数理解析研究所講究録
第380巻 1980年 1-15

推定尤度関数と棱線推定量の一般化

大阪大学基礎工学部　稲垣宣生

§1. 推定関数と推定量

推定関数（*Estimating Function*）の考えは最尤推定法と関係して古くからある（Wilks[10]）．推定関数 $\xi_n(\theta)=\xi_n(\theta,x)$ に基く推定量を $\hat{\theta}_n=\hat{\theta}_n(x)$ とする．すなわち，$\hat{\theta}_n$ は推定方程式（Estimating Equation）の解である：

$$\xi_n(\hat{\theta}_n)=0。$$

普通，推定関数は評点関数（Score Function）の和として書かれるので典型的な例を挙げる：

$$\xi_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\eta(X_i,\theta)\text{，独立同一分布観測，}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\eta(X_{i-1},X_i;\theta)\text{，マルコフ観測，}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\eta_i(X_i,\theta)\text{，独立非同一分布観測，}$$

$$=\frac{1}{n_1}\frac{1}{n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\eta(X_i,Y_j;\theta)\text{，2標本(観測)．}$$

$$n=n_1+n_2$$

とくに評点関数が尤度に基くとき，$\overbrace{\eta(x,\theta)}^{\text{たとえば}} = \frac{\partial}{\partial\theta}\log f(x,\theta)$ のようなとき，推定尤度関数（Likelihood Estimating Function）という．しかし，ここでは推定関数の具体的形は問題にしないで，一般的に推定関数と推定量（とくに解推定量）との関連に注目する．推定関数がある性質をみたすとき解推定量も対応する同様の性質をみたすであろうか：たとえば

推定関数 $\xi_n(\theta)$	性質	解推定量 $\hat{\theta}_n$
$E_\theta\{\xi_n(\theta)\} = 0$	不偏性	$E_\theta\{\hat{\theta}_n\} = \theta,$
$\xi_n(\theta) \xrightarrow{P_\theta} 0$	一致性	$\hat{\theta}_n \xrightarrow{P_\theta} \theta,$
$\sqrt{n}\,\xi_n(\theta) \xrightarrow{L_\theta} N(0, D(\theta))$	漸近正規性	$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L_\theta} N(0, D'(\theta)).$

その他にも可測性，不変性，有界性，漸近十分性，漸近有効性などが問題になるであろう．いわゆる"正則条件"の下でこれらの問題に対し肯定的な解答が与えられているが，しかし正則条件は確固不動のものがなかなかない．

§2．推定関数の漸近的可微分性

正則条件の下で推定関数の漸近正規性は一般に成立つ：

(2.1) $\sqrt{n}\,\xi_n(\theta) \longrightarrow N(0, D(\theta))$ in Law.

推定量の可測性，一致性から議論を始めるべきであろうがここではそれを仮定する：

(2.2) $\xi_n(T_n) \longrightarrow 0$ in $P_\theta \Rightarrow T_n \longrightarrow \theta$ in P_θ.

この逆(⇐)は以下で明らかになる。そのまえに推定関数のパラメタθに推定量T_nを代入する操作が不安になるかもしれないがここでは問題にしない(LeCam[7] Appendix).推定関数の漸近可微分性から次のことが成立つ.

定理 2.1 (Huber[3], Inagaki[5]).

一致推定量T_nに対して,

$$\frac{\xi_n(T_n)-\xi_n(\theta)-\Lambda(\theta)(T_n-\theta)}{n^{-1/2}+|\Lambda(\theta)(T_n-\theta)|}\longrightarrow 0 \text{ in } P_\theta \tag{2.3}$$

ただし,$\Lambda(\theta)$は$\xi_n(\theta)$の漸近微分である.

分母の$n^{-1/2}$は$T_n-\theta=O(n^{-1/2})$なる収束速度のオーダを中心にした調整項である.この定理から次のことが成立つ.(2.1)の逆(⇐)が成立つことも明らかである.

定理 2.2. (Inagaki [5])

(i) $\{\sqrt{n}\,\xi_n(T_n)\}$が確率有界であれば$\{\sqrt{n}(T_n-\theta)\}$も確率有界である.

(ii) $\{\sqrt{n}(T_n-\theta)\}$が確率有界であれば,

$$\sqrt{n}\,\xi_n(T_n)-\sqrt{n}\,\xi_n(\theta)-\Lambda(\theta)\sqrt{n}(T_n-\theta)\longrightarrow 0 \text{ in } P_\theta.$$

(iii) $\sqrt{n}\,\xi_n(T_n)\longrightarrow 0$ in P_θ ならば

$$\sqrt{n}(T_n-\theta)\longrightarrow N(0,\tilde{D}(\theta)) \text{ in } L_\theta\ ,\quad \tilde{D}(\theta)=\vec{\Lambda}(\theta)D(\theta)\vec{\Lambda}(\theta)'.$$

とくに解推定量$\hat{\theta}_n$に対して

$$\sqrt{n}(\hat{\theta}_n-\theta)\longrightarrow N(0,\tilde{D}(\theta)) \text{ in } L_\theta.$$

定理 2.3.

$\{P_\theta^n\}$ と $\{P_{\theta+h/\sqrt{n}}^n\}$ が接近している(Contiguous)とき,

$$\mathcal{L}[\sqrt{n}\,\xi_n(\theta+h/\sqrt{n})\,|\,P_{\theta+h/\sqrt{n}}] \to \mathcal{L}_\theta(\xi),\quad h \text{ に indep.}$$

$$\Leftrightarrow \mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta-h/\sqrt{n})\,|\,P_{\theta+h/\sqrt{n}}] \to \mathcal{L}_\theta(\hat{\theta}),\quad h \text{ に indep.}$$

すなわち $\hat{\theta}_n$ は正則(regular)である.

推定関数の漸近可微性によって,$T_n-\theta$ の性質は $\xi_n(T_n)$, $\xi_n(\theta)$ から決定されることを上の2つの定理は示している.

§3. 類似度関数

分布族を添字付けられた密度関数の集合として表わす:

$$\mathcal{G} = \{g(x,\theta)\}_{\theta\in\Theta}.$$

観測 X の分布が分布族 $\mathcal{F}=\{f(x,\theta)\}_{\theta\in\Theta}$ に属するとき,

$$L(\theta) = f(x,\theta)$$

を尤度関数(Likelihood Function)と呼ぶことに対して,

$$R(\theta) = g(x,\theta)$$

を類似度関数(Resemblance Function)と呼ぶことにしよう.あえて新語を作る冒険をしたのは最近分布に対する考え方が拡張されて分布型=模型というようになる様があるからである.真の分布と仮の分布;真の尤度と仮の尤度;真の母数と仮の母数等々言うよりもこの際,分布型と模型;尤度と類似度;母数と類似数等々と言った方がはっきりするのではない

だろうか。Strasser [7]; Prakasa Rao [8] はいずれの場合も区別を行わず Bayes 推定量，最尤推定量と呼んでいるが，分布族 G を使うときには Bayes 類似推定量，最似推定量などという別の呼び方をした方がよいのではなかろうか．この段になって真顔で尤度関数と類似度関数には違いがあるのかと問われれば，やはり相違があると断定したい(!).

推定関数 $\xi_n(\theta)$ は対数類似度の微分であるようなものに限定して考えよう．たとえば，

$$\xi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log g(x_i,\theta). \tag{3.1}$$

一方，推定尤度関数は尤度関数を使って次のように書ける：

$$\xi_n^{\circ}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i,\theta). \tag{3.2}$$

これらの推定関数，解推定量（最尤推定量と最似推定量）における性質に差が現われるのは，漸近十分性や漸近有効性の段階からで，それまでの一致性，漸近正規性等々には差がない。次の定理は尤度関数に関してのみ成立する。

定理 3.1 (Hajeck [1], Inagaki [4], [5]).

正則な推定量 T_n ($\mathcal{L}[\xi_n^{\circ}(T_n)]\,P_{\theta+h/\sqrt{n}}] \to \mathcal{L}(\xi(T))$ h に indep. ということと同値）に対して，

(i) $\xi_n^{\circ}(T_n)$ と $\xi_n^{\circ}(\theta)$ は漸近独立である．

(ii) $\lim \mathcal{L}[\sqrt{n}(T_n-\theta)] = \lim \mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta)] * \lim \mathcal{L}[\sqrt{n}(T_n-\hat{\theta}_n)]$.

すなわち最尤推定量の漸近有効性を示している．

§4. 稜線推定量の一般化.

線形回帰模型の場合の稜線推定量（Ridge Estimator）の導出法において Hoerl & Kennard [2] が示唆しているように，一般の場合にも稜線推定量を考えることができる．すなわち母数 $\theta=(\theta_1,\cdots,\theta_k)'$ と観測 $x=(x_1,\cdots,x_n)$ に対する密度関数が $f_n(x,\theta)$ であるとき，稜線推定量 $\hat{\theta}_{n\alpha}$ は次のように定義する：

(4.1) $\quad |\hat{\theta}_{n\alpha}|\leqq\alpha$ かつ $f_n(x,\hat{\theta}_{n\alpha})=\sup\{f_n(x,\theta);\ |\theta|\leqq\alpha\}$.

そのときの正規方程式は，可微分なノルム

(4.2) $\quad |\theta|^2=\theta_1^2+\cdots+\theta_k^2$

の下でラグランジュ乗数 ν を用いた関数

(4.3) $\quad -2\log f_n(x,\theta)+\nu(\theta'\theta-\alpha^2)$

を微分して得られる：

(4.4) $\quad \frac{\partial}{\partial\theta}\log f_n(x,\theta)=\nu\theta,\qquad \theta'\theta=\alpha^2.$

第一の方程式を稜線推定方程式，その解を稜線推定量と呼ぶことにする．推定尤度関数（3.2）の記号を使えば

(4.5) $\quad \xi_n^0(\theta)=\frac{\nu}{n}\theta\ .$

尤度関数の代りに類似度関数を用いて以上の議論を行えば直に稜線型推定量なるものが定義できる：

(4.1′) $\quad |\hat{\theta}_{n\alpha}|\leqq\alpha$ かつ $g_n(x,\hat{\theta}_{n\alpha})=\sup\{g_n(x,\theta);\ |\theta|\leqq\alpha\}$,

(4.5′) $\quad \xi_n(\theta)=\frac{\nu}{n}\theta\ .$

核線型推定量 $\hat{\theta}_n(\nu)$ はその導出法からみてもとの解推定量 $\hat{\theta}_n$(= $\hat{\theta}_n(0)$ である！) を原点方向へ縮小していることがわかる．縮小係数が ν である．なぜ原点方向に縮小するのであろうかという疑問にはパラメタの座標のとり方から論じなければ答えられない．しかしパラメタの決定的部分を除去した場合には原点方向への縮小は不自然でない．今，決定的部分が β であったとして，まだパラメタからこの部分を除去していないとして β 方向へ縮小したとすれば核線推定方程式は次のように変形される：

(4.6) $$\xi_n(\theta) = \frac{\nu}{n}(\theta - \beta).$$

§5. 偏りのある推定関数（推定方程式）

核線推定方程式は，偏りのない推定関数 $\xi_n(\theta)$ を使って偏りのある推定方程式 (4.5), (4.5′) 又は (4.6) に変形したものである．ここではもっと一般の偏りのある推定方程式について考えよう：

(5.1) $$\xi_n(\theta, x) = b_n(\theta),$$ 但し，$b_n(\theta)$ は確実関数．

簡単のために，$b_n(\theta)$ は

$$b_n(\theta) = b(\theta)/n, \quad b(\theta)/\sqrt{n} \quad \text{又は} \quad b(\theta)$$

なる形のものだけを考えることにする．解推定量として一致性をもつもののみ考えることにすれば，

$$\ell(T_n) \doteqdot \ell(\theta) + \ell'(\theta)(T_n - \theta) = \ell + BT_n$$

により $\ell(\theta)$ が θ の一次式である場合が重要になる．ここで $\ell = -B\beta$ と書直せば

$$\ell(T_n) = B(T_n - \beta)$$

なる場合が重要になる．すなわち

(5.2) $\quad \xi_n(\theta) = B(\theta - \beta)/n,$

(5.3) $\quad \xi_n(\theta) = B(\theta - \beta)/\sqrt{n},$

(5.4) $\quad \xi_n(\theta) = B(\theta - \beta).$

最後の式（5.4）の解推定量は一致性をもたないので除外することにする．ここで β, B は定数（ベクトルと行列）である．Bとして簡単な行列

$$B = \nu I \quad \text{又は} \quad \mathit{Diagonal}(\nu_1, \cdots, \nu_k)$$

をとったとき稜線型又は一般稜線型推定量が得られる．

βに関してこれ以上何の情報もないときには $\theta - \beta = \tau$ を新しいパラメタと考えれば θ を β 方向に縮小することは τ を原点方向に縮小することになるので $\beta = 0$ として議論を続けても一般性を失なわない．もし2つの推定関数 $\xi_n^1(\theta)$ と $\xi_n^2(\theta)$ の解 $\hat{\theta}_n^1$ と $\hat{\theta}_n^2$ を稜線によって結合するとすれば，

$$\xi_n^1(\theta) = B(\theta - \hat{\theta}_n^2) \quad \text{又は} \quad \xi_n^2(\theta) = B(\theta - \hat{\theta}_n^1)$$

なる方程式の解として求まる．その時には別の問題が起ってくるであろう．以降では稜線推定量のみを議論する．

§6. 核線推定方程式の解について

ここでは具体例によって核線推定方程式とその解を調べてみよう.

例 1.(正規分布型)

$$g_n(x,\theta)=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

のとき,

$$\xi_n(\theta)=\frac{1}{n}\frac{\partial}{\partial\theta}\log g_n(x,\theta)=\frac{1}{\sigma^2}(\bar{x}-\theta)$$

$$=\frac{\nu}{n}\theta \quad とおく.$$

これより解推定量 $\hat{\theta}_n$ と核線推定量 $\hat{\theta}_n(\nu)$ は

$$\hat{\theta}_n=\bar{x},$$

$$\hat{\theta}_n(\nu)=\bar{x}\Big/\left(1+\frac{\nu}{n}\sigma^2\right).$$

例 2.(ポアソン分布型)

$$g_n(x,\theta)=e^{-n\theta}\cdot\theta^{\Sigma x_i}/(x_1!\cdots x_n!)$$

$$\xi_n(\theta)=-1+\bar{x}/\theta \underset{おく}{=} \frac{\nu}{n}\theta$$

これより, $\hat{\theta}_n=\bar{x}$ であり, $\hat{\theta}_n(\nu)$ は方程式

$$\frac{\nu}{n}\theta^2+\theta-\bar{x}=0$$

の解である. $\theta>0$ なる解をもとめると

$$\hat{\theta}_n(\nu)=\bar{x}\left(1-2\nu\frac{\bar{x}}{n}\Big/\left(1+\sqrt{1+4\nu\bar{x}/n}+2\nu\bar{x}/n\right)\right).$$

例 3.(2 項分布型)

$$g_n(x,\Theta)=\theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$$

$$\xi_n(\theta)=\bar{x}/\theta-(1-\bar{x})/(1-\theta)$$

ここで $\xi_n(\theta) = \frac{\nu}{n}\theta$ とおくと方程式は θ の 3次式になるので θ の代りに $\hat{\theta} = \bar{x}$ を使い $\xi_n(\theta) = \frac{\nu}{n}\bar{x}$ を解けば

$$\tilde{\theta}_n(\nu) = 2\bar{x} \Big/ \left(1 + \frac{\nu}{n}\bar{x} + \sqrt{\left(1+\frac{\nu}{n}\bar{x}\right)^2 - 4\frac{\nu}{n}\bar{x}^2}\right).$$

例4. (両側指数分布型)

$$g_n(x,\theta) = 2^{-n} \exp\left\{-\sum_{i=1}^{n} |x_i - \theta|\right\}$$

$$\xi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{sgn}(x_i - \theta) \underset{おく}{=} \frac{\nu}{n}\theta$$

解は数値的にしかもとまらない。$\hat{\theta}_n = \mathrm{median}(x_1, \cdots, x_n)$ である。

§7. 稜線推定量の性質

~~解~~推定量 $\hat{\theta}_n$ は仮定(2.2)式より一致性をもつ。今, T_n も一致推定量とすれば, (2.3)式より

$$\xi_n(\hat{\theta}_n) - \xi_n(\theta) - \Lambda(\theta)(\hat{\theta}_n - \theta) \longrightarrow 0 \quad \text{in } P_\theta$$

$$\xi_n(T_n) - \xi_n(\theta) - \Lambda(\theta)(T_n - \theta) \longrightarrow 0 \quad \text{in } P_\theta$$

を得る。両式を引算して $\xi_n(\hat{\theta}_n) \longrightarrow 0$ in P_θ に注意すれば, 次の結果を得る。普通 $-\Lambda(\theta)$ が正定値行列である。

定理 7.1.

一致推定量 T_n に対して,

$$\xi_n(T_n) + \Gamma(\theta)(T_n - \hat{\theta}_n) \longrightarrow 0 \quad \text{in } P_\theta$$

ただし $\Gamma(\theta) = -\Lambda(\theta)$ とおく。

このことから稜線推定量 $\hat{\theta}_n(\nu)$ が一致性をもつとして次のような漸近的表現を得る。

となるので定理 2.2 (ii) から

(7.4) $\Gamma(\theta)\sqrt{n}\,(T_n(\nu)-\theta) \overset{A.D.}{=} \nu\theta + \sqrt{n}\,\xi_n(\theta) \overset{A.D.}{=} N_k(\nu\theta,\ D(\theta))$

を得る．この場合，$\sqrt{n}(T_n(\nu)-\theta)$ は $\sqrt{n}\,(\hat{\theta}_n-\theta)$ と漸近分散は同じであるが漸近平均は異なり偏り $\Gamma(\theta)^{-1}\nu\theta$ をもつ．もし ν が定数のときには，$T_n(\nu)$ は regular であるけれども，ν が確率変数の場合，たとえば尤度比 $\Lambda[P_{\theta+h/\sqrt{n}}, P_\theta]$ によって決定される（予備検定される）場合には ν は P_θ の下でと $P_{\theta+h/\sqrt{n}}$ の下でと異なる分布をするので定理2.3，定理3.1. で示唆されるように $\{\sqrt{n}\xi_n(T_n(\nu))\}$ が regular にならず したがって $T_n(\nu)$ は regular でない．したがって定理3.1 (ii) は成立しない．

定理 7.4.

擬線方程式

(7.5) $\xi_n(\theta) = \frac{\nu}{\sqrt{n}}\,\theta$

の解推定量 $T_n(\nu)$ が一致性をもつとき，

(7.6) $\sqrt{n}\,(T_n(\nu)-\theta) \longrightarrow N_k(\nu\Gamma(\theta)^{-1}\theta,\ \Gamma(\theta)^{-1}D(\theta)\Gamma(\theta)^{-1})$,

であり，$\{T_n(\nu)\}$ は正則である．

ν の影響は平均にはでるが分散には出ないというようなことがあっていいものであろうか．(7.1) 式においては有限標本数の段階で平均にも分散にも現われているのに (7.5),(7.6) では漸近分散には影響して来ない．漸近術にまどわされる ?!

定理 7.2.

一致稜線推定量 $\hat{\theta}_n(\nu)$ は次のような漸近的表現をもつ.

(7.1) $$\tilde{\theta}_n(\nu) = \left(\Gamma(\hat{\theta}_n) + \frac{\nu}{n} I\right)^{-1} \Gamma(\hat{\theta}_n)\, \hat{\theta}_n$$

とおくとき,

(7.2) $$\hat{\theta}_n(\nu) - \tilde{\theta}_n(\nu) \longrightarrow 0 \quad \text{in } P_\theta .$$

たしかに $\tilde{\theta}_n(\nu)$ もしたがって $\hat{\theta}_n(\nu)$ も一致性をもつ. 同様の議論から

(7.3) $$\sqrt{n}\,(\hat{\theta}_n(\nu) - \tilde{\theta}_n(\nu)) \longrightarrow 0 \quad \text{in } P_\theta$$

も示され, 漸近正規性をもつことが示される.

定理 7.3.

一致稜線推定量は漸近正規性をもつ.

$$\sqrt{n}\,(\hat{\theta}_n(\nu) - \theta) \longrightarrow N_k(0,\ \Gamma(\theta)^{-1} D(\theta)\, \Gamma(\theta)^{-1}) \quad \text{in law}.$$

証明

(7.3) 式からも出るが, 定理 2.2 を使って示そう.

$$\sqrt{n}\,\xi_n(\hat{\theta}_n(\nu)) = \frac{\nu}{\sqrt{n}}\, \hat{\theta}_n(\nu) \ , \quad \hat{\theta}_n(\nu) \longrightarrow \theta$$

によって,

$$\sqrt{n}\ \xi_n(\hat{\theta}_n(\nu)) \longrightarrow 0 \quad \text{in } P_\theta$$

ゆえに定理 2.2 (iii) から定理の結果を得る. ◁

ところで (5.3) 式のようなオーダ (n の代りに $\sqrt{n}$) を使えば,

$$\sqrt{n}\ \xi_n(T_n(\nu)) = \nu\, T_n(\nu) \longrightarrow \nu\,\theta \quad \text{in } P_\theta$$

§8. 縮小パラメタ ν の決定法

予測誤差を用いて稜線推定量の縮小パラメタνを決定する方法について考えよう．但し，以下の式における等号は漸近的な意味での等号であることに注意されたい．

稜線推定量の漸近的表現（7．1）式から，$X=\Gamma(\hat{\theta}_n)^{\frac{1}{2}}$として

$$(8.1)\quad X\cdot\hat{\theta}_n(\nu)=X\cdot(X'X+\tfrac{\nu}{n}I)^{-1}\cdot X'X\,\hat{\theta}_n$$

$$=Q\cdot X\hat{\theta}_n \qquad 但し\quad Q=X(X'X+\tfrac{\nu}{n}I)^{-1}X'.$$

であるので，（推定尤度関数を使ったとして $D(\theta)=\Gamma(\theta)$ とする）

$$(8.2)\quad n\cdot E|X\hat{\theta}_n(\nu)-X\theta|^2=nE|QX\hat{\theta}_n-QX\theta+(Q-I)X\theta|^2$$

$$=nE\,(X\hat{\theta}_n-X\theta)Q^2(X\hat{\theta}_n-X\theta)+n(X\theta)'(Q-I)^2(X\theta)$$

$$=\mathrm{trace}\;Q^2+n(X\theta)'(Q-I)^2(X\theta)$$

$$\underset{おく}{=}\;D(\nu)+B(\nu)$$

$B(\nu)$は偏り部分で，$D(\nu)$は分散部分である（Ingaki [6]）．ところが $B(\nu)$は未知母数 θを含むので，$B(\nu)$の不偏推定量でおきかえる．

$$n\cdot E(X\hat{\theta})'(I-Q)^2(X\hat{\theta})=n(X\theta)'(I-Q)^2(X\theta)+\mathrm{tr}\,(I-Q)^2$$

すなわち，不偏推定量は，

$$(8.3)\quad \hat{B}(\nu)=n\cdot(X\hat{\theta})'(I-Q)^2(X\hat{\theta})-\mathrm{trace}\,(I-Q)^2\;.$$

ゆえに予測誤差の推定量は

$$(8.4)\quad \hat{E}(\nu)=n\cdot(X\hat{\theta})(I-Q)^2(X\hat{\theta})+2\,\mathrm{trace}\;Q-k$$

を得る．一方，

(8.5) $-2\log\{f_n(x,\hat\theta_n(\nu))/f_n(x,\hat\theta_n)\}$

$$= -2\underbrace{\frac{\partial}{\partial\theta}\log f_n(x,\hat\theta_n)}_{\overset{\shortparallel}{n\,\xi_n^0(\hat\theta_n)=0}}(\hat\theta_n(\nu)-\hat\theta_n)$$

$$+\sqrt{n}(\hat\theta_n(\nu)-\hat\theta_n)'\left[-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log f_n(x,\hat\theta_n)\right]\sqrt{n}(\hat\theta_n(\nu)-\hat\theta_n)$$

(8.1) から

$$=\sqrt{n}(X(\hat\theta_n(\nu)-\hat\theta_n))'\sqrt{n}(X(\hat\theta_n(\nu)-\hat\theta_n))$$

$$=n\cdot(X\hat\theta_n)(I-Q)^2(X\hat\theta_n)$$

結局，予測誤差の推定量は(8.4)，(8.5)式から次のようになる．

定理8.1.

稜線推定量の予測誤差の推定量は

(8.6) $$\hat E_n(\nu)=-2\log\{f_n(x,\hat\theta_n(\nu))/f_n(x,\hat\theta_n)\}$$
$$+2\,\mathrm{trace}\{(\Gamma(\hat\theta_n)+\tfrac{\nu}{n}I)^{-1}\Gamma(\hat\theta_n)\}-k.$$

ゆえに，予測誤差の推定を最小にするνを求めるには

(8.7) $$E_n(\nu)=-2\log f_n(x,\hat\theta_n(\nu))+2\sum_{j=1}^{k}\frac{\lambda_i}{\lambda_i+\frac{\nu}{n}}-k$$

を最小にするνを求めればよい．ただし，$\lambda_1,\ldots,\lambda_k$は$\Gamma(\hat\theta_n)$（情報行列に推定量$\hat\theta_n$を代入したもの）の固有値である．

（以上）

参考文献

[1] Hajek,J.(1970). A characterization of limiting distributions of regular estimates, Z.Wahrsch.verw.Geb., 14, 323-330.

[2] Hoerl,A.E. and Kennard,R.W.(1970). Ridge regression: Biased estimation for non-orthogonal problems, Technometrics, 12, 55-67.

[3] Huber,P.J.(1967). The behavior of maximum likelihood estimators under nonstandard conditions, Proc. Fifth Berkeley Symp. Math.Stat.Prob., 1, 221-233.

[4] Inagaki,N.(1970). On the limiting distribution of a sequence of estimators with uniformity property, Ann.Inst.Stat.Math., 22, 1-13.

[5] __________(1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimators, Ann.Inst.Stat.Math.,25, 1-26.

[6] __________(1977). Two errors in statistical model fitting, Ann.Inst. Stat.Math., 29, 131-152.

[7] LeCam,L.(1960). Locally asymptotically normal families of distributions, Univ.California Publ. Stat., 3, 37-98.

[8] Prakasa Rao,B.L.S.(1980). Bounds for the equivalence of (modified) Bayes and maximum likelihood estimators for Markov processes, to appear.

[9] Strasser,H.(1977). Improved bounds for equivalence of Bayes and muximum likelihood estimation, Theory of Prob. and its Appli.,XXII,349-361.

[10] Wilks,S.(1962). Mathematical Statistics, John Wiley & Sons, Inc., New York.