$

分布族  $\mathcal{P}$  が指定されれば,  $W_i$  も具体的に表現される。例えば, 位置及び尺度母数を持つ正規分布族を  $\mathcal{P}$  として採用すれば

$W_i = x_i$ ,  $1 \leq i \leq m$  となる。このとき、大数の強法則により、各標本の大きさ  $n_h$ ,  $1 \leq h \leq l$ , が十分大きいとき、不等式 (6.8) が確率 1 で成立する。このことと定理 5.1 を合わせると, Rao [4] の結果が正規分布族の場合に対して証明されたこととなる。

### 付録

推定法	$S(z)$	$A(p, z)$	$K(A)$	$d(p)$
最小カイ2乗法	$z$	$\frac{(z-p)^2}{z}$	$\infty$	$p$
最尤法	1	$p \log \frac{p}{z}$	$\infty$	$p$
修正最小カイ2乗法	$z$	$\frac{(z-p)^2}{p}$	1	$p$
最小カルバツク・ライプラー法	1	$z \log \frac{z}{p}$	0	$p$
最小2乗法	1	$(z-p)^2$	1	$p^2$
最小ハリング距離法	1	$-\sqrt{pz}$	0	$p$

### References

- [1] Nakamura, T. (1984). Existence theorems of a maximum likelihood estimate from a generalized censored data sample, Ann. Inst. Statist. Math., 37.

- [2] \_\_\_\_\_(1984). The probability contents inner boundary of an interval-censored data sample for families of distributions (submitted for publication).
- [3] \_\_\_\_\_(1984). Existence theorems of a minimum contrast estimate from a pooled grouped sample, Okayama Statisticians Group, Tech. Rep. No. 9.
- [4] Rao, C. R. (1955). Theory of the method of estimation by minimum chi-square, Bull. Inst. Inter. Statist., 35, 25 - 32.