数理解析研究所講究録
第538巻 1984年 24-37

プールされたグループト標本にもとづく 最小コントラスト推定値の存在について

川崎医大 中村忠 (Tadashi Nakamura)

1. 問題設定

(T', T) は $R=(-\infty, \infty)$ 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 Θ は空でない任意の集合; $\{x_{hi}\}(h=1,\cdots,\ell)$ は (r_h+2) 個の数から成る集合で， $x_{h0}=T'<x_{h1}<\cdots<x_{hr_h}<x_{hr_h+1}=T$ をみたす; $X_{h1},\cdots, X_{hg_h}$ 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}, $1\le h\le \ell$; $1\le i\le g_h$, 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，$\mathcal{C}_{hi}\in\{[x_{h0}, x_{h1}), \cdots, [x_{hr_h}, x_{hr_h+1})\}$. このとき，$\mathcal{C}\equiv\{\mathcal{C}_{hi} ; 1\le h\le \ell, 1\le i\le g_h\}$ は プールされたグループト標本と呼ばれる。 $\mathcal{C}_{hj}=[x_{hk}, x_{hk+1})$ となる $\mathcal{C}_{hj}$, $1\le j\le g_h$, の個数を n_{hk} で表わす。プールされたグループト標本 $\mathcal{C}$ 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) $\inf\{\sum_{h=1}^{\ell} S(q_h)\sum_{k=0}^{r_h} A(n_{hk}/q_h,\ P_\theta([x_{hk}, x_{h\,k+1})) ;\ \theta\in\Theta\}$ を見つけよ。
ここに $S(z)$ は $(0, \infty)$ 上で定義された正値関数， $A(p, z)$ 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，$\lim_{z\to 0} A(p,z)=K(A)\,d(p)$.

(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。

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

$(\pm\infty)+(\pm\infty)=\pm\infty$，空集合上の和は 0 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・ライブラー分離度推定値，最小２乗推定値，最小ヘリン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，Θ 上で $P_\theta(\mathscr{C}_{hk})\equiv 0$ 又は 1 となる $\mathscr{C}_{hk}$ 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, …, x_m, x_{m+1} を決定する。

(i) $x_0 = T' < x_1 < \cdots < x_m < x_{m+1} = T$ 。

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

(iii) $\mathcal{C}_{hk}$, $1 \leq h \leq \ell$; $1 \leq k \leq g_h$ 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 $\mathcal{C}_{hk}$, $1 \leq h \leq \ell$; $1 \leq k \leq g_h$ の端点と $\mathcal{P}$-同値である。

ステップ2. 非負の整数 g_{hij}, $1 \leq h \leq \ell$; $0 \leq i < j \leq m+1$, 写像 $\mathbf{F}: \Theta \to \mathcal{Z} \equiv \{(z_1, \cdots, z_m) \in R^m;\ 0 \leq z_1 \leq \cdots \leq z_m \leq 1\}$, 関数 $\mathbb{L}: \mathcal{Z} \to \overline{R} = [-\infty, \infty]$ を次のように定義する。

g_{hij} = 端点が 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。

$$\mathbf{F}(\theta) = (P_\theta((T', x_1)), \cdots, P_\theta((T', x_m)))\text{。}$$

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

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

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

<u>ステップ4.</u> 次の条件が成立するための十分条件を求める。

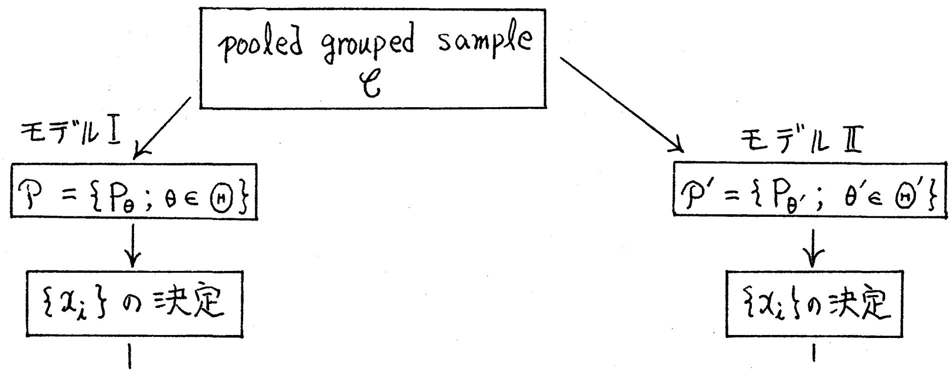
(2.1) $L(\mathbf{z}') \leq M_b \equiv \inf\{L(\mathbf{z})\,;\ \mathbf{z} \in \partial\mathbf{F}(\Theta)\}$ となる $\mathbf{z}' \in \partial\mathbf{F}(\Theta)$ が存在する。

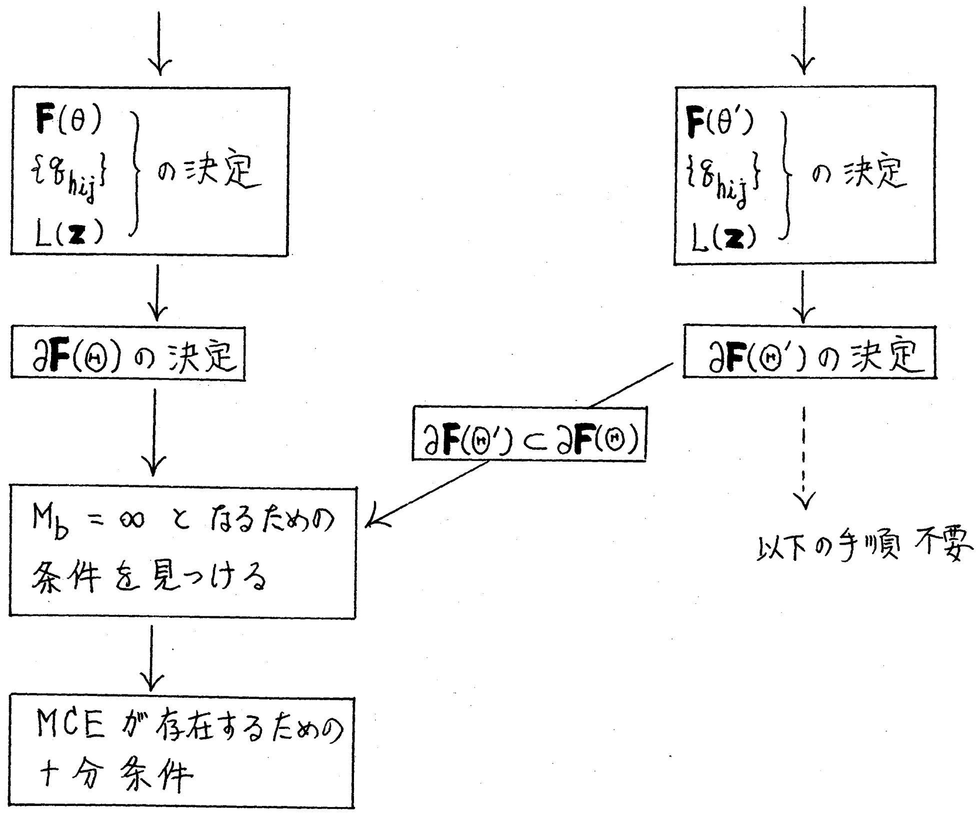
条件(2.1)はMCEが存在するための必要・十分条件である(cf.[3])。従ってステップ4で求めた十分条件はMCEが存在するための十分条件になっている。 ステップ1~4 を実行して MCEが存在するための十分条件を求める方法を<u>境界確率解析法</u>という。

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。

<u>境界確率解析法の流れ図</u>(K(A) = ∞ 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，$\partial\mathbf{F}(\Theta')\subset\partial\mathbf{F}(\Theta)$ 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，X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$(\mathcal{P},\mathcal{C})$が k-正則である とは次の条件が成立するときにいう。ただし，kは $1\le k\le m$ なる整数である。

(3.1)　各 $\mathbf{z} \in \partial\mathbf{F}(\Theta)$ に対し，$\mathbf{z}$ の成分で 0,1 でないものの内，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。

$M_b = \infty$ となる条件を求めるために次の記号を導入する。

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

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

$$\inf\{\textstyle\sum_{h=1}^{\ell} \sum_{(i,j)\in N''(h;\mathbf{z})} q_{hij}\ ;\ \mathbf{z} \in \partial\mathbf{F}(\Theta)\} > 0$$

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 $S(z)$，$A(p,z)$ 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). $S(z)=z$, $A(p,z)=(z-p)^2/z$ 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). $S(z)=1$, $A(p,z)=p\log(p/z)$ 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。

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

ここに $\mathbf{a}_i=(\overbrace{0,\cdots,0}^{i},1,\cdots,1)$.

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

$$g_{hi\cdot}=\Sigma_{j=i+1}^{m+1}g_{hij}, \quad i=0,\cdots,m; \quad g_{\cdot i\cdot}=\Sigma_{h=1}^{l}g_{hi\cdot}, \quad i=0,\cdots,m.$$

同様にして $g_{h\cdot j}$, $g_{\cdot\cdot j}$ 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。

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

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. $\mathbf{F}(\Theta) \subset (0,1)^m$ かつすべての $z \in (0,1)$ 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$(\mathcal{P},\mathcal{C})$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。

組$(\mathcal{P},\mathcal{C})$が2-正則である $\iff$ $\partial\mathbf{F}(\Theta) \subset \{\mathbf{a}_0, \cdots, \mathbf{a}_m\} \cup \mathcal{A} \cup (\cup_{i=1}^m \mathcal{A}_i)$,

ここに，$\mathbf{a}_i(z) = (\overbrace{0, \cdots, 0}^{i-1}, z, 1, \cdots, 1)$,

$\mathcal{A} = \{z\mathbf{1}\,;\; 0 < z < 1\}$,

$\mathcal{A}_i = \{\mathbf{a}_i(z)\,;\; 0 < z < 1\}$.

定理5.1. 条件(4.1)及び次の条件がみたされればMCEは存在する。

(5.1) $\partial\mathbf{F}(\Theta) \cap \mathcal{A} = \emptyset$ 又は $\sum_{1 \le i < j \le m} g_{\cdot ij} \ne 0$.

(5.2) $\partial\mathbf{F}(\Theta) \cap \mathcal{A}_k = \emptyset$ 又は $\sum_{j=1}^{k-1} g_{\cdot\cdot j} + \sum_{i=k+1}^{m} g_{\cdot i \cdot} \ne 0$, $k=1,\cdots,m$.

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。

<u>定理 5.2.</u> $l=1$, $\mathscr{S}_i=\{z_i;(z_1,\cdots,z_m)\in \mathbf{F}(\Theta)\}$, $1\le i\le m$ とする。次の条件がみたされているものとする。

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

(5.4) $\partial\mathbf{F}(\Theta)\cap \mathcal{A}\ne\emptyset$ なる限り $\{z\mathbf{1};z\in\mathscr{S}_1\}\subset\partial\mathbf{F}(\Theta)$ 又は $\{z\mathbf{1};z\in\mathscr{S}_m\}\subset\partial\mathbf{F}(\Theta)$。

(5.5) $\partial\mathbf{F}(\Theta)\cap \mathcal{A}_i\ne\emptyset$ なる限り，$\{\mathbf{a}_i(z);z\in\mathscr{S}_i\}\subset\partial\mathbf{F}(\Theta)$, $i=1,\cdots,m$.

(5.6) $\{z_i^*\mathbf{1},\mathbf{a}_i(z_i^*)\}\cap\mathscr{S}_i=\emptyset$, $i=1,m$. ここに $z_1^*=\sup\{z;z\in\mathscr{S}_1\}$ かつ $z_m^*=\inf\{z;z\in\mathscr{S}_m\}$。

(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。

<u>定理 5.3.</u> $l\ge 2$, $M<\infty$, $\sum_{1\le i<j\le m}\delta_{\bullet ij}\ne 0$ とする。条件 (5.3) 及び (5.5)−(5.7) がみたされているとき，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は最尤法に適用出きても最小カイ2乗法には適用出きない。後者の方に適用出きるものとして次がある。

<u>定理5.4.</u>　$l=1$とする。条件(5.3)及び次の条件がみたされているものとする。

(5.8)　$\partial\mathbf{F}(\Theta)\cap\mathcal{A}\neq\emptyset$ ならば $\mathcal{A}\subset\partial\mathbf{F}(\Theta)$。

(5.9)　$\partial\mathbf{F}(\Theta)\cap\mathcal{A}_k\neq\emptyset$ ならば $\mathcal{A}_k\subset\partial\mathbf{F}(\Theta)$，　$k=1,\cdots,m$。

(5.10)　固定した $p\in(0,1]$ に対し，不等式 $A(p,p)<A(p,z)$ がすべての $z\in(0,1]-\{p\}$ 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，$\partial\mathbf{F}(\Theta)\cap\mathcal{A}\neq\emptyset$ かつ プールされたグループト標本 $\mathcal{E}$ 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。プー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，$r_h=1$ のときに言う．　この場合必然的に $\Sigma_{1\leq i<j\leq m}\,g_{\cdot ij}=0$ となる。次の条件を考える。

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

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

(6.1) $\lim_{t\to 0} \mathbf{F}(\rho(t)) = z\mathbf{1}$ 。

(6.2) $P_{\rho(t)}((T', x_i))$, $1 \le i \le m$, は $(0, t_0)$ 上で微分可能であって $W(x_i; z) \equiv \lim_{t\to 0} w(t)\, d\, P_{\rho(t)}((T', x_i))/dt$ が存在して有限である。

<u>定理6.1.</u> 条件(P.1)と次の条件がみたされているものとする。

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

(6.4) 固定した $p \in (0, 1]$ に対し, $A(p, z)$ 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) $$\Sigma_{h=1}^{l} S(q_h) \Sigma_{1\le j\le m;\, p_{h0j}\ne 0} A'(p_{h0j}, z^*) W_j$$

$$< \Sigma_{h=1}^{l} S(q_h) \Sigma_{1\le i\le m;\, p_{h\,i\,m+1}\ne 0} A'(p_{h\,i\,m+1}, 1-z^*) W_i$$

ここに $A'(p_{hij}, z) = dA(p_{hij}, z)/dz$, $W_i = W(x_i; z^*)$ 。

推定方法が指定されたとき不等式(6.5)がどのようにな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). $\partial\mathbf{F}(\Theta)\cap\mathcal{A}=\mathcal{A}$, $g_{\cdot 0\cdot}>0$, $g_{\cdot\cdot m+1}>0$ を仮定する。$\partial\mathbf{F}(\Theta)\cap\mathcal{A}=\mathcal{A}$ 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。

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

ここに $a=\Sigma_{h=1}^{l} g_h\Sigma_{j=1}^{m} p_{h0j}^2$, $b=\Sigma_{h=1}^{l} g_h\Sigma_{i=1}^{m} p_{h i\, m+1}^2$, $c=\Sigma_{h=1}^{l} g_h(\Sigma_{j=1}^{m} C(p_{h0j})-\Sigma_{i=1}^{m} C(p_{h i\, m+1}))$, $C(x)$ 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。

$$(6.7)\qquad \Sigma_{h=1}^{l} g_h\Sigma_{i=1}^{m}(C(p_{h0i})-C(p_{h i\, m+1}))W_i$$
$$<\frac{c}{b}+\frac{a}{z^{*2}}\Big(\frac{1}{a}\Sigma_{h=1}^{l} g_h\Sigma_{j=1}^{m} p_{h0j}^2 W_j-\frac{1}{b}\Sigma_{h=1}^{l} g_h\Sigma_{i=1}^{m} p_{h i\, m+1}^2 W_i\Big).$$

特に各 h, $1\le h\le l$, 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。

$$(6.8)\qquad \frac{\Sigma_{h=1}^{l} g_h\Sigma_{i=1}^{m} p_{h i\, m+1}^2 W_i}{\Sigma_{h=1}^{l} g_h\Sigma_{i=1}^{m} p_{h i\, m+1}^2}<\frac{\Sigma_{h=1}^{l} g_h\Sigma_{j=1}^{m} p_{h0j}^2 W_j}{\Sigma_{h=1}^{l} g_h\Sigma_{j=1}^{m} p_{h0j}^2}.$$

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

$W_i = x_i$, $1 \le i \le m$ となる。 このとき，大数の強法則により，各標本の大きさ g_h, $1 \le h \le l$, 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になる。

付録

推定法	$S(z)$	$A(p,z)$	$K(A)$	$d(p)$
最小カイ2乗法	z	$\frac{(z-p)^2}{z}$	∞	p
最尤法	1	$p\log\frac{p}{z}$	∞	p
修正最小カイ2乗法	z	$\frac{(z-p)^2}{p}$	1	p
最小カルバック・ライブラー法	1	$z\log\frac{z}{p}$	0	p
最小2乗法	1	$(z-p)^2$	1	p^2
最小ヘリンガー距離法	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.