セミパラメトリックモデルの統計的漸近理論

今野 良彦
千葉大学大学院自然科学研究科

はじめに

このノートは統計的推測理論の漸近理論をまとめたものである．

目 次

第I部

情報限界と有効推定

第1章　極限定理、弱収束および Tightnesss

1.1　確率変数の収束

定義 1.1 ：Ω の部分集合からなる族 $\boldsymbol{F}$ が **$\boldsymbol{\sigma}-$ 加法族**とは

(i) $\Omega \in \boldsymbol{F}$,

(ii) $A \in \boldsymbol{F}$ ならば、$A^c = \Omega \setminus A \in \boldsymbol{F}$,

(iii) $A_n \in \boldsymbol{F} \quad (n = 1, 2, \ldots)$ ならば、$\bigcup_{n=1}^{\infty} A_n \in \boldsymbol{F}$

を満たすときをいう．

定義 1.2 ：P が $(\Omega, \boldsymbol{F})$ 上の**確率測度**とは

(i) $P : \boldsymbol{F} \mapsto [0, 1]$ かつ $P(\phi) = 0, P(\Omega) = 1.$

(ii) $A_n \in \boldsymbol{F}(n = 1, 2, \ldots)$ が互いに素ならば $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

のときをいう．

問い 1.1 ：$A_1 \supset A_2 \supset \cdots$ に対して、$P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$ が成立することを示せ．

補題 1.1 （Borel – Cantelli の補題）：$A_1, A_2, \ldots$ を事象の列とする．
(i). $\sum_{n=1}^{\infty} P(A_n) < \infty$ ならば、$P(\limsup_{n\to\infty} A_n) = 0$ ．
(ii). $A_1, A_2, \ldots$ が独立で、$\sum_{n=1}^{\infty} P(A_n) = \infty$ ならば、 $P(\limsup_{n\to\infty} A_n) = 1$ ．

証明：(i). $P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k)$ より、

$$0 \leq P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

(ii). $A_1, A_2, \ldots$ が独立ならば、$A_1^c, A_2^c, \ldots$ も独立．$m \geq n$ で

$$P(\bigcap_{k=n}^{\infty} A_k^c) = \lim_{m\to\infty} P(\bigcap_{k=n}^{m} A_k) = \lim_{m\to\infty} \prod_{k=n}^{m} P(A_k^c)$$

を得る．ここで、$1-x \le e^{-x}(x \ge 0)$ から、

$$\lim_{m\to\infty}\prod_{k=n}^{m} P(A_k^c) = \lim_{m\to\infty}\prod_{k=n}^{m}\{1-P(A_k)\} \le \exp\left(-\sum_{k=n}^{m} P(A_k)\right) = 0$$

つまり、

$$0 = \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k^c) = P(\liminf_{n\to\infty} A_n^c)$$

が成立．これより、

$$1 = P(\{\liminf_{n\to\infty} A_n^c\}^c) = P(\limsup_{n\to\infty} A_n)$$

がわかる．

定義 1.3 （法則収束）: 確率変数の列 $\{X_n\}$ が確率変数 X に**法則収束**するとは、すべての有界連続関数 g に対して、

$$\boldsymbol{E}[g(X_n)] \to \boldsymbol{E}[g(X)], \qquad n \to \infty$$

となることをいう．これを $X_n \xrightarrow{d} X$ または $\boldsymbol{L}(X_n) \longrightarrow \boldsymbol{L}(X)$ と記す．

注意 1.1 ：特に、X_n が $\boldsymbol{R}-$ 値確率変数ならば、X の分布関数 F のすべての連続点 x に対して、

$$F_n(x) \equiv P\{X_n \le x\} \longrightarrow F(x), \qquad n \to \infty$$

が成立することと同値である．

定義 1.4 （確率収束）: 確率変数の列 $\{X_n\}$ が確率変数 X に**確率収束**するとは、

$$P\{|X_n - X| > \epsilon\} \longrightarrow 0, \qquad n \to \infty$$

がすべての $\epsilon > 0$ に対して成立することである．

定義 1.5 （r 次平均収束）: 確率変数の列 $\{X_n\}$ が確率変数 X に$\boldsymbol{r}-$ **次平均収束**するとは、

$$\boldsymbol{E}|X|^r < \infty, \qquad \boldsymbol{E}|X_n|^r < \infty, \qquad n = 1, 2, \ldots,$$

であって、

$$\boldsymbol{E}|X_n - X|^r \longrightarrow 0, \qquad n \to \infty$$

となることである．

定義 1.6 （概収束）: 確率変数の列 $\{X_n\}$ が確率変数 X に**概収束**するとは、$P(A) = 0$ を満たす Ω の部分集合 A があって、

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \qquad \omega \in \Omega \backslash A$$

となることである．

定義 1.7 （一様可積分）: 確率変数の列 $\{X_n\}$ が**一様可積分**であるとは、

$$\lim_{a\to\infty}\sup_n \boldsymbol{E}[|X_n|1\{|X_n|\geq a\}]$$

となることである．

定理 1.1 ：確率変数の列 $\{X_n\}$ が一様可積分となるための必要十分条件はつぎのふたつである．

(i) $\sup \boldsymbol{E}|X_n| < \infty$ ．

(ii) 任意の $\epsilon > 0$ に対し、$\delta = \delta(\epsilon) > 0$ が存在し、$P(A) < \delta$ となる任意の A に対し、

$$\boldsymbol{E}|X_n|1_A < \epsilon, \qquad n = 1, 2, \ldots$$

をみたすことができる．

定義 1.8 ：関数 $u : [0,\infty) \to [0,\infty)$ が**一様可積分性判定関数**であるとは、u は増加、凸[1] かつ

$$\lim_{x\to\infty}\frac{u(x)}{x} = \infty$$

となることである．

たとえば、$u(x) = x^p (p > 1)$ は判定関数であるが、$u(x) = x$ はそうでない．

定理 1.2 ：関数族 $\{f_j\}_{j\in J}$ は

$$\sup_{j\in J}\left\{\int u(|f_j|)dP\right\} < \infty$$

を満たす判定関数 u が存在する；とき、およびそのときに限り一様可積分となる

定理 1.3 （有界収束定理）: $X_n \xrightarrow{P} X$ かつ $|X_n| \leq Y$ （Y は可積分）ならば、X は可積分で、

$$\boldsymbol{E}X_n \longrightarrow \boldsymbol{E}X, \qquad n \to \infty$$

である．

定理 1.4 （単調収束定理）: $\{X_n\}$ を非負確率変数の増大列とするならば、$X_n \xrightarrow{a.e.} X$ ならば、$\boldsymbol{E}X_n \longrightarrow \boldsymbol{E}X, (n \to \infty)$ である．

定理 1.5 （*Lebesgue-Fatou* の *lemma*）: $\{X_n\}$ を非負確率変数列とすれば、

$$\boldsymbol{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \boldsymbol{E}[X_n]$$

である．

---

[1] すなわち、すべての $x, y \in [0,\infty), \lambda \in [0,1]$ に対し、

$$u(\lambda x + (1-\lambda)y) \leq \lambda u(x) + (1-\lambda)u(y)$$

となることである．

定理 1.6 （Fubini の定理）: 直積測度空間 $(\boldsymbol{X}, \mathcal{A}, \mu) = (\boldsymbol{X}_1 \times \boldsymbol{X}_2, \mathcal{A}_1 \times \mathcal{A}_2, \mu_1 \times \mu_2)$ 上の可測関数 $f(x_1, x_2)$ が $f \geq 0$ または $\int |f| d\mu < \infty$ ならば、

$$\int_X f d\mu = \int_{X_1} d\mu_1 \int_{X_2} f(x_1, x_2) d\mu_2$$

が成立する．

定理 1.7 ：f_n と f は μ – 可測関数でつぎを満足するものとする．
(i)．μ に関してほとんどいたるところで $f_n \to f$．
(ii)．ある $p \geq 1$ に対し、

$$\limsup_{n\to\infty} \int |f_n|^p d\mu \leq \int |f|^p d\mu < \infty.$$

このとき、

$$\int |f_n - f|^p d\mu \to 0$$

が成立する．

証明．$a, b \geq 0$ に対して、$|a-b|^p \leq 2^p|a|^p + 2^p|b|^p$ が成立することに注意すれば、ほとんどいたるところで

$$0 \leq 2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p \to 2^{p+1}|f|^p$$

が成り立つ．Fatou の補題から

$$\begin{aligned}
\int 2^{p+1}|f|^p d\mu &= \int \liminf_{n\to\infty}(2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p) d\mu \\
&\leq \liminf_{n\to\infty} \int (2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p) d\mu \\
&\leq \limsup_{n\to\infty} \int 2^p|f_n|^p d\mu + \int 2^p|f|^p d\mu - \limsup_{n\to\infty} \int |f_n - f|^p d\mu \\
&\leq 2^{p+1} \int |f|^p d\mu - \limsup_{n\to\infty} \int |f_n - f|^p d\mu
\end{aligned}$$

がわかる．よって

$$\limsup_{n\to\infty} \int |f_n - f|^p d\mu \leq 0$$

から

$$\int |f_n - f|^p d\mu \to 0$$

がわかる． □

1.2 種々の収束の関係

定理 1.8 ：$\{X_n\}$ を確率変数の列、X を確率変数とする．このとき、つぎのことが成立する．
A. $\{X_n\}$ が X に概収束するならば、$\{X\}$ は X に確率収束する．
B. $\{X_n\}$ が X に確率収束するならば、$\{X\}$ は X に法則収束する．
C. $\{X_n\}$ が X に２次平均収束するならば、$\{X\}$ は X に法則収束する．

注意 1.2：上の定理の逆は一般には成立しない．

定理 1.9：$X_n \xrightarrow{d} X$ とする．このとき、つぎが成立する．
$A.$ $\{X_n\}$ が一様可積分ならば、$\boldsymbol{E}X_n \longrightarrow \boldsymbol{E}X$ である．
$B.$ $\{X_n\}$ が非負確率変数列で、X は可積な非負の確率変数とする．このとき、$\boldsymbol{E}X_n \longrightarrow \boldsymbol{E}X$ ならば、$\{X_n\}$ は一様可積分である．

証明．Serfling の page 14 を参照のこと． □

定理 1.10：$r > 0$ とする．$X_n \overset{r-\text{th}}{\longrightarrow} X$ となるための必要十分条件はつぎのふたつである．

(i) $X_n \xrightarrow{P} X.$

(ii) $\boldsymbol{E}|X_n|^r \longrightarrow \boldsymbol{E}|X|^r.$

証明．(⇐)．c_r- 不等式[2] より、

$$|X_n - X|^r \leq c_r|X_n|^r + c_r|X|^r \equiv Y_n \geq 0$$

となる．(i) と (ii) より、

$$Y_n \xrightarrow{P} 2c_r|X|^r \quad \text{かつ } \boldsymbol{E}Y_n \longrightarrow 2c_r\boldsymbol{E}|X|^r \tag{1.1}$$

となる．$Y = 2c_r|X|^r$ とおき、」(1.1) と有界収束定理をもちいれば

$$\begin{aligned}\boldsymbol{E}[Y_n 1\{|Y_n| \geq a\}] &= \boldsymbol{E}Y_n - \boldsymbol{E}[Y_n 1\{|Y_n| < a\}] \\ &\longrightarrow \boldsymbol{E}Y - \boldsymbol{E}[Y 1\{|Y_n| < a\}] \\ &= \boldsymbol{E}[Y 1\{|| \geq a\}] \downarrow 0 \qquad a \to \infty \text{ のとき}\end{aligned}$$

となり、$\{Y_n\}$ は一様可積分であるので、$|X_n^X|^r$ も一様可積分であることがわかる． □

定理 1.11 (Portmanteau)：X_n と X を任意の確率ベクトルとしたとき、つぎの条件は同値である．
(i)．$P(X_n \leq x) \to P(X \leq x)$ が $x \mapsto P(X \leq x)$ のすべての連続点に関して成立する．ただし、上の式の $X \leq x$ は成分ごとの比較である．
(ii)．すべての有界連続関数 f に対し、

$$\boldsymbol{E}[f(X_n)] \to \boldsymbol{E}[f(X)]$$

が成立する．
(iii)．すべての有界 Lipschitz [3] 関数に対し、

$$\boldsymbol{E}[f(X_n)] \to \boldsymbol{E}[f(X)]$$

[2] Loéve の page 153
[3] 関数 f が Lipschitz であるとは、ある $L > 0$ が存在し、すべての x と y に対し、

$$| f(x) - f(y) | \leq L \mid x - y \mid$$

が成立することである．

が成立する．
(iv)．すべての非負連続関数 f に対し、

$$\liminf \boldsymbol{E}[f(X_n)] \geq \boldsymbol{E}[f(X)]$$

が成立する．
(v)．すべての開集合 G に対し、

$$\liminf P(X_n \in G) \geq P(X \in G)$$

が成立する．
(vi)．すべての閉集合 F に対し、

$$\limsup P(X_n \in F) \leq P(X \in F)$$

が成立する．
(vii)．すべての $P(X \in \delta B) = 0$ なる Borel 集合 B に対し、

$$P(X_n \in B) \to P(X \in B)$$

が成立する．ただし、δB は B の境界、すなわち、B の閉包と内点との差である．

証明：(i) $\Rightarrow$ (ii). どんな x に対しても、ある $A > 0$ が存在して $|f(x)| \leq A$ とできる．また、どんな $\epsilon > 0$ に対しても、ある $B > 0$ が存在して、$P(|X| \geq B) \leq \epsilon/(2A)$ ともできる．h を実数値連続関数とし、$0 \leq h(x) \leq 1$ でかつ

$$h(x) = \begin{cases} 0 & (\,|x| \geq B+1\,) \\ 1 & (\,|x| \leq B+1\,) \end{cases}$$

を満足するものとする．このとき、

$$\begin{aligned} |\boldsymbol{E}[f(X_n)] - \boldsymbol{E}[f(X)]| &\leq |\boldsymbol{E}[f(X_n)] - \boldsymbol{E}[f(X_n)h(X_n)]| \\ &\quad + |\boldsymbol{E}[f(X_n)h(X_n)] - \boldsymbol{E}[f(X)h(X)]| + |\boldsymbol{E}[f(X)h(X)] - \boldsymbol{E}[f(X_n)]| \\ &\leq \frac{\epsilon}{2} + |\boldsymbol{E}[f(X_n)h(X_n)] - \boldsymbol{E}[f(X)h(X)]| \end{aligned}$$

となる．したがって、コンパクトな台 I をもつ連続関数 g に対して、

$$|\boldsymbol{E}[g(X_n)] - \boldsymbol{E}[g(X)]| < \frac{\epsilon}{2}$$

が成立することを示せばよい．I はコンパクトで、g は一様連続なので、有限個の矩形領域 I_j でその上での g の変動が $\epsilon/12$ 以下になり、$I \subset \cup_j I_j$ とできる．それぞれの I_j から一点 x_j を取り出し、$g_\epsilon(x) = \sum_j g(x_j) 1_{I_j}(x)$ とする．すると、

$$|\boldsymbol{E}[g(X_n)] - \boldsymbol{E}[g_\epsilon(X_n)]| \leq \frac{\epsilon}{12} + P(X_n \notin I) \leq \frac{\epsilon}{6}$$
$$|\boldsymbol{E}[g(X)] - \boldsymbol{E}[g_\epsilon(X)]| \leq \frac{\epsilon}{12} + P(X \notin I) \leq \frac{\epsilon}{6}$$

となる[4]．さらに、n を十分おおきくとれば、

$$|\boldsymbol{E}[g_\epsilon(X_n)] - \boldsymbol{E}[g_\epsilon(X)]| \le \sum_j |P(X_n \in I_j) - P(X \in I_j)||g(x_j) \le \frac{\epsilon}{6}$$

とできることからわかる．

あとは、Van der Vaart (1999) の pages 6–7 を参照されたし． □

定義 1.9 ：確率変数 X が tight であるとは、すべての $\epsilon > 0$ に対し、ある定数 $M > 0$ が存在して、

$$P(\|X\| > M) < \epsilon$$

を満足することである．また、確率変数の集まり $\{X_\alpha : \alpha \in A\}$ が**一様に** tight　dであるとは、すべての $\epsilon > 0$ に対し、ある定数 $M > 0$ が存在し、

$$\sup_\alpha P(\|X\| > M) < \epsilon$$

を満足することである．

定理 1.12 （Prohorov）：X_n は k 次元確率変数とする．

(i)．$X_n \xrightarrow{d} X$ （ある確率変数 X ）ならば、$\{X_n : n \in \boldsymbol{N}\}$ は一様に tight である．

(ii)．$\{X_n : n \in \boldsymbol{N}\}$ は一様に tight ならば、ある部分列 $\{X_{n_j}\}$ が存在し、$j \to \infty$ のとき、

$$X_{n_j} \xrightarrow{d} X \qquad \text{（ある確率変数 } X\text{）}$$

を満足する．

証明：Van der Vaart (1999) の pages 8 – 9 を参照されたい． □

定理 1.13 ：

A. $C(g) = \{x \in \boldsymbol{R}^d; g : \boldsymbol{R}^d \mapsto \boldsymbol{R}^k$ は x で連続$\}$ とし、$X_n \in \boldsymbol{R}^d$ とする．更に、$P\{X \in C(g)\} = 1$ を仮定する．このとき、

(i) $X_n \xrightarrow{a.e.} X \Rightarrow g(X_n) \xrightarrow{a.e.} g(X)$.

(ii) $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$.

(iii) $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$.

B. $X_n \xrightarrow{d} X$ かつ $(X_n - Y_n) \xrightarrow{P} 0$ ならば、

$$Y_n \xrightarrow{d} X.$$

C. $X_n \in \boldsymbol{R}^d$、$Y_n \in \boldsymbol{R}^k$ とする．このとき、$X_n \xrightarrow{d} X$ かつ $Y_n \xrightarrow{d} c$ 定数 ならば、

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ c \end{bmatrix}.$$

証明．A は Serfling (1989) の page 24 を参照．B と C については Ferguson (1996) の page 41 を参照． □

[4] 法則収束することより、十分おおきな n をとれば、$P(X_n \in I) < \epsilon/12$ とできることがわかる

1.3　対数の法則と中心極限定理

定理 1.14 ：$X_1, X_2, \ldots, X_n$ は独立に同一の分布に従うとする．また、

$$\boldsymbol{E}|X_1| < \infty, \qquad \boldsymbol{E}[X_1] = \mu$$

であるとする．このとき、

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

となる．

証明. 略．

系 1.1 ：$X_1, X_2, \ldots, X_n$ は独立に同一の分布 P に従うとする．x を固定して、$p = P1\{X_1 \leq x\}$ とおいたとき、

$$\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} \xrightarrow{a.s.} p$$

となる．**証明**．$Z_i = 1\{X_i \leq x\} - p$ とおく．すると、$Z_1, Z_2, \ldots, Z_n$ は独立な確率変数の列となる．また、

$$\begin{aligned} g(t) &= \log \boldsymbol{E}[\exp(tZ_i)] \\ &= -pt + \log(pe^t(1-p)) \end{aligned}$$

となる．いま、任意の $\epsilon > 0$ をとり、

$$B_n = \left\{\omega \in \Omega \;\middle|\; \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} > p + \epsilon\right\}$$

とおく．すると、$t > 0$ に対して、

$$\begin{aligned} P(B_n) &= P\left(\sum_{i=1}^{n} tZ_i > n\epsilon \cdot t\right) \\ &= P\left(\Pi_{i=1}^{n} \exp(tZ_i) > \exp(n\epsilon \cdot t)\right) \\ &\leq \exp(-n\epsilon \cdot t)\boldsymbol{E}[\Pi_{i=1}^{n} \exp(tZ_i)] && \text{（マルコフの不等式）} \\ &= \exp(-n\epsilon \cdot t)\Pi_{i=1}^{n}\boldsymbol{E}[\exp(tZ_i)] && \text{（独立性）} \\ &= \{\exp(-\epsilon t + g(t))\}^n \end{aligned}$$

となる．容易にわかるように、

$$g(0) = 0, \qquad g'(0) = 0$$

であるので、$t > 0$ が十分小さければ、

$$a = -\epsilon t + g(t) < 0$$

となる．これより、

$$\sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} e^{na} < \infty$$

を得る．Borel - Cantelli の定理から

$$P\left(\bigcap_{i=1}^{\infty}\left(\bigcup_{j=i}^{\infty} B_j\right)\right) = 0$$

がわかる．これをいいかえれば、

$$P\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} \leq p + \epsilon\right) = 1$$

を意味している．$\epsilon > 0$ は任意だったので、これより

$$P\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} \leq p\right) = 1$$

を得る．同様にして、

$$P\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} \geq p\right) = 1$$

を得ることができる．したがって、

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\} = p\right) = 1$$

を得る．

定理 1.15 ：$\{X_n\}_{n=1}^{\infty}$ を独立な確率変数の列とし、それぞれの X_n の分布関数を $F_n(\cdot)$、平均を μ_n、分散を $\sigma_n^2(0 < \sigma_n^2 < \infty)$ とする．$S_n = \sum_{i=1}^{n} X_i$、$B_n^2 = \sum_{i=1}^{n} \sigma_i^2$ とおく．すべての $\epsilon > 0$ で

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{i=1}^{n}\int_{[|x-\mu_i|>\epsilon B_n]}(x-\mu_i)^2 dF_i(x) = 0 \tag{1.2}$$

ならば、

$$\lim_{n\to\infty}\max_{1\leq i\leq n}\frac{\sigma_i^2}{B_n^2} = 0 \tag{1.3}$$

かつ

$$\frac{S_n - \boldsymbol{E}(S_n)}{B_n} \xrightarrow{d} N(0,1) \tag{1.4}$$

である．

逆に、(1.3) と (1.4) が成立するならば、(1.2) が成り立つ．

1.4　距離空間上の弱収束と Tightness

$(\boldsymbol{M}, d)$ を可分な距離空間とし、$\mathcal{M}_B$ を M の Borel-集合体[5] とする．$\boldsymbol{C}_b(M)$ によって、M 上の有界連続な実数値関数の全体をあらわす．

定理 1.16 ：P を $(M, \mathcal{M}_B)$ 上の確率測度とする．このとき、任意の $B \in \mathcal{M}_B$ に対し、

$$P(B) = \sup_{F \subset B}\{P(F); F \text{ は閉集合 }\} = \inf_{B \subset G}\{P(G); G \text{ は開集合 }\} \tag{1.5}$$

が成り立つ．

証明：

$$\mathcal{U} = \{B \in \mathcal{M}_B : (1.5) \text{ が成立 }\}$$

とすれば、$\mathcal{U}$ は Borel-集合体であることが示せる．[6]

いま、G を開集合として、

$$F_n = \{x : d(x, G^c) \geq \frac{1}{n}\}$$

とおけば[7] 、F_n は単調増大な閉集合族で、$\bigcup F_n = G$ となる．すると、

$$P(G) = \lim_{n \to \infty} P(F_n)$$

となること（すなわち、G は (1.5) を満足）より、$G \in \mathcal{U}$ となる．$\mathcal{U}$ は開集合を含むことより、$\mathcal{M}_B \subset \mathcal{U}$ となる．よって、$\mathcal{U} = \mathcal{M}_B$ が成立．

定理 1.17 ：P, Q をそれぞれ $(M, \mathcal{M}_B)$ 上の確率測度とする．もし、

$$\int_M f(x)dP(x) = \int_M f(x)dQ(x)$$

がすべての $f \in \boldsymbol{C}_b$ について成立すれば、$P = Q$ である．

証明：上の定理より、任意の閉集合 F に対して、$P(F) = Q(F)$ を示せば十分．ところで、$f_n \in \boldsymbol{C}_b, n = 1, 2, \ldots$ で

$$0 \leq f_n \leq 1 \quad \text{かつ} \quad \lim f_n(x) = 1_F(x)$$

なるものが存在する[8] ．すると、有界収束定理より、

$$P(F) = \lim \int_M f_n(x)dP(x) = \lim \int_M f_n(x)dQ(x) = Q(F)$$

が成立することがわかる．

[5] $\boldsymbol{M}$ のすべての開集合を含む最小の σ – 加法族

[6] 確率微分方程式（渡辺、産業図書）の page 155 を参照

[7] $d(x, G^c) = \inf\{d(x, y) : y \in G^c\}$

[8] 実際、

$$\varphi(t) = \begin{cases} 1 & (t \leq 0) \\ 1 - t & (0 \leq t \leq 1) \\ 0 & (t \geq 1) \end{cases}$$

とし、$f_n(x) = \varphi(nd(x, F))$ とすればよい．

定義 1.10 ：$P_n (n = 1, 2, \ldots), P$ がそれぞれ $(\boldsymbol{M}, \mathcal{M}_B)$ 上の確率測度の列とする．このとき P_n が P に**弱収束**するとは、任意の $f \in \boldsymbol{C}_b(\boldsymbol{M})$ に対し、つぎのようになることである．

$$\int_M f dP_n \to \int_M f dP \qquad n \to \infty$$

これを記号で $P_n \longrightarrow_d P$ とあらわす．

定義 1.11 ：$X_n (n = 1, 2, \ldots,), P$ をすべて $\boldsymbol{M}$ - 値の *random elements* とし、任意の $f \in \boldsymbol{C}_b(\boldsymbol{M})$ に対し、

$$\boldsymbol{E}[f(X_n)] \to \boldsymbol{E}[f(X)] \qquad n \to \infty$$

を満足するとき、X_n は X に**法則収束**するといい、$X_n \longrightarrow_d X$ と記す．

命題 1.1 ：すぎの４条件は同値である．
A. $P_n \longrightarrow_d P$ ．
B. 任意の閉集合 F に対して、$\limsup_{n\to\infty} P_n(F) \le P(F)$ ．
C. 任意の開集合 G に対して、$\liminf_{n\to\infty} P_n(G) \ge P(G)$ ．
$A \in \mathcal{M}_B$ で $P(\partial A) = 0$ なるものに対して、$\lim_{n\to\infty} P_n(A) = P(A)$.

証明：$A \Rightarrow B$ はつぎのようにしてわかる．関数 $f_k(x) = \varphi(kd(x, F))$ は $\boldsymbol{C}_b(\boldsymbol{M})$ に属し、F が閉集合ならば、$f_k(x)$ は $k \to \infty$ のとき $1_F(x)$ に各点収束することより、

$$\limsup_{n\to\infty} P_n(F) \le \lim_{n\to\infty} \int_M f_k dP_n = \int_M f_k dP$$

を得る．ここで $k \to \infty$ とすると有界収束定理より右辺は $P(F)$ に収束する．$B \Leftrightarrow C$ は互いに補集合をとればよい．つぎに $C \Rightarrow A$ を示す．$f \in \boldsymbol{C}_b(\boldsymbol{M})$ としたとき、一般性を失わず $0 \le f \le 1$ とできる．各整数 k に対して、

$$\sum_{i=1}^{k} \frac{i-1}{k} P\left\{x; \frac{i-1}{k} \le f(x) < \frac{i}{k}\right\} \le \int_M f dP \le \sum_{i=1}^{k} \frac{i}{k} P\left\{x; \frac{i-1}{k} \le f(x) < \frac{i}{k}\right\}$$

である．$F_i = \{x; i/k \le f(x)\}$ とおけば、

$$\frac{1}{k} \sum_{i=1}^{k} P(F_i) \le \int_M f dP \le \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} P(F_i)$$

と書ける．同様にして、

$$\frac{1}{k} \sum_{i=1}^{k} P_n(F_i) \le \int_M f dP_n \le \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} P_n(F_i)$$

を得る．この２式と B より、

$$\limsup_{n\to\infty} \int_M f dP_n \le \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \limsup_{n\to\infty} P_n(F_i) \le \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} P(F_i)$$

を得、これより

$$\limsup_{n\to\infty}\int_M fdP_n \leq \int_M fdP$$

がわかる．f を $1-f$ と置き換えると

$$\liminf_{n\to\infty}\int_M fdP_n \geq \int_M fdP$$

となり、したがって、A が成立する．

注意 1.3 ：上で定義した $(\boldsymbol{M},\mathcal{M}_B)$ 上の確率測度の弱収束はある距離に関する収束となる．すなわち、$\boldsymbol{P}$ を $(\boldsymbol{M},\mathcal{M}_B)$ 上の確率測度の全体とすれば、$\boldsymbol{P}$ は距離化可能である．その距離として、$P,Q\in\boldsymbol{P}$ とすれば、Prohorov metric

$$d_P(P,Q)=\inf\{\epsilon>0: P(A)\leq Q(A^\epsilon)+\epsilon \quad \text{すべての } A\in\mathcal{M}_B \text{ に対して}\}$$

や bounded Lipschitz metric

$$d_{BL}(P,Q)=\sup_{f\in BL_1}\left|\int fdp-\inf fdQ\right|$$

がある．ここで、$A^\epsilon=\{x\in\boldsymbol{M}:\inf_{y\in A}d(x,y)\leq\epsilon$、$BL_1$ は $\parallel f\parallel_\infty=1$ で Lipsshipz 関数[9] の集まりである．詳しい議論は Huber (1981) の 2 章を参照されたし．

定義 1.12 ：$\boldsymbol{P}$ を $(\boldsymbol{M},\mathcal{M}_B)$ 上の確率測度の集まりとする．このとき、$\boldsymbol{P}$ が **tight**（または**一様に tight** ともいう）であるとはつぎのことが成り立つことである．どんな $\epsilon>0$ に対してもコンパクト集合 $K\subset\boldsymbol{M}$ が存在して、

$$P(K)\geq 1-\epsilon$$

がすべての $P\in\boldsymbol{P}$ に対して成り立つ．すなわち、

$$\inf_{P\in\boldsymbol{P}}P(K)\geq 1-\epsilon$$

が成り立つ．

定義 1.13 ：$\boldsymbol{P}$ を $(\boldsymbol{M},\mathcal{M}_B)$ 上の確率測度の集まりとする．$\boldsymbol{P}$ が**相対コンパクト**（*relative compact*）であるとは、$\boldsymbol{P}$ のすべての列 $\{P_n\}$ が弱収束する部分列を含むことである．よって、$\{P_n\}\subset\boldsymbol{P}$ はつぎを満足する部分列 $\{P_{n'}$ を含む．

$$P_{n'}\longrightarrow_d \text{ ある } Q \text{（}\boldsymbol{P} \text{ に含まれるとは限らない）}$$

[9] ある $K<\infty$ が存在して、任意の $x,y\in\boldsymbol{M}$ に対して、

$$|f(x)-f(y)|\leq Kd(x,y)$$

が成立することである．

命題 1.2 ：
A.（Le Cam） $P_n \longrightarrow_d P$ ならば、$\{P_n\}$ は tight である．
B. $P_n \longrightarrow_d P$ ならば、$\{P_n\}$ は相対コンパクトである．
C. $\{P_n\}$ が相対コンパクトで、任意の部分列 $\{P_{n'}\} \subset \{P_n\}$ に対して、 $\{P_{n'}\}$ に含まれる弱収束する部分列 $\{P_{n''}\}$ の収束先が 1 点 P ならば、

$$P_n \longrightarrow_d P \qquad n \to \infty$$

が成立する．

定理 1.18 （Prohorov）: $\boldsymbol{P}$ を $(\boldsymbol{M}, \mathcal{M}_B)$ 上の確率測度の集まりとする．
A. $\boldsymbol{P}$ が *tight* ならば、$\boldsymbol{P}$ は相対コンパクトである．
B. $(\boldsymbol{M}, \mathcal{M}_B)$ が可分[10] で完備な距離空間[11] のとき、$\boldsymbol{P}$ が相対コンパクトならば、$\boldsymbol{P}$ は *tight* である．

証明: A については渡辺 (1975, 付録 I, pages 161–162) を参照．B については Billingsley (1968, section 8, pages 35–40) を参照されたい．

定理 1.19 （Skorohod）: $(\boldsymbol{M}, d)$ は可分、完備な距離空間とし、$P_n (n = 1, 2, \ldots), P$ を $(\boldsymbol{M}, \mathcal{M}_B)$ 上の確率測度で、$P_n \xrightarrow{d} P (n \to \infty)$ なるのもとする．このとき、適当な確率空間 $(\Omega^*, \boldsymbol{A}^*, P^*)$ 上の $\boldsymbol{M}$ – 値 *random elements* $X_n^* (n = 1, 2, \ldots), X^*$ をつぎのように構成できる．

(i) $P_{X_n^*} = P_n (n = 1, 2, \ldots), \qquad P_{X^*} = P$ ．

(ii) X_n^* は X^* に概収束する．すなわち、

$$P^*\{\omega \in \Omega^* * d(X_n^*(\omega), X^*(\omega)) \to 0 \quad (n \to \infty)\} = 1$$

が成立．

とくに、$\boldsymbol{M}$ – 値 *random elements* $X_n (n = 1, 2, \ldots), X$ に対し、

$$X_n \xrightarrow{d} X \qquad (n \to \infty)$$

ならば、適当な確率空間上に $X_n^* (n = 1, 2, \ldots), X^*$ を

$$\boldsymbol{L}(X_n) = \boldsymbol{L}(X_n^*), \quad \boldsymbol{L}(X) = \boldsymbol{L}(X^*) \quad \text{かつ} \quad X_n^* \xrightarrow{a.s.} X^* \qquad (n \to \infty)$$

となるように構成できる．すなわち、法則収束を概収束で表現できる．

---

[10] norm 空間 $\boldsymbol{M}$ が可分（separable）であるとは、$\boldsymbol{M}$ の部分集合 S でつぎの条件を満足するものが存在することである．

(i) $\bar{S} = \boldsymbol{M}$、すなわち、S は $\boldsymbol{M}$ で稠密．

(ii) S は可算集合．

すなわち、$\boldsymbol{M}$ の中の可算集合 $S = \{s_1, s_2, \ldots\}$ を選んで、任意の $m \in \boldsymbol{M}$ が S の触点になるようにできることである．

[11] 任意の Cauchy 列は収束する．

証明：渡辺 (1975, 付録 I, pages 163–166) を参照されたい．
Skorohod の定理よりつぎの結果を得る．

定理 1.20 （連続写像定理）：$\boldsymbol{M}$ を可分とし、$(\boldsymbol{M}', d')$ を別の距離空間とする． $X_n(n = 1, 2, \ldots), X$ を $\boldsymbol{M}$ – 値 *random elements* とし、写像 $g : \boldsymbol{M} \mapsto \boldsymbol{M}'$ を $P = P_X$ に関してほとんどいたるところ連続[12] とする．このとき、$g(X_n) \xrightarrow{d} g(X)$ が成立する．

[12] すなわち、

$$C(g) = \{x \in \boldsymbol{M} :\ g \text{ は } x \text{ において連続 }\}$$

とおいたとき、$P(X \in C(g)) = 1$ が成立する．

第2章　接触性（ Contiguity ）定理と Le Cam の３つの補題

2.1　接触性の定義

次のような統計的問題を考える．

測度空間：　$(\boldsymbol{X}_n, \boldsymbol{A}_n, \mu_n)$

確率測度：　$P_n \ll \mu_n, \quad Q_n \ll \mu_n$

確率密度関数：　$p_n \equiv \frac{dP_n}{d\mu_n}, \quad q_n \equiv \frac{dQ_n}{d\mu_n}$

尤度比：　$L_n = \begin{cases} q_n/p_n & (p_n > 0) \\ 1 & (p_n = q_n = 0) \\ \infty & (q_n > 0 = p_n) \end{cases}$

定義 2.1 ：確率分布の列 $\{Q_n\}$ が $\{P_n\}$ に対して**接触 （ contigious ）**であるとは、任意の $\{B_n\}$ （ $B_n \in \boldsymbol{A}_n$ ）に対して $P_n(B_n) \to 0$ ならば、$Q_n(B_n) \to 0$ が成立することである．これを $\{Q_n\} \lhd \{P_n\}$ と記す．また、もし、$\{Q_n\} \lhd \{P_n\}$ と $\{P_n\} \lhd \{Q_n\}$ が共に成立するとき、$\{P_n\}$ と $\{Q_n\}$ は**互いに接触**するといい、$\{P_n\} \lhd \rhd \{Q_n\}$ と記す．

注意 2.1 $\{P_n\}$ に対する $\{Q_n\}$ の接触性は *dominant measure* に関して、$\{Q_n\}$ は $\{P_n\}$ に対して漸近的に絶対連続であることを意味する．

接触性の逆の概念として、以下のものがある．

定義 2.2 ：確率分布の列 $\{Q_n\}$ が $\{P_n\}$ に対して**漸近的に直交** であるとは、ある列 $\{B_n\}$ （ $B_n \in \boldsymbol{A}_n$ ）が存在して、

$$Q_n(B_n) \to 1 \quad \text{かつ} \quad P_n(B_n) \to 0$$

が成立することである．

2.2　Le Cam の補題

補題 2.1 （*Le Cam* の第 1 補題）:

$$\boldsymbol{L}(L_n \mid P_n) \xrightarrow{d} \boldsymbol{L}(L) \quad \text{かつ} \quad \boldsymbol{E}(L) = 1$$

ならば、$\{Q_n\} \lhd \{P_n\}$ が成立する，

証明：$P(B_n) \to 0$ となる $B_n \in \boldsymbol{A}_n$ を選ぶ．すると、P_n のもとで

$$1\{\boldsymbol{X}_n - B_n\} \to 1$$

となるので、$1\{\boldsymbol{X}_n - B_n\}$ は 1 に分布収束することがわかるので、定理 1.11(i) より一様に tight となる．同様に、L_n も一様に tight となる．このことより、$(1\{\boldsymbol{X}_n - B_n\}, L_n)$ は一様に tight となるので、定理 1.11(ii) より、ある $\{n_j\}$ の部分列が存在し、$j \to \infty$ としたとき、

$$(1\{\boldsymbol{X}_{n_j} - B_{n_j}\}, L_{n_j}) \xrightarrow{d} (1, L)$$

となる． 以後、表記を簡単にするために n_j を n と記すことにする． 関数 $(v,t) \mapsto vt$ は $[0,\infty) \times \{0,1\}$ 上で非負で連続であるので、定理 1.10(iv) を使えば、

$$\begin{aligned}
&\liminf Q_n(\boldsymbol{X}_n - B_n) \\
&\quad = \liminf \left[\int_{p_n \neq 0} 1\{\boldsymbol{X}_n - B_n\} dQ_n + \int_{p_n = 0} 1\{\boldsymbol{X}_n - B_n\} dQ_n \right] \\
&\quad \geq \liminf \int_{p_n \neq 0} 1\{\boldsymbol{X}_n - B_n\} L_n dP_n \geq \boldsymbol{E} 1 \cdot L = 1
\end{aligned}$$

を得る．これより、$n \to \infty$ のとき、

$$Q_n(B_n) = 1 - Q_n(\boldsymbol{X}_n - B_n) \to 0$$

となることより、補題は証明された． □

系 2.1 ：もし、

$$\boldsymbol{L}(\log L_n \mid P_n) \xrightarrow{d} \boldsymbol{L}(\log L) = N(-\frac{1}{2}\sigma^2, \sigma^2)$$

ならば、$\{Q_n\} \lhd \{P_n\}$ である．

証明：Z を標準正規分布に従う確率変数とする．

$$\boldsymbol{L}(L) = \boldsymbol{L}(e^{\sigma Z - (1/2)\sigma^2})$$

となることに注意すれば、

$$\begin{aligned}
\boldsymbol{E}(L) &= \boldsymbol{E}(e^{\sigma Z - (1/2)\sigma^2}) \\
&= \frac{1}{\sqrt{2\pi}} \int e^{\sigma Z - (1/2)\sigma^2} \cdot e^{-(1/2)z^2} dz \\
&= \frac{1}{\sqrt{2\pi}} \int e^{-(1/2)(z-\sigma)^2} dz = 1
\end{aligned}$$

となるので、補題 2.1 の仮定が満たされることがわかる．

□

いま、

$$(\boldsymbol{X}_n, \boldsymbol{A}_n, \mu_n) = (\boldsymbol{X}, \boldsymbol{A}, \mu)^n \quad \text{かつ} \quad (X_{n1}, X_{n2}, \ldots, X_{nn}) \in \boldsymbol{X}_n$$

とし、

$$p_n(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{ni}(x_{ni}), \quad q_n(\boldsymbol{x}_n) = \prod_{i=1}^{n} g_{ni}(x_{ni}),$$

とおき、

$$\Lambda_n = \sum_{i=1}^{n} \log\left[\frac{g_{ni}}{f_{ni}}(X_{ni})\right] = \begin{cases} < \infty & a.s. \quad P_n \\ > -\infty & a.s. \quad Q_n \end{cases} \tag{2.1}$$

と定義する．

定義 2.3 ：*(2.1)* の和が**一様漸近 neglibility** （*UAN*） 条件を満たすとは、任意の $\epsilon > 0$ に対して、

$$\lim_{n\to\infty} \sup_{1\le i\le n} P_n\left\{\left|\frac{g_{ni}}{f_{ni}}(X_{ni}) - 1\right| > \epsilon\right\} = 0 \tag{2.2}$$

を満足することである．

有限の分散（古典的な中心極限定理が適応可能なように）をもつ確率変数を得るために

$$W_n = 2\sum_{i=1}^{n}\left\{\sqrt{\frac{g_{ni}}{f_{ni}}(X_{ni})} - 1\right\} \tag{2.3}$$

とおく．P_n のもとで、

$$\boldsymbol{var}\left[\sqrt{\frac{g_{ni}}{f_{ni}}(X_{ni})}\right] \le \boldsymbol{E}\left[\frac{g_{ni}}{f_{ni}}(X_{ni})\right] = \int 1\{f_{ni} > 0\} g_{ni} d\mu \le 1$$

である．

補題 2.2 （ Le Cam の第 2 補題）: *UAN* 条件 *(2.2)* が成立するとし、

$$\boldsymbol{L}(W_n \mid P_n) \xrightarrow{d} N(-\frac{1}{4}\sigma^2, \sigma^2)$$

が成立すると仮定する．このとき、P_n のもとで、

$$\log L_n - (W_n - \frac{1}{4}\sigma^2) = o_P(1) \tag{2.4}$$

が成立する．よって、

$$\boldsymbol{L}(L_n \mid P_n) \xrightarrow{d} N(-\frac{1}{2}\sigma^2, \sigma^2) \tag{2.5}$$

が成立する．

証明：2 回連続微分可能な関数 h に対し、部分積分の公式[1] を利用すれば、

$$h(x) = h(x_0) + (x - x_0)h'(x_0) + \frac{1}{2}(x - x_0)^2 \int_0^1 2(1-\lambda)h''(x_0 + \lambda(x - x_0))d\lambda$$

を得る．これを $h(x) = \log(1+x)$ に対して適応すれば、

$$\log(1+x) = x - \frac{1}{2}x^2 \int_0^1 \frac{2(1-\lambda)}{(1+\lambda x)^2} d\lambda$$

を得る．ここで、

$$T_{ni} = 2\left\{\sqrt{\frac{g_{ni}}{f_{ni}}(X_{ni})} - 1\right\}$$

として、上式に代入すれば、

$$\log\left[\frac{g_{ni}}{f_{ni}}(X_{ni})\right] = 2\log\left(1 + \frac{1}{2}T_{ni}\right) = T_{ni} - \frac{1}{4}T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{1 + \frac{1}{2}\lambda T_{ni}^2} d\lambda \tag{2.7}$$

を得る．更に、和をとれば、

$$\log L_n = W_n - \frac{1}{4}\sum_{i=1}^n T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{1 + \frac{1}{2}\lambda T_{ni}^2} d\lambda \tag{2.8}$$

となる．よって、(2.4) を示すために、

$$\sum_{i=1}^n T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{(1 + (1/2)\lambda T_{ni})^2} d\lambda = \sigma^2 + o_P(1)$$

を示せばよいことがわかる．そのために、$1 > \delta > 0$ に対し、

$$T_{ni}^\delta = T_{ni}1\{T_{ni} > \delta\}$$

とおけば、$\boldsymbol{L}(W_n|P_n) \xrightarrow{d} N(-(1/4)\sigma^2, \sigma^2)$ と (2.2) が成立するための必要十分条件は

$$\begin{aligned} \sum_{i=1}^n P_n\{T_{ni} > \delta\} &\rightarrow 1 && (2.9) \\ \sum_{i=1}^n \boldsymbol{E}_{P_n}(T_{ni}^\delta) &\rightarrow -\frac{1}{4}\sigma^2 && (2.10) \\ \sum_{i=1}^n \boldsymbol{var}_{P_n}(T_{ni}^\delta) &\rightarrow \sigma^2 && (2.11) \end{aligned}$$

---

[1]

$$\begin{aligned} \int_0^\infty 2(1-\lambda)h''(x_0 + \lambda(x - x_0))d\lambda &= \frac{1}{x - x_0}\left[2(1-\lambda)h'(x_0 + \lambda(x - x_0))\right]_0^1 \\ &\quad + \frac{2}{x - x_0}\int_0^1 h'(x_0 + \lambda(x - x_0))d\lambda \\ &= \frac{2}{x - x_0}h'(x_0) + \frac{2}{(x - x_0)^2}\left[h(x_0 + \lambda(x - x_0))\right]_0^1 \end{aligned} \tag{2.6}$$

よりわかる．

であることが、Normal convergence criterion [2] よりわかる．$\int_0^1 2(1-\lambda)d\lambda = 1$ と (2.9) から

$$P_n\{\max_{1\le i\le n}|T_{ni}| > \delta\} \le \sum_{i=1}^n P_n\{|T_{ni}| > \delta\} \to 1$$

となることに注意する．$S_n \equiv \{\max_{1\le i\le n}|T_{ni}| \le \delta\}$ とおく．任意の $0 < \eta < 1$ に対し、ある $N = N(\eta)$ が存在し、$n > N$ なる n に対し、 $P_n(S_n) < \eta$ とできる．さらに、S_n 上では[3]、

$$\sup_\lambda \max_{1\le i\le n}\left|\frac{1}{\left(1+\frac{1}{2}\lambda T_{ni}\right)^2} - 1\right| < \delta$$

を得る．これから、

$$\max_{1\le i\le n}\left|\int_0^1 \frac{2(1-\lambda)}{(1+(1/2)\lambda T_{ni})^2}d\lambda - 1\right| \le 1$$

がわかる．また、S_n 上では、$T_{ni} = T_{ni}^\delta, i = 1, 2, \ldots n$ なので、

$$\left|\sum_{i=1}^n T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{(1+(1/2)\lambda T_{ni})^2}d\lambda - \sum_{i=1}^n T_{ni}^2\right| \le \delta\sum_{i=1}^n T_{ni}^2 = \delta\sum_{i=1}^n (T_{ni}^\delta)^2$$

とより、

$$\left|\sum_{i=1}^n T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{(1+(1/2)\lambda T_{ni})^2}d\lambda \Big/ \sum_{i=1}^n (T_{ni}^\delta)^2 - 1\right| \le \delta \qquad S_n \text{ 上において}$$

から、

$$\sum_{i=1}^n T_{ni}^2 \int_0^1 \frac{2(1-\lambda)}{(1+(1/2)\lambda T_{ni})^2}d\lambda = \sum_{i=1}^n (T_{ni}^\delta)^2 + o_P(1)$$

がわかる．よって、補題を示すためには、P_n のもと、$n \to \infty$ としたとき、

$$\sum_{i=1}^n (T_{ni}^\delta)^2 = \sigma^2 + o_P(1) \tag{2.12}$$

であることを示せばよい．任意の $\epsilon > 0$ に対し、

$$|\sum_{i=1}^n (T_{ni}^\delta)^2 - \sigma^2| \le |\sum_{i=1}^n (T_{ni}^\delta)^2 - \sum_{i=1}^n \boldsymbol{E}(T_{ni}^\delta)^2| + |\sum_{i=1}^n \boldsymbol{E}(T_{ni}^\delta)^2 - \sigma^2|$$

であることと上式の第 1 項目に Chebychev の不等式を用いれば、(2.12) を示すためには、

$$\sum_{i=1}^n \boldsymbol{E}(T_{ni}^\delta)^2 \quad\to\quad \sigma^2 \tag{2.13}$$

$$\liminf_{\delta\to\infty}\limsup_{n\to\infty}\sum_{i=1}^n \boldsymbol{var}(T_{ni}^\delta)^2 \quad=\quad 0 \tag{2.14}$$

---

[2] Loéve, Probability Theory I, pages 328 を参照されたい．

[3]
$$1-\delta \le \frac{1}{(1+\delta/2)^2} \le \frac{1}{(1+(1/2)\lambda T_{ni})^2} \le \frac{1}{(1-\delta/2)^2} \le 1+\delta$$

に注意せよ．

を示せばよいことがわかる．しかし、(2.11) より、(2.13) は

$$\sum_{i=1}^{n}[\boldsymbol{E}T_{ni}^{\delta}]^2 \to 0 \tag{2.15}$$

と同値なことがわかる．

はじめに (2.15) （すなわち (2.13) ）を示す．まず $\delta > 2$ として[4] 、(2.15) を示そう．T_{ni} の定義より、$T_{ni} \geq -2$ なので、ほとんどいたるところで $T_{ni}^{\delta} \leq T_{ni}$ が成立する．よって、

$$\boldsymbol{E}T_{ni}^{\delta} \leq \boldsymbol{E}T_{ni} = -2\left(1 - \boldsymbol{E}[\sqrt{\frac{g_{ni}}{f_{ni}}(X_{ni})}]\right) = -d_H^2(P_{ni}, Q_{ni}) \leq 0$$

が成立する[5] ．これに注意すれば、

$$\sum_{i=1}^{n}(-\boldsymbol{E}T_{ni}^{\delta})^2 \leq \max_{1\leq n\leq}(-\boldsymbol{E}T_{ni}^{\delta}\sum_{i=1}^{n}(-\boldsymbol{E}T_{ni}^{\delta} \to 1$$

が得る．上の式の右辺がゼロに収束することは、(2.10) と UAN 条件 (2.2) より、

$$\sum_{i=1}^{n}(-\boldsymbol{E}T_{ni}^{\delta}) \to \frac{1}{4}\sigma^2$$

と

$$\max_{1\leq i\leq n}(-\boldsymbol{E}T_{ni}^{\delta}\quad) \to 0$$

よりわかる．

つぎに、$\delta > 2$ に対して (2.15) が成立するならば、$\delta > 0$ に対して (2.15) が成立することを示す．$\delta' < \delta$ ならば、

$$\sum_{i=1}^{n}\boldsymbol{E}[(T_{ni}^{\delta'})^2] \leq \sum_{i=1}^{n}\boldsymbol{E}[(T_{ni}^{\delta})^2]$$

であるある．また、(2.11) より

$$\boldsymbol{var}[T_{ni}^{\delta'}] \to \sigma^2 \qquad \boldsymbol{var}[T_{ni}^{\delta}] \to \sigma^2$$

が成り立する．これらより、

$$\begin{aligned}
\sum_{i=1}^{n}[\boldsymbol{E}T_{ni}^{\delta'}]^2 &= \sum_{i=1}^{n}[\boldsymbol{E}[(T_{ni}^{\delta'})^2] - \boldsymbol{var}(T_{ni}^{\delta'})] \\
&\leq \sum_{i=1}^{n}[\boldsymbol{E}[(T_{ni}^{\delta})^2] - \boldsymbol{var}(T_{ni}^{\delta}) + \boldsymbol{var}(T_{ni}^{\delta}) - \boldsymbol{var}(T_{ni}^{\delta'})] \\
&= \sum_{i=1}^{n}[\sum_{i=1}^{n}[\boldsymbol{E}T_{ni}^{\delta'}]^2 + \boldsymbol{var}(T_{ni}^{\delta}) - \boldsymbol{var}(T_{ni}^{\delta'})] \\
&\to 0 + \sigma^2 - \sigma^2
\end{aligned}$$

[4] とりあえず、$\delta > 2$ として、(2.15) を示し、これを用いて、任意の $\delta > 0$ に対して、(2.15) が成立することを示す．

[5]

$$d_n(P_{ni}, Q_{ni}) = \int(\sqrt{f_{ni}} - \sqrt{g_{ni}})^2 d\mu = 2\int(1 - \sqrt{f_{ni}g_{ni}})d\mu = 1 - \boldsymbol{E}_{P_{ni}}[\sqrt{\frac{g_{ni}}{f_{ni}}(X_{ni})}]$$

より、(2.15) が示せた．

最後に (2.14) を示す．

$$\sum_{i=1}^n \boldsymbol{var}[(T_{ni}^\delta)^2] \le \sum_{i=1}^n \boldsymbol{E}[(T_{ni}^\delta)^4] \le \delta^2 \sum_{i=1}^n \boldsymbol{E}[(T_{ni})^2]$$

に注意すれば、(2.13) より、

$$\limsup_{n\to\infty} \sum_{i=1}^n \boldsymbol{var}[(T_{ni}^\delta)^2] \le \delta^2\sigma^2$$

より、(2.14) がわかる． □

系 2.2 （補題 1.2 の系）: $\{f_n\}$ を確率密度関数の列、f を確率密度関数とし、

$$\| \sqrt{n}(\sqrt{f_n} - \sqrt{f}) - \delta \|_{L_2(\mu)} \longrightarrow 0 \quad (n \to \infty)$$

を満足するものとする．ただし、$\delta \in L_2(\mu)$ とする．更に、

$$p_n(\boldsymbol{x}) = \prod_{i=1}^n f(x_i), \quad q_n(\boldsymbol{x}) = \prod_{i=1}^n f_n(x_i)$$

とおく．このとき、P_n のもとで、

$$\log L_n - \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta}{\sqrt{f}}(X_i) - 2 \| \delta \|_{L_2(\mu)}^2 \right] = o_P(1) \tag{2.16}$$

が成立する．よって、$\sigma^2 = 4 \| \delta \|_{L_2(\mu)}^2$ としたとき、

$$\boldsymbol{L}(\log L_n \mid P_n) \xrightarrow{d} N(-\frac{1}{2}\sigma^2, \sigma^2) \tag{2.17}$$

が成立する．

証明：系の仮定[6] より、

$$n\|\sqrt{f_n} - \sqrt{f}\|^2 \quad \to \quad \|\delta\| \tag{2.18}$$

$$\|\sqrt{f_n} - \sqrt{f}\| \quad \to \quad 0 \tag{2.19}$$

となる．これに注意すれば[7] 、

$$\begin{aligned} 1 &= \int f_n d\mu = \int \{\sqrt{f_n} + \frac{\delta}{\sqrt{n}} + \sqrt{f_n} - \frac{\delta}{\sqrt{n}}\}^2 d\mu \\ &= 1 + \frac{1}{\sqrt{n}} \int \delta\sqrt{f} d\mu + o(n^{-1/2}) \end{aligned}$$

[6] $(|x| - |y|)^2 \le (x-y)^2$ に注意すればよい．

[7]

$$\begin{aligned} \int \sqrt{f}(\sqrt{f} - \sqrt{f_n} - \frac{\delta}{\sqrt{n}}) d\mu &\le \sqrt{\int f d\mu}\sqrt{\int (\sqrt{f} - \sqrt{f_n} - \frac{\delta}{\sqrt{n}})^2 d\mu} \\ &\le \sqrt{\frac{1}{n}\|\sqrt{n}(\sqrt{f} - \sqrt{f_n}) - \delta\|^2} = o(n^{-1/2}) \end{aligned}$$

を利用すればよい．

であることがわかる．これから、直交関係

$$\int \delta\sqrt{f}d\mu = o(1) \tag{2.20}$$

を得る．また、ある $\epsilon > 0$ に対し、

$$\begin{aligned}
\epsilon P_n\left\{\left|\frac{f_n}{f}(X_i)-1\right| \geq \epsilon\right\} &\leq \boldsymbol{E}\left|\frac{f_n}{f}(X_i)-1\right| \qquad \text{（Markov の不等式）}\\
&= \boldsymbol{E}\left\{\left|\sqrt{\frac{f_n}{f}(X_i)}-1\right|\cdot\left|\sqrt{\frac{f_n}{f}(X_i)}+1\right|\right\}\\
&= \int(\sqrt{f_n}-\sqrt{f})(\sqrt{f_n}+\sqrt{f})d\mu\\
&\leq \|\sqrt{f_n}-\sqrt{f}\|\|\sqrt{f_n}+\sqrt{f}\| \to 0
\end{aligned}$$

が (2.19) より $1 \leq i \leq n$ について一様に成立することから、UAN 条件 (2.2) が成立することがわかる. 更に、に注意して、(2.20) と (2.18) より、

$$\begin{aligned}
&\boldsymbol{E}\left\{W_n - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{\sqrt{f}}(X_i)+\frac{1}{2n}\|\delta\|^2\right\}^2\\
&= 4n\boldsymbol{E}\left\{\sqrt{\frac{f_n}{f}(X_1)}-1-\frac{1}{\sqrt{n}}\frac{\delta}{\sqrt{f}}(X_1)+\frac{1}{2n}\|\delta\|^2\right\}^2\\
&+n(n-1)\left[\boldsymbol{E}\left\{2\sqrt{\frac{f_n}{f}(X_1)}-2-\frac{2}{\sqrt{n}}\frac{\delta}{\sqrt{f}}(X_1)+\frac{1}{n}\|\delta\|^2\right\}\right]^2\\
&= 4\int_0^1\{\sqrt{n}(\sqrt{f_n}-\sqrt{f})-\delta\}^2 d\mu\\
&+\frac{4}{n}\|\delta\|^2\int_0^1\{\sqrt{n}(\sqrt{f_n}-\sqrt{f})-\delta\}d\mu+\frac{1}{n}\|\delta\|\\
&+\left(1-\frac{1}{n}\right)\left[-n\|\sqrt{f_n}-\sqrt{f}\|^2-2\sqrt{n}\int\delta\sqrt{f}d\mu+\|\delta\|^2\right]\to 0
\end{aligned} \tag{2.21}$$

であることが、

$$\|\sqrt{f_n}-\sqrt{f}\|^2 = -2\boldsymbol{E}_{P_n}\left[\sqrt{\frac{f_n}{f}}-1\right]$$

に注意して計算すればわかる．よって、

$$W_n - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{\sqrt{f}}(X_i)+\|\delta\|^2 = o_P(1) \tag{2.22}$$

を得る．(2.20) から

$$\boldsymbol{var}\left(\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{\sqrt{f}}(X_i)\right) = 4\|\delta\|^2+4\left(\int\delta\sqrt{f}d\mu\right)^2 = 4\|\delta\|+o(1)$$

となるので、

$$\boldsymbol{L}(\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{\sqrt{f}}(X_i)|P_n) \xrightarrow{d} N(0,4\|\delta\|^2)$$

となる．これと (2.22) から

$$\boldsymbol{L}(W_n|P_n) \xrightarrow{d} N(-\|\delta\|^2, 4\|\delta\|^2)$$

を得る．$\sigma^2 = 4\|\delta\|^2$ として、補題 2.2 を使えば、

$$\begin{aligned}\log L_n &= W_n - \|\delta\|^2 + o_P(1)\\ &= \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{\sqrt{f}}(X_i) - 2\|\delta\|^2 + o_P(1)\end{aligned}$$

となり、系が示せた． □

例 2.1 ：

$$\boldsymbol{P} = \{P_\theta \ll \mu : \theta \in \Theta \subset \boldsymbol{R}^k\}$$

を「正則」母数モデル[8] とし、

$$p(x;\theta) = \frac{dP_\theta}{d\mu}, \quad \dot{l}_\theta(x) = \frac{\partial}{\partial\theta}\log p(x;\theta)$$

とし、系 *1.2* において

$$f(x) = p(x;\theta_0), \quad f_n(x) = p(x;\theta_n), \quad \theta_n = \theta_0 + \frac{t}{\sqrt{n}}, \quad t \in \boldsymbol{R}^k$$

とおく．このとき、正則性より

$$\begin{aligned}&\int \left[\sqrt{n}\{\sqrt{p(x;\theta_0)} - \sqrt{p(x;\theta_n)}\} - t^T\delta(x)\right]^2 d\mu(x)\\ &= \int \left[\frac{\sqrt{p(x;\theta_0)} - \sqrt{p(x;\theta_n)}}{1/\sqrt{n}} - t^T\delta(x)\right]^2 d\mu(x) \to 0\end{aligned}$$

となる．ただし、

$$\dot{l}_{\theta_0}(x) = \frac{2}{\sqrt{p(x;\theta_0)}}\frac{\partial}{\partial\theta}\sqrt{p(x;\theta)}\Big|_{\theta=\theta_0} = \frac{2\delta(x)}{f(x)}$$

となる．また、

$$\sigma^2 = 4 \parallel \delta \parallel^2_{L_2(\mu)} = 4 \parallel \frac{\dot{l}_{\theta_0} f}{2} \parallel_{L_2(\mu)} = \boldsymbol{E}(\dot{l}_{\theta_0})^2$$

より、

$$\boldsymbol{L}(L_n \mid P_n) \xrightarrow{d} N(-\frac{1}{2}\sigma^2, \sigma^2)$$

となる．

補題 1.2b（Hall and Loynes）：$\{Q_n\} \lhd \{P_n\}$ の必要十分条件はつぎのふたつが成立することである．

[8] モデルの「正則性」の定義については第２章を参照

(i) $\{P_n\}$ について、L_n は u.i. である．

(ii)
$$Q_n\{p_n = 0\} \to 0.$$

証明：任意の $B_n \in \boldsymbol{A}$ に対し、

$$\begin{aligned} Q_n(B_n) &= \int 1_{B_n} dQ_n \\ &= \int 1\{B_n \cap \{p_n = 0\}\} dQ_n + \int 1\{B_n \cap \{p_n > 0\}\} dQ_n \\ &= \int 1\{B_n \cap \{p_n = 0\}\} dQ_n + \int 1_{B_n} L_n dP_n \\ &\geq \int_{B_n} L_n dP_n \end{aligned} \tag{2.23}$$

と

$$\begin{aligned} Q_n(B_n) &= \int 1\{B_n \cap \{p_n =\}\} dQ_n + \int_{B_n} L_n dP_n \\ &\leq Q_n\{p_n = 0\} + \int_{B_n} L_n dP_n \end{aligned}$$

が成立する． これらに注意すれば、L_n が一様可積分で $Q_n\{p_n = 0\} \to 0$ が成立するとき、$P_n(B_n) \to 0$ ならば、$Q_n(B_n) \to 0$ が成立することがわかる．したがって、$\{Q_n\} \lhd \{P_n\}$ である．

逆に、$\{Q_n\} \lhd \{P_n\}$ が成立するとする．$P_n(B_n) \to 0$ ならば、(2.23) から、

$$\int L_n 1_{B_n} dP_n \to 0$$

が成立する．よって、定理 1.1(ii) が満たされる．また、

$$\boldsymbol{E}_{P_n}(L_n) = \int L_n dP_n = \int 1\{p_n > 0\} dQ_n \leq 1$$

から定理 1.1(i) が成立することもわかる．よって、L_n は一様可積分であることが示せた．更に、接触性より、$P_n\{p_n = 0\} = 0$ から $Q_n\{p_n = 0\} \to 0$ がわかる．よって、補題は証明された． □

補題 2.3 （Le Cam の第３補題）：

$$\boldsymbol{L}\left(\left[\begin{array}{c} T_n \\ \log L_n \end{array}\right] \Big| P_n\right) \xrightarrow{d} \boldsymbol{L}\left(\left[\begin{array}{c} T \\ \Lambda \end{array}\right]\right) = N\left(\left[\begin{array}{c} \mu \\ -\frac{\sigma^2}{2} \end{array}\right], \left[\begin{array}{cc} \tau^2 & c \\ c & \sigma^2 \end{array}\right]\right)$$

ならば

$$\boldsymbol{L}\left(\left[\begin{array}{c} T_n \\ \log L_n \end{array}\right] \Big| Q_n\right) \xrightarrow{d} \boldsymbol{L}\left(\left[\begin{array}{c} T + c \\ \Lambda + \sigma^2 \end{array}\right]\right) = N\left(\left[\begin{array}{c} \mu + c \\ \frac{\sigma^2}{2} \end{array}\right], \left[\begin{array}{cc} \tau^2 & c \\ c & \sigma^2 \end{array}\right]\right)$$

が成立する．

証明：系 2.1 より、$\{Q_n\} \lhd \{P_n\}$ がわかる．よって、補題 2.2b を使えば、L_n は一様可積分で $Q_n\{p_n = 0\} \to 0$ が成立することがわかる．

いま、$f : \boldsymbol{R}^2 \mapsto \boldsymbol{R}$ を有界連続関数とし、$\Lambda_n = \log L_n, \Lambda = \log L$ とおく．このとき、

$$
\begin{aligned}
\boldsymbol{E}_{Q_n}[f(T_n, \Lambda_n)] &= \boldsymbol{E}_{Q_n}[f(T_n, \Lambda_n)(1\{p_n > 0\} + 1\{p_n = 0\})] \\
&= \boldsymbol{E}_{P_n}[f(T_n, \Lambda_n)L_n] + \boldsymbol{E}_{Q_n}[f(T_n, \Lambda_n)1\{p_n = 0\}] \\
&\to \boldsymbol{E}[f(T, \Lambda)L] \qquad (2.24) \\
&= \boldsymbol{E}[f(T + c, \Lambda + \sigma^2)] \qquad (2.25)
\end{aligned}
$$

が成立する．

(2.24) が成立することは、

$$|\boldsymbol{E}_{Q_n}[f(T_n, \Lambda_n)1\{p_n = 0\}]| \le \|f\|_\infty Q_n\{p_n = 0\} \to 0$$

と、L_n は一様可積分で f は有界であることから fL_n は一様可積分にであることに注意すれば

$$\lim_{n\to\infty} \boldsymbol{E}_{P_n}[f(T_n, \Lambda_n)L_n] = \boldsymbol{E}[f(T, \Lambda)L]$$

が成立することからわかる．

(2.25) を示すために、U, V, W を独立に $N(0, 1)$ に従う確率変数とする．

$$L = \exp(\sigma U - \frac{\sigma^2}{2})$$

は $N(\sigma, 1)$ と $N(0, 1)$ の尤度比統計量とみなすこと[9] ができることに注意する．これより、任意の有界可測関数 g に対し、

$$
\begin{aligned}
\boldsymbol{E}[g(U, V)L] &= \int g(u, v)LdN(0, 1)dN(0, 1) \\
&= \int g(u, v)dN(\sigma, 1)dN(0, 1) \\
&= \int g(w + \sigma, v)dN(0, 1)dN(0, 1) \\
&= \boldsymbol{E}[g(W + \sigma, V)] \qquad (2.26)
\end{aligned}
$$

が成立する．また、$(T, \Lambda)'$ は

$$(\rho\tau U + (1 - \rho^2)^{1/2}\tau V + \mu, \sigma U - \frac{\sigma^2}{2})'$$

と同じ分布をもつことに注意する．ただし、$\rho\tau\sigma = c$ である．

$$f(\rho\tau U + (1 - \rho^2)^{1/2}\tau V + \mu, \sigma U - \frac{\sigma^2}{2}) = g(U, V)$$

---

[9]

$$\frac{dN(\sigma, 1)}{dN(0, 1)} = \exp[-\frac{1}{2}(u - \sigma)^2]/\exp[-\frac{1}{2}u^2] = \exp[\sigma u - \frac{u^2}{2}$$

に注意せよ．

とおいて、(2.26) を利用すれば、

$$
\begin{aligned}
\boldsymbol{E}[f(T,\Lambda)L] &= \boldsymbol{E}[f(\rho\tau U+(1-\rho^2)^{1/2}\tau V+\mu, \sigma U-\frac{\sigma^2}{2})] \\
&= \boldsymbol{E}[g(U,V)L] = \boldsymbol{E}[g(W+\sigma, V)] \\
&= \boldsymbol{E}[f(\rho\tau(W+\sigma)+(1-\rho^2)^{1/2}\tau V+\mu, \sigma(W+\sigma)-\frac{\sigma^2}{2})] \\
&= \boldsymbol{E}[f(\rho\tau W+(1-\rho^2)^{1/2}\tau V+\mu+\rho\tau\sigma, \sigma W+\frac{\sigma^2}{2})] \\
&= \boldsymbol{E}[f(T+c, \Lambda+\sigma^2)]
\end{aligned}
$$

が成立することがわかる．よって、補題は証明された． □

第3章　母数モデルの漸近推測理論

3.1　同一独立標本における正則母数モデル

母数モデル

$$\boldsymbol{P} = \{P_\theta \ll \mu : \theta \in \Theta \subset \boldsymbol{R}^k\}$$

を考える．ただし、

(i) Θ は「手頃な」 $\boldsymbol{R}^k$ の部分集合

(ii) 対応 $\theta \longrightarrow P_\theta$ は「なめらか」

と仮定する．また、記号を

$$p(\theta) = p(\cdot;\theta) = \frac{dP_\theta}{d\mu}(\cdot), \quad \boldsymbol{l}(\theta) = \log p(\theta), \quad s(\theta) = \sqrt{p(\theta)}$$

で定義する．

定義 3.1 ：θ_0 は写像 $\theta \longrightarrow P_\theta$ の**正則点**であるとは、

(i) Θ から $L_2(\mu)$ への写像 $\theta \longrightarrow s(\theta)$ が θ_0 において Fréchet differentiable [1] である．

(ii) Fisher 情報行列の正則性：$k \times k$ の行列[2]

$$\int \dot{s}(\theta_0)\dot{s}^T(\theta_0)d\mu$$

が正則である．

定義 3.2 ：写像 $\theta \longrightarrow P_\theta$ の**正則**であるとは、

(i) Θ の任意の点が正則．

(ii) Θ から $L_2(\mu)$ の写像 $\theta \longrightarrow \dot{s}_i(\theta)$ が連続（$i = 1, 2, \ldots, k$ ）．

---

[1] $\dot{s}(\theta_0) = (\dot{s}_1(\theta_0), \ldots, \dot{s}_k(\theta_0))^T (\dot{s}_i \in L_2(\mu), i = 1, 2, \ldots, k)$ なるものが存在し、

$$\| s(\theta_0 + h) - s(\theta_0) - \dot{s}^T(\theta)h \|_{L_2(\mu)} = o(| h |)$$

が成立することである．ただし、$h \in \boldsymbol{R}^k, | h |^2 = \sum h_i^2$ である．

[2] Fisher 行列はこの行列の４倍である．

であることである．

θ の **Fisher 情報行列**を

$$I(\theta) = 4\int \dot{s}(\theta)\dot{s}^T(\theta)d\mu$$

で定義し、モデル $\boldsymbol{P}$ が定義 2.2 の意味で正則写像をもつとき、$\boldsymbol{P}$ を**正則母数モデル**とよぶ．また、**スコア関数** $\dot{\boldsymbol{l}}$ を

$$\dot{\boldsymbol{l}} = 2\frac{\dot{s}(\theta)}{s(\theta)}1\{s(\theta) > 0\} = \frac{\dot{p}(\theta)}{p(\theta)}1\{p(\theta) > 0\}$$

で定義する．ただし、$\dot{p}(\theta) = 2s(\theta)\dot{s}(\theta)$ である．θ が正則点ならば、$\dot{\boldsymbol{l}} \in L_2(P_\theta)$ となり、Fisher 情報行列は

$$I(\theta) = \int \dot{\boldsymbol{l}}(\theta)\dot{\boldsymbol{l}}^T(\theta)dP_\theta$$

と表現される．また、$h = (0, \ldots, 0, h_i, 0, \ldots, o)$ と Fréchet differentiability の定義において取ることにより、

$$\boldsymbol{E}\dot{l}_i(\theta) = 2\langle \dot{s}_i(\theta), s(\theta)\rangle_{L_2(\mu)} = 0$$

が成立することがわかる．

正則母数モデルと局所漸近正規性（LAN）の関係を議論するためにつぎの補題をしめす．

補題 3.1 ：母数モデル

$$\boldsymbol{P} = \{P_\theta \ll \mu : \theta \in \Theta \subset \boldsymbol{R}^k\}$$

は正則とし、

$$T_n \equiv 2\left\{\frac{s(\theta + t/\sqrt{n})}{s(\theta)}(X) - 1\right\}, \quad \theta \in \boldsymbol{R}^k$$

とおいたとき、

$$\begin{aligned}
\boldsymbol{E}_\theta T_n &= -\frac{1}{n}\frac{1}{4}t^T I(\theta)t + o(n^{-1}) && (3.1)\\
\boldsymbol{E}_\theta T_n^2 &= \frac{1}{n}t^T I(\theta)t + o(n^{-1}) && (3.2)\\
\boldsymbol{E}_\theta\Big(T_n - \frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\Big)^2 &= o(n^{-1}) && (3.3)\\
\text{任意の } \epsilon > 0 \text{ に対して、} P_\theta(|\, T_n \,| \geq \epsilon) &= o(n^{-1}) && (3.4)
\end{aligned}$$

が成立する．

証明：まず、(3.3) を示す．コンパクト集合 C 上で、正則性の仮定より $\theta \mapsto \dot{s}(\theta)$ は一様連続となることに注意する．$h \downarrow 0$ のとき、

$$\begin{aligned}
&\sup_{\theta\in C}\frac{1}{|h|^2}\|s(\theta+h) - s(\theta) - h^T\dot{s}(\theta)\|^2 = \sup_{\theta\in C}\frac{1}{|h|^2}\|h^T\{\int_0^1 \dot{s}(\theta+uh)du - \dot{s}(\theta)\}\|^2\\
&\quad\leq \int_0^1 \sup_{\theta\in C}\|\dot{s}(\theta+uh) - \dot{s}(\theta)\|^2 du \to 0 \qquad (3.5)
\end{aligned}$$

を得る．これより

$$
\begin{aligned}
&\boldsymbol{E}_\theta\left(T_n - \frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right)^2 \\
&= \int\left\{2\left(\frac{s(\theta+t/\sqrt{n})}{s(\theta)}\right) - 2\frac{t^T}{\sqrt{n}}2\frac{\dot{s}(\theta)}{s(\theta)}1\{s(\theta)>0\}\right\}^2 s^2(\theta)d\mu \\
&= 4\|s(\theta+\frac{t}{\sqrt{n}}) - s(\theta) - \frac{t^T}{\sqrt{n}}\dot{s}(\theta)\|^2_{L_2(\mu)} = o(n^{-1})
\end{aligned}
$$

となり、(3.3) は示された．

(3.4) と

$$
\boldsymbol{E}_\theta\left(\frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right)^2 = \frac{1}{n}t^T I(\theta)t
$$

から

$$
\begin{aligned}
|\boldsymbol{E}_\theta T_n^2 - \frac{1}{n}t^T I(\theta)t| &\leq \sqrt{\boldsymbol{E}_\theta\left(T_n - \frac{t}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right)^2}\sqrt{\boldsymbol{E}_\theta\left(T_n + \frac{t}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right)^2} \\
&= o(n^{-1/2})o(n^{-1/2}) = o(n^{-1})
\end{aligned}
$$

がわかる[3]．したがって、(3.2) が示せた．

$$
\begin{aligned}
&\int_{s(\theta)=0} s^2(\theta+\frac{t}{\sqrt{n}})d\mu \leq \int_{s(\theta)=0}\left(s(\theta+\frac{t}{\sqrt{n}}) + s(\theta-\frac{t}{\sqrt{n}})\right)^2 d\mu \\
&\leq 2\int_{s(\theta)=0}\left(s(\theta+\frac{t}{\sqrt{n}}) - s(\theta)\right)^2 d\mu + \int_{s(\theta)=0}\left(s(\theta-\frac{t}{\sqrt{n}}) - s(\theta)\right)^2 d\mu \\
&= 4\left(\frac{t}{n}\int \dot{s}^2(\theta)d\mu + o(n^{-1})\right) = o(n^{-1}) \qquad (3.6)
\end{aligned}
$$

に注意すれば、

$$
\boldsymbol{E}_\theta T_n^2 = 4\int\left\{\frac{s(\theta+t/\sqrt{n})}{s(\theta)}(x) - 1\right\}^2 s(\theta)^2 1\{s(\theta)>0\}d\mu
$$

---

[3]

$$
\boldsymbol{E}_\theta T_n = 4\|s(\theta+\frac{f}{\sqrt{n}}) - s(\theta)\|_{L_2(\mu)} = 4\|\frac{t^T}{\sqrt{n}}\dot{s}(\theta) + o(n^{-1/2})\|_{L_2(\mu)} = o(n^{-1})
$$

と

$$
\boldsymbol{E}_\theta\left\{T_n\frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}\right\} \leq \sqrt{\boldsymbol{E}_\theta T_n^2}\sqrt{\frac{1}{n}\boldsymbol{E}_\theta(t^T\dot{\boldsymbol{l}}^2)^2} = o(n^{-1})
$$

から

$$
\sqrt{\boldsymbol{E}_\theta\left(T_n + \frac{t}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right)^2} = o(n^{-1/2})
$$

になることに注意せよ．

$$
\begin{aligned}
&= 4\int s^2(\theta+\frac{t}{\sqrt{n}})1\{s(\theta)>0\}d\mu \\
&\quad -8\int \frac{s(\theta+t/\sqrt{n})}{s(\theta)}s^2(\theta)1\{s(\theta)>0\}d\mu+4 \\
&= -4\int 2\left(\frac{s(\theta+t/\sqrt{n})}{s(\theta)}-1\right)^2 s^2(\theta)1\{s(\theta)>0\}d\mu \\
&\quad -4\int 1\{s(\theta)=0\}s^2(\theta+\frac{t}{\sqrt{n}})d\mu \\
&= -4\boldsymbol{E}_\theta T_n + o(n^{-1})
\end{aligned}
$$

がわかる．

最後に、(3.4) を示す．(3.3) に注意すれば、

$$
\begin{aligned}
P_\theta(|T_n|\geq\epsilon) &\leq P_\theta\left(\left|T_n-\frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right|\geq\frac{1}{2}\epsilon\right) \\
&\quad +P_\theta\left(\left|\frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right|\geq\frac{1}{2}\epsilon\right) \\
&\leq \frac{4}{\epsilon^2}\boldsymbol{E}_\theta\left|T_n-\frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X,\theta)\right|^2 \\
&\quad +\frac{|t|^2}{n\epsilon^2}\boldsymbol{E}_\theta\left[|\dot{\boldsymbol{l}}(X,\theta)|^2 1\{|\dot{\boldsymbol{l}}(X,\theta)|\geq\frac{\epsilon\sqrt{n}}{2|t|}\}\right]=o(n^{-1})
\end{aligned}
$$

がわかる． □

つぎの命題よりモデルの正則性より LAN を導くことができることがわかる．

命題 3.1 ：母数モデル

$$\boldsymbol{P}=\{P_\theta\ll\mu:\theta\in\Theta\subset\boldsymbol{R}^k\}$$

は正則とし、

$$\log\Pi_{i=1}^n\frac{p(X_i;\theta+t/\sqrt{n})}{p(X_i;\theta)}=t^TW_n(\theta)-\frac{1}{2}t^TI(\theta)t+R_n(\theta,t) \tag{3.7}$$

とおく．ただし、

$$W_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\dot{\boldsymbol{l}}(X_i,\theta)$$

とする．このとき、任意の $\epsilon>0$、$0<M<\infty$、コンパクト集合 $C\subset\boldsymbol{R}^k$ に対して、$n\to\infty$ とすれば、

$$\sup_{|t|\leq M}\sup_{\theta\in C}P_\theta(\mid R_n(t,\theta)\mid>\epsilon)\longrightarrow 0 \tag{3.8}$$

が成立する．更に、

$$\boldsymbol{L}(S_n(\theta))\xrightarrow{d}N(0,I(\theta)) \tag{3.9}$$

が、$\theta\in C$ に対し一様に成立する．また、$\theta\in C$ と $\mid t\mid\leq M$　なる θ に対して一様に

$$W_n(\theta+\frac{t}{\sqrt{n}})-W_n(\theta)+I(\theta)t\xrightarrow{P}0 \tag{3.10}$$

が成立する．

証明：はじめに

$$T_{ni} = 2\left\{\frac{s(\theta + t/\sqrt{n})}{s(\theta)}(X_i) - 1\right\}$$

とおいたとき、

$$\sum_{i=1}^{n} T_{ni} = t^T W_n(\theta) - \frac{1}{4}t^T I(\theta)t + o_P(1) \tag{3.11}$$

が成立することを示す．モデルの正則性より

$$\int \left[s(\theta + \frac{t}{\sqrt{n}}) - s(\theta) - \frac{1}{2}\frac{t}{\sqrt{n}}\dot{\boldsymbol{l}}s(\theta)\right]^2 d\mu = o(n^{-1})$$

が成立することに注意すれば、

$$\begin{aligned}\boldsymbol{var}\left(\sum_{i=1}^{n} T_{ni} - t^T W_n(\theta)\right) &= \sum_{i=1}^{n} \boldsymbol{var}\left(T_{ni} - \frac{t^T}{\sqrt{n}}\dot{\boldsymbol{l}}(X_i;\theta)\right) \\ &\leq \boldsymbol{E}\left\{\sqrt{n}(T_{ni} - t^T\dot{\boldsymbol{l}}(X_1;\theta))\right\}^2 \to 0\end{aligned}$$

となる．また、有界収束定理より

$$\begin{aligned}\boldsymbol{E}[\sum_{i=1}^{n} T_{ni}] &= 2n\left(\int s(\theta + \frac{t}{\sqrt{n}})s(\theta)d\mu - 1\right) \\ &= -n\int [s(\theta + \frac{t}{\sqrt{n}}) - s(\theta)]^2 d\mu \to -\frac{1}{4}t^T I(\theta)t^T\end{aligned}$$

となる．これらより (3.11) が成立することがわかる．

つぎに、尤度比を $\sum_{i=1}^{n} T_{ni}$ で近似できることを示す．そのために、$0 < \epsilon < 1$ に対し、

$$A_n = \{\max_{1 \leq i \leq n} |T_{ni}| < \epsilon\}$$

とおく．A_n 上で Taylor 展開をすれば、

$$\begin{aligned}\log \Pi_{i=1}^{n} \frac{p(X_i;\theta + t/\sqrt{n})}{p(X_i;\theta)} &= 2\sum_{i=1}^{n} \log\left(1 + \frac{1}{2}T_{ni}\right) \\ &= \sum_{i=1}^{n} T_{ni} - \frac{1}{4}\sum_{i=1}^{n} T_{ni}^2 + \frac{1}{6}\alpha_{ni}|T_{ni}|^3 \\ &= t^T W_n(\theta) - \frac{1}{2}t^T I(\theta)t + \sum_{i=1}^{n} T_{ni} \\ &\quad -\left(t^T W_n(\theta) - \frac{1}{4}t^T I(\theta)t\right) \\ &\quad -\frac{1}{4}\left(\sum_{i=1}^{n} T_{ni}^2 - t^T I(\theta)t\right) + \frac{1}{6}\sum_{i=1}^{n} \alpha_{ni}|T_{ni}|^3\end{aligned}$$

を得る．ただし、A_n 上では $|\alpha_{ni}| < 1$ である．したがって、(3.7) と (3.8) を示すためには、$\theta \in C$ （ $C \subset \boldsymbol{R}^k$ はコンパクト集合）と $|\,t\,| < M$ （M はある定数）に対し、

$$\sum_{i=1}^{n} T_{ni}^2 - t^T I(\theta) t \quad \xrightarrow{P_\theta} \quad 0 \tag{3.12}$$

$$\sum_{i=1}^{n} \alpha_{ni} |T_{ni}|^3 \quad \xrightarrow{P_\theta} \quad 0 \tag{3.13}$$

$$P_\theta(A_n^c) \quad \rightarrow \quad 0 \tag{3.14}$$

が成立することをしめせば十分である．

まず、(3.14) を示す．(3.4) より

$$P_\theta(A_n^c) \leq \sum_{i=1}^{n} P_\theta(|T_{ni}| \geq \epsilon) = nP_\theta(|T_{ni}| \geq \epsilon) = o(1)$$

がわかる．

つぎに、(3.12) を示す．(3.12) の右辺を変形すれば

$$\sum_{i=1}^{n} \left\{ T_{ni}^2 - \frac{1}{n}(t^T \dot{\boldsymbol{l}}(X_i;\theta))^2 \right\} + \frac{1}{n}\sum_{i=1}^{n}(t^T \dot{\boldsymbol{l}}(X_i;\theta))^2 - t^T I(\theta) t \tag{3.15}$$

となる．Cauchy – Schwarz の不等式と (3.4) を用いれば、

$$\begin{aligned}
&\boldsymbol{E}_\theta \left| \sum_{i=1}^{n} \{ T_{ni}^2 - \frac{1}{n}(t^T \dot{\boldsymbol{l}}(X_i;\theta))^2 \} \right| \\
&\leq n \sqrt{\boldsymbol{E}_\theta \left| T_{ni} - \frac{1}{n} t^T \dot{\boldsymbol{l}}(X_i;\theta) \right|^2} \sqrt{\boldsymbol{E}_\theta \left| T_{ni} + \frac{1}{n} t^T \dot{\boldsymbol{l}}(X_i;\theta) \right|^2} \\
&= no(\frac{1}{n})O(1) = o(1)
\end{aligned}$$

を得る．したがって、Markov の不等式を (3.15) の第 1 項目に用い、弱対数の法則から

$$\frac{1}{n}\sum_{i=1}^{n}(t^T \dot{\boldsymbol{l}}(X_i;\theta))^2 \xrightarrow{P_\theta} t^T I(\theta) t$$

であることより、(3.12) は確率収束することがわかる．

つぎに、(3.13) を示す．

$$\begin{aligned}
P_\theta(\sum_{i=1}^{n} |T_{ni}|^3 > \epsilon) \quad &\leq \quad P_\theta(\max_{1\leq i\leq n} \sum_{i=1}^{n} |T_{ni}|^3 > \epsilon) \\
&= \quad P_\theta(\max_{1\leq i\leq n} |T_{ni}| \sum_{i=1}^{n} T_{ni}^2 > \epsilon, \sum_{i=1}^{2} T_{ni}^2 \leq 1 + t^T I(\theta) t) \\
&\quad\; + P_\theta(\max_{1\leq i\leq n} |T_{ni}| \sum_{i=1}^{n} T_{ni}^2 > \epsilon, \sum_{i=1}^{2} T_{ni}^2 > 1 + t^T I(\theta) t) \\
&\leq \quad P_\theta(\max_{1\leq i\leq n} |T_{ni}| > \frac{\epsilon}{1 + t^T I(\theta) t}) + P_\theta(\sum_{i=1}^{n} T_{ni}^2 > 1 + t^T I(\theta) t) \rightarrow 0
\end{aligned}$$

が (3.12) と (3.14) よりわかる．

(3.9) を示すために、任意の有界連続な関数 $f: \boldsymbol{R}^k \mapsto \boldsymbol{R}$ とコンパクト集合 $C \subset \Theta$ に対し、

$$\sup_{\theta \in C} |\boldsymbol{E}_\theta f(S_n(\theta)) - \boldsymbol{E}_\theta f(Z)| \tag{3.16}$$

が成立することを示せばよい．ただし、$Z \sim N(0, I(\theta))$ である．(3.16) が成立しないと仮定する．ある列 $\{\theta_n\}$ （$\theta_n \in C$ ）が存在し

$$\limsup_{n\to\infty} |\boldsymbol{E}_{\theta_n} f(S_n(\theta_n)) - \boldsymbol{E}_{\theta_n} f(Z)| > 0 \tag{3.17}$$

のようにできる．C はコンパクトなので、部分列をとれば、ある点 $\theta \in C$ が存在し、$j \to \infty$ のとき、$\theta_{n_j} \to \theta$ できる．

$$Y_i = t^T \dot{\boldsymbol{l}}(X_i; \theta_{n_j})/\sigma_j^2, \qquad \sigma_j^2 = t^T I(\theta_{n_j}) t$$

とおけば、

$$\boldsymbol{L}(t^T W_n(\theta_{n_j})/\sigma_j^2 \mid P_{\theta_{n_j}}) \xrightarrow{d} N(0,1)$$

となる．更に、$I(\theta)$ の連続性より、

$$\boldsymbol{L}(t^T W_n(\theta_{n_j}) \mid P_{\theta_{n_j}}) \xrightarrow{d} N(0, t^T I(\theta) t)$$

となり、(3.17) と矛盾する．したがって、(3.9) が示せた．

最後に (3.9) を示す．

$$K_n(\theta) = \sum_{i=1}^n \log p(X_i; \theta)$$

とおく．(3.9) と $I(\theta)$ の連続性より

$$\begin{aligned}
0 &= K_n(\theta + \frac{t}{\sqrt{n}}) - K_n(\theta) + K_n(\theta + \frac{t+h}{\sqrt{n}}) - K_n(\theta + \frac{t}{\sqrt{n}}) \\
&\quad - \left\{ K_n(\theta + \frac{t+h}{\sqrt{n}}) - K_n(\theta) \right\} \\
&= t^T W_n(\theta) - \frac{1}{2} t^T I(\theta) t + h^T W_n(\theta + \frac{t+h}{\sqrt{n}}) - \frac{1}{2} h^T I(\theta + \frac{t}{\sqrt{n}}) h \\
&\quad - (t+h)^T W_n(\theta) + \frac{1}{2}(t+h)^T I(\theta)(t+h) + o_P(1) \\
&= h^T W_n(\theta + \frac{t}{\sqrt{n}}) - t^T W_n(\theta) + \frac{1}{2} h^T I(\theta) t \\
&\quad - \frac{1}{2} h^T I(\theta + \frac{t}{\sqrt{n}}) + \frac{1}{2} h^T I(\theta) h + o_P(1) \\
&= h^T \{ W_n(\theta + \frac{t}{sqrtn}) - W_n(\theta) + I(\theta) t \} + o_P(1)
\end{aligned}$$

からわかる． □

モデルの正則性を示すために次の補題は有用である．

補題 3.2 ：$\Theta \subset \boldsymbol{R}^k$ を開集合とする．すべての $\theta \in \Theta$ と n に対し p_θ を μ に関する確率密度関数とする．さらに、すべての x について写像 $\theta \mapsto s_\theta(x) = \sqrt{p_\theta(x)}$ は連続微分可能とする．このとき、

$$I_\theta = 4\int \left(\frac{\dot{p}_\theta}{p_\theta}\right)\left(\frac{\dot{p}_\theta}{p_\theta}\right)^T p_\theta d\mu$$

の各成分が定義され、θ に関して連続ならば、関数 $\dot{\boldsymbol{l}}_\theta$ が存在して

$$\int \left[\sqrt{p_{\theta h}} - \sqrt{p_\theta} - \frac{1}{2}h^T \dot{\boldsymbol{l}}_\theta \sqrt{p_\theta}\right]^2 d\mu = o(|h|)$$

が成立する．更に、$\dot{\boldsymbol{l}}_\theta = \dot{p}_\theta / p_\theta$ で与えられる．

証明：仮定から写像 $\theta \mapsto p_\theta(x) = s_\theta^2(x)$ は微分可能で微分 $\dot{p}_\theta = 2s_\theta \dot{s}_\theta$ をもつ．また、s_θ は非負なので $s_\theta = 0$ を満足する x において、$s_\theta = 0$ が成立する．よって、

$$\dot{s}_\theta = \frac{1}{2}\left(\frac{\dot{p}_\theta}{p_\theta}\right)\sqrt{p_\theta}$$

と書ける．また、$p_\theta = 0$ のときは、$\dot{p}_\theta$ は任意にすればよい．さらに、仮定より写像 $\theta \mapsto I_\theta = 4\int \dot{s}_\theta \dot{s}_\theta^T d\mu$ は連続となる．

写像 $\theta \mapsto s_\theta(x)$ は連続微分可能なので、

$$s_{\theta+h}(x) - s_\theta = \int_0^1 h^T \dot{s}_{\theta+uh}(x) du \tag{3.18}$$

と書ける．Chauchy – Schwarz の不等式を使えば、

$$\left\{\int_0^1 h^T \dot{s}_{\theta+uh}(x) du\right\}^2 \le \int_0^1 (h^T \dot{s}_{\theta+uh}(x))^2 du$$

となる．(3.18) と Fubini の定理を用いれば、

$$\int \left(\frac{s_{\theta+th_t} - s_\theta}{t}\right)^2 d\mu = \int\int_0^1 (h_t^T \dot{s}_{\theta+uth_t})^2 du d\mu = \frac{1}{4}\int_0^1 h_t^T I_{\theta+uth_t} du$$

を得る．$\tilde{u} \in [0,1]$ が存在して、

$$\int_0^1 h_t^T I_{\theta+uth_t} h_t du = h_t^T I_{\theta+\tilde{u}th_t} h_t$$

と書けるので、$t \to 0$ と $h_t \to h$ とすれば、$\theta \mapsto I_\theta$ の連続性より

$$\frac{s_{\theta+th_t} - s_\theta}{t} - h^T \dot{s}_\theta \to 0$$

を得る．よって、定理 1.6 を用いれば、

$$\int \left[\frac{s_{\theta+th_t} - s_\theta}{t} - h^T \dot{s}_\theta\right]^2 d\mu \to 0$$

が成立することがわかる． □

3.2 正則推定量

$\boldsymbol{P}$ をモデルとし、$\nu : \boldsymbol{P} \mapsto \boldsymbol{R}^m (m \le k)$ を $\boldsymbol{P}$ から Euclid 空間への写像とする．$X_1, X_2, \ldots, X_n$ を P_θ からの無作為標本とし、$T_n(X_1, X_2, \ldots, X_n) \in \boldsymbol{R}^m$ を $(X_1, X_2, \ldots, X_n)$ に基づく ν の推定量とする．ここで、推定量の列 $\{T_n\}_{n\ge 1}$ の極限の振る舞いを調べて行く．

推定量の望ましい性質として、**一致性**がある．すなわち、すべての $P_0 \in \boldsymbol{P}$ と $\epsilon > 0$ に対して、$n \to \infty$ のとき、

$$P(\mid T_n - \nu(P_0) \mid > \epsilon) \longrightarrow 0$$

を満足することである．

一致性よりも強い性質として $\sqrt{n}$– 一致性がある．すなわち、

$$T_n - \nu(P_0) = O_P\left(\frac{1}{\sqrt{n}}\right)$$

を満足することである．

定義 3.3 ：$T = \{T_n\}$ が θ_0 において**局所正則（locally regular）**であるとは、$\sqrt{n}|\theta_n - \theta_0|$ が有界となる θ_n に対して、

$$\boldsymbol{L}_{P_{\theta_n}}(\sqrt{n}(T_n - \nu(P_{\theta_n})) \xrightarrow{d} Z$$

を満足することである．ここで、$\boldsymbol{L}(Z)$ は $\{\theta_n\}$ に依存しない分布である．

定義 3.4 ：$T = \{T_n\}$ が ν の**漸近線形推定量（asymptotically linear estimator）**であるとは、

$$\psi : \boldsymbol{X} \times \boldsymbol{P} \mapsto \boldsymbol{R}^m$$

が存在し、すべての $P \in \boldsymbol{P}$ に対して、

$$\psi(\cdot; P) \in L_2(P) \quad \text{かつ} \quad \int \psi(x; P) dP(x) = 0$$

を満足し、

$$T_n = \nu(P) + \frac{1}{n}\sum_{i=1}^{n} \psi(X_i; P) + o_P(\frac{1}{n})$$

となる．よって、

$$\sqrt{n}(T_n - \nu(P)) \xrightarrow{d} N(0, \Sigma)$$

となる．ここで、

$$\Sigma = \int \psi(x; P)\psi^T(x; P) dP(x)$$

である．

注意 3.1 ：関数 $\psi(\cdot; P)$ のことを**影響関数（influence function）**とよぶ．

3.3　たたみこみ定理

正則母数モデル $\boldsymbol{P} = \{P_\theta \ll \mu; \theta \in \Theta \subset \boldsymbol{R}^k\}$ を考える．写像 $q : \Theta \mapsto \boldsymbol{R}^m (m \leq k)$ を全微分可能とし、ν を

$$\nu(P_\theta) = q(\theta)$$

で定義する．ここで、

$$\dot{q} : m \times k = \left(\frac{\partial q_i}{\partial \theta_j}\right)_{i=1,2,\ldots,m, j=1,2,\ldots,k}$$

とする．

ν に対する**情報限界（informaiton bound）**を

$$I^{-1}(P_\theta \mid \nu, \boldsymbol{P}) = \dot{q}(\theta) I^{-1}(\theta) \dot{q}^T(\theta) \tag{3.19}$$

で定義する．また、ν に対する**有効影響関数（efficient influence function）**を

$$\tilde{\boldsymbol{l}}(\cdot, P_\theta \mid \nu, \boldsymbol{P}) = \dot{q}(\theta) I^{-1}(\theta) \dot{\boldsymbol{l}}(\theta) \tag{3.20}$$

で定義する．

定理 3.1 （たたみこみ定理）：$\boldsymbol{P}$ を正則母数モデルとし、$X_1, X_2, \ldots, X_n$ を $P_\theta \in \boldsymbol{P}$ に従う独立同一標本とする．更に、$\{T_n\}$ を θ における $\nu(P_\theta) = q(\theta)$ の正則推定量とする．すなわち、$\theta_n = \theta + \frac{t_n}{\sqrt{n}}, t_n \in \boldsymbol{R}^k$ とし、$n \to \infty$ のとき、

$$\sqrt{n}(T_n - \nu(P_{\theta_n})) \xrightarrow{d} Z, \quad (P_{\theta_n} \text{ のもとで })$$

が成立する．このとき、

$$Z = Z_\theta + \Delta_\theta$$

が成立する．ただし、$Z_\theta \perp \Delta_\theta$ であり、

$$Z_\theta \sim N(0, I^{-1}(P_\theta \mid \nu, \boldsymbol{P}))$$

である．より一般に、

$$\boldsymbol{L}\begin{bmatrix} \sqrt{n}(T_n - \nu(P_\theta)) - (1/\sqrt{n}) \sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i, P_\theta \mid \nu, \boldsymbol{P}) \\ (1/\sqrt{n}) \sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i, P_\theta \mid \nu, \boldsymbol{P}) \end{bmatrix} \xrightarrow{d} \boldsymbol{L}\begin{bmatrix} \Delta_\theta \\ Z_\theta \end{bmatrix}$$

が成立する．ただし、$\Delta_\theta \perp Z_\theta$ である．

証明：θ を固定する．$\{T_n\}$ の正則性とモデルの正則性より、

$$Z_n \equiv \sqrt{n}(T_n - \nu(P)) \qquad V_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\boldsymbol{l}}(X_i, \theta)$$

はそれぞれ分布収束（それぞれの収束先を Z と $V \sim N(0, I^{-1}(\theta))$ とかくことにする）する．したがって、(Z_n, V_n) は一様に tight となる[4]．任意の部分列 $\{n'\}$ をひとつ固定する．すると、Prohorov の定理より、$\{n'\}$ の中にある部分列 $\{n''\}$ が存在し、

$$(Z_{n''}, V_{n''}) \xrightarrow{d} (Z, V)$$

を満足するようにとれる．以後、n'' を n と記すことにする．

$$W_n = \log \Pi_{i=1}^n \frac{p(X_i; \theta + t/\sqrt{n})}{p(X_i; \theta)}$$

とおく．命題 3.1 より、

$$(Z_n, W_n) \xrightarrow{d} (Z, t^T V - \frac{1}{2} t^T I(\theta) t)$$

となる．更に、$\{T_n\}$ の正則性より、$\theta_n = \theta + t/\sqrt{n}$ のもとで

$$\sqrt{n}(T_n - q(\theta + \frac{t}{\sqrt{n}})) \xrightarrow{d} Z$$

である．ただし、$q(\theta) = \nu(P_\theta)$ と記した．更に、ν のなめらかさより

$$\sqrt{n}(q(\theta + \frac{t}{\sqrt{n}}) - q(\theta)) = \dot{q}(\theta)t + o(n^{-1/2})$$

となることに注意すれば、P_{θ_n} のもとで

$$\begin{aligned} Z_n &= \sqrt{n}(T_n - q(\theta)) = \sqrt{n}(T_n - q(\theta + \frac{t}{\sqrt{n}}) + \sqrt{n}(q(\theta + \frac{t}{\sqrt{n}}) - q(\theta)) \\ &\to Z + \dot{q}(\theta)t \end{aligned} \tag{3.21}$$

を得る．よって、(3.21) より

$$\boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)] \to \exp(ia^T \dot{q}(\theta)t) \boldsymbol{E}[\exp(ia^T Z)] \tag{3.22}$$

となる．

一方、(3.6) から

$$\boldsymbol{E}_{\theta_n}[1\{W_n = \infty\}] \leq n \boldsymbol{E}_{\theta_n}[1\{p(\theta) = 0\}] = o(1)$$

---

[4] A をコンパクト集合とする．Portneantou の定理より、任意の $\epsilon > 0$ に対し、

$$\limsup_{n \to \infty} P(Z_n \in A^c) \leq P(Z \in A^c) \leq \epsilon$$

できる．よって、$\{Z_n\}$ は一様に tight となる．更にコンパクト集合 B に対し、

$$P((Z_n, V_n) \in A \times B) \leq P(Z_n \in A) \leq \epsilon$$

とできるので、(Z_n, V_n) も tight であることがわかる．

に注意すれば、

$$\begin{aligned}\boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)] &= \boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)1\{W_n<\infty\}+\exp(ia^T Z_n)1\{W_n=\infty\}] \\ &= \boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)1\{p(\theta)>0\}]+\boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)1\{W_n=\infty\}] \\ &= \boldsymbol{E}_{\theta}[\exp(ia^T Z_n+W_n)]+o(1) \qquad (3.23)\end{aligned}$$

がわかる．また、W_n は分布収束するので、

$$\{|\exp(ia^T Z_n+W_n)|\}=\{\exp(W_n)\}$$

は一様可積分であることがわかる[5]．よって、(3.22) と (3.23) を使えば、

$$\begin{aligned}\boldsymbol{E}[\exp(ia^T Z+t^T V-\frac{1}{2}t^T I(\theta)t)] &= \lim_{n\to\infty}\boldsymbol{E}_{\theta}[\exp(ia^T Z_n+W_n)] \\ &= \boldsymbol{E}_{\theta_n}[\exp(ia^T Z_n)]+o(1) \\ &= \exp(ia^T\dot{q}(\theta)t)\boldsymbol{E}[\exp(ia^T Z)]+o(1) \qquad (3.24)\end{aligned}$$

を得る．(3.24) の両辺を t の関数として、複素平面上で正則であるので、実数軸上で一致していれば、解析接続から複素平面上でも一致することがわかる．特に、$t^T=-i(a-b)^T\dot{q}(\theta)I^{-1}(\theta)$ ととれば、(3.24) は

$$\begin{aligned}&\boldsymbol{E}[\exp\{ia^T(Z-\dot{q}(\theta)I^{-1}(\theta)V)+ib^T\dot{q}(\theta)I^{-1}(\theta)\}] \\ &\quad=\boldsymbol{E}[\exp\{ia^T Z+\frac{1}{2}a^T\dot{q}(\theta)I^{-1}(\theta)\dot{q}(\theta)a\}]\exp\{-\frac{1}{2}b^T\dot{q}(\theta)I^{-1}(\theta)\dot{q}(\theta)b\} \qquad (3.25)\end{aligned}$$

がわかる．(3.25) は $(Z_{n''}-\dot{q}(\theta)I^{-1}(\theta)V_{n''},\dot{q}(\theta)V_{n''})$ の特性関数の極限[6] となり、部分列 $\{n'\}$ の取り方に依存しないので、$(Z_n-\dot{q}(\theta)I^{-1}(\theta)V_n,\dot{q}(\theta)V_n)$ の特性関数の極限が (3.25) になることがわかる．更に、(3.25) において $b=0$ とおけば、

$$\boldsymbol{E}[\exp\{ia^T(Z-\dot{q}(\theta)I^{-1}(\theta)V)\}]=\boldsymbol{E}[\exp\{ia^T Z+\frac{1}{2}a^T\dot{q}(\theta)I^{-1}(\theta)\dot{q}(\theta)a\}] \qquad (3.26)$$

を得る．(3.25) と (3.26) を合わせば、

$$\begin{aligned}&\boldsymbol{E}[\exp\{ia^T(Z-\dot{q}(\theta)I^{-1}(\theta)V)+ib^T\dot{q}(\theta)I^{-1}(\theta)V\}] \\ &\quad=\boldsymbol{E}[\exp\{ia^T(Z-\dot{q}(\theta)I^{-1}(\theta)V\}]\exp\{-\frac{1}{2}b^T\dot{q}(\theta)I^{-1}(\theta)\dot{q}(\theta)b\} \qquad (3.27)\end{aligned}$$

がわかる．(3.27) において $b=a$ とおけば、

$$\boldsymbol{L}(T_n)\xrightarrow{d}\boldsymbol{L}(Z_\theta+\Delta_\theta)$$

がわかる．よって、定理は証明された． □

[5] Contineous mapping 定理を使えば、$\exp(W_n)\xrightarrow{d}\exp(t^T V-(1/2)t^T I(\theta)t)$ がわかる． 更に、$V\sim N(0,I^{-1}(\theta))$ から $\boldsymbol{E}\exp(t^T V-(1/2)t^T I(\theta)t)=1$ と $E_\theta\exp(W_n)=1$ が成立するので、BKRW の Lemma 2b(pages 469) より $\{W_n\}$ は一様可積分であることがわかる．

[6] すなわち、

$$\lim_{n''\to\infty}\boldsymbol{E}[\exp\{a^T(Z_{n''}-\dot{q}(\theta)I^{-1}(\theta)V_{n''})+b^T\dot{q}(\theta)V_{n''}\}]$$

が (3.25) の右辺に一致する．

3.4　漸近ミニマック定理

この節では、たたみこみ定理から得られる漸近ミニマック定理について述べる．この定理を証明するときに必要な補題をまず述べる．

補題 3.3 （Anderson の補題）：X と Y を独立な d – 次元確率ベクトルとし、X は原点に関して対称でひと山の確率密度関数（すなわち、$f_X(x) = f_X(-x)$ が成立する）をもつものとする．更に、C を原点に関して対称な、$\boldsymbol{R}^d$ の凸集合とする．このとき、

$$P(X + Y \in C) \leq P(X \in C)$$

が成立する．

証明：Ibragimov and Has'minskii (1981) の pages 155 – 156 を参照されたい． □

定義 3.5 ：関数 $l : \boldsymbol{R}^m \mapsto \boldsymbol{R}^+$ が **bowl – shaped** であるとは、

(i) ．任意の $x \in \boldsymbol{R}^m$ に対し、$l(x) = l(-x)$ を満たす．

(ii) ．任意の定数 $c \geq 0$ に対し、$\{x : l(x) \leq c\}$ が凸集合になる

を満足することである．

l を損失関数としたとき、推定量 $T = \{T_n\}$ の**危険関数**を

$$\boldsymbol{E}_\theta l(\sqrt{n}(T_n - q(\theta)))$$

で定義する．ただし、$q(\theta) = \nu(P_\theta)$ とした．

定理 3.2 （Asymptotic optimal theorem）：$T = \{T_n\}$ を正則な推定量とし、l を bowl – shaped な損失関数とする．このとき、

$$\liminf_{n \to infty} \boldsymbol{E}_\theta(l(\sqrt{n}(T_n - q(\theta))) \geq \boldsymbol{E}(Z_\theta)$$

を満足する．ただし、$Z_\theta \sim N(0, I^{-1}(P_\theta|\nu, \boldsymbol{P})$ である．

証明：$C_{ik} \equiv \{y : l(y) \leq i2^{-k}\}$ とし、

$$l_k(x) = \frac{1}{2^k} \sum_{i=1}^{k2^k} (1 - 1_{C_{ik}}(x))$$

とおく．$k \to \infty$ のとき、$l_k \uparrow l$ で、l_k は bowl – shaped である．定理 3.1 （たたみこみ定理）から

$$\sqrt{n}(T_n - q(\theta)) \xrightarrow{d} Z_\theta + \Delta_\theta$$

となる．l_k は非負で Lebeague 測度ゼロの集合を除いて、ほとんどいたるところで連続であるので、Portanteau の補題[7] を使えば、

$$\liminf_{n\to\infty} \boldsymbol{E}_\theta l(\sqrt{n}(T_n - q(\theta))) \geq \liminf_{n\to\infty} \boldsymbol{E}_\theta l_k(\sqrt{n}(T_n - q(\theta))) \geq \boldsymbol{E} l_k(Z_\theta + \Delta_\theta)$$

を得る．更に、補題 3.3（Anderson の補題）を用いれば、

$$\boldsymbol{E} l_k(Z_\theta + \Delta_\theta) \geq \boldsymbol{E} l_k(Z_\theta)$$

がわかる．最後に単調収束定理を用いれば、

$$\liminf_{n\to\infty} \boldsymbol{E}_\theta l(\sqrt{n}(T_n - q(\theta))) \geq \lim_{k\to\infty} \boldsymbol{E} l_k(Z_\theta) = \boldsymbol{E} \lim_{k\to\infty} l_k(Z_\theta) = \boldsymbol{E} l(Z_\theta)$$

を得る．

□

[7] Vaart (1999) の pages 6 の Lemma 2.2(iv) を参照されたい．

3.5 幾何的考察と擾乱母数

$\theta = (\nu^T, \eta^T)^T, \nu \in \boldsymbol{N} \subset \boldsymbol{R}^m, \eta \in \boldsymbol{H} \subset \boldsymbol{R}^{k-m}$ とし、ν を興味のある母数をする．モデルは正則とし、$\theta_0^T = (\nu_0^T, \eta_0^T)^T \quad \in \Theta$ におけるスコア関数を $\dot{\boldsymbol{l}}$ と記し、$\boldsymbol{P}$ の中で P_{θ_0} における θ_0 に対する有効影響関数を $\tilde{\boldsymbol{l}} = I^{-1}(\theta_0)\dot{\boldsymbol{l}}$ と書く．

いま、

$$\dot{\boldsymbol{l}} = \begin{bmatrix} \dot{\boldsymbol{l}}_1 \\ \dot{\boldsymbol{l}}_2 \end{bmatrix}, \quad \tilde{\boldsymbol{l}} = \begin{bmatrix} \tilde{\boldsymbol{l}}_1 \\ \tilde{\boldsymbol{l}}_2 \end{bmatrix}$$

なる分割を考える．ただし、$\dot{\boldsymbol{l}}_1, \tilde{\boldsymbol{l}}_1$ は m 次元ベクトル、$\dot{\boldsymbol{l}}_2, \tilde{\boldsymbol{l}}_2$ は $(k-m)$ 次元ベクトルとする．更に、これに対応して、$I(\theta_0)$ を分割し、

$$I(\theta_0) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} \quad \text{（ただし、} I_{11} \text{ は } m \times m\text{）}$$

と

$$I^{-1}(\theta_0) = \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix} = \begin{bmatrix} I_{11\cdot 2}^{-1} & -I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} \\ -I_{22\cdot 1}^{-1} I_{21} I_{11}^{-1} & I_{22\cdot 1}^{-1} \end{bmatrix}$$

と記す．

ただし、$I_{11\cdot 2} = I_{11} - I_{12}I_{22}^{-1}I_{21}, I_{22\cdot 1} = I_{22} - I_{21}I_{11}^{-1}I_{12}$ である．$q(\theta) = \nu$ とおけば、$\dot{q} = (I_m, 0)$ になることに注意すれば、(3.19) と (3.20) より、ν に対する情報限界は $I^{11} = I_{11\cdot 2}^{-1}$ となり、$\boldsymbol{P}$ における ν に対する有効影響関数は

$$\tilde{\boldsymbol{l}}_1 = I^{11}\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2 = I_{11\cdot 2}^{-1}(\dot{\boldsymbol{l}}_1 - I_{12}I_{22}^{-1}\dot{\boldsymbol{l}}_2)$$

となる．また、

$$I_{11\cdot 2} = \boldsymbol{E}_{\theta_0}[(\dot{\boldsymbol{l}}_1 - I_{12}I_{22}^{-1}\dot{\boldsymbol{l}}_2)(\dot{\boldsymbol{l}}_1 - I_{12}I_{22}^{-1}\dot{\boldsymbol{l}}_2)^T] = \boldsymbol{E}_{\theta_0}[\boldsymbol{l}_1^*(\boldsymbol{l}_1^*)^T]$$

である．$q(\theta) = \nu(P_\theta)$ として定理 2.1 を用いると $\dot{q}(\theta) = (I_m, 0)$ となり、

$$I^{-1}(P_{\theta_0} \mid \nu, \boldsymbol{P}) = \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{q}^T(\theta_0) = I_{11\cdot 2}^{-1}$$

となる．よって、$\boldsymbol{l}_1^*$ を $\boldsymbol{P}$ における ν に対する有効スコア関数とよび、$I_{11\cdot 2}$ を ν に対する情報量とよぶ．

他方、$\eta = \eta_0$ が既知として、

$$\boldsymbol{P}_1(\eta_0) = \{P_\theta : \eta = \eta_0, \nu \in \boldsymbol{N}\}$$

における ν の情報限界は I_{11}^{-1} となる．

$$I^{11} = I_{11\cdot 2}^{-1} = I_{11}^{-1} + I^{12}(I^{22})^{-1}I^{21}$$

に注意すれば、η 未知であることにより、$I^{12}(I^{22})^{-1}I^{21}$ だけ情報限界が増加したことになる．

命題 3.2 ：*A.* 有効スコア関数 $\boldsymbol{l}_1^*(\cdot, P_{\theta_0} \mid \nu, \boldsymbol{P})$ は $L_2(P_{\theta_0})$ 空間における $[\dot{\boldsymbol{l}}_2]$ の直交補空間へスコア関数 $\dot{\boldsymbol{l}}_1$ を射影したものである．
B. 有効影響関数 $\tilde{\boldsymbol{l}}(\cdot, P_{\theta_0} \mid \nu, \boldsymbol{P}_1(\eta_0))$ は $L_2(P_{\theta_0})$ 空間における $[\dot{\boldsymbol{l}}_1]$ へ有効影響関数 $\tilde{\boldsymbol{l}}_1(\cdot, P_{\theta_0} \mid \nu, \boldsymbol{P})$ を射影したものである．

証明：A. $\boldsymbol{l}^*$ の定義より明らかである．
B. $\tilde{\boldsymbol{l}}_1 - \tilde{\boldsymbol{l}}$ が $\dot{\boldsymbol{l}}_1$ に直交していることを示せばよい．実際、

$$\langle \tilde{\boldsymbol{l}}_1 - \tilde{\boldsymbol{l}}, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0}} = \langle I^{11}\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0}} = I^{11}I_{11} + I^{12}I_{21} - \mathrm{id} = 0$$

よりわかる． □

注意 3.2 ：

$$\begin{aligned} \tilde{\boldsymbol{l}}_1 - \tilde{\boldsymbol{l}} &= I^{11}\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2 - I_{11}^{-1}\dot{\boldsymbol{l}}_1 \\ &= (I_{11}^{-1} + I^{12}(I^{22})^{-1}I^{21})\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2 - I_{11}^{-1}\dot{\boldsymbol{l}}_1 \\ &= I^{12}(I^{22})^{-1}(I^{21}\dot{\boldsymbol{l}}_1 + I^{22}\dot{\boldsymbol{l}}_2) = I^{12}(I^{22})^{-1}\dot{\boldsymbol{l}}_2. \end{aligned}$$

命題 3.3 ：$m = 1$ とする．T_n を ν の漸近線形推定量とし、その影響関数を ψ とする．このとき、以下が成立する．
A. T_n が *Gaussian* 正則であるための必要十分条件は

$$\psi - \tilde{\boldsymbol{l}}_1 \perp \dot{\boldsymbol{P}} = [\dot{\boldsymbol{l}}_1, \dot{\boldsymbol{l}}_2] \tag{3.28}$$

である．更に、*(3.28)* を書き直せば、

$$\langle \psi, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0})} = 1 \tag{3.29}$$

かつ

$$\psi \perp [\dot{\boldsymbol{l}}_2] \tag{3.30}$$

である．
B. T_n が正則のとき、次は同値である．

(i) $\psi \in \dot{\boldsymbol{P}} = [\dot{\boldsymbol{l}}_1, \dot{\boldsymbol{l}}_2]$.

(ii) $\psi = \tilde{\boldsymbol{l}}_1$.

証明：(3.28) が (3.29) かつ (3.30) と同値であることだけを示し、残りは命題 3.4 の証明で示す．

$$\begin{aligned} 0 &= \langle \Psi - \tilde{\boldsymbol{l}}, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0})} = \langle \Psi, \dot{\boldsymbol{l}} \rangle_{L_2(P_{\theta_0})} - \langle I^{11}\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0})} \\ &= \langle \Psi, \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_{\theta_0})} - 1 \\ 0 &= \langle \Psi - \tilde{\boldsymbol{l}}_1, \dot{\boldsymbol{l}}_2 \rangle_{L_2(P_{\theta_0})} = \langle \Psi, \dot{\boldsymbol{l}}_2 \rangle_{L_2(P_{\theta_0})} - \langle I^{11}\dot{\boldsymbol{l}}_1 + I^{12}\dot{\boldsymbol{l}}_2, \dot{\boldsymbol{l}}_2 \rangle_{L_2(P_{\theta_0})} = 0 \end{aligned}$$

からわかる．
□

命題 3.4 ：q を Θ から $\boldsymbol{R}^m$ への写像とする．T_n は θ_0 における $\nu(P_\theta)=q(\theta)$ の漸近線形推定量とし、その影響関数を ψ とする．このとき、以下が成立する．
A. T_n が *Gaussian* 正則であるための必要十分条件は
(i)　$q(\theta)$ は θ_0 で微分可能で導関数を $\dot{q}(\theta_0)$ とし、

$$\tilde{\boldsymbol{l}}=\tilde{\boldsymbol{l}}(\cdot,P_{\theta_0}\mid\nu,\boldsymbol{P})=\dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}}(\theta_0)$$

とおいたとき、

$$\psi-\tilde{\boldsymbol{l}}\perp\dot{\boldsymbol{P}}[\dot{\boldsymbol{l}}_1,\dot{\boldsymbol{l}}_2] \tag{3.31}$$

が成立することである．更に、(3.31) は

$$\boldsymbol{E}_{\theta_0}[\psi\dot{\boldsymbol{l}}^T]=\dot{q}(\theta_0) \tag{3.32}$$

と同値である．
B. T_n が正則のとき、$\psi\in(\dot{\boldsymbol{P}})^m$ であるための必要十分条件は

$$\psi=\tilde{\boldsymbol{l}}=\dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}}(\theta_0) \tag{3.33}$$

である．

証明：A. $\{T_n\}$ の漸近線形性と命題 3.1 から、$t_n\to t,n\to\infty$ のとき

$$\boldsymbol{L}_{\theta_0}\left[\begin{array}{c}\sqrt{n}(T_n-q(\theta_0))\\ \log\Pi_{i=1}^n\{p(X_i;\theta_0+t_n/\sqrt{n})/p(X_i;\theta_0)\}\end{array}\right]\xrightarrow{d}N\left(\left[\begin{array}{c}0\\-\Sigma_{22}/2\end{array}\right],\Sigma\right) \tag{3.34}$$

となる．ただし、

$$\Sigma=\left[\begin{array}{cc}\Sigma_{11}&\Sigma_{12}\\ \Sigma_{21}&\Sigma_{22}\end{array}\right],\Sigma_{11}=\boldsymbol{E}[\Psi\Psi^T],\Sigma_{12}=\boldsymbol{E}[\Psi\dot{\boldsymbol{l}}]t,\Sigma_{22}=t^TI(\theta_0)t=t^T(\boldsymbol{E}[\dot{\boldsymbol{l}}\dot{\boldsymbol{l}}^T])t$$

である．補題 2.3（Le Cam の第 3 補題）を使えば、

$$\boldsymbol{L}_{\theta_0+t_n/\sqrt{n}}(\sqrt{n}T_n-q(\theta_0))\xrightarrow{d}N(\Sigma_{12},\Sigma_{11}) \tag{3.35}$$

を得る．一方、$\{T_n\}$ の正則性より

$$\boldsymbol{L}_{\theta_0+t_n/\sqrt{n}}(\sqrt{n}T_n-q(\theta_0+\frac{t_n}{\sqrt{n}}))\xrightarrow{d}N(0,\Sigma_{11}) \tag{3.36}$$

がわかる．したがって、(3.35) と (3.36) から

$$\sqrt{n}(q(\theta_0+\frac{t_n}{\sqrt{n}})-q(\theta_0))\to\Sigma_{12}=\boldsymbol{E}_{\theta_0}[\Psi\dot{\boldsymbol{l}}^T]t \tag{3.37}$$

となり、これから q は可微分で微分 $\dot{q}(\theta_0)=\boldsymbol{E}[\Psi\dot{\boldsymbol{l}}^T]$ をもつことがわかる．
　一方、q が可微分で、(3.32) が成立するとき、(3.35) ならば、(3.36) であることがわかる．したがって、$\{T_n\}$ は正則である．
B. A から $\dot{q}(\theta_0)$ と $I(\theta_0)$ は $\mathrm{well-defined}$ である．更に、$\tilde{\boldsymbol{l}}\in\dot{\boldsymbol{P}}$ と (3.31) から、$\Psi\in\dot{\boldsymbol{P}}$ であるための必要十分条件は $\Psi-\tilde{\boldsymbol{l}}=0$ であることがわかる．
□

第4章　無限次元モデルにおける Euclid 母数に対する情報限界

4.1　導入と概観

$\boldsymbol{P} \subset \boldsymbol{M}$ を（一般）のモデルとし、母数 $\nu : \boldsymbol{P} \mapsto \boldsymbol{R}^m$ を推定の目標とする．ここで真のモデルを $P_0 \in \boldsymbol{M}$ とする．また、

$$\boldsymbol{M} = \{\ \mu \text{ に支配される } (\boldsymbol{X}, \boldsymbol{A}) \text{ 上の確率測度のすべて } \}$$

であった．

定義 4.1 ：$\boldsymbol{Q}$ を $\boldsymbol{P}$ の部分集合で正則母数モデルのとき、$\boldsymbol{Q}$ を $\boldsymbol{P}$ の**正則母数部分モデル**とよぶ．P_0 が $\boldsymbol{P}$ の任意の正則母数部分モデルに属するならば、$\boldsymbol{P}$ は P_0 において **正則**とよぶ．

定義 4.2 ：推定量 T が $\boldsymbol{P}$ のすべての正則母数部分モデルで（局所）正則のとき、T を $\boldsymbol{P}$ 上で（局所）正則であるという．

注意 4.1 ：局所または一様 Gaussian 正則や線形性も定義 *3.2* と同様に定める．

$m = 1$ と仮定する．$\boldsymbol{P}$ 上で ν の推定量 T の分散が $\Sigma(P_0, T)$ で、P_0 を含む正則母数部分モデルを $\boldsymbol{Q}$ とする．ν を $\boldsymbol{Q}$ の関数とみたとき、$I^{-1}(P_0 \mid \nu, \boldsymbol{Q})$ が定義できるのに十分な「なめらかさ」があるならば、

$$\Sigma(P_0, T) \geq I^{-1}(P_0 \mid \nu, \boldsymbol{Q})$$

である．このとき、

$$I^{-1}(P_0 \mid \nu, \boldsymbol{P}) \equiv \sup \left\{ I^{-1}(P_0 \mid \nu, \boldsymbol{Q}) : \text{ 任意の正則母数部分モデル } \quad \boldsymbol{Q} \subset \boldsymbol{P} \ \right\}$$

と定義する．

例 4.1 （確率密度関数の推定）: $\boldsymbol{P}$ を連続型確率密度関数 $p(\cdot)$ をもつ $\boldsymbol{R}$ 上の分布からなる集合とし、

$$\nu(P) = p(0)$$

とする．このとき、

$$I^{-1}(P_0 \mid \nu, \boldsymbol{P}) = \infty$$

となることが以下からわかる．実数 $|\eta|<1$ に対して、

$$p(x;\eta)=p(x)(1+\eta h(x))$$

とおく．ただし、

$$\| h \|_\infty \le 1 \qquad と \qquad \int h(x)p(x)d\mu(x)=0$$

とする．$p(x;\eta)$ の分布関数を $P\eta$ と記せば、

$$\nu(P_\eta)=p(0)(1+\eta h(0))\equiv q(\eta)$$

となる．このとき、$\dot{q}(0)=p(0)h(0)$ である．$\boldsymbol{Q}=\{P_\eta\}$ は正則なことが容易にわかる．また、$s_\eta=\sqrt{p(\cdot;\eta)}$ とおけば、簡単な計算から、

$$\begin{aligned}
\boldsymbol{\dot{l}} &= \left.\frac{2\dot{s}_\eta}{s_\eta}\right|_{\eta=0} = \left.\frac{p(x)h(x)}{\sqrt{p(x)(1+\eta h(x))}}\Big/\sqrt{p(x)(1+\eta h(x))}\right|_{\eta=0}\\
&= \left.\frac{p(x)h(x)}{p(x)(1+\eta h(x))}\right|_{\eta=0},\\
I &= \boldsymbol{E}_{\eta=0}[\boldsymbol{\dot{l}\dot{l}}]=\int h^2(x)p(x)d\mu(x)
\end{aligned}$$

がわかる．よって、

$$I^{-1}(P_0\mid\nu,\{P_\eta\})=\dot{q}(0)I^{-1}\dot{q}(0)=\frac{(p(0)h(0))^2}{\int h^2(x)p(x)d\mu(x)}$$

である．ここで、

$$g_j(x)=\begin{cases}0 & (x\ge 1/j)\\ -jx+1 & (1/j>x\ge 0)\\ jx+1 & (0>x\ge -1/j)\\ 0 & (x<-1/j)\end{cases}$$

とし、

$$h_j(x)=g_j(x)-\int g_j(x)d\mu(x)$$

とおけば、$\| h_j \|_\infty=1$ かつ $\int h_j(x)d\mu(x)=0$ となる．しかし、

$$\int h_j^2(x)p(x)d\mu(x)\longrightarrow 0 \quad (j\to\infty)$$

になることがわかる．故に、

$$I^{-1}(P_0\mid\nu,\boldsymbol{P})=\infty$$

となる．

問い 4.1 ：$\int h_j^2(x)p(x)d\mu(x)\to 0$ を示せ．

4.2　接空間

$\boldsymbol{P}$ をノンパラメトリックまたはセミパラメトリックモデルとする．

定義 4.3 ：母数モデル $\boldsymbol{Q} = \{Q_\theta : \theta \in \boldsymbol{R}^k\}$ が Hellinger 微分可能で導関数 $\dot{\boldsymbol{l}} \in L_2(Q_\theta)$ をもつとは、$| h | \to 0$ （$h \in \boldsymbol{R}^k$）のとき

$$\left\{\int [\sqrt{q_{\theta+h}} - \sqrt{q_\theta} - \frac{1}{2} h \dot{\boldsymbol{l}}_\theta^T \sqrt{q_\theta}]^2 d\mu(x)\right\}^{1/2} = o(h)$$

が成立することである．ただし、$\boldsymbol{Q}$ は μ に支配された確率測度の集合であり、$q_\theta \equiv dQ_\theta/d\mu$ とおいた．

定義 4.4 ：$P_0 \in \boldsymbol{P}$ をひとつ固定する．P_0 を含む Hellinger 微分可能な 1 次元部分母数モデル $\boldsymbol{Q}$ のすべてに対して、P_θ における接空間を考える．その集合

$$\{\dot{\boldsymbol{l}}(\cdot; \boldsymbol{Q}) : \text{任意の Hellinger 微分可能な 1 次元母数部分モデル } \boldsymbol{Q} \subset \boldsymbol{P}\} \equiv \dot{\boldsymbol{P}}^0$$

を P_0 におけるモデル $\boldsymbol{P}$ の**接集合**とよぶ．更に、$\dot{\boldsymbol{P}}^0$ の closed linear span space

$$\dot{\boldsymbol{P}} = \overline{lin(\dot{\boldsymbol{P}}^0)}$$

を P_0 におけるモデル $\boldsymbol{P}$ の**接空間**とよぶ．

例 4.2 ：モデルを

$$\boldsymbol{P} = \{P \ll \mu : (\boldsymbol{X}, \boldsymbol{A}) \text{ 上の } \mu \text{ に支配されたすべての確率測度 }\}$$

とし、真のモデルを P_0 と書くことにする．このとき、

$$\dot{\boldsymbol{P}} = L_2^0(P_0) \qquad \text{かつ} \qquad \dot{\boldsymbol{P}}^0 = L_2^0(P_0)$$

となる．

$\dot{\boldsymbol{P}}^0 \subset L_2^0(P_0)$ は自明なので、$\dot{\boldsymbol{P}}^0 \supset L_2^0(P_0)$ を示せばよい．任意の $h \in L_2^0(P_0)$ をとる．すなわち、

$$\int h(x) dP_0(x) = 0 \qquad \text{かつ} \qquad \int h^2(x) dP_0 < \infty$$

である． 更に、$\psi : \boldsymbol{R} \mapsto [0, \infty)$ を有界、連続微分可能な関数で、その導関数 $\dot{\psi}$ は有界、$\psi(0) = \dot{\psi}(0) = 1$、かつ $\dot{\psi}/\psi$ は有界なものとする[1]．いま、$p_0 = dP_0/d\mu$ とし、

$$p_\eta^* = \frac{p_0(x)\psi(\eta h(x))}{\int p_0(x)\psi(\eta h(x)) d\mu(x)}$$

[1] 例として、$\psi(x) = 2/(1 + \exp(-2x))$ がある．

とおく[2]．このとき、

$$\log p_\eta^* \;=\; \log(\psi(\eta h(x))) - \log[\int p_0(x)\psi(\eta h(x))d\mu(x)] \\ + \; constant \;（\eta \text{ に依存しない項}）$$

となる．さらに、$\dot\psi$ の定義から積分記号と微分記号の順序交換が保障されることに注意して、

$$\begin{aligned} \dot{\boldsymbol{l}} &= h(x)\frac{\dot\psi}{\psi}(\eta h(x))\bigg|_{\eta=0} - \frac{\int p_0(x)\dot\psi(\eta h(x))h(x)d\mu(x)}{\int p_0(x)\psi(\eta h(x))d\mu(x)}\bigg|_{\eta=0} \\ &= h(x) - \int h(x)p_0(x)d\mu(x) = h(x) \end{aligned}$$

となることから、$h \in \dot{\boldsymbol{P}}^0$ となる．よって、$\dot{\boldsymbol{P}}^0 \supset L_2^0(P_0)$．

定義 4.5 ：$\boldsymbol{P}$ が有限次元ではない接空間 $\dot{\boldsymbol{P}}(P)$ を持ち、それが $L_2^0(P)$ の真部分集合ならば、$\boldsymbol{P}$ は**セミパラメトリックモデル**という．

例 4.3 ：

$$\boldsymbol{P} = \{P \ll \mu : \ P \text{ は } \theta_0 \text{ について対称}\}$$

とする．一般性を失わず、$\theta_0 = 0$ としてよい．$p(\cdot;\eta)$ を $p(\cdot;0) = p_0 = dP_0/d\mu$ を通る1次元母数部分モデルとし、$\eta \to 0$ のとき、

$$\int \left[\frac{s(\cdot;\eta) - s(\cdot;0)}{\eta} - \dot s\right]^2 d\mu \to 0$$

を満足すると仮定すれば、$\dot s$ も原点に関して対称となる．したがって、

$$\dot{\boldsymbol{P}}^0 \subset \{h \in L_2^0(P_0) : h(x) = h(-x), a.e.P_0\}$$

となる．例 4.2 と同様にすれば[3]、$\dot{\boldsymbol{P}}^0 = \dot{\boldsymbol{M}}_s$ がわかる．したがって、$\dot{\boldsymbol{P}}^0$ は $L_2^0(P_0)$ の真部分集合となることがわかる．

例 4.4 （対称位置母数モデル）：g を原点に関して対称で

$$I_g = \int \frac{(g')^2}{g}(x)dx < \infty$$

[2] $\eta = 0$ のとき、$p_0(x) = p_\eta^*(x)\big|_{\eta=0}$ になることに注意．

[3] $\dot{\boldsymbol{P}}^0 \subset \dot{\boldsymbol{M}}_s$ は自明なので、$\dot{\boldsymbol{P}}^0 \supset \dot{\boldsymbol{M}}_s$ を示せばよい．
$h \in \dot{\boldsymbol{M}}_s$ とし、

$$p_\eta^* = \frac{p_0(x)\psi(\eta h(x))}{\int p_0(x)\psi(\eta h(x))d\mu(x)}$$

とおけば、$p_\eta^*(-x) = p_\eta^*(x)$ になることより

$$\{p_\eta^*(x) : |\eta| < 1\} \subset \boldsymbol{P}$$

は p_0 を通る部分モデルになる．あとは、例 4.2 と同じ．

なる関数とし、

$$\boldsymbol{P} = \left\{ P_{\theta,g} \ll \mu : \frac{dP_{\theta,g}}{d\mu} = g(x-\theta), \theta \in \boldsymbol{R} \right\}$$

とおく．ここで、

$$\boldsymbol{P}_1 = \{ P_{\theta,g} \in \boldsymbol{P} : \ g \text{ は固定} \}$$

とする．すると

$$\dot{\boldsymbol{l}}_\theta = -\frac{g'}{g}(x-\theta)$$

となることより、$\dot{\boldsymbol{P}}_1 = [\dot{\boldsymbol{l}}_\theta]$ は奇関数となる．

一方、

$$\boldsymbol{P}_2 = \{ P_{\theta,g} \in \boldsymbol{P} : \ \theta \text{ は固定} \}$$

とすれば、

$$\dot{\boldsymbol{P}}_2 = \left\{ h \in L_2^0(P_{\theta,g}) : h(x-\theta) = h(-x+\theta), \ \theta \text{ は固定} \right\}$$

となる．したがって、定義 4.4 より

$$\dot{\boldsymbol{P}} \supset \dot{\boldsymbol{P}}_1 \bigcup \dot{\boldsymbol{P}}_2$$

となる．

注意 4.2 ：更に、$\dot{\boldsymbol{P}}_1 \bigcup \dot{\boldsymbol{P}}_2 \supset \dot{\boldsymbol{P}}$ を示すこともできる．この証明については BKRW の pages 56 – 57 を参照されたい．

4.3　たたみこみ定理

定義 4.6 ：$m=1$ とし、$\{s(\eta) : \eta \in \boldsymbol{R}\} \subset \boldsymbol{S}$ を $s_0 = s(0)$ を通る任意の path で

$$\frac{s(\eta) - s(0)}{\eta} \longrightarrow t \qquad (\eta \to 0)$$

を満足するものとする．ν が **pathwise に微分可能**であるとは、有界線形汎関数 $\dot{\nu}(s_0) : \dot{\boldsymbol{S}} \mapsto \boldsymbol{R}$ が存在して、$\eta \to 0$ としたとき、

$$\nu(s(\eta)) = \nu(s_0) + \eta\dot{\nu}(s_0)(t) + o(|\ \eta\ |) \tag{4.1}$$

が成立することである．

$\dot{\boldsymbol{P}}$ 上の有界線形汎関数を

$$\dot{\nu}(P_0)(h) = \dot{\nu}(s_0)(\frac{1}{2}hs_0)$$

で定義すれば、(4.1) が成立するための必要十分条件は、$\{s(\eta)\}$ に対応する $\boldsymbol{P}$ の中の曲線 $\{P_\eta\}$ と $h = 2t/s_0$ に対して、$\eta \to 0$ のとき、

$$\nu(P_\eta) = \nu(P_0) + \eta\dot{\nu}(P_0)(h) + o(|\ \eta\ |) \tag{4.2}$$

が成立することである．(4.2) が成立するとき、 $\boldsymbol{P}$ 上の汎関数 ν は P_0 において**経路ごとに（pathwise）微分可能**であるという．

記号：以後、$\dot{\nu}(P_0)(h)$ を $\dot{\nu}(h)$ と記し、$\dot{\nu}(s_0)(t)$ を $\dot{\nu}(t)$ と記すことにする．

注意 4.3 ：有界な線形汎関数[4] $\dot{\nu}$ が $\dot{\boldsymbol{P}}^0$ 上で定義されているならば、$\dot{\boldsymbol{P}}$ を定義域とする有界な線形汎関数に一意的に拡張できる[5]．また、*Riesz* の表現定理[6] より、$\tilde{\nu} \in \dot{\boldsymbol{P}}$ が一意的に定まって、すべての $h \in \dot{\boldsymbol{P}}$ に対して、

$$\dot{\nu}(h) = \langle \tilde{\nu}, h \rangle_{L_2(P_0)} = \int \tilde{\nu}(x) h(x) dP_0(x)$$

と表現できる．混乱がなければ、$\tilde{\nu}$ のことを $\dot{\nu}$ とも記すことにする．

定理 4.1 ：$m = 1$ とし、ν を $\boldsymbol{P}$ 上の汎関数とし、P_0 において経路ごとに微分可能とする．任意の正則母数部分モデル $\boldsymbol{Q}$ で $I^{-1}(P_0 \mid \nu, \boldsymbol{Q})$ が定義できるものに対して、有効影響関数 $\tilde{\boldsymbol{l}}(\cdot, P_0 \mid \nu, \boldsymbol{Q})$ は

$$\tilde{\boldsymbol{l}}(\cdot, P_0 \mid \nu, \boldsymbol{Q}) = \Pi_0(\dot{\nu}(P_0) \mid \dot{\boldsymbol{Q}}) \tag{4.3}$$

で与えられる．ただし、$\Pi_0(\cdot \mid \dot{\boldsymbol{Q}})$ は内積 $\langle \cdot, \cdot \rangle_{P_0}$ に関する $\dot{\boldsymbol{Q}}$ への射影作用素とする．更に、

$$I^{-1}(P_0 \mid \nu, \boldsymbol{Q}) = \| \Pi_0(\dot{\nu}(P_0) \mid \dot{\boldsymbol{Q}}) \|^2_{L_2(P_0)} \leq \| \dot{\nu}(P_0) \|^2_{L_2(P_0)} \tag{4.4}$$

が成立し、等号成立は

$$\dot{\nu}(P_0) \in \dot{\boldsymbol{Q}}$$

の場合のみである．

証明：$P_{\theta_0} \in \boldsymbol{P}$ を固定する．$P_0 \equiv p_{\theta_0}$ を通る k – 次元の任意の正則母数部分モデル $\boldsymbol{Q} = \{P_\theta : \theta \in \boldsymbol{R}^k\}$ で $I^{-1}(P_\theta | \nu, \boldsymbol{Q})$ を定義できるものを考える．$\boldsymbol{Q}$ に対し、

$$\nu(P_\theta) = q(\theta) : \boldsymbol{R}^k \mapsto \boldsymbol{R}$$

としたとき、q が θ_0 において全微分可能で微分 $\dot{q} : 1 \times k$ をもつならば、ν は経路ごとに微分可能となる．

　一方、$a \in \boldsymbol{R}^k$ に対し、$\boldsymbol{Q}$ の中の曲線を

$$P(\eta) = P(\theta_0 + \eta a), \qquad P_0 = P(0)$$

[4] 線形汎関数が有界ではるとは

$$\| T \| = \sup_{\|x\| \leq 1} \frac{| T(x) |}{\| x \|} < \infty$$

を満足することである．また、

$$T \text{ が有界} \Leftrightarrow T \text{ が連続}$$

である．

[5] 関数解析（千葉克裕、培風館）pp. 92 の定理 5 を参照．

[6] 関数解析（千葉克裕、培風館）pp. 167 または Functional Analsysis (Rudin, W., McGall-Hill) pp. 80 を参照．

とすると、$\{P(\eta\})$ は P_0 において接ベクトル $a^T\dot{\boldsymbol{l}}$ をもつ．ただし、

$$\dot{\boldsymbol{l}}^T = (\dot{l}_1, \ldots, \dot{l}_k) = \frac{\dot{\boldsymbol{s}}^T}{\sqrt{s}}, \quad \dot{\boldsymbol{s}}^T = (\dot{s}_1, \ldots, \dot{s}_k), \quad \dot{s}_i = \frac{\partial s}{\partial \theta_i} \quad (i = 1, \ldots, k)$$

である．$I^{-1}(\theta_0)\int \dot{\boldsymbol{l}}\dot{\boldsymbol{l}}^T dP_\theta = \mathrm{id}$ に注意すれば、

$$\begin{aligned} \nu(P(\eta)) &= q(\theta_0 + \eta a) = q(\theta_0) + \eta\dot{q}(\theta_0)a + o(\eta) \\ &= q(\theta_0) + \eta\dot{q}(\theta_0)I^{-1}(\theta_0)\int \dot{\boldsymbol{l}}\dot{\boldsymbol{l}}^T dP_\theta + o(\eta) \\ &= q(\theta_0) + \eta\langle \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}}, \dot{\boldsymbol{l}}\rangle_{L_2(P_0)} + o(\eta) \end{aligned} \tag{4.5}$$

が成立する．

一方、$\dot{\boldsymbol{l}}^T a \in \dot{\boldsymbol{Q}}$ は $\{P(\eta)\}$ の接ベクトルであることより

$$\begin{aligned} \nu(P(\eta)) &= \nu(P_0) + \eta\dot{\nu}(P_0)(\dot{\boldsymbol{l}}a) + o(\eta) \\ &= \nu(P_0) + \eta\langle \dot{\nu}, \dot{\boldsymbol{l}}a\rangle_{L_2(P_0)} + o(\eta) \end{aligned} \tag{4.6}$$

となる．(4.5) と (4.6) を比較すれば、

$$\langle \dot{\nu}, \dot{\boldsymbol{l}}^T a\rangle_{L_2(P_0)} = \langle \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}}, \dot{\boldsymbol{l}}^T a\rangle_{L_2(P_0)} \tag{4.7}$$

となり、$\dot{\boldsymbol{Q}}$ 上で

$$\dot{\nu} = \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}} = \tilde{\boldsymbol{l}}(\cdot, P_0|\mu, \boldsymbol{Q})$$

となる．更に、

$$\Pi_0(\dot{\nu}(P_0)|\dot{\boldsymbol{Q}}) = (\langle \dot{\nu}, \dot{l}_1\rangle_{L_2(P_0)}, \ldots, \langle \dot{\nu}, \dot{l}_k\rangle_{L_2(P_0)})I^{-1}(\theta_0)\dot{\boldsymbol{l}}$$

であるので、(4.7) を用いれば、

$$\begin{aligned} \Pi_0(\dot{\nu}(P_0)|\dot{\boldsymbol{Q}}) &= \dot{q}(\theta_0)I^{-1}(\theta_0)\boldsymbol{E}_{P_0}[\dot{\boldsymbol{l}}\dot{\boldsymbol{l}}^T]I^{-1}(\theta_0)\dot{\boldsymbol{l}} \\ &= \dot{q}(\theta_0)I^{-1}(\theta_0)\dot{\boldsymbol{l}} \end{aligned}$$

となる．更に、$\dot{\boldsymbol{Q}} \subset \dot{\boldsymbol{P}} \subset L_2(P_0)$ より、ピタゴラスの定理を用いれば、(4.4) がわかる． □

定理 3.1 の結果は $m \geq 1$ の場合に拡張できる．そのため、

$$\tilde{\boldsymbol{l}}_j = \dot{\nu}_j(P_0), \quad \tilde{\boldsymbol{l}} = (\tilde{\boldsymbol{l}}_1, \ldots, \tilde{\boldsymbol{l}}_m)^T, \qquad j = 1, 2, \ldots, m \tag{4.8}$$

とおく．

系 4.1 ： つぎを仮定する．

(i) ν は経路ごとに微分可能である．

(ii) $\boldsymbol{Q}$ は $I^{-1}(P_0 \mid \nu, \boldsymbol{Q})$ が定義できるような正則母数部分モデルである．

このとき、有効影響関数は定義され、(4.3) によって与えられる．更に、

$$I^{-1}(P_0 \mid \nu, \boldsymbol{Q}) = \parallel \Pi_0(\dot{\nu}(P_0) \mid \dot{\boldsymbol{Q}}) \parallel_{P_0} \leq \langle \tilde{\boldsymbol{l}}, \tilde{\boldsymbol{l}}^T \rangle_{P_0}$$

が成立する．

証明：任意の定数 $a_1, \ldots, a_m$ を取り、定理 4.1 を $\nu(P) = \sum_{j=1}^m a_j \nu_j(P)$ に対し適用すると

$$\Pi_0(\sum_{j=1}^m a_j \nu_j(P_0) | \dot{\boldsymbol{Q}}) = \sum_{j=1}^m a_j \Pi_0(\dot{\nu}_j(P_0) | \boldsymbol{Q})$$

から系は証明された． □

仮定 4.1 ： 定理 3.1 と系 3.1 のように ν を経路ごとに微分可能とし、更に

$$[\dot{\nu}_j(P) : j = 1, 2, \ldots, m] \subset \overline{\dot{\boldsymbol{P}}^0}$$

と仮定する．

この仮定のもとで、ν に対する有効影響関数と情報限界が定義される．

定義 4.7 ：(4.8) により定義された $\tilde{\boldsymbol{l}} = \tilde{\boldsymbol{l}}(\cdot, P_0 \mid \nu, \boldsymbol{P})$ を ν の**有効影響関数**という．すなわち、

$$\tilde{\boldsymbol{l}} = \tilde{\boldsymbol{l}}(\cdot, P_0 \mid \nu, \boldsymbol{P}) = \dot{\nu}(P_0)$$

である．また、$\boldsymbol{P}$ の中での P_0 における ν に対する**情報** $I(P_0 \mid \nu, \boldsymbol{P})$ を

$$I^{-1}(P_0 \mid \nu, \boldsymbol{P}) = \langle \tilde{\boldsymbol{l}}, \tilde{\boldsymbol{l}} \rangle_{L_2(P_0)} = \boldsymbol{E}_{P_0}[\tilde{\boldsymbol{l}}\ \tilde{\boldsymbol{l}}^T]$$

の逆行列で定義する．

定義 4.8 ：もし、$\hat{\nu}$ が ν の局所 Gaussian 正則推定量で漸近分散が $\Sigma(P_0 \mid \hat{\nu}) = I^{-1}(P_0 \mid \nu, \boldsymbol{P})$ を満足するならば、$\hat{\nu}$ は P_0 において**有効（efficient）**という．$\hat{\nu}$ がすべての正則モデル P_0 において有効ならば、**有効**という．

定理 4.2 ：$\nu : \boldsymbol{P} \mapsto \boldsymbol{R}^m$ を P_0 において経路ごとに微分可能とし、仮定 3.1 が成立するものとする．T_n を ν の推定量とし、その極限分布を $\boldsymbol{L}(Z)$ とする．このとき、
A. $\boldsymbol{L}(Z)$ は分布 $N(0, I^{-1}(P_0 \mid \nu, \boldsymbol{P}))$ と $\boldsymbol{R}^m$ 上の他の分布とのたたみこみとして表現できる．一般的には

$$\boldsymbol{L}_{P_0}\left[\begin{array}{c} \sqrt{n}(T_n - \nu(P_0)) - (1/\sqrt{n})\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i, P_0 \mid \nu, \boldsymbol{P}) \\ (1/\sqrt{n})\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i, P_0 \mid \nu, \boldsymbol{P}) \end{array}\right] \xrightarrow{d} \boldsymbol{L}\left[\begin{array}{c} \Delta_0 \\ Z_0 \end{array}\right]$$

が成立する．ただし、$Z_0 \sim N(0, I^{-1}(P_0 \mid \nu, \boldsymbol{P}))$ で、 Δ_0 は Z_0 と独立である．より一般的には、もし、$h \in \overline{\dot{\boldsymbol{P}}^0}^l$ ならば、

$$\boldsymbol{L}_{P_0}\left[\begin{array}{c} \sqrt{n}(T_n - \nu(P_0)) - (1/\sqrt{n})\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i, P_0 \mid \nu, \boldsymbol{P}) \\ (1/\sqrt{n})\sum_{i=1}^n h(X_i) \end{array}\right] \xrightarrow{d} \boldsymbol{L}\left[\begin{array}{c} \Delta_0 \\ W_0 \end{array}\right]$$

が成立する．ただし、$W_0 \sim N(0, \boldsymbol{E}_{P_0}[hh^T])$ で、Δ_0 は W_0 と独立である．
B. $\Delta_0 = 0$ であるための必要十分条件は、T_n が影響関数 $\tilde{\boldsymbol{l}}$ をもつ漸近線形推定量であることである．このとき、T_n は有効推定量になる．

命題 4.1 ：$\nu : \boldsymbol{P} \mapsto \boldsymbol{R}^m$ とする．T_n を P_0 において影響関数 ψ をもつ ν の漸近線形推定量とする．このとき、
A. T_n は P_0 において正則であるための必要十分条件は、ν は経路ごとに微分可能で微分 $\dot{\nu}$ をもち、

$$\psi - \dot{\nu} = \psi - \tilde{\boldsymbol{l}} \perp \dot{\boldsymbol{P}} \tag{4.9}$$

が成立する．
B. T_n を P_0 において正則とし、仮定 3.1 が成立するとする．T_n が P_0 において有効であるための必要十分条件は、$\psi = \tilde{\boldsymbol{l}}$ である．

証明：証明は命題 3.3 と同様にできる．
A. $\{P_\eta\}$ を $\eta = 0$ におていスコア関数 $h \in (\overline{\dot{\boldsymbol{P}}^0})^l$ をもつ l 次元正則母数部分モデルとする．$n \to \infty$ のとき、$t_n \to t$ とし、$\eta_n = t_n/\sqrt{n}$ 、$P_n = P_{\eta_n}$ とおく．P_0 において $\{T_n\}$ は漸近線形であることと命題 3.1 から

$$\boldsymbol{L}_{P_0}\left[\begin{array}{c} \sqrt{n}(T_n - \nu(P_0)) \\ \log \Pi_{i=1}^n \{dP_{\eta_n}/dP_0\}(X_i) \end{array}\right] \xrightarrow{d} N\left(\left[\begin{array}{c} 0 \\ -\frac{1}{2}\Sigma_{22} \end{array}\right], \Sigma\right)$$

が成立する．ただし、

$$\Sigma = (\Sigma_{ij}), \quad \Sigma_{11} = \boldsymbol{E}[\Psi\Psi^T], \quad \Sigma_{12} = \boldsymbol{E}[\Psi h^T]t, \quad \Sigma_{22} = t^T \boldsymbol{E}[hh^T]t$$

である．補題 2.3 （Le Cam の第 3 補題）から

$$\boldsymbol{L}_{P_n}(\sqrt{n}(T_n - \nu(P_0)) \xrightarrow{d} N(\Sigma_{12}, \Sigma_{11}) \tag{4.10}$$

を得る．一方、$\{T_n\}$ の正則性より

$$\boldsymbol{L}_{P_n}(\sqrt{n}(T_n - \nu(P_n)) \xrightarrow{d} N(0, \Sigma_{11}) \tag{4.11}$$

となる．(4.10) と (4.11) を比較すれば、

$$\sqrt{n}(\nu(P_n) - \nu(P_0)) \to \Sigma_{12} = \boldsymbol{E}[\Psi h^T]t \tag{4.12}$$

が成立することがわかる．(4.12) を書き直せば、任意の $\{\eta_n\}$ に対し

$$\nu(P_n) = \nu(P_0) + \langle \Psi, h \rangle_{L_2(P_0)} \eta_n + o(|\eta_n|) \tag{4.13}$$

が成立することがわかる． (4.13) より、ν は経路ごとに微分可能で、微分 $\dot{\nu}$ は、すべての $h \in \dot{\boldsymbol{P}}$ に対し、

$$\dot{\nu}(h) = \langle \Psi, h \rangle_{L_2(P_0)} \tag{4.14}$$

を満足することがわかる．Reisz の表現定理を用いれば、

$$\dot{\nu}(h) = \langle \dot{\nu}, h \rangle_{L_2(P_0)} \tag{4.15}$$

となる．(4.14) と (4.15) から

$$\langle \dot{\nu} - \Psi, h \rangle_{L_2(P_0)} = 0$$

となるので、(4.9) がわかる．

逆に、ν が経路ごとに微分可能で (4.9) が成立するならば、(4.12) が成立し、(4.10) と合わせれば、(4.13) が成立することから証明さらる．

B. $\{T_n\}$ が有効とする．定義 4.8 と定理 4.1.A から $\Lambda_0 = 0$ となり、$\{T_n\}$ は有効影響関数 $\Psi = \tilde{\boldsymbol{l}}$ をもつことが定理 4.1.B からわかる．

つぎに、$\Psi \in \dot{\boldsymbol{P}}^m$ を仮定する．このとき、$\Psi - \tilde{\boldsymbol{l}} \in \dot{\boldsymbol{P}}^m$ となり、(4.9) から、$\Psi = \tilde{\boldsymbol{l}}$ となって、$\{T_n\}$ は有効であることがわかる．□

4.4　スコア関数と情報限界

4.4.1　ノンパラメトリックモデルの観点から

系 4.2 ：g をある実数値関数とする．$\nu : \boldsymbol{P} \mapsto \boldsymbol{R}$ が

$$\nu(P) = \int g(x) dP(x)$$

で与えられるとする．もし、

$$\sup_{P \in \boldsymbol{P}} \boldsymbol{E}_P[g^2(X)] < \infty$$

ならば、ν は $P_0 \in \boldsymbol{P}$ において経路ごとに微分可能で微分 $\dot{\nu} \in L_2^0(P_0)$ をもち、

$$\dot{\nu} = \dot{\nu}(P_0) = g(x) - \boldsymbol{E}_{P_0}[g(X)] \tag{4.16}$$

で与えられる．

証明：$\{s(\eta)\}$ を $\boldsymbol{S} = \{s = \sqrt{dP/d\mu} : P \in \boldsymbol{P}\}$ の中の曲線とし、接ベクトル $t \in \dot{\boldsymbol{S}} \subset L_2(\mu)$ をもつものとする．すなわち、

$$\left\{ \int [s(\eta) - s(0) - \eta t]^2 d\mu \right\}^{1/2} = o(\eta)$$

である．ただし、$s_0 = s(0)$ とした．このとき、Riesz の定理と定義 4.3 から

$$\langle \dot{\nu}(s_0), t \rangle_{L(\mu)} = \dot{\nu}(s_0)(t) = \dot{\nu}(P_0)\left(\frac{2t}{s_0}\right) = \langle \dot{\nu}(P_0), \frac{2t}{s_0} \rangle_{L_2(P_0)}$$

となる．したがって、

$$\nu(s(\eta)) - \nu(s_0) - \eta \langle \dot{\nu}(s_0), t \rangle_{L_2(\mu)} = \nu(s(\eta)) - \nu(s_0) - \eta \langle \dot{\nu}(P_0), \frac{2t}{s_0} \rangle_{L_2(P_0)}$$

から、(4.16) を証明するために

$$\nu(s(\eta)) - \nu(s_0) - \eta \langle g - \boldsymbol{E} g(X), \frac{2t}{s_0} \rangle_{L_2(P_0)} = o(\eta) \tag{4.17}$$

を示せばよい．

ν の定義より、(4.17) の左辺は

$$\begin{aligned}
&\int g\{s^2(\eta) - s^2(0) - 2\eta t s_0\}d\mu - 2\eta \boldsymbol{E}g(X)\int t s_0 d\mu \\
&= \int g\{s^2(\eta) - s^2(0) - 2\eta t s_0 d\mu + o(\eta) \\
&= \int g\{s(\eta) - s(0) - \eta t\}\{s(\eta) + s(0)\}d\mu + \eta \int g\{s(\eta) - s(0)\}d\mu \\
&\equiv \eta\{A(\eta) + B(\eta)\}
\end{aligned}$$

と変形できる[7]．しかし、

$$\begin{aligned}
|A(\eta)| &\le \sqrt{\int g^2\{s(\eta)+s(0)\}^2 d\mu} \cdot \eta \sqrt{\int \left\{\frac{s(\eta)-s(0)}{\eta} - \eta\right\}^2 d\mu} \\
&\le \eta\sqrt{2\{\boldsymbol{E}_\eta g^2 + \boldsymbol{E}_0 g^2\}} o(1) = \eta O(1) o(1)
\end{aligned}$$

となる．一方、

$$\begin{aligned}
|B(\eta)| &\le \left|\int_{\{|g|\le 1/\sqrt{n}\}} gt(s(\eta)-s(0))d\mu\right| + \left|\int_{\{|g|>1/\sqrt{n}\}} gt(s(\eta)-s(0))d\mu\right| \\
&\le \sqrt{\frac{1}{\eta}\int t^2 d\mu \int \{s(\eta)-s(0)\}^2 d\mu} + \sqrt{\int_{\{|g|\ge 1/\sqrt{n}\}} t^2 d\mu 2(\boldsymbol{E}_\eta g^2 + \boldsymbol{E}_0 g^2)} \\
&= o(1) + o(1)
\end{aligned}$$

となる[8]． □

---

[7] 第１番目の等式は

$$\lim_{\eta\to 0} \eta \int t s_0 d\mu = \lim_{\eta\to 0} \eta \langle t, s_0\rangle_{L_2(\mu)} = \lim_{\eta\to 0} \langle s(\eta) - s(0) + o(\eta), s_0\rangle_{L_2(\mu)}$$

からわかる．

[8] $t \in L_2(\mu)$ に注意して

$$\begin{aligned}
\frac{1}{\eta^2}\int (s(\eta)-s(0))^2 d\mu &\le \int \left\{\frac{s(\eta)-s(0)}{\eta} - t\right\}^2 d\mu + \int t^2 d\mu \\
&\le \int t^2 d\mu + o(\eta)
\end{aligned}$$

より、$\eta \to 0$ に対し、

$$\frac{1}{\eta}\int t^2 d\mu \int (s(\eta)-s(0))^2 d\mu \le \eta[\int t^2 d\mu]^2 + o(\eta)$$

となる．一方、$\boldsymbol{E}g^2 < \infty$ より $P1\{|g| > 1/\sqrt{n}\} \to 0 (\eta \to 0)$ に注意すれば、$\eta \to 0$ に対し、

$$\int_{\{|g|>1/\sqrt{n}\}} t^2 d\mu \to 0$$

となることからわかる．

例 4.5 ：モデル

$$\boldsymbol{P} = \Big\{ \boldsymbol{P} \text{ 上の確率測度 } P \ll \mu : \ median\,(P) = 0 \ \text{（固定）}, \boldsymbol{E}_P \mid X \mid^{2+\delta} < \infty \Big\}$$

とする．ただし、$\delta > 0$ である．ここで、$\nu(P) = \boldsymbol{E}_P[X]$ の推定を考える．接空間 $\dot{\boldsymbol{P}}$ を求めるために、$P \in \boldsymbol{P}$ に対して、

$$\gamma(P) = \boldsymbol{E}_P\{1_{(-\infty,0]}(X) - \frac{1}{2}\} = 0$$

とおく．任意の正則部分モデル $\{P_\eta\}$ に対して、

$$\boldsymbol{E}_{P_\eta} 1_{(-\infty,0]}(X) = \frac{1}{2}$$

を満足する．よって、$\dot{\boldsymbol{l}}_\eta = (\partial/\partial\eta)\log p_\eta$ とおけば、

$$\int \{1_{(-\infty,0]}(x) - \frac{1}{2}\} \dot{\boldsymbol{l}}_\eta(x) p_\eta(x) d\mu(x) = 0$$

が成立する．故に、

$$\dot{\boldsymbol{P}}(P_0) \subset \Big\{ h \in L_2^0(P_0) : h \perp \{1_{(-\infty,0]}(x) - \frac{1}{2}\} \Big\}$$

となる．実際、

$$\dot{\boldsymbol{P}}(P_0) = \Big\{ h \in L_2^0(P_0) : h \perp \dot{\gamma}(P_0) \Big\}$$

が成立する[9]．これを示すために、

$$p_\eta(x) = \frac{(1 + \eta h(x)) p_0(x)}{\int (1 + \eta h(x)) p_0(x) d\mu(x)}$$

とおく．ただし、$p_0 = dP_0/d\mu$ である．このとき、 $median\,(P_0) = 0$ と h の性質より

$$\boldsymbol{E}_{P_\eta}\{1_{(-\infty,0]}(X) - \frac{1}{2}\} = 0$$

かつ

$$\dot{\boldsymbol{l}} = h(x)$$

が成立する．命題 3.2 を $\nu(P) = \boldsymbol{E}_P X$ に適応すれば、

$$\dot{\nu}(P_0) = x - \boldsymbol{E}_{P_0} X$$

となる．$\dot{\boldsymbol{P}} \perp \dot{\gamma}(P_0)$ より、

$$\tilde{\boldsymbol{l}}(\cdot, P_0 \mid \nu, \boldsymbol{P}) = \Pi_0(\dot{\nu}(P_0) \mid \dot{\boldsymbol{P}}) = \dot{\nu}(P_0) - \frac{\langle \dot{\nu}(P_0), \dot{\gamma}(P_0) \rangle_{L_2(P_0)}}{\| \dot{\gamma}(P_0) \|_{L_2(P_0)}} \dot{\gamma}(P_0)$$

[9] $\dot{\gamma}(P_0) = 1_{(-\infty,0]}(x) - \frac{1}{2}$ である．

となる．更に、

$$\begin{aligned}\boldsymbol{E}\{\tilde{\boldsymbol{l}}(X)\}^2 &= \boldsymbol{E}(X-\boldsymbol{E}X)^2 - \frac{\langle\dot{\nu}(P_0),\dot{\gamma}(P_0)\rangle^2_{L_2(P_0)}}{\|\dot{\gamma}(P_0)\|^2_{L_2(P_0)}} \\ &= \sigma^2(P_0)\left(1-\frac{\langle\dot{\nu}(P_0),\dot{\gamma}(P_0)\rangle^2_{L_2(P_0)}}{\sigma^2(P_0)\|\dot{\gamma}(P_0)\|^2_{L_2(P_0)}}\right) \\ &= \sigma^2(P_0)\left(1-\frac{\langle\dot{\nu}(P_0),\dot{\gamma}(P_0)\rangle^2_{L_2(P_0)}}{\|\dot{\nu}(P_0)\|^2_{L_2(P_0)}\|\dot{\gamma}(P_0)\|^2_{L_2(P_0)}}\right)\end{aligned}$$

となる．ここで、

$$\langle\dot{\nu}(P),\dot{\gamma}(P)\rangle_{L_2(P)} = \boldsymbol{E}\left[(1_{(-\infty,0]}(X)-\frac{1}{2})(X-\boldsymbol{E}X)\right]$$

と

$$\|\dot{\gamma}(P_0)\|^2_{L_2(P_0)} = \frac{1}{4}$$

である．

例 4.6 ：μ は区間 $[-M,M]$ に集中すると仮定し、制約モデル

$$\boldsymbol{P} = \left\{P : \frac{\boldsymbol{var}(X)}{\nu^2(P)} = c_0, \quad c_0 \text{ は固定}\right\}$$

を考える．ただし、$\nu(P)=\boldsymbol{E}(X)$ である．ここで、$\nu(P)$ の推定を考える．モデルの制約は

$$\gamma(P) = \int x^2 dP(x) - (1+c_0)\nu^2(P) = 0$$

と表現できる．任意の正則部分モデル $\{P_\eta\}$ は

$$\boldsymbol{E}_{P_\eta}\{X^2-(1+c_0)\nu^2(P_\eta)\} = 0$$

を満足し、上の式の左辺を η について微分すれば、

$$\begin{aligned}&\int x^2\dot{\boldsymbol{l}}_\eta(x)p_\eta(x)d\mu(x) - 2(1+c_0)\nu(P_\eta)\int x\dot{\boldsymbol{l}}_\eta(x)p_\eta(x)d\mu(x) \\ &= \int\{x^2-2(1+c_0)\nu(P_\eta)x\}\dot{\boldsymbol{l}}_\eta(x)p_\eta(x)d\mu(x) = 0\end{aligned}$$

が成立することがわかる．これより、

$$\int\left\{x^2-2(1+c_0)\nu(P_0)x-\int\{x^2-2(1+c_0)\nu(P_0)\}dP_0(x)\right\}\dot{\boldsymbol{l}}(x)p_0(x)d\mu(x) = 0$$

となる．よって、

$$\dot{\boldsymbol{P}} \subset \left\{h \in L_2^0(P_0) : h \perp \{x^2 - \int x^2 dP_0(x) - 2(1+c_0)\nu(P_0)(x-\nu(P_0))\}\right\}$$

となる．実際、

$$\dot{\boldsymbol{P}} = \dot{\gamma}(P_0)^{\perp} \tag{4.18}$$

である．(4.18) を示すために、

$$p_\eta = \frac{(1+\eta h(x))p_0(x)}{\int(1+\eta h(x))p_0(x)d\mu(x)}$$

とおく．このとき、$h(x)$ の定義より

$$\begin{aligned}
&\boldsymbol{E}_{P_\eta}\left\{X^2 - \int x^2 dP_0(x) - 2(1+c_0)\nu(P_0)(X-\nu(P_0))\right\} \\
&\quad = \frac{\eta}{\int(1+\eta h(x))p_0(x)d\mu(x)} \\
&\qquad \times \int h(x)\{x^2 - \int x^2 dP_0(x) - 2(1+c_0)(x-\nu(P_0))\}dP_0(x) = 0
\end{aligned}$$

となり、$\{P_\eta\} \subset \boldsymbol{P}$ となる．また、

$$\tilde{\boldsymbol{l}} = h(x)$$

である．ここで、

$$\dot{\gamma}(P) = x^2 - \int x^2 dP_0(x) - 2(1+c_0)\nu(P_0)(x-\nu(P_0))$$

とおくと、$P \mapsto \dot{\gamma}(P)$ は P_0 で連続となる．$\dot{\boldsymbol{P}} \perp \dot{\gamma}(P_0)$ なので、$\dot{\nu}(P_0) = X - \boldsymbol{E}X$ より、

$$\begin{aligned}
\tilde{\boldsymbol{l}}(x, P_0 \mid \nu, \boldsymbol{P}) &= \Pi_0(\dot{\nu} \mid \dot{\boldsymbol{P}}) = \dot{\nu}(P_0) - \frac{\langle \dot{\nu}(P_0), \dot{\gamma}(P_0)\rangle_{L_2(P_0)}}{\| \dot{\gamma}(P_0) \|_{L_2(P_0)}}\dot{\gamma}(P_0) \\
&= [1 + 2(1+c_0)(\boldsymbol{E}X)a(P_0)](x - \boldsymbol{E}X) - a(P_0)(x^2 - \boldsymbol{E}X^2)
\end{aligned}$$

となる．ただし、

$$a(P_0) = \frac{\boldsymbol{E}[\dot{\nu}(P_0)\dot{\gamma}(P_0)]}{\boldsymbol{var}[\dot{\gamma}(P_0)]} = \frac{\boldsymbol{cov}[X - \boldsymbol{E}X, X^2 - 2(1+c_0)(\boldsymbol{E}X)X]}{\boldsymbol{var}[X^2 - 2(1+c_0)(\boldsymbol{E}X)X]}$$

となる．

注意 4.4 ：例 3.5 と 3.6 の計算で情報限界を求めることができた．これを達成する正則推定量の構成については Haberman (1984), Sheeby (1987, 1988) を参照のこと．

4.4.2　セミパラメトリックモデルの観点から

$\boldsymbol{N} \subset \boldsymbol{R}^m$ を開集合とし、$\boldsymbol{G}$ は適当な空間とする．

$$\boldsymbol{P} = \left\{P_{(\nu,G)} \ll \mu : \nu \in \boldsymbol{N}, G \in \boldsymbol{G}\right\}$$

とし、$p(\cdot;\nu,G)=dP_{(\nu,G)}/d\mu$、 $P_0=P_{(\nu_0,G_0)}$ を固定する．また、

$$\dot{\boldsymbol{l}}=\frac{\partial}{\partial\nu}p(\cdot;\nu,G)\bigg|_{\nu=\nu_0,G=G_0}$$

とする．これはモデル

$$\boldsymbol{P}_1=\left\{P_{(\nu,G_0)}:\nu\in\boldsymbol{N}\right\}$$

に対するスコア関数である．更に、

$$\boldsymbol{P}_2=\left\{P_{(\nu_0,G)}:G\in\boldsymbol{G}\right\}$$

とする．

定理 4.3 ：つぎを仮定する．

(i) $\boldsymbol{P}_1$ は正則である．

(ii) ν は１次元母数とする．

ここで、

$$\boldsymbol{l}_1^*=\dot{\boldsymbol{l}}_1-\Pi_0(\dot{\boldsymbol{l}}_1\mid\dot{\boldsymbol{P}}_2)$$

とおく．このとき、
$A.$ $\{G_\gamma:\gamma\in\Gamma\subset\boldsymbol{R}\}\subset\boldsymbol{G}$ とする．

$$\boldsymbol{Q}=\left\{P_{(\nu,G_\gamma)}:\nu\in\boldsymbol{N},\gamma\in\Gamma\right\}$$

が $P_0\in\boldsymbol{Q}$ において $\boldsymbol{P}$ の正則部分モデルならば、

$$I(P_0\mid\nu,\boldsymbol{Q})\geq\|\,\boldsymbol{l}_1^*\,\|^2_{L_2(P_0)}=\boldsymbol{E}_{P_0}(\boldsymbol{l}_1^*)^2$$

が成立する．等号成立は $\boldsymbol{l}_1^*\in\dot{\boldsymbol{Q}}$ の場合のみである．
$B.$ $\dot{\boldsymbol{P}}=\dot{\boldsymbol{P}}_1+\dot{\boldsymbol{P}}_2$ が成立し、$\boldsymbol{l}_1^*\neq 0$ で、定理 3.1 における $\boldsymbol{P}$ と ν の仮定が成立するならば、ν に対する有効影響関数は

$$\tilde{\boldsymbol{l}}=\|\,\boldsymbol{l}_1^*\,\|^{-2}_{L_2(P_0)}\,\boldsymbol{l}_1^*\equiv I^{-1}(P_0\mid\nu,\boldsymbol{P})\boldsymbol{l}_1^*$$

で与えられる．

証明：
A. 命題 3.3 を用いれば、任意の正則母数部分モデル $\boldsymbol{Q}$ に対し、

$$\begin{aligned}I(P_0|\nu,\boldsymbol{Q}) &= \|\dot{\boldsymbol{l}}_1-\Pi_0(\dot{\mathbf{i}}_1|\dot{\boldsymbol{Q}}_2)\|^2_{L_2(P_0)}\\ &= \|\dot{\boldsymbol{l}}_1-\Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)+\Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)-\Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{Q}}_2)\|_{L_2(P_0)}\\ &= \|\boldsymbol{l}_1^*\|_{L_2(P_0)}+\|\Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{Q}}_2)-\Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)\|_{L_2(P_0)}\end{aligned}$$

となる．最後の等式は $\boldsymbol{l}_1^* \perp \dot{\boldsymbol{P}}_2$ （$\boldsymbol{l}_1^*$ の定義）であることと

$$\dot{\boldsymbol{Q}}_2 = \left\{ \frac{\partial}{\partial \eta} \log P_{(\nu, G_\gamma)} : \gamma \in \Gamma \right\} \subset \dot{\boldsymbol{P}}_2$$

なので、$\Pi_0(\dot{\boldsymbol{l}}_1 |\ \boldsymbol{P}\ _2)$ と $\Pi_0(\dot{\boldsymbol{l}}_1 |\ \boldsymbol{Q}\ _2)$ は $\dot{\boldsymbol{P}}_2$ に含まれることよりわかる．

B. はじめに、

$$\langle \dot{\nu}(P_0), \dot{\boldsymbol{l}} \rangle_{L_2(P_0)} = 1 \tag{4.19}$$

と

$$\dot{\nu}(P_0) \perp \dot{\boldsymbol{P}}_2 \tag{4.20}$$

を示す．(4.19) を示すために、$\boldsymbol{P}_1$ に制約したモデルを考える．ν の定義から $P_{(\nu, G_0)} \equiv P_\nu \in \boldsymbol{P}_1$ に対し、

$$\nu(P_{(\nu,G_0)}) - \nu(P_0) = \nu - \nu_0 \equiv t \tag{4.21}$$

となる．一方、ν が経路ごとに微分可能であることより

$$\begin{aligned} \nu(P_{(\nu,G_0)}) &= \nu(P_0) + t\dot{\nu}(P_0)(\dot{\boldsymbol{l}}_1) + o(|t|) \\ &= \nu(P_0) + t\langle \dot{\nu}(P_0), \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_0)} + o(|t|) \end{aligned} \tag{4.22}$$

となる．すべての t について、(4.21) と (4.22) は等しくなるので

$$1 = \langle \dot{\nu}(P_0), \dot{\boldsymbol{l}}_1 \rangle_{L_2(P_0)}$$

が成立することがわかる．

(4.20) を示すために、$h \in \dot{\boldsymbol{P}}_2^0$ をとる．すると、h を接ベクトルとしてもつ 1 次元正則母数部分モデル

$$\boldsymbol{Q}_2 = \{ P_{(\nu_0, G_\gamma)} : \gamma \in \Gamma \}$$

が存在する．$\boldsymbol{Q}_2$ に制約した ν を考えると

$$\begin{aligned} 0 &= \nu(P_{(\nu_0,G_\gamma)}) - \nu(P_0) = \gamma\dot{\nu}(P_0)(h) + o(\gamma) \\ &= \gamma\langle \dot{\nu}(P_0), h \rangle_{L_2(P_0)} + o(\gamma) \end{aligned}$$

より、任意の $h \in \dot{\boldsymbol{P}}_2^0$ に対し、

$$\langle \dot{\nu}(h), h \rangle_{L_2(P_0)} = 0$$

が成立する．仮定 $\dot{\boldsymbol{P}} = \dot{\boldsymbol{P}}_1 + \dot{\boldsymbol{P}}_2$ で $\boldsymbol{P}_1$ が正則であることと $\dot{\nu}(P_0) \in \dot{\boldsymbol{P}}$ に注意すれば、$\tilde{h} \in \dot{\boldsymbol{P}}_2$ が存在して、

$$\begin{aligned} \dot{\nu}(P_0) &= a\dot{\boldsymbol{l}}_1 + b\tilde{h} = a\boldsymbol{l}_1^* + b\tilde{h} + a\Pi_0(\dot{\boldsymbol{l}}_1 | \dot{\boldsymbol{P}}_1) \\ &= a\boldsymbol{l}_1^* + bh \end{aligned}$$

と書ける．ただし、$h \in \dot{\boldsymbol{P}}_2$ である．このとき、(4.20) より

$$\dot{\nu}(P_0) = a\boldsymbol{l}_1^*$$

となり、(4.19) より

$$\begin{aligned} 1 &= a\langle \boldsymbol{l}_1^*, \dot{\boldsymbol{l}}_1\rangle_{L_2(P_0)} = a\langle \boldsymbol{l}_1^*, \boldsymbol{l}_1^*\rangle_{L_2(P_0)} - \langle \boldsymbol{l}_1^*, \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)\rangle_{L_2(P_0)} \\ &= a\|\boldsymbol{l}_1^*\|_{L_2(P_0)} \end{aligned}$$

となる．したがって、

$$\dot{\nu}(P_0) = \frac{1}{\|\boldsymbol{l}\|_{L_2(P_0)}}\boldsymbol{l}_1^* = I^{-1}(P_0|\nu, \boldsymbol{P})\boldsymbol{l}_1^*$$

となる．□

定義 4.9 ：$\boldsymbol{l}_1^*$ を $\boldsymbol{P}$ における ν の**有効スコア関数**といい、$\boldsymbol{l}_1^*(\cdot, P_0 \mid \nu, \boldsymbol{P})$ と記す．

ここで、ν を m 次元ベクトル $(m \geq 1)$ とし、スコア関数を

$$\boldsymbol{l}_1^* = (\boldsymbol{l}_{11}^*, \ldots, , \boldsymbol{l}_{1m}^*)$$

と定める．ただし、$\boldsymbol{l}_{11}^*, \ldots, , \boldsymbol{l}_{1m}^*$ は ν の成分に対するスコア関数とし、

$$\boldsymbol{l}_{1i}^* = \dot{\boldsymbol{l}}_{1i} - \Pi_0(\dot{\boldsymbol{l}}_{1i} \mid \dot{\boldsymbol{P}}_2), \quad \dot{\boldsymbol{l}}_{1i} = \left.\frac{\partial}{\partial \nu_i}\log p(\cdot; \nu, G_0)\right|_{\nu=\nu_0}, \quad i = 1, 2, \ldots, m$$

で与えられる．

系 4.3 ：$\boldsymbol{P}_1$ を正則と仮定する．
A. $\boldsymbol{Q}$ をある正則母数部分モデルとし、$I^{-1}(P_0 \mid \nu, \boldsymbol{Q})$ が定まるとする．このとき、

$$I(P_0 \mid \nu, \boldsymbol{Q}) \geq \boldsymbol{E}_{P_0}[\boldsymbol{l}_1^*(\boldsymbol{l}_1^*)^T]$$

である．等号成立は $[\boldsymbol{l}_1^*]$ が $\dot{\boldsymbol{Q}}$ に含まれるときである．
B. $\dot{\boldsymbol{P}} = \dot{\boldsymbol{P}}_1 + \dot{\boldsymbol{P}}_2$ と定理 *3.1* の条件が成立するとする．このとき、有効影響関数は

$$\boldsymbol{l}_1^* = I^{-1}(P_0 \mid \nu, \boldsymbol{P})\boldsymbol{l}_1^*$$

で与えられる．ただし、

$$I(P_0 \mid \nu, \boldsymbol{P}) = \boldsymbol{E}[\boldsymbol{l}_1^*(\boldsymbol{l}_1^*)^T]$$

である．

第5章　無限次元母数にたいする情報限界

5.1　共役作用素

定義 5.1 ：線形空間 $\boldsymbol{X}$ から $[0,\infty)$ への写像 $\|\cdot\|_X$: $\boldsymbol{X} \longrightarrow [0,\infty)$ がつぎの性質 (i), (ii), (iii) をもつとき、$\boldsymbol{X}$ を**ノルム空間（normed linear spaec）**とよび、$\|\cdot\|_{\boldsymbol{X}}$ を $\boldsymbol{X}$ 上の**ノルム**という：

(i) $\| x \|_X = 0 \Leftrightarrow x = 0$

(ii) $\| x+y \|_X \leq \| x \|_X + \| y \|_X$ 　　　　（任意の $x, y \in \boldsymbol{X}$ ）

(iii) $\| cx \|_X = c \| x \|_X$ 　　　　c は任意のスカラー

定義 5.2 ：ノルム空間 $\boldsymbol{X}$ が距離空間として完備なとき、すなわち $\boldsymbol{X}$ の中の任意のコーシー列も $\boldsymbol{X}$ の元に収束するとき、$\boldsymbol{X}$ を **Banach 空間**という．

定義 5.3 ：$\boldsymbol{X}$ が $\boldsymbol{R}$ 上のノルム空間のとき、$\boldsymbol{X}$ 上の有界な線形汎関数の全体がつくるノルム空間を $\boldsymbol{X}$ の**双対空間（dual space）**とよび、これを $\boldsymbol{X}^*$ で表すことにする．

定義 5.4 ：$\boldsymbol{X}$ 上の線形空間 $\boldsymbol{X}$ がつぎの性質をもつ内積 $\langle \cdot, \cdot \rangle_X : \boldsymbol{X} \times \boldsymbol{X} \longrightarrow \boldsymbol{R}$ をもつとき、実内積空間という．

(i) $\langle x, y \rangle_X = \langle y, x \rangle_X$

(ii) $\langle x+y, z \rangle_X = \langle x, z \rangle_X + \langle y, z \rangle_X$ 　　　　（$x, y, z \in \boldsymbol{X}$）

(iii) $\langle x, x \rangle_X \geq 0$
とくに、$\langle x, x \rangle_X = 0$ となるのは $x = 0$ のときにかぎる．

$\boldsymbol{R}$ 上の内積空間 $\{\boldsymbol{X}, \langle \cdot, \cdot \rangle_X\}$ の任意の元 $x \in \boldsymbol{X}$ に対して、

$$\| x \|_X = \sqrt{\langle x, x \rangle_X}$$

とおくと、$\|\cdot\|_X$ は $\boldsymbol{X}$ 上のノルムとなる．よって、内積空間をこの意味でノルム空間と考える．

定義 5.5 ：$\boldsymbol{X}$ 上の内積空間であって、その内積から導かれるノルムに関して完備な距離空間すなわち *Banach* 空間になるとき、$\boldsymbol{X}$ を $\boldsymbol{R}$ 上の **Hilbert 空間**とよぶ．

いま、随伴作用素の定義を述べる．$\boldsymbol{X}$ と $\boldsymbol{Y}$ をノルム空間とし、それぞれの双対空間を $\boldsymbol{X}^*, \boldsymbol{Y}^*$ とする．$A : \boldsymbol{X} \longrightarrow \boldsymbol{Y}$ を $\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素[1] とする．このとき、A によって導

[1] $\| A \| = \sup\{| Ax | : x \in \boldsymbol{X}$かつ　$\| x \|_{\boldsymbol{X}} \leq 1\} < \infty$
また、有界 $\Leftrightarrow$ 連続

かれた $\boldsymbol{Y}^*$ から $\boldsymbol{X}^*$ への作用素 $A^T : \boldsymbol{Y}^* \longrightarrow \boldsymbol{X}^*$ を

$$y^* \longmapsto y^* \circ A \qquad (y^* \in \boldsymbol{Y}^*)$$

で定義し、A^T を**随伴作用素（adjoint operator）**とよぶ．ここで、$\boldsymbol{X}^* \times \boldsymbol{X}^*$ から $\boldsymbol{R}$ への写像を

$$(x, x^*) \mapsto \langle x, x^* \rangle_X = x^*(x)$$

で定義することにする．このとき、任意の $x \in \boldsymbol{X}$ と $y^* \in \boldsymbol{Y}^*$ に対して、

$$\langle Ax, y^* \rangle_Y = \langle x, A^T y^* \rangle_X$$

が成立する．

有界線形作用素に対する随伴線形作用素についての定理が成立する．

定理 5.1 ：
A. I を $\boldsymbol{X}$ 上の恒等作用素とすれば、$I^T = I$
B. A, B を $\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素とし、$c \in \boldsymbol{R}$ とする．このとき、

(i) $(A+B)^T = A^T + B^T$

(ii) $(cA)^T = cA^T$

C. A を $\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素とし、B を $\boldsymbol{Y}$ から $\boldsymbol{Z}$ への有界線形作用素とする．このとき、

$$(BA)^T = A^T B^T$$

D. A を $\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素とし、有界な逆作用素[2] をもつとする．このとき、

$$(A^{-1})^T = (A^T)^{-1}$$

E. $\| A^T \| = \| A \|$
F. A を $\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素とする．このとき、

$$(A^T)^T = A$$

$\boldsymbol{X}$ から $\boldsymbol{Y}$ への有界線形作用素に対して、

$$\text{Kernel }(A) = \{x \in \boldsymbol{X} : Ax = 0\}$$

を**零化空間**とよび、A の像を

$$\text{Image }(A) = \{y \in \boldsymbol{Y} : \text{ ある } x \in \boldsymbol{X} \text{ に対して、} y = Ax \quad \text{をみたす}\}$$

[2] $A : \boldsymbol{X} \longrightarrow \boldsymbol{Y}$ が単射の場合、$y \in \{Ax : x \in \boldsymbol{X}\}$ に対して、$Ax = y$ となる元 x を対応させる写像を逆写像とよぶ．

と記す．また、簡単に

$$\begin{aligned}\text{Kernel }(A^T) &= \{\text{ Image }(A)\}^{\perp}\\ \text{Kernel }(A) &= \{\text{ Image }(A^T)\}^{\perp}\end{aligned}$$

がわかる．なぜならば、$y^* \in$ Kernel (A^T) と $y = Ax \in$ Image (A) に対して、

$$\langle y^*, y\rangle_Y = \langle y^*, Ax\rangle_Y = \langle A^T x, x\rangle_X = 0$$

よりわかる．

5.2　スコア作用素

欠損値データについてのモデルを含む多くのセミパラメトリックモデルに対して、スコア作用素を考えると便利である．欠損値データの場合については、スコア作用素は実際に与えられたデータのスコア作用素の条件付き期待値として表現できる．

5.2.1　欠損値データモデルのスコア作用素

$\boldsymbol{Q} \equiv \{Q_\theta : \theta \in \Theta\}$ を母数モデルとして、$X^0 \sim Q_\theta$ と仮定する．$T : (\boldsymbol{X}^0, \boldsymbol{A}^0) \longrightarrow (\boldsymbol{X}, \boldsymbol{A})$ を可測関数とし、$X = T(X^0)$ を観測とする．ここで、T によって導出されたモデルを $\boldsymbol{P}_T = \boldsymbol{Q}T^{-1} = \{Q_\theta T^{-1} : \theta \in \Theta\}$ と記す ことにする．$\boldsymbol{P}_T = \boldsymbol{Q}T^{-1}$ と $\boldsymbol{Q}$ のスコア関数を関係づける命題を導出するために、つぎの補題を用いる．

補題 5.1 ：$\boldsymbol{Q}$ を正則モデルとし、$Q_\theta \ll \mu$ とする．$s(\theta) = \sqrt{dQ_\theta/d\mu}$ とおいたとき、

$$\| \, s(\theta+h) - s(\theta) - \frac{2\dot{s}^T(\theta)hs(\theta)}{s(\theta+h)+s(\theta)} \, \|_{L_2(\mu)} = o(h)$$

が成立する．

証明：

$$\frac{|s(\theta+h)-s(\theta)|}{s(\theta+h)+s(\theta)} \le 1$$

に注意すれば、$i = 1, \ldots, k$ とある $K > 0$ に対し、

$$\begin{aligned}
\left\| \dot{s}_i(\theta) - \frac{2\dot{s}_i(\theta)s(\theta)}{s(\theta+h)+s(\theta)} \right\|_{L_2(\mu)} &= \left\| \frac{\dot{s}_i(\theta)\{s(\theta+h)-s(\theta)\}}{s(\theta+h)+s(\theta)} \right\|_{L_2(\mu)} \\
&\le \left\{ \int_{|\dot{s}_i(\theta)|>Ks(\theta)} \left[\frac{\dot{s}_i(\theta)\{s(\theta+h)-s(\theta)\}}{s(\theta+h)+s(\theta)} \right]^2 d\mu \right\}^{1/2} \\
&\quad + \left\{ \int_{|\dot{s}_i(\theta)|\le Ks(\theta)} \left[\frac{\dot{s}_i(\theta)\{s(\theta+h)-s(\theta)\}}{s(\theta+h)+s(\theta)} \right]^2 d\mu \right\}^{1/2} \\
&\le \left\{ \int_{|\dot{s}_i(\theta)|>Ks(\theta)} \dot{s}_i^2(\theta) d\mu \right\}^{1/2} + K^2 \left\{ \int \{s(\theta+h)-s(\theta)\}^2 d\mu \right\}^{1/2}
\end{aligned}$$

となる．$\dot{s}_i(\theta)/s(\theta) \in L_2(P_\theta)$ と $s(\cdot)$ が $\|\cdot\|_{L_2(\mu)}$ について連続[3] であることに注意すれば、(5.1) の最右辺の第一項目は、$K \uparrow \infty$ とすれば、一様可積分性よりゼロになり、第二項目は $\|h\| \to 0$ のとき、ゼロとなる．

つぎに、モデルの正則性から

$$\|s(\theta+h) - s(\theta) - \dot{s}^T(\theta)h\|_{L_2(\mu)} = o(|h|)$$

であることと (5.1) を用いると

$$\begin{aligned}
&\left\| s(\theta+h) - s(\theta) - \frac{2\dot{s}^T(\theta)hs(\theta)}{s(\theta+h)+s(\theta)} \right\|_{L_2(\mu)} \\
&\quad \leq \|s(\theta+h) - s(\theta) - \dot{s}^T(\theta)h\|_{L_2(\mu)} + \left\| \dot{s}^T(\theta)h - \frac{2\dot{s}^T(\theta)hs(\theta)}{s(\theta+h)+s(\theta)} \right\|_{L_2(\mu)} \\
&\quad = o(|h|)
\end{aligned} \tag{5.1}$$

がわかる．したがって、補題は示された． □

補題 5.2 ：

$$p_{T,\theta} = \frac{dP_{T,\theta}}{d\mu T^{-1}}$$

とし、$s_T(\theta) = \sqrt{p_{T,\theta}}$ とする．このとき、

$$s_T^2(\theta) = \boldsymbol{E}_\mu(s^2(\theta)|T) \qquad a.s. \quad \mu T^{-1} \tag{5.2}$$

が成立する．

証明：任意の $A \in \boldsymbol{A}$ に対し、

$$\begin{aligned}
P_T(A) &= QT^{-1}(A) = Q\{T(X^0) \in A\} = \int_{T(x^0)\in A} \frac{dQ}{d\mu}(x^0)d\mu(x^0) \\
&= \boldsymbol{E}_\mu\left[1\{T(X^0)\in A\}\boldsymbol{E}_\mu\left\{\frac{dQ}{d\mu}(X^0)\middle|T\right\}\right] \\
&= \int_{T(x^0)\in A} \boldsymbol{E}_\mu\left\{\frac{dQ}{d\mu}\middle|T = T(x^0)\right\}d\mu(x^0) \\
&= \int_A \boldsymbol{E}_\mu\left\{\frac{dQ}{d\mu}\middle|T = x\right\}d\mu T^{-1}(x)
\end{aligned}$$

より補題は証明された． □

補題 5.3 ：(X, Y, Z) を 3 変量確率変数とし、$X, Y \geq 0, a.s.$ かつ $\boldsymbol{E}X, \boldsymbol{E}Y, \boldsymbol{E}\mid Z\mid$ は有限とする．このとき、

$$\left(\sqrt{\boldsymbol{E}X} - \sqrt{\boldsymbol{E}Y} - \frac{\boldsymbol{E}Z}{\sqrt{\boldsymbol{E}X}+\sqrt{\boldsymbol{E}Y}}\right)^2 \leq \boldsymbol{E}\left(\sqrt{X} - \sqrt{Y} - \frac{Z}{\sqrt{X}+\sqrt{Y}}\right)^2$$

が成立する．

[3] $s(\theta)$ は Fréchet 微分可能であることから、$\|\cdot\|_{L_2(\mu)}$ について $s(\theta)$ は連続となることがわかる．

証明：BKRW の pages 462 を参照されたい． □

命題 5.1 ：$\boldsymbol{Q} = \{Q_\theta : \theta \in \Theta \subset \boldsymbol{R}^k\}$ を正則モデルとし、スコア関数

$$a(x^0) = a(x^0, Q_\theta \mid \theta, \boldsymbol{Q}) = \frac{\partial}{\partial\theta} \log p(x^0, \theta)$$

をもつとする．このとき、

$$\boldsymbol{P}_T = \{Q_\theta T^{-1} : Q_\theta \in \boldsymbol{Q}\}$$

は正則モデルでスコア関数

$$\dot{\boldsymbol{l}}(x, P_{T,\theta}|\theta, \boldsymbol{P}_\theta) = \boldsymbol{E}[a(X^0)|T(X^0) = x] \equiv \dot{\boldsymbol{l}}_T(x, P_{T,\theta}) \tag{5.3}$$

をもつ．更に、

$$\begin{aligned} I(\theta, P_{T,\theta}) &= \boldsymbol{E}[\dot{\boldsymbol{l}}(X, P_{T,\theta})\dot{\boldsymbol{l}}^T(X, P_{T,\theta})] \\ &= \boldsymbol{E}\Big\{\boldsymbol{E}[\dot{\boldsymbol{l}}(X^0, Q_\theta)|T(X^0) = x]\boldsymbol{E}[\dot{\boldsymbol{l}}^T(X^0, Q_\theta)|T(X^0) = x]\Big\} \\ &\leq \boldsymbol{E}[\dot{\boldsymbol{l}}(X^0, Q_\theta)\dot{\boldsymbol{l}}^T(X^0, Q_\theta)] = I(\theta, Q_\theta) \end{aligned}$$

が成立する．

証明：

$$\dot{s}_T(\theta) = \frac{\boldsymbol{E}_\mu(\dot{q}(\theta)|T)}{2s_T(\theta)} \tag{5.4}$$

とおく．補題 5.1、5.2、5.3 を用いれば、

$$\begin{aligned} &\left\| s_T(\theta+h) - s_T(\theta) - \frac{2\dot{s}_T^T(\theta)hs_T(\theta)}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &\leq \Bigg\| \sqrt{\boldsymbol{E}\{s^2(\theta+h)|T\}} - \sqrt{\boldsymbol{E}\{s^2(\theta)|T\}} \\ &\quad - \frac{\boldsymbol{E}\{\dot{q}^T(\theta)h|T\}}{\sqrt{\boldsymbol{E}\{s^2(\theta+h)|T\}} + \sqrt{\boldsymbol{E}\{s^2(\theta)|T\}}} \Bigg\|_{L_2(\mu)} \\ &\leq \left\| s(\theta+h) - s(\theta) - \frac{\dot{p}^T(\theta)h}{s(\theta+h) + s(\theta)} \right\|_{L_2(\mu)} \\ &= \left\| s(\theta+h) - s(\theta) - \frac{2\dot{s}^T(\theta)hs(\theta)}{s(\theta+h) + s(\theta)} \right\|_{L_2(\mu)} = o(|h|) \end{aligned} \tag{5.5}$$

が成立することがわかる．更に、補題 5.2 を用いて、

$$\begin{aligned} (\dot{s}_T)_i^2(\theta) &= \frac{\{\boldsymbol{E}(\dot{q}_i(\theta)|T)\}^2}{4s_T^2(\theta)} \\ &= \frac{\{\boldsymbol{E}(2\dot{s}_i(\theta)s(\theta)|T)\}^2}{4s_T^2(\theta)} \\ &\leq \frac{\boldsymbol{E}(\dot{s}_i^2(\theta)|T)\boldsymbol{E}(s^2(\theta)|T)}{\boldsymbol{E}(s^2(\theta)|T)} \\ &= \boldsymbol{E}(\dot{s}_i^2(\theta)|T) \end{aligned} \tag{5.6}$$

が成立することがわかる．(5.6) と $\dot{s}_i(\theta)/s(\theta) \in L_2(\mu)$ [4] を用いれば、

$$\begin{aligned}\int \left(\frac{(\dot{s}_T)_i(\theta)}{s_T(\theta)}\right)^2 dQ_\theta T^{-1} &= \int \frac{(\dot{s}_T)_i(\theta)}{s_T^2(\theta)} s_T^2 d\mu T^{-1} \\ &= \int (\dot{s}_T)_i^2(\theta) d\mu T^{-1} \leq \int \boldsymbol{E}_\mu(\dot{s}_i^2(\theta)|T) d\mu T^{-1} \\ &= \int \boldsymbol{E}_\mu(\dot{s}_i^T(\theta)|T) d\mu \leq \int \dot{s}_i^2(\theta) d\mu < \infty\end{aligned}$$

が成立するので、

$$\frac{(\dot{s}_T)_i(\theta)}{s_T(\theta)} \in L_2(Q_\theta T^{-1})$$

となる．(5.5) と補題 5.1 から

$$\|s_T(\theta+h) - s_T(\theta) - \dot{s}_T^T(\theta)h\|_{L_2(\mu)} = o(|h|) \tag{5.7}$$

が成立し、これより $s_T(\theta)$ は θ について連続であることがわかる．

つぎに、$\dot{s}_T$ の連続性を示す．そのために、(5.4) を用いれば、

$$\begin{aligned}&\|(\dot{s}_T)_i(\theta+h) - (\dot{s}_T)_i(\theta)\|_{L_2(\mu)} \\ &= \Bigg\| \frac{\{2s_T(\theta+h) - (s_T(\theta+h) - s_T(\theta))\}(\dot{s}_T)_i(\theta+h)}{s_T(\theta+h) + s_T(\theta)} \\ &\quad - \frac{\{2s_T(\theta) - (s_T(\theta+h) - s_T(\theta)\})(\dot{s}_T)_i(\theta)}{s_T(\theta+h) + s_T(\theta)} \Bigg\|_{L_2(\mu)} \\ &\leq \left\| \frac{2s_T(\theta+h)(\dot{s}_T)_i(\theta+h) - 2s_T(\theta)(\dot{s}_T)_i(\theta)}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &\quad + \left\| \frac{(\dot{s}_T)_i(\theta+h)(s_T(\theta+h) - s_T(\theta))}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &\quad + \left\| \frac{(\dot{s}_T)_i(\theta)(s_T(\theta+h) - s_T(\theta))}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &= \left\| \frac{\boldsymbol{E}_\mu(\dot{q}_i(\theta+h) - \dot{q}_i(\theta)|T)}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &\quad + \left\| \frac{(\dot{s}_T)_i(\theta+h)(s_T(\theta+h) - s_T(\theta))}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)} \\ &\quad + \left\| \frac{(\dot{s}_T)_i(\theta)(s_T(\theta+h) - s_T(\theta))}{s_T(\theta+h) + s_T(\theta)} \right\|_{L_2(\mu)}\end{aligned} \tag{5.8}$$

となることがわかる．$\dot{s}(\theta)$ の連続性[5] と $\dot{s}(\theta) \in L_2(\mu)$ であることより、$\dot{s}_i(\theta+h)/s(\theta+h)$ は一様2乗可積分となる．したがって、$(\dot{s}_T)_i(\theta+h)/s(\theta+h)$ も一様2乗可積分[6] となる．このこと

[4] 正則性より $\dot{s}_i(\theta)/s(\theta) \in L_2(P_\theta)$ であるので、これは明らか．

[5] すなわち、$|h| \to 0$ のとき

$$\|\dot{s}(\theta+h) - \dot{s}(\theta)\|_{L_2(\mu)} \to 0$$

が成立する．

[6] K.L. Chung の pages 97 から $\dot{s}_i(\theta+h)/s(\theta+h) \in L_2(P_{\theta+h})$ ならば $(\dot{s}_T)_i(\theta+h)/s_T(\theta+h) \in L_2(P_{\theta+h}T^{-1})$ がわかる．

と補題 5.1 を用いれば、ある $K>0$ を取れば、(5.8) の最右辺の第二項目は、$|h|\to 0$ のとき、

$$\begin{aligned}
&\left\|\frac{(\dot{s}_T)_i(\theta+h)(s_T(\theta+h)-s_T(\theta))}{s_T(\theta+h)+s_T(\theta)}\right\|^2_{L_2(\mu)} \\
&\quad\le \int_{|(\dot{s}_T)_i(\theta+h)|>Ks_T(\theta+h)}\left(\frac{(\dot{s}_T)_i(\theta+h)}{s_T(\theta+h)}\right)^2 dP_{\theta+h}T^{-1} \\
&\quad\quad +K^2\int\{s_T(\theta+h)-s_T(\theta)\}^2 d\mu T^{-1}\to 0
\end{aligned} \tag{5.9}$$

となることがわかる．また、Cauchy – Schwarz の不等式と Jensen の不等式を利用すれば、第一項目については

$$\begin{aligned}
&\{\boldsymbol{E}_\mu(\dot{q}_i(\theta+h)-\dot{q}_i(\theta)|T)\}^2 \\
&\quad\le \boldsymbol{E}_\mu\left[\left\{\frac{\dot{q}_i(\theta+h)-\dot{q}_i(\theta)}{s(\theta+h)+s(\theta)}\right\}^2\middle|T\right]\boldsymbol{E}_\mu(s(\theta+h)+s(\theta))^2 \\
&\quad\le \boldsymbol{E}_\mu\left[\left\{\frac{2(\dot{s}_i(\theta+h)s(\theta+h)-\dot{s}_i(\theta)s(\theta))}{s(\theta+h)+s(\theta)}\right\}^2\middle|T\right]O(1) \\
&\le \boldsymbol{E}_\mu\left[\left\{\frac{s(\theta+2h)+s(\theta+h)}{s(\theta+h)+s(\theta)}\right\}^2\dot{s}_i^2(\theta+h)\middle|T\right]O(1) \\
&\quad\quad +\boldsymbol{E}_\mu[\{\dot{s}_i(\theta+h)-\dot{s}_i(\theta)\}^2|]+o(|h|)
\end{aligned} \tag{5.10}$$

のように評価できる．最後の不等式は $\theta=\theta+h$ として補題 5.1 の証明中の (5.1) を用いると

$$2\dot{s}_i(\theta+h)s(\theta+h)=\dot{s}_i(\theta+h)(s(\theta+2h)+s(\theta+h))+o(|h|)$$

となることに注意すれば、

$$\frac{2\{\dot{s}_i(\theta+h)s(\theta+h)-\dot{s}_i(\theta)s(\theta)\}}{s(\theta+h)+s(\theta)}=\frac{\dot{s}_i(\theta+h)\{s(\theta+2h)+s(\theta+h)\}}{s(\theta+h)+s(\theta)}$$

となることよりわかる．$\dot{s}_i(\theta)$ の連続性[7] を用いれば、(5.10) の最右辺は、$|h|\to 0$ のとき、ゼロに収束することがわかる．したがって、(5.9) と (5.10) より、$(\dot{s}_T)_i$ の θ に関する連続性が示せた． □

5.2.2 Mixture モデルのスコア関数

まず、

$$\boldsymbol{Q}=\{Q_{\theta,\eta}\ll\mu:\theta\in\Theta\subset\boldsymbol{R}^r,\eta\in\boldsymbol{H}\subset\boldsymbol{R}^q\}$$

[7] すなわち、$|h|\to 0$ のとき、

$$\|\dot{s}_i(\theta+h)-s_i(\theta)\|_{L_2(\mu)}\to 0$$

が成立する．

を正則母数モデルとし、確率密度関数を $f(\cdot,\theta,\eta)=dQ_{\theta,\eta}/d\mu$ と記す．$\boldsymbol{H}$ 上のすべての分布の適当な部分集合 $\boldsymbol{G}$ [8] に対して、新たなモデルを

$$\boldsymbol{P}=\left\{P_{(\theta,G)}=\int Q_{(\theta,\eta)}dG(\eta):\theta\in\Theta,G\in\boldsymbol{G}\right\}$$

で定義する．また、$p(x;\theta,G)=dP_{(\theta,G)}/d\mu$ と書く．更に、

$$\begin{aligned}\boldsymbol{P}_1 &= \left\{P_{(\theta,G_0)}:G_0\ \text{は固定}\right\}\subset\boldsymbol{P}\\ \boldsymbol{P}_2 &= \left\{P_{(\theta_0,G)}:\theta_0\ \text{は固定}\right\}\subset\boldsymbol{P}\end{aligned}$$

とおき、

$$J(\theta,\eta)=\int\frac{|\,\dot{f}(x;\theta,\eta)\,|^2}{f(x;\theta,\eta)}d\mu(x)$$

とおく．ただし、$\dot{f}$ は θ に関する f の導関数とする．$\boldsymbol{Q}$ は正則なので、写像 $\theta\mapsto J(\theta.\eta)$ は連続である．

命題 5.2 ：つぎを仮定する．

(i) $\boldsymbol{Q}$ は正則．

(ii) すべての θ について

$$\int J(\theta,\eta)dG(\eta)<\infty$$

で、すべての $G\in\boldsymbol{G}$ について

$$\theta\mapsto\int J(\theta,\eta)dG(\eta)$$

は連続とする．

(iii) G を固定してモデル $\boldsymbol{P}_1(G)$ における θ の情報行列

$$I(\theta)=I(P_0\mid\theta,\boldsymbol{P}_1(G))=\int\frac{\dot{p}\dot{p}^T}{p}(x;\theta,G)d\mu$$

は正則とする．

このとき、$\boldsymbol{P}_1(G_0)$ は正則モデルとなり、

$$\begin{aligned}\dot{\boldsymbol{l}}_1(X,\theta,G) &= \frac{\int\dot{f}(X;\theta,\eta dG(\eta)}{\int f(X;\theta,\eta)dG(\eta)}\\ &= \frac{\int(\dot{f}/f)(X;\theta,\eta)f(X;\theta,\eta)dG(\eta)}{\int f(X;\theta,\eta)dG(\eta)}\\ &= \boldsymbol{E}[\boldsymbol{l}(X,\theta,\dot{U})\mid X]\end{aligned}$$

[8] $\boldsymbol{G}=\{G\ll\nu\}$ とする．

が成立する．ただし、

$$\dot{\boldsymbol{l}}(X,\theta,\eta)=\frac{\dot{f}(X,\theta,\eta)}{f(X,\theta,\eta)}$$

である．もし、

$$\left\{P_{(\theta,G_\gamma)}:\mid\gamma\mid<1,\frac{dG_\gamma}{dG_0}=g_\gamma,g_0=1\right\}$$

が $\boldsymbol{P}_2$ の正則母数部分モデルならば、

$$\frac{\partial}{\partial\gamma}\boldsymbol{l}(X,\theta,G_\gamma)\bigg|_{\gamma=0}=\boldsymbol{E}[a(U)\mid X]$$

が成立する，ただし、

$$a(\eta)=\frac{\partial}{\partial\gamma}\log g_\gamma\bigg|_{\gamma=0}$$

である．更に、$\boldsymbol{G}$ が十分豊かならば、

$$\dot{\boldsymbol{P}}_2\supset\left\{\boldsymbol{E}[a(U)\mid X]:a\in L_2^0(G)\right\}\tag{5.11}$$

が成立する．

証明：(i)　より

$$\begin{aligned}
&\frac{1}{|\Delta|}\|\sqrt{f(\cdot,\theta+\Delta,\eta)}-\sqrt{f(\cdot,\theta,\eta)}\|_{L_2(\mu)}\\
&=\frac{1}{|\Delta|}\left\|\frac{1}{2}\int_0^1\Delta^T\frac{\dot{f}}{\sqrt{f}}(\cdot,\theta+\xi\Delta,\eta)d\xi\right\|_{L_2(\mu)}\\
&\leq\frac{1}{|\Delta|}\sup_{0\leq\xi\leq1}\left\|\frac{1}{2}\Delta^T\frac{\dot{f}}{\sqrt{f}}(\cdot,\theta+\xi\Delta,\eta)\right\|_{L_2(\mu)}\\
&\leq\frac{1}{2}\sup_{0\leq\xi\leq1}\sqrt{J(\theta+\xi\Delta,\eta)}
\end{aligned}\tag{5.12}$$

となる．固定した $G\in\boldsymbol{G}$ について

$$s(x,\theta,G)=\sqrt{\int f(x,\theta,\eta)dG(\eta)}$$

とおく．補題 5.3 を用いれば、

$$\begin{aligned}
&\int\left\{\frac{s(x,\theta+\lambda\Delta,G)-s(x,\theta,G)}{\lambda}-\frac{\Delta^T\int\dot{f}(x,\theta,\eta)dG(\eta)}{s(x,\theta+\lambda\Delta,G)+s(x,\theta,G)}\right\}^2d\mu\\
&\leq\int\int\left\{\frac{\sqrt{f(x,\theta+\lambda\Delta,\eta)}-\sqrt{f(x,\theta,\eta)}}{\lambda}\right.\\
&\qquad\left.-\frac{\Delta^T\dot{f}(x,\theta,\eta)}{\sqrt{f(x,\theta+\lambda\Delta,\eta)}-\sqrt{f(x,\theta,\eta)}}\right\}^2d\mu(x)dG(\eta)
\end{aligned}\tag{5.13}$$

を得る．(5.12) と補題 5.1 を用いれば、

$$\begin{aligned}&\left\|\frac{\sqrt{f(x,\theta+\lambda\Delta,\eta)}-\sqrt{f(x,\theta,\eta)}}{\lambda}-\frac{\Delta^T\dot{f}(x,\theta,\eta)}{\sqrt{f(x,\theta+\lambda\Delta,\eta)}-\sqrt{f(x,\theta,\eta)}}\right\|_{L_2(\mu)}\\&\le\frac{1}{2}\sup_{0\le\xi\le1}\sqrt{J(\theta+\xi\Delta,\eta)}+\left\|\frac{\Delta^T\dot{f}(x,\theta,\eta)}{f(x,\theta,\eta)}\right\|_{L_2(\mu)}\\&\le\frac{1}{2}\sup_{0\le\xi\le1}\sqrt{J(\theta+\xi\Delta,\eta)}+\sqrt{J(\theta,\eta)}|\Delta| \qquad (5.14)\end{aligned}$$

を得る．(5.14) の最右辺は任意の $\lambda\in[0,1]$ に対し、G で可積分であることに注意して、有界収束定理を用いれば、$\lambda\to0$ のとき、(5.13) の最右辺はゼロに収束することがわかる．したがって、$U\sim G$ とし、

$$\dot{s}(x,\theta,G)=\frac{\boldsymbol{E}[\dot{f}(X,\theta,U)|X]}{2s(x,\theta,G)} \qquad (5.15)$$

とおけば、

$$\left\|s(x,\theta+\lambda\Delta,G)-s(x,\theta,G)-\lambda\Delta^T\frac{2\dot{s}(x,\theta,G)s(x,\theta,G)}{s(x,\theta+\lambda\Delta,G)+s(x,\theta,G)}\right\|_{L_2(\mu)}=o(|\lambda|) \qquad (5.16)$$

がわかる．(5.16) と補題 5.1 の証明中の (5.1) を用いれば、

$$\|s(x,\theta+\lambda\Delta,G)-s(x,\theta,G)-\lambda\Delta^T\dot{s}(x,\theta,G)\|_{L_2(\mu)}=o(|\lambda|) \qquad (5.17)$$

が成立することがわかる．

つぎに、$\dot{s}_T(\theta)$ の θ に関する連続性を示す．以後、簡略して $s_G(\theta)=s(x,\theta,G)$ と記すことにする．(5.12) から

$$\begin{aligned}&\|(\dot{s}_G)_i(\theta+h)-(\dot{s}_G)_i(\theta)\|_{L_2(\mu)}\\&\quad=\left\|\frac{\{2s_G(\theta+h)-(s_G(\theta+h)-s_G(\theta))\}(\dot{s}_G)_i(\theta+h)}{s_G(\theta+h)+s_G(\theta)}\right.\\&\qquad\left.-\frac{\{2s_G(\theta)-(s_G(\theta+h)-s_G(\theta))\}(\dot{s}_G)_i(\theta)}{s_G(\theta+h)+s_G(\theta)}\right\|_{L_2(\mu)}\\&\quad\le\left\|\frac{\boldsymbol{E}(\dot{f}(X,\theta+h,U)-\dot{f}(X,\theta,U))}{s_G(\theta+h)+s_G(\theta)}\right\|_{L_2(\mu)}\\&\qquad+\left\|\frac{(\dot{s}_T)_i(\theta+h)(s_G(\theta+h)-s_G(\theta))}{s_G(\theta+h)+s_G(\theta)}\right\|_{L_2(\mu)}\\&\qquad+\left\|\frac{(\dot{s}_T)_i(\theta)(s_G(\theta+h)-s_G(\theta))}{s_G(\theta+h)+s_G(\theta)}\right\|_{L_2(\mu)}\end{aligned}$$

となることがわかる． $\dot{f}(x,\theta,\eta)$ の連続性と $\dot{f}_i(x,\theta,\eta)/f(x,\theta,\eta)$ が一様 2 乗可積分であることから、$(\dot{s}_G)_i(\theta+h)/s_G(\theta+h)$ は一様 2 乗可積分であることがわかるので、ある $K>0$ を取れ

ば、$|h| \to 0$ のとき、

$$\begin{aligned}&\left\|\frac{(\dot{s}_T)_i(\theta+h)(s_G(\theta+h)-s_G(\theta))}{s_G(\theta+h)+s_G(\theta)}\right\|_{L_2(\mu)}^2\\&\quad\le\int_{|(\dot{s}_G)_i(\theta+h)|>Ks_G(\theta+h)}\left(\frac{(\dot{s}_G)_i(\theta+h)}{s_G(\theta+h)}\right)^2dP_{(\theta+h,G)}\\&\qquad+K^2\int\{s_G(\theta+h)-s_G(\theta)\}^2d\mu\to 0\end{aligned}\tag{5.18}$$

となることがわかる．一方、$s(x,\theta,u)=\sqrt{f(x.\theta,u)}$ とおき、Cauchy – Schwarz の不等式を用いて、命題 5.1 の証明中の (5.10) と同様の変形を行えば、

$$\begin{aligned}&\{\boldsymbol{E}[\dot{f}_i(X,\theta+h,U)-\dot{f}_i(X,\theta,U)|X]\}^2\\&\quad\le\boldsymbol{E}\left[\left\{\frac{s(X,\theta+2h,U)+s(X,\theta+h,U)}{s(X,\theta+h,U)+s(X,\theta,U)}-1\right\}^2\dot{s}_i(X,\theta+h,U)\middle|X\right]\\&\qquad+\boldsymbol{E}[\{\dot{s}_i(X,\theta+h,U)-\dot{s}_i(X,\theta,U)\}^2]O(1)+o(|h|)\end{aligned}\tag{5.19}$$

を得る．$\dot{s}_i$ の連続性を用いれば、(5.19) の右辺はゼロに収束することがわかる．(5.18) と (5.19) から、$\dot{s}_G$ の連続性が示せた．

最後に、(5.11) を示すために、h を有界関数で $\boldsymbol{E}h(U)=0$ を満足するものとする．十分小さい γ に対し、

$$g_\gamma=\frac{dG_\gamma}{dG_0},\qquad g_\gamma(\eta)=\frac{e^{\gamma h(\eta)}}{\int e^{\gamma h(\eta)}dG(\eta)}$$

とおく．まず、$|h|\to 0$ のとき、

$$\begin{aligned}&\left\|\sqrt{f(x,\theta,\eta)g_\gamma(\eta)\frac{dG_\gamma}{d\nu}}-\sqrt{f(x,\theta,\eta)\frac{dG_0}{d\nu}}-\frac{h}{2}\sqrt{f(x,\theta,\eta)\frac{dG_0}{d\nu}}\right\|_{L_2(\mu\times\nu)}\\&\quad\le\int\{\sqrt{g_\gamma(\eta)}-1-\frac{h}{2}\}^2dG(\eta)\to 0\end{aligned}$$

となることに注意する．

$$\sqrt{f(x,\theta,\eta)g_\gamma(\eta)\frac{dG_\gamma}{d\nu}}=s(x,\theta,\eta,G,\gamma)=s(\gamma)$$

とおけば、

$$\dot{s}(\gamma)=\frac{h}{2}s(0)$$

となる．更に、

$$s_{(\theta,G)}(\gamma)=\sqrt{\int s^2(\gamma)dG_\gamma}$$

と

$$\dot{s}_{(\theta,G)}(\gamma)=\boldsymbol{E}[\frac{h}{2}|X]s_{(\theta,G)}(\gamma)$$

とおく．

$$\left\| s_{(\theta,G)}(\gamma) - s_{(\theta,G)}(0) - \frac{2\dot{s}_{(\theta,G)}(0)s_{(\theta,G)}(0)}{s_{(\theta,G)}(\gamma)+s_{(\theta,G)}(0)} \right\|_{L_2(\mu)}$$
$$\leq \left\| s(\gamma) - s(0) - \frac{2\dot{s}(\gamma)s(\gamma)}{s(\gamma)+s(0)} \right\|_{L_2(\mu)} = o(\gamma)$$

となることがわかる．

あとは $\dot{s}_{(\theta,G)}(\gamma)$ の γ に関する連続性を同様な議論を用いて示せば、補題は証明される．

□

5.2.3　スコア作用素と情報限界

(5.2) における条件付き期待値は $L_2^0(\boldsymbol{Q})$ の中の関数 $a(\cdot) = \dot{\boldsymbol{l}}(\cdot, Q|\theta, \boldsymbol{Q})$ を $L_2^0(\boldsymbol{P}_T)$ の元 $b \equiv \dot{\boldsymbol{l}}_T(\cdot, P_T|\theta, \boldsymbol{P}_T)$ に写す．このことより、(5.3) におけるスコア関数 $\dot{\boldsymbol{l}}_T$ の左辺を

$$\dot{\boldsymbol{l}}a$$

と書くことにする．ただし、$\dot{\boldsymbol{l}} = \boldsymbol{E}[\cdot|T(X^0)]$ は単に条件付き期待値とする．すると $\dot{\boldsymbol{l}}$ は $L_2^0(\boldsymbol{Q})$ から $L_2^0(\boldsymbol{P}_T)$ への有界線形写像となる．更に、$\dot{\boldsymbol{l}}$ の随伴写像を

$$\dot{\boldsymbol{l}}^T : L_2^0(\boldsymbol{P}_T) \longrightarrow L_2^0(\boldsymbol{Q})$$

と書くことにする．これは、すべての $a \in L_2^0(\boldsymbol{P}_T)$ と $b \in L_2^0(\boldsymbol{Q})$ に対して、

$$\langle b, \dot{\boldsymbol{l}}a \rangle_{L_2(P_T)} = \langle \dot{\boldsymbol{l}}^T b, a \rangle_{L_2(Q)}$$

を満足するものとである．このとき、

$$\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} : \dot{\boldsymbol{Q}} \longrightarrow \dot{\boldsymbol{Q}}$$

を**情報作用素**とよぶ．$\dot{\boldsymbol{l}}$ が条件付き期待値作用素として定義されるときには、$\dot{\boldsymbol{l}}^T$ も条件付き期待値作用素、すなわち

$$\dot{\boldsymbol{l}}^T = \boldsymbol{E}[b(X)|X^0 = x^0]$$

として定義される．

例 5.1 ：$\Theta \subset \boldsymbol{R}^k$ とする．スコア作用素

$$\dot{\boldsymbol{l}} : \Theta \longrightarrow L_2^0(P_\theta)$$

をすべての $h \equiv (h_1, h_2, \ldots, h_k) \in \boldsymbol{R}^k$ に対して、

$$\dot{\boldsymbol{l}}h = \dot{\boldsymbol{l}}_\theta h^T = \sum_{i=1}^k \dot{\boldsymbol{l}}_{\theta i} h_i$$

で定義する．ただし、$\dot{\boldsymbol{l}}_\theta = (\dot{\boldsymbol{l}}_{\theta 1}, \ldots, \dot{\boldsymbol{l}}_{\theta k})$ で

$$\dot{\boldsymbol{l}}_{\theta i} = \frac{\partial}{\partial \theta_i} \log p_\theta, \qquad i = 1, 2, \ldots, k$$

とする．このとき、

$$\left\{ \int (\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h \dot{\boldsymbol{l}}_\theta^T \sqrt{p_\theta})^2 d\mu \right\}^{1/2} = o(|\ h\ |)$$

が成立することがわかる．随伴作用素 $\dot{\boldsymbol{l}}^T : L_2(P_\theta) \longrightarrow \dot{\Theta} = \boldsymbol{R}^k$ は、すべての $b \in L_2^0(P_\theta)$ に対して、

$$\dot{\boldsymbol{l}}^T = \boldsymbol{E}_\theta[\dot{\boldsymbol{l}}_\theta^T b] \in \boldsymbol{R}^k$$

で定義され、情報作用素 $\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} : \boldsymbol{R}^k \longrightarrow \boldsymbol{R}^k$ は、任意の $h \in \boldsymbol{R}^k$ に対して、

$$\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} h = \dot{\boldsymbol{l}}^T (\dot{\boldsymbol{l}}_\theta h^T) = \boldsymbol{E}_\theta[(\dot{\boldsymbol{l}}_\theta^T \dot{\boldsymbol{l}}_\theta) h^T] = I(\theta) h^T$$

で定義される．

例 5.2 （Interval censoring）：$X^0 = (Y, Z) \in [0, \infty) \times [0, \infty)$ とし、Y と Z は独立な確率変数で分布 $F(\cdot)$ と $G(\cdot)$ をもつとする．$Q = F \times G$ とし、$\boldsymbol{Q} = \{Q :\ F$ と G は μ と λ に関して絶対連続 $\}$ と記すことにする．

$$X = T(X^0) = (Z, 1\{Y \le Z\}) = (Z, \delta)$$

を観測したとする．このとき、モデルは

$$\boldsymbol{P}_T = \left\{ P_T = QT^{-1} : Q = F \times G, F \ll \mu, G \ll \lambda \right\}$$

となる．よって、$P_T = P_{F,G}$ は確率密度関数

$$p_{F,G}(x) = p_{F,G}(z, \delta) = (F(z))^\delta (1 - F(z))^{1-\delta} g(z) \tag{5.20}$$

をもつことがわかる．ただし、$g(z) = dG/d\lambda, \delta = \{0, 1\}$ である．G を固定し、既知とすれば、$\dot{\boldsymbol{Q}}$ と $L_2^0(F)$ を同一視でき、モデルのスコア関数は、$a \in L_2^0(F)$ に対し

$$\begin{aligned} (\dot{\boldsymbol{l}} a)(x) &= \boldsymbol{E}[a(Y) | T(X^0) = x] \\ &= \delta \frac{\int_0^z a(y) dF}{F(z)} + (1 - \delta) \frac{\int_z^\infty a(y) dF}{1 - F(z)} \end{aligned} \tag{5.21}$$

で与えられる[9]．

[9] $x = (z, \delta)$ が与えられたときの Y の条件付き確率密度関数が

$$f(y|x) = \begin{cases} f(y) 1_{(0,z)}(y) / F(z) & (\delta = 1) \\ f(y) 1_{(z,\infty 9}(y)(1 - F(y)) & (\delta = 0) \end{cases}$$

で与えられるｋｔｏｎｉ注意．ここで、$f = dF/d\mu$ とした．

つぎに、随伴作用素を求める．任意の $b \in L_2^0(P_{F,G})$ に対し

$$
\begin{aligned}
\langle b, \boldsymbol{i}a\rangle_{L_2(P_{F,G})} &= \int_0^\infty b(x)\{(\boldsymbol{i}a)(x)\}dP_{F,G}(x) \\
&= \int_0^\infty b(z,\delta)\left[\delta\frac{\int_0^z a(y)dF}{F(z)} + (1-\delta)\frac{\int_z^\infty a(y)dF}{1-F(z)}\right]dP_{F,G}(x) \\
&= \int_0^\infty b(z,1)\frac{\int_0^z a(y)dF}{F(z)}F(z)dG(z) \\
&\quad + \int_0^\infty b(z,0)\frac{\int_z^\infty a(y)dF}{1-F(z)}(1-F(z))dG(z) \\
&= \int_0^\infty a(y)\int_0^\infty [1\{y \le z\}b(z,1) + 1\{y > z\}b(z,0)]dG(z)dF(y) \\
&= \int_0^\infty a(y)\int_0^\infty [(1\{y \le z\} - F(z))b(z,1) \\
&\quad +(1\{y > z\} - 1 + F(z))b(z,0)]dG(z)dF(y) \\
&= \langle a, \int_0^\infty [(1\{y \le z\} - F(z))b(z,1) \\
&\quad +(1\{y > z\} - 1 + F(z))b(z,0)]dG(z)\rangle_{L_2(F)}
\end{aligned}
$$

となる．最後から2番目の等式は、$a \in L_2^0(F)$ に注意すれば、

$$
\int_0^\infty a(y)\int_0^\infty F(z)b(z,1)dG(z)dF(y) = \int_0^\infty F(z)b(z,1)dG(z)\int_0^\infty a(y)dF(y) = 0
$$

となることよりわかる．よって、随伴作用素 $\boldsymbol{i}^T$ は

$$
\begin{aligned}
(\boldsymbol{i}^T b)(y) &= \int_0^\infty [(1\{y \le z\} - F(z))b(z,1) \\
&\quad +(1\{y > z\} - 1 + F(z))b(z,0)]dG(z)
\end{aligned} \tag{5.22}
$$

で与えられる．また、情報作用素は

$$
\begin{aligned}
\boldsymbol{i}^T\boldsymbol{i}a &= \int_0^\infty [(1\{y \le z\} - F(z))\frac{\int_0^z a(t)dF(t)}{F(Z)} \\
&\quad +(1\{y > z\} - 1 + F(z))\frac{\int_z^\infty a(t)dF(t)}{1-F(z)}]dG(z) \\
&= \int_0^\infty a(t)K(y,t)dF(t)
\end{aligned} \tag{5.23}
$$

を得る．ただし、

$$
K(y,t) = \int_0^{y\wedge z}\frac{dG(z)}{1-F(z)} + \int_{y\vee t}^\infty \frac{dG(z)}{F(z)} - 1
$$

である．(5.22) の最後の式は等式

$$
\begin{aligned}
\int_0^\infty 1\{y \le z\}\frac{\int_0^\infty a(t)dF(t)}{F(z)}dG(z) &= \int_0^\infty\int_0^\infty 1\{y \le z\}1\{t \le z\}\frac{a(t)}{F(z)}dG(z)dF(t) \\
&= \int_0^\infty a(t)\int_{y\vee t}^\infty \frac{dG(z)}{F(z)}dF(t)
\end{aligned}
$$

と

$$\int_0^\infty \int_0^z a(t)dF(t)dG(z) + \int_0^\infty \int_z^\infty a(t)dF(t)dG(z) = \int_0^\infty a(t) \int_0^\infty dG(Z)dF(t)$$
$$= \int_0^\infty a(t)dF(t) = 0$$

に注意すればわかる．

後で観るように、(5.23) で与えられる情報作用素は有界な逆作用素をもたないことから、任意の F に対する情報限界をもたないことがわかる．

例 5.3 （Convolution モデル）: $X^0 = (Z, W) \in \boldsymbol{R}^2$ とし、Z は未知の分布 F に従うとし、W は固定した分布 G に従い、それぞれ独立とする．$Q = F \times G$ とおく．いま、$X = T(X^0) = Z + W$ を考える．このとき、モデルは

$$\boldsymbol{P} = \left\{ P_{\theta,G} = QT^{-1} : Q = F \times G,\ F \text{ と } G \text{ は } \mu \text{ と } \lambda \text{ に関し絶対連続} \right\}$$

で与えられるとする．よって、$P = P_{\theta,G}$ の確率密度関数は

$$p_{\theta,G} = \int g_\theta(x - z)dF(z) \tag{5.24}$$

となる．ただし、$g = dG/d\lambda$ である．G が既知のとき、$\dot{\boldsymbol{Q}}$ は $L_2^0(F)$ と同一視できる．このとき、

$$Z|X \sim \frac{g_\theta(x-z)}{p_{\theta,G}(x)} \frac{dF}{d\mu}$$

となることに注意すれば、このモデルに対するスコア関数は

$$\begin{aligned} (\boldsymbol{i}a)(x) &= \boldsymbol{E}[a(Z)|T(X^0) = x] \\ &= \frac{\int a(z)g_\theta(x-z)dF(z)}{p_{\theta,G}(x)} \end{aligned}$$

で与えられる．つぎに随伴作用素 $\boldsymbol{i}^T$ を求める．任意の $b \in L_2^0(P_{\theta,G})$ に対して、

$$\begin{aligned} \langle b, \boldsymbol{i}a \rangle_{L_2^0(P_{\theta,G})} &= \int b(x)(\boldsymbol{i}a)(x)dP_{\theta,G}(x) \\ &= \int b(x) \left[\frac{1}{p_{\theta,g}} \int a(z)g_\theta(x-z)dF(z) \right] dP_{\theta,G}(x) \\ &= \int a(z) \left[\int \frac{b(x)g_\theta(x-z)}{p_{\theta,G}} dP_{\theta,G}(x) \right] dF(z) \\ &= \langle a, \int b(x)g_\theta(x-z)d\mu \rangle_{L_2^0(F)} \end{aligned} \tag{5.25}$$

となるなることより、随伴作用素 $\boldsymbol{i}^T : L_2^0(P_{\theta,G}) \longrightarrow L_2^0(F)$ は

$$(\boldsymbol{i}^T b)(z) = \int b(x)g_\theta(x-z)d\mu(x)$$

で与えられることがわかる．更に、情報作用素は

$$\begin{aligned}(\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}z)(z) &= \int(\dot{\boldsymbol{l}}^T a)(x)g_\theta(x-z)d\mu(x) \\ &= \int\frac{\int a(z)g_\theta(x-z')dF(z')}{p_{\theta,G}(x)}g_{\theta,G}(x-z)d\mu(x) \\ &= \int a(z')\left[\int\frac{g_\theta(x-z')g(x-z)}{p_{\theta,G}(x)}d\mu(x)\right]dG(z')\end{aligned}$$

$$\int a(z')K(z,z')dG(z') \tag{5.26}$$

で与えられる．ただし、

$$K(z,z') = \int\frac{g_\theta(x-z)g_\theta(x-z')}{p_{\theta,G}(x)}d\mu(x)$$

である．

注意 5.1 例 5.2 と同様、任意の F に対する情報限界を求めることはできない．これは、分布関数 F の推定は $1/\sqrt{n}$ のオーダーでは不可能なことを暗示する．

例 5.4 （Has'minskii – Ibragimov モデル）: 確率変数 $Y_1,\dots,Y_n$ が

$$f_{\theta,G}(y) = \int q_\theta(y|z')dG(z')$$

に独立同一に従うものとする．更に、これとは独立に G に独立同一に従う確率変数 $Z_1,\dots,Z_n$ を観測したとする．$G \ll \lambda$ のとき、$g(z)=dG/d\lambda$ と記せば、(Y_1,Z_1) の確率密度関数は

$$p_{\theta,G}(y,z) = g(z)\int q_\theta(y|z')dG(z')$$

で与えられる．いま、

$$\boldsymbol{P} = \{g(z)q_\theta(y|z')g(z') : \theta\in\Theta, G\in\boldsymbol{G}\}$$

と

$$\boldsymbol{P}_2 = \{g(z)q_{\theta_0}(y|z')g(z') :\ \theta_0 \text{ は固定}, G\in\boldsymbol{G}\}$$

とおく．このとき、任意の1次元部分モデル $\{g_\eta\}$ に対し、

$$\frac{\partial}{\partial\eta}\log\{g_\eta(z)q_{\theta_0}(y|z')g_\eta(z')\}\Big|_{\eta=0} = \frac{\dot g(z)}{g(z)}+\frac{\dot g(z')}{g(z')}$$

を得る．したがって、

$$\begin{aligned}&\boldsymbol{E}\left[\left\{\frac{\dot g(z)}{g(z)}+\frac{\dot g(z')}{g(z')}\right\}\Big| T(X^0)=x=(z,y)\right] \\ &= \int\left\{\frac{\dot g(z)}{g(z)}+\frac{\dot g(z')}{g(z')}\right\}\frac{g(z)q_{\theta_0}(y|z')g(z')}{g(z)p_{\theta,G}(x)}d\mu(z') \\ &= \frac{\dot g(z)}{g(z)}+\frac{1}{p_{\theta,G}}\int a(z')q_{\theta_0}(y|z')g(z')d\mu(z')\end{aligned}$$

となることより、スコア作用素 $\dot{\boldsymbol{l}}_g : L_2^0(G) \longrightarrow L_2^0(P_{\theta,G})$ は

$$(\dot{\boldsymbol{l}}_g a)(y,z) = a(z) + \frac{1}{p_{\theta,G}} \int a(z')q_{\theta_0}(y|z')dG(z')$$

で与えられる．

つぎに、随伴作用素 $\dot{\boldsymbol{l}}_g^T$ を求める． $P_{\theta,G} \ll \mu \times \lambda$ とする． 任意の $b(y,z) \in L_2^0(P_{\theta,G})$ と $a(z) \in L_2^0(G)$ に対し、

$$\begin{aligned}
\langle b, \dot{\boldsymbol{l}}_g a\rangle_{L_2^0(P_{\theta,G})} &= \int b(x)(\dot{\boldsymbol{l}}_g a)(x)dP_{\theta,G} \\
&= \int b(y,z)a(z)dP_{\theta,G}(x) + \int b(y,z)\frac{\int a(z')q_{\theta_0}(y|z')dG(z')}{p_{\theta,G}(x)}dP_{\theta,G}(x) \\
&= \int a(z)b(y,z)d\mu(y)dG(z)dG(z') + \int a(z')b(y,z)q_{\theta_0}(y|z')d\mu(y)dG(z)dG(z') \\
&= \int a(z)\left[\int b(y,z)d\mu(y) + \int b(y,z')q_{\theta_0}(y|Zz)d\mu(y)\right]dG(z)dG(z') \\
&= \int a(z)\left[\int b(y,z)d\mu(y) + \int b(y,z')q_{\theta_0}(y|z')d\mu(y)dG(z')\right]dG(z) \\
&= \langle a, \int b(y,z)d\mu(y) + \int b(y,z')q_{\theta_0}(y|z)d\mu(y)dG(z')\rangle_{L_2(G)}
\end{aligned}$$

を得る．したがって、随伴作用素 $\dot{\boldsymbol{l}}_g^T : L_2^0(P_{\theta,G}) \longrightarrow L_2^0(G)$ は

$$(\dot{\boldsymbol{l}}_g^T b)(z) = \int b(y,z)d\mu(y) + \int b(y,z')q_{\theta_0}(y|z)d\mu(y)dG(z')$$

で与えられる．更に、情報作用素 $\dot{\boldsymbol{l}}_g^T\dot{\boldsymbol{l}}_g : L_2^0(G) \longrightarrow L_2^0(G)$ は

$$\begin{aligned}
(\dot{\boldsymbol{l}}_g^T\dot{\boldsymbol{l}}_g a)(z) &= \int(\dot{\boldsymbol{l}}_g a)(y,z)d\mu(y) \\
&\quad + \int\int(\dot{\boldsymbol{l}}_g a)(y,z')q_{\theta_0}(y|z)d\mu(y)dG(z') \\
&= \int\Bigg[a(z) + \frac{1}{p_{\theta,G}(y,z)}\int a(z')q_{\theta_0}(y|z')dG(z') + a(z')q_{\theta_0}(y|z) \\
&\quad + \frac{1}{p_{\theta,G}(y,z)}\int a(z'')q_{\theta_0}(y|z')q_{\theta_0}(y|z'')dG(z')dG(z'')\Bigg]d\mu(y)
\end{aligned}$$

となることがわかる．

これは有界逆作用素をもつので、$1/\sqrt{n}$ のオーダーで推定可能であることを暗示する．

5.3 Van der Vaart の可微分性定理

モデル $\boldsymbol{P} = \{P = P_{F,G}\}$ で与えられたとき、推定したい母数が

$$\nu(P_{F,G}) = \psi(F) \tag{5.27}$$

のように陰に表現されている場合を考えよう．ここで、ψ は F の平均なり Quantile $F(t_0)$ のようなものと考えればよい．

(5.27) のように陰に定義された汎関数がどのような場合に道ごとに微分可能かを調べるのがこの節の目的である．仮に ν が道ごとに微分可能であるならば、ν の推定問題に関する畳み込み定理および情報限界を後でみるように求めることができ、ν は $n^{-1/2}$ のオーダーで推定可能できる可能性を示す．一方、ν が微分可能でなければ、ν の推定は収束が $n^{-1/2}$ のオーダーよりも遅い rate の推定量のみ推定できることが予想される．

$\boldsymbol{H}$ を Hilbert 空間として、モデル $\boldsymbol{P}$ は $\boldsymbol{H}$ の部分集合を母数空間としてもつモデルとする．すなわち、

$$\boldsymbol{P} = \{P_g : g \in \boldsymbol{G} \subset \boldsymbol{H}\}$$

である．固定した元 $g \in \boldsymbol{G}$ に対して、$\boldsymbol{G}_g$ を g を通過し $\eta \in \boldsymbol{R}$ に対して、ある $h \in \boldsymbol{H}$ が存在して、

$$g_\eta = g + \eta h + o(|\eta|) \in \boldsymbol{H} \tag{5.28}$$

を満足する曲線 $\{g_\eta : \eta \in \boldsymbol{R}\}$ の集まりとする．このとき、

$$\dot{\boldsymbol{G}}^0 = \bigcup \{h :\ h \text{ はある曲線 } \{g_\eta\} \in \boldsymbol{G}_g \text{ に対応するもの }\}$$

とする．

議論を簡単にするために $\boldsymbol{P}$ のすべての元 P はある測度 μ に関して絶対連続とし、

$$p_g = \frac{dP_g}{d\mu}, \qquad s(g) = s_g = \sqrt{p_g}$$

と記すことにする．つぎのような仮定をモデルについておく．ひとつは接集合 $\dot{\boldsymbol{G}}^0$ に関するものであり、一方はモデル $\boldsymbol{P}$ を g の関数と観たときの微分可能性である．

(A1)．$\dot{\boldsymbol{G}}^0$ は $\boldsymbol{H}$ の閉かつ線形部分空間とする．これは、

$$\dot{\boldsymbol{G}} = [\overline{\dot{\boldsymbol{G}}^0}] = \dot{\boldsymbol{G}}^0$$

を意味する．

(A2)．有界線形作用素

$$\dot{\boldsymbol{l}}_g = \dot{\boldsymbol{l}} : \dot{\boldsymbol{G}}^0 \longrightarrow L_2^0(P_g)$$

が存在して、(5.28) を満足する $\boldsymbol{G}_g$ の中の任意の曲線 $\{g_\eta; \eta \in \boldsymbol{R}\} \in \boldsymbol{G}_g$ に対して、

$$s(g_\eta) - s(g) - \eta\left(\frac{1}{2}\dot{\boldsymbol{l}}h\right)s(g) = o(\eta) \tag{5.29}$$

が $L_2(\mu)$ の意味[10]で成り立つとする．

---

[10]
$$\sqrt{\int\left\{s(g_\eta) - s(g) - \eta\left(\frac{1}{2}\dot{\boldsymbol{l}}h\right)s(g)\right\}^2 d\mu} = 0(\eta)$$

(5.29) が成立するとき、モデル $\boldsymbol{P}$ は $\{P_{g_\eta}\}$ に沿って P_g において **Hellinger 微分可能**という．また、$\dot{\boldsymbol{l}}$ を g に対する**スコア作用素**とよぶ．

注意 5.2 ：$\mathrm{Image}(\dot{\boldsymbol{l}}) \subset \dot{\boldsymbol{P}}^0$ である．

つぎに推定したい母数の仮定である．$\boldsymbol{B}$ を Banach 空間とし、$\psi : \boldsymbol{G} \longrightarrow \boldsymbol{B}$ を次の意味で道ごとに微分可能とする．(A1) が成立し、有界線形作用素

$$\dot{\psi}_g \equiv \dot{\psi} : \dot{\boldsymbol{G}}^0 \longrightarrow \boldsymbol{B}$$

が存在して、(5.28) を満足する $\boldsymbol{G}_g$ の中の任意の曲線 $\{g_\eta\} \in \boldsymbol{G}_g$ に対して、

$$\left\| \eta^{-1}(\psi(g_\eta) - \psi(g)) - \dot{\psi}(h) \right\|_B = o(1)$$

が成立するとする．ただし、$\| \cdot \|_B$ は $\boldsymbol{B}$ の norm とする．

(A3) 母数 $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{G}$ は

$$\nu(P_g) = \psi(g) \tag{5.30}$$

で与えられる．

定義 5.6 ：ν が道ごとに微分可能で微分 $\dot{\nu}$ をもつとき、モデル $\boldsymbol{P}$ における ν の推定問題にたいする**有効影響作用素**

$$\tilde{\boldsymbol{l}}(P|\nu, \boldsymbol{P}) \equiv \tilde{\boldsymbol{l}}_\nu : \boldsymbol{B}^* \longrightarrow \dot{\boldsymbol{P}}$$

を

$$\tilde{\boldsymbol{l}}_\nu(b^*) \equiv \dot{\nu}^T b^* = \dot{\nu}_{b^*}$$

で定義する．

定理 5.2 （Van der Vaart ）*:* (A1) – (A3) が成立するとする．
A. もし、(5.30) で与えられる $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ が $P_g \in \boldsymbol{P}$ において微分可能ならば、

$$\mathrm{Image}(\dot{\psi}) \subset \mathrm{Image}(\dot{\boldsymbol{l}}^T) \tag{5.31}$$

が成立する．(5.31) と同値条件は

$$\mathrm{Kernel}(\dot{\boldsymbol{l}}) \subset \mathrm{Kernel}(\dot{\psi}) \tag{5.32}$$

である．ただし、$\dot{\psi}^T$ と $\dot{\boldsymbol{l}}^T$ は $\dot{\psi}$ と $\dot{\boldsymbol{l}}$ の随伴作用素である．
B. もし、(5.28) が成立し、$\dot{\boldsymbol{P}} = \overline{\mathrm{Image}(\dot{\boldsymbol{l}})}$ ならば、ν は道ごとに微分可能であり、ν を推定するときの有効影響作用素 $\tilde{\boldsymbol{l}}_\nu$ は $\boldsymbol{B}^*$ から $\dot{\boldsymbol{P}}$ への作用素で一意的に存在し、

$$\dot{\psi}^T = \dot{\boldsymbol{l}}^T \tilde{\boldsymbol{l}}_\nu \tag{5.33}$$

を満足する．

である．また、

$$\| f \|_{L_2(\mu)} = \sqrt{\int f^2 d\mu}$$

と記すことにする．

証明：ν を道ごとに微分可能と仮定する．すなわち、

$$\sqrt{p_{g_\eta}} = \sqrt{p_g} + \frac{1}{2}(\dot{\boldsymbol{l}}h)\sqrt{p_g} + o(\eta)$$

を満足するような曲線 $\{P_{g_\eta}\}$ に対し

$$\left\|\frac{\nu(P_{g_\eta}) - \nu(P_g)}{\eta} - \dot{\nu}(\dot{\boldsymbol{l}}h)\right\|_B = o(\eta)$$

を満足することである．ψ が道ごとに微分可能であることと仮定 (A2) を用いれば、

$$\begin{aligned} \dot{\nu}(\dot{\boldsymbol{l}}h) &= \lim_{\eta\to 0}\frac{\nu(P_{g_\eta}) - \nu(P_g)}{\eta} \\ &= \lim_{\eta\to 0}\frac{\psi(g_\eta) - \psi(g)}{\eta} = \dot{\psi}(h) \end{aligned} \tag{5.34}$$

が成立する．よって、連続線形写像 $\dot{\psi} : \dot{\boldsymbol{G}} \longrightarrow \boldsymbol{B}$ は

$$\dot{\psi} = \dot{\nu}\dot{\boldsymbol{l}} \tag{5.35}$$

をみたすことがわかる．ただし、$\dot{\boldsymbol{l}} : \dot{\boldsymbol{G}} \longrightarrow \dot{\boldsymbol{P}}$ と $\dot{\nu} : \dot{\boldsymbol{P}} \longrightarrow \dot{\boldsymbol{B}}$ である．定理 5.1.C より

$$\dot{\psi}^T = \dot{\boldsymbol{l}}^T \dot{\nu}^T \tag{5.36}$$

を得る．

逆に、(5.31) が成立するとする．写像 $\dot{\nu} : \mathrm{Image}(\dot{\boldsymbol{l}}) \longrightarrow \boldsymbol{B}$ をすべての $h \in \dot{\boldsymbol{G}}$ に対し

$$\dot{\nu}(\dot{\boldsymbol{l}}h) = \dot{\psi}(h)$$

で定義すると well–defined である．なぜなら、$\dot{\boldsymbol{l}}h_1 = \dot{\boldsymbol{l}}h_2$ としたとき、(5.32) より

$$h_1 - h_2 \in \mathrm{Kernel}(\dot{\boldsymbol{l}}) \subset \mathrm{Kernel}(\dot{\psi})$$

が成立することから、$\dot{\psi}(h_1) = \dot{\psi}(h_2)$ となり、$\dot{\nu}$ は well–defined なことがわかる．

$h_1, h_2 \in \dot{\boldsymbol{G}}$ と スカラー c_1, c_2 に対し

$$\dot{\nu}\{\dot{\boldsymbol{l}}(c_1h_1 + c_2h_2)\} = c_1\dot{\psi}(h_1) + c_2\dot{\psi}(h_2) = c_1\dot{\nu}(\dot{\boldsymbol{l}}h_1) + c_2\dot{\nu}(\dot{\boldsymbol{l}}h_2)$$

であることに注意すれば、$\dot{\nu}$ は $\mathrm{Image}(\dot{\boldsymbol{l}})$ 上で線形であることがわかる．

つぎに $\dot{\nu}$ の連続性を示す．(5.31) から、任意の $b^* \in \boldsymbol{B}^*$ に対して、ある $a \in L_2^0(P_g)$ が存在して、$\dot{\psi}^T b^* = \dot{\boldsymbol{l}}^T a$ を満足することに注意すれば、

$$b^*\dot{\nu}(\dot{\boldsymbol{l}}h) = \langle b^*, \dot{\psi}h\rangle_B = \langle \dot{\psi}^T b^*, h\rangle_H = \langle \dot{\boldsymbol{l}}^T a, h\rangle = \langle a, \dot{\boldsymbol{l}}h\rangle_{L_2(P_g)}$$

となり、任意の $b^* \in \boldsymbol{B}^*$ に対し、$b^*\dot{\nu}$ は $\mathrm{Image}(\dot{\boldsymbol{l}})$ 上の連続線形作用素となることがわかる．また、共鳴定理[11] より、$\dot{\nu} : \mathrm{Image}(\dot{\boldsymbol{l}}) \longrightarrow \boldsymbol{B}$ の有界性がわかる．更に、連続性から $\dot{\nu}$ は $\overline{\mathrm{Image}(\dot{\boldsymbol{l}})}$ 上の作用素へ拡張できる．よって、(5.33) は (5.36) と 定義 5.6 よりわかる．

[11] Banach–Steinhaus 定理とも呼ばれる．たとえば、関数解析（千葉克裕著、培風館）pages 114 を参照．

最後に一意性を示す．$\dot{\boldsymbol{l}}_\nu^{(1)}$ と $\dot{\boldsymbol{l}}_\nu^{(2)}$ をともに (5.33) の解と仮定し、ある $b^* \in \boldsymbol{b}$ に対し

$$d \equiv \dot{\boldsymbol{l}}_\nu^{(1)} - \dot{\boldsymbol{l}}_\nu^{(2)} \neq 0$$

とする．定義 5.6 より $d \in \dot{\boldsymbol{P}}$ となる．このとき、(5.33) を用いれば、

$$\dot{\boldsymbol{l}}^T(d) = \dot{\boldsymbol{l}}^T(\dot{\boldsymbol{l}}_\nu^{(1)}) - \dot{\boldsymbol{l}}^T(\dot{\boldsymbol{l}}_\nu^{(2)}) = \dot{\psi}(b^*) - \dot{\psi}(b^*) = 0$$

となる．このことより、任意の $h \in \dot{\boldsymbol{G}}$ に対し

$$0 = \langle \dot{\boldsymbol{l}}^T(d), h \rangle_H = \langle d, \dot{\boldsymbol{l}} h \rangle_{L_2(P_g)}$$

となるので、 d は $\mathrm{Image}(\dot{\boldsymbol{l}})$ と直交する．しかし、$\overline{\mathrm{Image}(\dot{\boldsymbol{l}})} \subset \dot{\boldsymbol{P}}$ であることより、$d = 0$ となり、一意性が示せた． □

5.4 有効スコア作用素と有効影響作用素の計算

ここでは、体系的に有効スコア作用素と有効影響作用素を導出する計算方法を考える．

$\boldsymbol{G}$ を Hilbert 空間 $(\boldsymbol{H}, \langle \cdot, \cdot \rangle_H)$ の部分集合とし、$\boldsymbol{G}$ をパラメータにもつ Hellinger 微分可能なモデル

$$\boldsymbol{P} = \{P_g : g \in \boldsymbol{G}\}$$

を考える．$\dot{\boldsymbol{l}} \equiv \dot{\boldsymbol{l}}_g : \dot{\boldsymbol{G}} \longrightarrow L_2^0(P_g)$ をスコア作用素とし、$\dot{\boldsymbol{l}}^T : L_2^0(P_g) \longrightarrow \dot{\boldsymbol{G}}$ をその随伴作用素とする．このとき、**情報作用素**を

$$\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} : \dot{\boldsymbol{G}} \longrightarrow \dot{\boldsymbol{G}}$$

で定義する．

5.4.1 可微分性定理の系

系 5.1 ：仮定 (A1) – (A3) が成り立つものとする．もし、

$$\mathrm{Kernel}(\dot{\boldsymbol{l}}) \subset \mathrm{Kernel}(\dot{\psi}) \tag{5.37}$$

かつ

$$\mathrm{Image}(\dot{\boldsymbol{l}}) \text{ は閉} \tag{5.38}$$

ならば、ν は $P_g \in \boldsymbol{P}$ において道ごとに微分可能である．

証明：定理 5.2 より、(5.37) と (5.38) より、(5.31) が成立することを示せばよいことがわかる．任意の $h \in \mathrm{Kernel}(\dot{\psi}) \subset \boldsymbol{G}$ に対し

$$\langle \dot{\psi}^T b^*, h \rangle_H = \langle b^*, \dot{\psi} h \rangle_B = 0$$

となる．よって、(5.37) より

$$\dot{\psi}^T b^* \in \{\mathrm{Kernel}(\dot{\psi})\}^T \subset \{\mathrm{Kernel}(\dot{\boldsymbol{l}})\}^T = \overline{\mathrm{Image}(\dot{\boldsymbol{l}}^T)} = \mathrm{Image}(\dot{\boldsymbol{l}}^T)$$

が成立する．最後から2番目の等式は Image($\dot{\boldsymbol{l}}$) が得非であるための必要十分条件は Image$\dot{\boldsymbol{l}}^T$) が閉であることからわかる．よって、(5.31) が成立する． □

系 5.2 ：仮定 (A1) − (A3) が成立するとする．
A. もし、$\overline{\mathrm{Image}(\dot{\boldsymbol{l}})} = \dot{\boldsymbol{P}}$ かつ

$$\mathrm{Kernel}(\dot{\boldsymbol{l}}) = \{0\} \tag{5.39}$$

で、(5.38) が成立するならば、$\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ は道ごとに微分可能である．
B. $\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$ が1対1で上への写像であるための必要十分条件は (5.38) と (5.39) が成立することである．
C. Image($\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$) ⊂ Image($\dot{\boldsymbol{l}}^T$) である．さらに、上の式で等号が成立するための必要十分条件は (5.38) が成立することである．

証明：はじめに A を示す．(5.39) が成立するならば、(5.37) が成立するので、系 5.1 よりわかる．

つぎに、C を示す．Image($\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$) ⊂ Image($\dot{\boldsymbol{l}}^T$) は明らかなので、(5.39) のときに、逆の包含関係も成立することを示せばよい．Image($\dot{\boldsymbol{l}}$) は閉なので、$h = \dot{\boldsymbol{l}}\alpha \in \mathrm{Image}(\dot{\boldsymbol{l}}^T)$ と仮定する．このとき、ある $h_1 \in \dot{\boldsymbol{G}}$ が存在して、$h = (\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}})h_1$ と書ける[12]．したがって、$h \in \mathrm{Image}(\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}})$ である．

逆に、(5.39) において等号が成立するとする．$\alpha \in \overline{\mathrm{Image}(\dot{\boldsymbol{l}})}$ とする．$\mathrm{Image}(\dot{\boldsymbol{l}}^T) = \mathrm{Image}(\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}})$ より、ある $h_0 \in \dot{\boldsymbol{G}}$ が存在して、$\dot{\boldsymbol{l}}^T\alpha = \dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}h_0$ と書ける．したがって、$\dot{\boldsymbol{l}}^T(\alpha - \dot{\boldsymbol{l}}h_0) = 0$ より、$\alpha - \dot{\boldsymbol{l}}h_0 \in \mathrm{Kernel}(\dot{\boldsymbol{l}}^T) = \mathrm{Image}(\dot{\boldsymbol{l}})^\perp$ となる．一方、$\alpha - \dot{\boldsymbol{l}}h_0 \in \overline{\mathrm{Image}(\dot{\boldsymbol{l}})}$ なので、$\alpha - \dot{\boldsymbol{l}}h_0 = 0$ となり、$\alpha \in \mathrm{Image}(\dot{\boldsymbol{l}})$ が成立する．したがって、Image($\dot{\boldsymbol{l}}$) は閉である．

最後に、B を示す．まず、$\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$ が1対1上への写像ならば、(5.38) と (5.39) が成立することを示す．仮定より、$\dot{\boldsymbol{l}}^T$ は上への写像なので、$\mathrm{Image}(\dot{\boldsymbol{l}}^T) = \dot{\boldsymbol{G}}^0$ となり、仮定 (A.1) より閉集合となる．したがって、Image($\dot{\boldsymbol{l}}$) も閉集合になる[13]．したがって、(5.38) が成立する．また、$\dot{\boldsymbol{l}}$ は1対1より、$\mathrm{Kernel}(\dot{\boldsymbol{l}}) = \{0\}$ となるので、(5.39) が成立する．

逆に、(5.38) と (5.39) が成立すると仮定する． (5.38) より、$\mathrm{Image}(\dot{\boldsymbol{l}}^T) = \overline{\mathrm{Image}(\dot{\boldsymbol{l}}^T)} = \mathrm{Kernel}(\dot{\boldsymbol{l}})^\perp = \{0\}$ より、$\dot{\boldsymbol{l}}^T$ は上への写像であることがわかる．さらに、C より $\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$ も上へ写像である．$\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$ が1対1を示すために、$\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}h = 0$ とすれば、$\dot{\boldsymbol{l}}h \in \mathrm{Kernel}(\dot{\boldsymbol{l}}^T) = \mathrm{Image}(\dot{\boldsymbol{l}})^\perp$ より、$h \in \mathrm{Image}(\dot{\boldsymbol{l}}) \cap \mathrm{Image}(\dot{\boldsymbol{l}})^\perp = \{0\}$ なので、$\dot{\boldsymbol{l}}h = 0$ となる．さらに、(5.39) から $h \in \mathrm{Kernel}(\dot{\boldsymbol{l}}) = \{0\}$ となるので、$\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}$ は1対1であることがわかる．

□

[12] 実際、$\mathrm{Image}(\dot{\boldsymbol{l}})^\perp = \mathrm{Kernel}(\dot{\boldsymbol{l}}^T)$ を使えば、

$$h = \dot{\boldsymbol{l}}^T\alpha = \dot{\boldsymbol{l}}^T\{\Pi(\alpha|\mathrm{Image}(\dot{\boldsymbol{l}})) + \Pi(\alpha|\mathrm{Image}(\dot{\boldsymbol{l}})^\perp)\} = \dot{\boldsymbol{l}}^T(\dot{\boldsymbol{l}}h_1 + \Pi(\alpha|\mathrm{Kernel}(\dot{\boldsymbol{l}}^T))) = \dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}}h_1$$

よりわかる．

[13] BKRW の pages 419 の命題 A.1.7.D を参照のこと．

5.4.2 母数モデルにおける有効影響関数

ここでは、$\boldsymbol{G} \subset \boldsymbol{R}^k$ と考える．第３章の記号と対応させるために $\boldsymbol{G} = \Theta, g = \theta$ とおく．$\boldsymbol{P}$ を正則母数モデルとし、

$$\boldsymbol{P} = \{P_\theta \ll \mu; \theta \in \Theta\}$$

と記す．

$\theta_0 \in \Theta$ を固定する．このとき、$\dot{\boldsymbol{G}}^0 = \dot{\boldsymbol{G}} = \boldsymbol{R}^k$ で $\dot{\boldsymbol{P}} = [\dot{l}_1, \ldots \dot{l}_k]$ となる．ただし、

$$\dot{l}_i = \frac{\partial}{\partial \theta_i} \log \frac{dP_\theta}{d\mu}\bigg|_{\theta=\theta_0}, \quad i = 1, \ldots, k$$

である．

いま、$\underline{\dot{\boldsymbol{l}}} = (\dot{l}_1, \ldots, \dot{l}_k)$ と記すこと[14] にする．スコア作用素

$$\dot{\boldsymbol{l}} : \dot{\boldsymbol{G}} \longrightarrow \dot{\boldsymbol{P}} \subset L_2^0(P_{\theta_0})$$

は、$\boldsymbol{t} = (t_1, \ldots, t_k)' \in \dot{\boldsymbol{G}} = \boldsymbol{R}^k$ に対し

$$\dot{\boldsymbol{l}}(\boldsymbol{t}) = \underline{\dot{\boldsymbol{l}}}\boldsymbol{t} = \sum_{i=1}^{k} t_i \dot{l}_i$$

で定義される．また、随伴作用素

$$\dot{\boldsymbol{l}}^T : L_2^0(P_{\theta_0}) \longrightarrow \dot{\boldsymbol{G}} = \boldsymbol{R}^k$$

は、任意の $h \in L_2^0(P_{\theta_0})$ に対し

$$\dot{\boldsymbol{l}}^T h = (\langle \dot{l}_1, h \rangle_{L_2(P_{\theta_0})}, \ldots, \langle \dot{l}_k, h \rangle_{L_2(P_{\theta_0})})' = \boldsymbol{E}[\underline{\dot{\boldsymbol{l}}}' h] \in \boldsymbol{R}k$$

で定義する．更に、情報作用素

$$\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} : \dot{\boldsymbol{G}} \longrightarrow \dot{\boldsymbol{G}}$$

は、任意の $\boldsymbol{t} \in \dot{\boldsymbol{G}} = \boldsymbol{R}^k$ に対し

$$\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}}(\boldsymbol{t}) = \boldsymbol{E}[\underline{\dot{\boldsymbol{l}}}' \underline{\dot{\boldsymbol{l}}} \boldsymbol{t}] = I(\theta_0)\boldsymbol{t}$$

となることがわかる．また、有効影響関数作用素

$$\dot{\boldsymbol{l}}_g : \dot{\boldsymbol{G}} \longrightarrow \dot{\boldsymbol{P}}$$

は

$$\dot{\boldsymbol{l}}_g(\boldsymbol{t}) = \dot{\boldsymbol{l}}(\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}})\boldsymbol{t} = \underline{\dot{\boldsymbol{l}}}\{\boldsymbol{E}_{P_{\theta_0}}(\underline{\dot{\boldsymbol{l}}}' \underline{\dot{\boldsymbol{l}}})\}^{-1}\boldsymbol{t} = \underline{\dot{\boldsymbol{l}}} I(\theta_0)\boldsymbol{t} = \underline{\tilde{\boldsymbol{l}}}\boldsymbol{t}$$

で表現される．ただし、$\underline{\tilde{\boldsymbol{l}}} = \underline{\dot{\boldsymbol{l}}} I(\theta_0)\boldsymbol{t}$ は θ_0 における有効影響関数である．

---

[14] 随伴作用素で転置の記号をもちいるので、混乱をさけるためにベクトルの転置の記号は $\underline{\dot{\boldsymbol{l}}}$ に対して用いない

5.4.3　直積で表現される母数空間をもつモデル

$(\boldsymbol{H}_i, \langle\cdot,\cdot\rangle_{H_i}), i=1,2$ を Hilbert 空間とし、$\boldsymbol{G}_i$ を $\boldsymbol{H}_i$ の部分空間とする．$\boldsymbol{G}=\boldsymbol{G}_1\times\boldsymbol{G}_2$ を母数空間し、モデル

$$\boldsymbol{P}=\left\{P_g \ll \mu : g=\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, g_i\in\boldsymbol{G}_i, i=1,2\right\}$$

を考える．ここで、 $\boldsymbol{G}_1$ と $\boldsymbol{G}_2$ は無限次元空間であってもよい．いま、 $\eta\to 0$ のとき、ある $h_i\in\boldsymbol{H}_i(i=1,2)$ が存在して、

$$g_i(\eta)=g_i+h_i\eta+o(\eta), \qquad i=1,2$$

となるような曲線 $\{g_i(\eta):\eta\in\boldsymbol{R}\}$ に対し

$$s(g_1,g_2)=\sqrt{\frac{dP_g}{d\mu}}$$

とおいたとき、ある有界線形作用素

$$\dot{\boldsymbol{l}}_i:\dot{\boldsymbol{G}}_i\longrightarrow L_2^0(P_g), \qquad i=1,2$$

が存在して、$L_2(\mu)$ の意味において

$$s(g_1(\eta),g_2(\eta))-s(g_1,g_2)-\frac{1}{2}(\dot{\boldsymbol{l}}_1h_1+\dot{\boldsymbol{l}}_2h_2)s(g_1,g_2)=o(\eta)$$

を満足するとする．このとき、

$$\dot{\boldsymbol{G}}=\dot{\boldsymbol{G}}_1\times\dot{\boldsymbol{G}}_2$$

として、有界線形作用素 $\dot{\boldsymbol{l}}:\dot{\boldsymbol{G}}\longrightarrow L_2^0(P_g)$ を

$$\dot{\boldsymbol{l}}\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}=\dot{\boldsymbol{l}}_1h_1+\dot{\boldsymbol{l}}_2h_2$$

で定義する．また、任意の $a\in L_2^0(P_g)$ に対し

$$\begin{aligned}
\langle h,\dot{\boldsymbol{l}}^Ta\rangle_H &= \langle\dot{\boldsymbol{l}}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix},a\rangle_{L_2(P_g)}=\langle\dot{\boldsymbol{l}}_1h_1+\dot{\boldsymbol{l}}_2h_2,a\rangle_{L_2(P_g)} \\
&= \langle\dot{\boldsymbol{l}}_1h_1,a\rangle_{L_2(P_g)}+\langle\dot{\boldsymbol{l}}_2h_2,a\rangle_{L_2(P_g)}=\langle h_1,\dot{\boldsymbol{l}}^Ta\rangle_{H_1}+\langle h_2,\dot{\boldsymbol{l}}^Ta\rangle_{H_2} \\
&= \langle\begin{bmatrix} h_1 \\ h_2 \end{bmatrix},\begin{bmatrix} \dot{\boldsymbol{l}}^Ta \\ \dot{\boldsymbol{l}}^Ta \end{bmatrix}\rangle_H=\langle h,\begin{bmatrix} \dot{\boldsymbol{l}}^Ta \\ \dot{\boldsymbol{l}}^Ta \end{bmatrix}\rangle_H
\end{aligned}$$

となることに注意すれば、随伴作用素 $\dot{\boldsymbol{l}}^T:L_2^0(P_g)\longrightarrow\dot{\boldsymbol{G}}$ は

$$\dot{\boldsymbol{l}}^Ta=\begin{bmatrix} \dot{\boldsymbol{l}}^Ta \\ \dot{\boldsymbol{l}}^Ta \end{bmatrix}\in\dot{\boldsymbol{G}}_1\times\dot{\boldsymbol{G}}_2$$

で定義される[15]．更に、

$$\boldsymbol{P}_1 = \{P_g : g_2 \in \boldsymbol{G}_2 \text{ は固定}\}, \qquad \boldsymbol{P}_2 = \{P_g : g_1 \in \boldsymbol{G}_2 \text{ は固定}\}$$

としたとき、

$$\boldsymbol{l}_1^* = (I - \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2))\dot{\boldsymbol{l}}_1, \qquad \boldsymbol{l}_2^* = (I - \Pi_0(\dot{\boldsymbol{l}}_2|\dot{\boldsymbol{P}}_1))\dot{\boldsymbol{l}}_2$$

で定義する．

系 5.3 ：つぎを仮定する．
(A1). $\boldsymbol{G}^0$ は $\boldsymbol{H}$ の閉かつ線形部分空間とする．すなわち、$\dot{\boldsymbol{G}} = \dot{\boldsymbol{G}}^0$ である．
(A2). 有界線形作用素 $\dot{\boldsymbol{l}} : \dot{\boldsymbol{G}} \longrightarrow L_2^0(P_g)$ が存在して、$\eta \to 0$ としたとき、$h_i \in \boldsymbol{H}_i (i = 1, 2)$ が存在し、

$$g_i(\eta) = g_i + \eta h_i + o(\eta), \qquad i = 1, 2 \tag{5.40}$$

を満足する $\boldsymbol{G}$ のなかの曲線

$$\{g_\eta : \eta \in \boldsymbol{R}\} = \{(g_1(\eta), g_2(\eta)) : \eta \in \boldsymbol{R}\}$$

に対し、$L_2(\mu)$ の意味で

$$s(g_\eta) - s(g) - \eta \frac{1}{2} \dot{\boldsymbol{l}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} s(g) = o(\eta) \tag{5.41}$$

を満足し、

$$\dot{\boldsymbol{l}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \dot{\boldsymbol{l}}_1 h_1 + \dot{\boldsymbol{l}}_2 h_2$$

と表現できると仮定する．
　更に、

$$\overline{\mathrm{Image}(\dot{\boldsymbol{l}})} = \dot{\boldsymbol{P}}, \qquad \overline{\mathrm{Image}(\dot{\boldsymbol{l}}_2)} = \dot{\boldsymbol{P}}_2 \tag{5.42}$$

を仮定する．
　いま、

$$\nu(P_g) = \chi(g_1)$$

とし、$\chi : \dot{\boldsymbol{G}} \longrightarrow \boldsymbol{B}$ は道ごとに微分可能[16] とする．このとき、

$$\mathrm{Image}(\dot{\chi}^T) \subset \mathrm{Image}(\dot{\boldsymbol{l}}^T) \tag{5.43}$$

[15] ここで、$\boldsymbol{H}_1 \times \boldsymbol{H}_2$ を、内積 $\langle \cdot, \cdot \rangle_{H_1 \times H_2}$ が任意の $g = (g_1, g_2), g' = (g_1', g_2') \in \boldsymbol{H}_1 \times \boldsymbol{H}_2$ に対し、

$$\langle g, g' \rangle_{H_1 \times H_2} = \langle g_1, g_1' \rangle_{H_1} + \langle g_2, g_2' \rangle_{H_2}$$

で定義される Hilbert 空間と考える

[16] すなわち、(5.40) を満足する任意の曲線 $\{g_1(\eta) : \eta \in \boldsymbol{R}\}$ に対し、

$$\left\| \frac{1}{\eta}(\chi(s_1(\eta)) - \chi(s_1)) - \dot{\chi}(h_1) \right\|_B = o(\eta)$$

を満足することである．

ならば、$\nu(P_g)$ は道ごとに微分可能である．この場合、ν に対する有効影響作用素 $\tilde{\boldsymbol{l}}_\nu$ は一意的に存在し、任意の $b^* \in \boldsymbol{B}^*$ に対し方程式

$$\dot{\chi}^T b^* = \boldsymbol{l}_1^* \tilde{\boldsymbol{l}}_\nu b^*, \qquad 0 = \boldsymbol{l}_2^* \tilde{\boldsymbol{l}}_\nu b^* \tag{5.44}$$

を満足する．

証明：いま、$\psi : \boldsymbol{G} \longrightarrow \boldsymbol{B}$ を

$$\psi \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \chi(g_1)$$

とすれば、$\dot{\psi} : \dot{\boldsymbol{G}} \longrightarrow \boldsymbol{B}$ が

$$\dot{\psi} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \dot{\chi}(h_1)$$

となるので、任意の $b^* \in \boldsymbol{B}^*$ に対し

$$\dot{\psi}^T b^* = \begin{bmatrix} \dot{\chi}^T b^* \\ 0 \end{bmatrix} \in \dot{\boldsymbol{G}}_1 \times \dot{\boldsymbol{G}}_2$$

となる．なぜならば、$b^* \in \boldsymbol{B}^*$ と $h = (h_1, h_2)^T \in \boldsymbol{H}_1 \times \boldsymbol{H}_2$ に対し、

$$\begin{aligned} \langle h, \dot{\psi}^T b^* \rangle_{H_1 \times H_2} &= \langle \dot{\psi} h, b^* \rangle_B = \langle \dot{\chi}(h_1), b^* \rangle_B = \langle h_1, \dot{\chi}^T b^* \rangle_{H_1} \\ &= \langle \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} \dot{\chi} b^* \\ 0 \end{bmatrix} \rangle_{H_1 \times H_2} = \langle h, \begin{bmatrix} \dot{\chi}^T b^* \\ 0 \end{bmatrix} \rangle_{H_1 \times H_2} \end{aligned}$$

となることよりわかる．よって、

$$\mathrm{Image}(\dot{\psi}^T) = \mathrm{Image}(\dot{\chi}^T) \times \{0\} \tag{5.45}$$

となる．一方、(5.42) から

$$\dot{\boldsymbol{P}}_2^\perp = \{\overline{\mathrm{Image}(\dot{\boldsymbol{l}}_2)}\}^\perp = \mathrm{Kernel}(\dot{\boldsymbol{l}}_2^T)$$

より、$a \in \dot{\boldsymbol{P}}_2^\perp$ ならば、$\dot{\boldsymbol{l}}_2^T a = 0$ になることに注意して、

$$\begin{aligned} \mathrm{Image}(\dot{\boldsymbol{l}}^T) &\supset \left\{ \begin{bmatrix} \dot{\boldsymbol{l}}_2^T a \\ 0 \end{bmatrix} : a \in \dot{\boldsymbol{P}}_2^\perp \right\} \\ &= \left\{ \begin{bmatrix} \dot{\boldsymbol{l}}_1 (I - \Pi_0(\cdot | \dot{\boldsymbol{P}}_2) a \\ 0 \end{bmatrix} : s \in \dot{\boldsymbol{P}}_2^* \perp \right\} \\ &= \mathrm{Image}([\boldsymbol{l}_1^*]^T) \times \{0\} \end{aligned} \tag{5.46}$$

となる．ただし、

$$(\boldsymbol{l}_1^*)^T = (\boldsymbol{l}_1^T (I - \Pi_0(\dot{\boldsymbol{l}} | \dot{\boldsymbol{P}}_2))$$

である．これは、任意の $h_1 \in \boldsymbol{H}_1$ と $a \in L_2^0(P_g)$ に対し

$$
\begin{aligned}
\langle h_1, (\boldsymbol{l}_1^*)^T a\rangle_{H_1} &= \langle \boldsymbol{l}_1^* h_1, a\rangle_{L_2(P_g)} = \langle (I - \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)\dot{\boldsymbol{l}}_1 h_1, a\rangle_{L_2(P_g)} \\
&= \langle \dot{\boldsymbol{l}} h_1, (I - \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)a)\rangle_{L_2(P_g)} = \langle h_1, \dot{\boldsymbol{l}}_1^T (I - \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2))\rangle_{H_1}
\end{aligned}
$$

よりわかる．しかし、(5.45)、(5.46)、(5.43) より

$$
\mathrm{Image}(\dot{\psi}^T) \subset \mathrm{Image}(\dot{\chi}^T) \times \{0\} \subset \mathrm{Image}[(\boldsymbol{l}_1^*)^T] \times \{0\} \subset \mathrm{Image}(\dot{\boldsymbol{l}}^T)
$$

となることより、定理 5.2 を用いいれば、$\nu(\cdot)$ は道ごとに微分可能であることがわかる．(5.44) についても (5.33) から

$$
\dot{\psi}^T b^* = \dot{\boldsymbol{l}}^T \tilde{\boldsymbol{l}}_\nu b^* = \begin{bmatrix} \dot{\boldsymbol{l}}_1^T \tilde{\boldsymbol{l}}_\nu b^* \\ \dot{\boldsymbol{l}}_2^T \tilde{\boldsymbol{l}}_\nu b^* \end{bmatrix}
$$

であることと

$$
\dot{\boldsymbol{l}}_1^T \tilde{\boldsymbol{l}}_\nu b^* = (\boldsymbol{l}_1^*)^T \tilde{\boldsymbol{l}}_\nu b^* \tag{5.47}
$$

からわかる．(5.47) については $\tilde{\boldsymbol{l}}_\nu^T = \dot{\nu}^T$ に注意すれば、任意の $a \in \dot{\boldsymbol{P}}_2 \subset L_2^0(P_g)$ に対し

$$
\langle \tilde{\boldsymbol{l}}_\nu b^*, a\rangle_{L_2(P_g)} = \langle b^*, \dot{\nu} a\rangle_B = 0
$$

となり、$\nu(P_g) = \chi(g_1)$ より、$\boldsymbol{P}_2$ の中の曲線に対して $\nu(\cdot)$ の値あｈ変化しないので、$a \in \dot{\boldsymbol{P}}_2$ に対して、$\dot{\nu} a = 0$ からわかる． □

つぎに、$\boldsymbol{G}_1 \equiv \Theta \subset \boldsymbol{R}^k$ を仮定したモデルについて考える．すなわち、

$$
\boldsymbol{P} = \left\{ P_g \ll \mu : g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, g_i \in \boldsymbol{G}_i \subset \boldsymbol{H}_i, i = 1, 2 \right\}
$$

である．ただし、$\boldsymbol{G}_1 = \Theta \subset \boldsymbol{R}^k$ 、$\boldsymbol{G}_2 \subset \boldsymbol{H}_2$ は無限次元とする．更に、(5.41) におてい、$h_1 \in \boldsymbol{R}^k$ に対し

$$
\dot{\boldsymbol{l}}_1 h_1 = \underline{\dot{\boldsymbol{l}}}_1 h_1
$$

が成立するものとする．ただし、

$$
\underline{\dot{\boldsymbol{l}}}_1 = (\dot{l}_{11}, \ldots, \dot{l}_{1k}), \qquad l_{ij} = \frac{\partial}{\partial \theta_j} \log \frac{dP_g}{d\mu}
$$

である．この場合に、おおくのモデルにおいて

$$
\overline{\mathrm{Image}(\dot{\boldsymbol{l}})} = \dot{\boldsymbol{P}}, \qquad \dot{\boldsymbol{P}} = \dot{\boldsymbol{P}}_1 + \dot{\boldsymbol{P}}_2 = [\dot{\boldsymbol{l}}_1] + \overline{\mathrm{Image}(\dot{\boldsymbol{l}}_2)} \tag{5.48}
$$

が成立する．これより、$\dot{\boldsymbol{G}} = \dot{\boldsymbol{G}}_1 \times \dot{\boldsymbol{G}}_2 = \boldsymbol{R}^k \times \dot{\boldsymbol{G}}_2$ に対し、

$$
\dot{\boldsymbol{l}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \underline{\dot{\boldsymbol{l}}}_1 h_1 + \dot{\boldsymbol{l}}_2 h_2
$$

がうまく対応する．また、随伴作用素 $\dot{\boldsymbol{l}}^T : L_2^0(P_g) \longrightarrow \boldsymbol{R}^k \times \boldsymbol{H}_2$ が、任意の $a \in L_2^0(P_g)$ に対し、

$$\dot{\boldsymbol{l}}^T a = \begin{bmatrix} \dot{\boldsymbol{l}}_1^T a \\ \dot{\boldsymbol{l}}_2^T a \end{bmatrix} \in \dot{\boldsymbol{G}}_1 \times \dot{\boldsymbol{G}}_2 \subset \boldsymbol{R}^k \times \boldsymbol{H}_2$$

で定義されることがわかる．更に、$g_1 = \theta$ に対する有効影響関数 $\boldsymbol{l}_1^*$ を

$$\boldsymbol{l}_1^* = \dot{\boldsymbol{l}}_1 - \Pi_0(\dot{\boldsymbol{l}}_1 | \dot{\boldsymbol{P}}_2)$$

で自然に定義できる．ただし、

$$\Pi_0(\dot{\boldsymbol{l}}_1 | \dot{\boldsymbol{P}}_2) = (\Pi_0(\dot{l}_{11} | \dot{\boldsymbol{P}}_2), \ldots, \Pi_0(\dot{l}_{1k} | \dot{\boldsymbol{P}}_2))$$

である．このとき、$g_1 = \theta$ を推定するときの有効情報行列は

$$I_*(\theta) = \boldsymbol{E}_{P_g}[(\boldsymbol{l}_1^*)^T \boldsymbol{l}_1^*] : k \times k$$

で与えられる．

系 5.4 ：系 *5.1* の条件が成立すると仮定する．更に、有効情報行列 $I_*(\theta)$ は正則とする．

$$\nu(P_g) = g_1 = \theta \in \boldsymbol{R}^k$$

のとき、$\nu(P_g)$ は $g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$ において、道ごとに微分可能である． また、有効影響関数 $\tilde{\boldsymbol{l}}_\nu$: $\boldsymbol{R}^k \longrightarrow \dot{\boldsymbol{P}}$ は一意的に存在し、任意の $h_1 \in \dot{\boldsymbol{G}}_1 = \boldsymbol{R}^k$ に対し

$$\tilde{\boldsymbol{l}}_\nu h_1 = \boldsymbol{l}_1^* I_*(\theta)^{-1} h_1$$

を満足する．

証明：この場合、$\boldsymbol{B} = \boldsymbol{B}^* = \boldsymbol{R}^k$ で

$$\psi \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = g_1 = \theta$$

から、$h_1 \in \boldsymbol{R}^k$ に対し

$$\dot{\psi} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = h_1$$

である．また、$h_1 \in \boldsymbol{B}^* = \boldsymbol{R}^k$ より、$\dot{\psi}^T : \boldsymbol{R}^k \longrightarrow \dot{\boldsymbol{G}}_1 \times \dot{\boldsymbol{G}}_2$ は

$$\dot{\psi} h_1 = \begin{bmatrix} h_1 \\ 0 \end{bmatrix} \in \dot{\boldsymbol{G}}_1 \times \dot{\boldsymbol{G}}_2$$

で与えられる．$I_*(\theta)$ が正則であることより、

$$\begin{aligned}
\mathrm{Image}(\dot{\boldsymbol{l}}^T) &\supset \mathrm{Image}(\dot{\boldsymbol{l}}_1^T) \times \{0\} \\
&= \{\dot{\boldsymbol{l}}_1^T(I - \Pi_0(\cdot|\dot{\boldsymbol{P}}_2))a : a \in \dot{\boldsymbol{P}}_2^{\perp}\} \times \{0\} \\
&\supset \{\dot{\boldsymbol{l}}_1(I - \Pi_0(\cdot|\dot{\boldsymbol{P}}_2))(I - \Pi_0(\cdot|\dot{\boldsymbol{P}}_2))\dot{\boldsymbol{l}}_1 h_1 : h_1 \in \boldsymbol{R}^k\} \times \{0\} \\
&= \{I_*(\theta)h_1 : h_1 \in \boldsymbol{R}^k\} \times \{0\} \\
&= \boldsymbol{R}^k \times \{0\} = \mathrm{Image}(\dot{\psi}^T)
\end{aligned}$$

となるので、定理 5.2 より、$\nu(\cdot)$ は道ごとに微分可能である．更に、(5.44) を適用すれば

$$h_1 = \dot{\psi}(h_1) = (\boldsymbol{l}_1^*)^T \tilde{\boldsymbol{l}}_\nu(h_1) \tag{5.49}$$

がわかる．また、$h_1, \tilde{h}_1 \in \boldsymbol{R}^k$ に対し

$$\begin{aligned}
\langle h_1 - (\boldsymbol{l}_1^*)^T \boldsymbol{l}_1^* I_*(\theta)^{-1} h_1, \tilde{h}_1 \rangle_{H_1} &= \langle h_1, \tilde{h}_1 \rangle_{H_1} - h_1'(\boldsymbol{E}_{P_g}[(\boldsymbol{l}_1^*)^T \boldsymbol{l}_1^T]) I_*^{-1}(\theta)\tilde{h}_1 \\
&= \langle h_1, \tilde{h}_1 \rangle_{H_1} - \langle h_1, \tilde{h}_1 \rangle_{H_1} = 0
\end{aligned}$$

より

$$h_1 = (\boldsymbol{l}_1^*)^T \boldsymbol{l}_1^* I_*^{-1}(\theta) h_1$$

となり、$\boldsymbol{l}_1^* I_*^{-1}(\theta)$ は方程式 (5.49) の解になることより

$$\tilde{\boldsymbol{l}}_\nu h_1 = \boldsymbol{l}_1^* I_*^{-1}(\theta) h_1$$

となる． □

5.4.4 可微分性定理の応用例

例 5.5 （例 *5.2* の続き）: 固定した点 t_0 における分布関数 $F(t_0)$ の推定を考える．すなわち、

$$\nu(P_{F,G}) = \psi(\sqrt{f}) = \int_0^{t_0} (\sqrt{f})^2 d\mu = F(t_0)$$

である．ただし、$f = dF/d\mu$ とした．モデルは

$$\boldsymbol{P} = \{P_{F,G} \ll \mu \times \lambda : G(\ll \lambda)\,\text{は固定}, F \ll \mu\}$$

であるから、$\boldsymbol{G} = \boldsymbol{P}$ となり、Hellinger 微分可能であることより、任意の曲線 $\{f_\eta : \eta \in \boldsymbol{R}\}$ に対して[17]、

$$\sqrt{f_\eta} = \sqrt{f} + \eta \frac{1}{2} h \sqrt{f} + o(\eta)$$

[17] 正確には、曲線 $\{f_\eta \times g : \eta \in \boldsymbol{R}\}$ に対し、$L_2(\mu \times \lambda)$ の意味で、

$$\sqrt{f_\eta \times g} = \sqrt{fg} + \eta \frac{1}{2} h \sqrt{fg} + o(\eta)$$

が成立することである．

が成立する．ただし、$h \in L_2(F)$ である．よって、

$$
\begin{aligned}
&\frac{\psi(\sqrt{f_\eta})-\psi(\sqrt{f})}{\eta}-\int_0^{t_0} h(\sqrt{f})^2 d\mu \\
&= \frac{1}{\eta}\Big[\int_0^{t_0}(\sqrt{f_\eta})^2 d\mu-\int_0^{t_0}(\sqrt{f})^2 d\mu-2\eta\int_0^{t_0}\frac{1}{2}h(\sqrt{f})^2 d\mu\Big] \\
&= \frac{1}{2}\int_0^{t_0} 2\{\sqrt{f_\eta}-\sqrt{f}-\eta\frac{1}{2}h\sqrt{f}\}\sqrt{f}d\mu+\eta\int_0^{t_0}\left\{\frac{\sqrt{f_\eta}-\sqrt{f}}{\eta}\right\}^2 d\mu \\
&\le 2\left\|\frac{\sqrt{f_\eta}-\sqrt{f}}{\eta}-\eta\frac{1}{2}\sqrt{f}\right\|_{L_2(\mu)}+\eta\left\|\frac{1}{2}h\sqrt{f}\right\|_{L_2(\mu)}+o(1)\to 0 \quad (\eta\to 0)
\end{aligned}
$$

となる．更に、$\int h(z)dF(z)=0$ より、

$$
\begin{aligned}
\int_0^{t_0} h(\sqrt{f})^2 d\mu &= \int 1_{[0,t_0)}(z)h(z)dF(z)-F(t_0)\int h(z)dF(z) \\
&= \int[1_{[0,t_0)}(z)-F(t_0)]h(z)dF(z)
\end{aligned}
$$

になることに注意すれば、

$$
\dot{\psi}h=\int_0^{\infty}[1_{[0,t_0)}(z)-F(t_0)]h(z)dF(z)
$$

となる．更に、$\dot{\psi}:\dot{\boldsymbol{P}}\longrightarrow\boldsymbol{B}=\boldsymbol{R}$ と $\boldsymbol{B}^*=\boldsymbol{R}$ であることに注意する．任意の $b^*\in\boldsymbol{R}$ に対し

$$
\begin{aligned}
\langle b^*,\dot{\psi}h\rangle_B &= b^*\int_0^{\infty}[1_{[0,t_0)}(z)-F(t_0)]h(z)dF(z) \\
&= \langle b^*[1_{[0,t_0)}(\cdot)-F(t_0),h\rangle_{L_2(F)}
\end{aligned}
$$

であることより、$\dot{\psi}^T:\boldsymbol{R}\longrightarrow\dot{\boldsymbol{P}}$ は

$$
(\dot{\psi}^T b^*)(z)=b^*[1_{[0,t_0)}(z)-F(t_0)]
$$

で与えられる．よって、任意の b^* に対し、$\dot{\psi}^T b^*$ は z の不連続関数になる．

一方、例 5.2 の (5.22) より $(\dot{\boldsymbol{l}})b(z)$ は連続関数となる．よって、$\nu(P_{F,G})=F(t_0)$ は道ごとに微分可能でないことがわかる．

つぎに、固定した汎関数

$$
\nu(P_{F,G})=\psi(\sqrt{f})=\int_0^M zdF(z) \qquad 0<M<\infty
$$

を考える．ただし、 F の台は有界閉集合 $[0,M]$ に含まれると仮定[18] する．$h\in L_2^0(F)$ が存在して、

$$
\sqrt{f_\eta}=\sqrt{f}+\frac{1}{2}\eta h\sqrt{f}+o(\eta)
$$

[18] この仮定は証明の上で技術的に必要となる．

を満足する $\boldsymbol{G}=\boldsymbol{P}$ の中の曲線 $\{\sqrt{f_\eta}:\eta\in\boldsymbol{R}\}$ に対し、

$$\begin{aligned}
&\frac{\psi(\sqrt{f_\eta})-\psi(\sqrt{f})}{\eta}-\int_0^M h(z)z(\sqrt{f})^2d\mu\\
&=\frac{1}{\eta}\left[\int_0^M z(\sqrt{f_\eta})^2d\mu-\int_0^M z(\sqrt{f})^2d\mu-2\eta\int_0^M z\frac{1}{2}h(z)(\sqrt{f})^2d\mu\right]\\
&=\frac{1}{\eta}\int_0^M 2\{\sqrt{f_\eta}-\sqrt{f}-\eta\frac{1}{2}h\sqrt{f}\}z\sqrt{f}d\mu+\eta\int_0^M\left\{\frac{\sqrt{f_\eta}-\sqrt{f}}{\eta}\right\}^2d\mu\\
&\le 2\left\|\frac{\sqrt{f_\eta}-\sqrt{f}}{\eta}-\frac{1}{2}h\sqrt{f}\right\|_{L_2(\mu)}\sqrt{\int_0^M z^2f(z)d\mu}\\
&\quad+\left\|\frac{1}{2}h\right\|_{L_2(F)}\eta\sqrt{\int_0^M z^2d\mu}\to 0 \qquad (\eta\to 0)
\end{aligned}$$

となる．更に、$h\in L_2(F)$ であることより、

$$0=\boldsymbol{E}(Z)\int_0^M h(z)dF(z)=\int_0^M \boldsymbol{E}(Z)h(z)dF(z)$$

であるから、

$$\dot{\psi}(h)=\int_0^M(z-\boldsymbol{E}(Z))h(z)dF(z)$$

となる．$\dot{\psi}:\dot{\boldsymbol{P}}\longrightarrow\boldsymbol{B}=\boldsymbol{R}$ と $\boldsymbol{B}^*=\boldsymbol{R}$ に注意する．任意の $b^*\in\boldsymbol{R}$ に対し、

$$\langle b^*,\dot{h}\rangle_B=\langle\dot{\psi}^Tb^*,h\rangle_{L_2(F)}$$

と

$$\begin{aligned}
\langle b^*,\dot{\psi}h\rangle_B &= b^*\int_0^M(z-\boldsymbol{E}(Z))h(z)dF(z)\\
&= \langle b^*(\cdot-\boldsymbol{E}(Z)),h\rangle_{L_2(F)}
\end{aligned}$$

を得る．よって、随伴作用素 $\dot{\psi}^T:\boldsymbol{R}\longrightarrow\dot{\boldsymbol{P}}$ は

$$(\dot{\psi}^Tb^*)(z)=b^*(z-\boldsymbol{E}(Z))$$

で与えられる．

つぎに、$\mathrm{Image}(\dot{\psi}^T)\subset\mathrm{Image}(\dot{\boldsymbol{l}}^T)$ を示すために、特に

$$\tilde{\boldsymbol{l}}_\nu=\left(-\delta\frac{1-F(z)}{g(z)}+(1-\delta)\frac{F(z)}{g(z)}\right)1\{g(z)>0\}$$

ととる．例 5.2 の (5.20) と (5.21) を用いれば、任意の $h\in L_2^0(F)$ に対し、

$$\langle\dot{\boldsymbol{l}}^T,h\rangle_{L_2(F)} = \langle\tilde{\boldsymbol{l}}\nu,\dot{\boldsymbol{l}}h\rangle_{L_2(P_{F,G})}$$

$$
\begin{aligned}
&= \int_0^M \left[-\delta\frac{1-F(u)}{g(u)}+(1-\delta)\frac{F(u)}{g(u)}\right]1\{g(u)>0\}(\dot{\boldsymbol{l}}h)(u)dP_{F,G}(u)\\
&= \boldsymbol{E}_\delta\Big[\int_0^M \left\{-\delta\frac{1-F(u)}{g(u)}+(1-\delta)\frac{F(u)}{g(u)}\right\}\\
&\quad \times\left\{\delta\frac{\int_0^u h(z)dF(z)}{F(u)}+(1-\delta)\frac{\int_u^M h(z)dF(z)}{1-F(u)}\right\}\\
&\quad \times\Big\{F(u)^\delta(1-F(u))^{1-\delta}g(u)\Big\}\,d\mu\Big]\\
&= \int_0^M \left[-(1-F(u))\int_0^u h(z)dF(z)+F(u)\int_u^M h(z)dF(z)\right]d\mu\\
&= \int_0^M\int_0^M[-(1-F(u))1\{z\le u\}+F(u)1\{u<z\le M\}]h(z)d\mu dF(z)\\
&= \int_0^M h(z)\int_0^M[1\{z\le u\}+F(u)]d\mu dF(z)\\
&= \int_0^M h(z)\left[-M+z\int_0^M F(u)d\mu\right]dF(z)\\
&= \int_0^M h(z)\left[-M+z+[uF(u)]_0^M-\int_0^M uf(u)d\mu\right]dF(z)\\
&= \int_0^M h(z)\left[-M+z+M-\boldsymbol{E}(Z)\right]dF(z)\\
&= \langle h,\cdot-\boldsymbol{E}(Z)\rangle_{L_2(F)}
\end{aligned}
$$

を得る．よって、

$$
(\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}})(z)=z-\boldsymbol{E}(Z)
$$

となる．

最後に、$\tilde{\boldsymbol{l}}_\nu\in\overline{\mathrm{Image}(\dot{\boldsymbol{l}})}$ を示すために

$$
h=\left\{-\left(\frac{F(1-F)}{g}\right)'\Big/f\right\}
$$

ととる．例 5.2 の (5.21) を用いれば、

$$
\begin{aligned}
(\dot{\boldsymbol{l}}^T\dot{\boldsymbol{l}})(z) &= -\frac{\delta}{F(z)}\int_0^z\left(\frac{F(1-F)}{g}\right)'\frac{dF}{f}+\frac{1-\delta}{1-F(z)}\int_z^M\left(\frac{F(1-F)}{g}\right)'\frac{dF}{f}\\
&= -\delta\frac{1-F}{g}+(1-\delta)\frac{F}{g}=\tilde{\boldsymbol{l}}_\nu
\end{aligned}
$$

を得る．よって、$\tilde{\boldsymbol{l}}_\nu$ の定義から

$$
\begin{aligned}
I_\nu^{-1} &= \boldsymbol{E}(\tilde{\boldsymbol{l}}_\nu^2)\\
&= \int\left[\delta\frac{(1-F(^2}{g^2}(z)+(1-\delta)\frac{F^2}{g^2}(z)\right]dP_{F,G}
\end{aligned}
$$

$$= \int_0^M \left[\frac{(1-F)^2F}{g}(z) + \frac{F^2(1-F)}{g}(z)\right]dz$$

$$= \int_0^M \frac{F(1-F)}{g}(z)dz \text{ を得る.} \qquad (5.50)$$

注意 5.3 :Huang and Wellner (1995) にあるように、情報限界は F の Nonpaarmetric Maximum Likelihood estimator $\hat{F}_n$ による推定量 $\int zd\hat{F}_n(z)$ によって達成される．

例 5.6 （例 5.3 の続き）: $X^0 \equiv (Z, W) \in \boldsymbol{R}^2$ とし、Z と W は独立で、Z は未知の分布 F に従い、W は標準正規分布に従うとする．このとき、$X = T(X^0) = Z + W$ の確率密度関数は

$$p_F(x) = \int g(x-z)dF(z)$$

で与えられる．ただし、$g(w) = (1/\sqrt{2\pi})\exp(-w^2/2)$ とする．命題 5.2 より、このモデル

$$\boldsymbol{P} = \{P_F(x) : p_F \ll \mu, F \ll \lambda\}$$

は、ある $h \in L_2^0(F)$ が存在して

$$\sqrt{f_\eta} = \sqrt{f} + \frac{1}{2}\eta h\sqrt{f} + o(\eta)$$

を満足する[19]．$\boldsymbol{P}$ の中の曲線 $\{f_\eta : \eta \in \boldsymbol{R}\}$ に対し、Hellinger 微分可能であり、そのスコア作用素は任意の $a \in \dot{\boldsymbol{G}}$ に対し

$$(\boldsymbol{l}^* a)(x) = \boldsymbol{E}[a(Z)|T(X^0) = x] = \frac{\int a(z)g(x-z)dF(z)}{\int g(x-z)dF(z)} \qquad (5.51)$$

となる．まず、$\dot{\boldsymbol{l}}^T : L_2^0(P_F) \longrightarrow \dot{\boldsymbol{G}}$ を求める．任意の $a \in \dot{\boldsymbol{G}}$ と $h^* \in L_2^0(P_F)$ に対し

$$\begin{aligned}
\langle \dot{\boldsymbol{l}}^T h^*, a\rangle_{L_2(F)} &= \langle h^*, \dot{\boldsymbol{l}} a\rangle_{L_2(F)} = \int h^*(x)\frac{\int a(z)g(x-z)dF(z)}{\int g(x-z)dF(z)}p_F(x)d\mu(x) \\
&= \int a(z)\int h^*(x)g(x-z)d\mu(x)dF(z) \\
&= \langle a, \int h^*(x)g(x-\cdot)d\mu(x)\rangle_{L_2(F)} \qquad (5.52)
\end{aligned}$$

を得る．また、$h^* \in L_2^0(P_F)$ に注意すれば

$$\begin{aligned}
\boldsymbol{E}_Z[\int h^*(x)g(x-z)d\mu(x)] &= \int h^*(x)\int g(x-z)dF(z)d\mu(x) \\
&= \int h^*(x)p_F(x)d\mu(x) = 0 \qquad (5.53)
\end{aligned}$$

がわかる．(5.52) と (5.53) から、$\dot{\boldsymbol{l}}^T : L_2^0(P_F) \longrightarrow L_2^0(F)$ は

$$(\dot{\boldsymbol{l}}^T h^*)(z) = \int h^*(x)g(x-z)d\mu(x) \qquad (5.54)$$

[19] $\boldsymbol{P}$ のパラメータ空間を $\boldsymbol{G} = \{\sqrt{f} : f = dF/d\lambda\}$ とみる．

で与えられることがわかる．更に、(5.51) と (5.54) から、任意の $a \in L_2^0(F)$ に対し、

$$\begin{aligned}
(\dot{\boldsymbol{l}}^T \dot{\boldsymbol{l}} a)(z) &= \dot{\boldsymbol{l}}^T \left[\frac{\int a(z')g(\cdot - z')dF(z')}{p_F(\cdot)} \right](z) \\
&= \int \frac{\int a(z')g(x-z')dF(z')}{p_F(x)} g(x-z)d\mu(x) \\
&= \int a(z') \int \frac{g(x-z')g(x-z)}{p_F(x)} d\mu(x)dF(z') \\
&= \int a(z')K(z,z')dF(z')
\end{aligned}$$

を得る．ただし、

$$K(z,z') = \int \frac{g(x-z)g(x-z')}{p_F(x)} d\mu(x)$$

で与えられる．

ここで、固定した点 t_0 における分布関数 F の値の推定を考えよう．すなわち、

$$\nu(P_F) = \psi(\sqrt{f}) = \int_{-\infty}^{t_0} (\sqrt{f})^2 d\lambda(x) = F(t_0)$$

である．例 5.5 より、ある $h \in L_2^0(F)$ が存在して

$$\sqrt{f_\eta} = \sqrt{f} + \frac{1}{2}\eta h \sqrt{f} + o(\eta)$$

を満足する曲線に対し、$\dot{\psi} : \dot{\boldsymbol{G}} = L_2(F) \longrightarrow \boldsymbol{B} = \boldsymbol{R}$ は任意の $h \in \dot{\boldsymbol{G}}$ をとれば、

$$\dot{\psi} h = \int [1_{(-\infty, t_0]}(z) - F(t_0)]h(z)dF(z)$$

となる．また、$\dot{\psi}^T : \boldsymbol{B}^* = \boldsymbol{R} \longrightarrow L_2(F)$ は任意の $b^* \in \boldsymbol{R}$ に対し

$$(\dot{\psi} b^*)(z) = b^* \{1_{(-\infty, t_0]}(z) - F(t_0)\}$$

で与えられる．

これより、すべての $b`*$ にたいし、$\dot{\psi}^T b^*$ は z の不連続関数となる．一方、$\dot{\boldsymbol{l}}^T h^*$ は連続関数なので、$\dot{\psi}^T \notin \mathrm{Image}(\dot{\boldsymbol{l}}^T)$ である．よって、定理 5.2 より ν は経路ごとの微分可能性を満足しないことがわかる．したがって、$F(t_0)$ の推定は $1/\sqrt{n}$ よりも遅いオーダーになることが予想される．

つぎに、$c(z)$ を固定し、$\boldsymbol{E}_F[c^2(Z)] < \infty$ を満足すると仮定したとき、

$$\psi(\sqrt{f}) = \int c(z)dF(z)$$

の推定を考える．例 5.5 と同様に、ある $h \in L_2^0(F)$ に対し、

$$\sqrt{f_\eta} = \sqrt{f} + \frac{1}{2}\eta h \sqrt{f} + o(\eta)$$

を満足する曲線 $\{f_\eta : \eta \in \boldsymbol{R}\}$ を考える．これに対し、$\eta \to \infty$ のとき、

$$\begin{aligned}
&\frac{\psi(\sqrt{f_\eta}) - \psi(\sqrt{f})}{\eta} - \int c(z)h(z)(\sqrt{f})^2 d\mu(z) \\
&= \frac{1}{\eta}\left[\int 2\{\sqrt{f_\eta} - \sqrt{f} - \eta\frac{1}{2}h\sqrt{f}\}c(z)\sqrt{f}d\mu(z)\right] \\
&\quad + \eta\int\left\{\frac{\sqrt{f_\eta} - \sqrt{f}}{\eta}\right\}^2 c(z)d\mu(z) \\
&\le 2\left\|\frac{\sqrt{f_\eta} - \sqrt{f}}{\eta} - \eta\frac{1}{2}h\sqrt{f}\right\|_{L_2(\mu)} \cdot \|c\|_{L_2(F)} + \eta\|\frac{1}{2}h\|_{L_2(F)}\|c\|_{L_2(F)} \to \infty
\end{aligned}$$

を得る．したがって、任意の $h \in \dot{\boldsymbol{G}} = L_2^0(F)$ に対し、

$$\begin{aligned}
\dot{\psi}h &= \int h(z)c(z)dF(z) \\
&= \int h(z)c(z)dF(z) - \boldsymbol{E}_F c(Z)\int h(z)dF(z) \\
&= \int h(z)\{c(z) - \boldsymbol{E}_F c(Z)\}dF(z)
\end{aligned}$$

となる．また、任意の $b^* \in \boldsymbol{B}^* = \boldsymbol{R}$ に対し、

$$\begin{aligned}
\langle \dot{\psi}^T b^*, h\rangle_{L_2(F)} &= \langle b^*, \dot{\psi}h\rangle_{B^*} = b^*\dot{\psi}h = \int b^* h(z)\{c(z) - \boldsymbol{E}_F c(Z)\}dF(z) \\
&= \langle a, b^*\{c - \boldsymbol{E}_F c(Z)\}\rangle_{L_2(F)}
\end{aligned}$$

より、$\dot{\psi}^T : \boldsymbol{R} \longrightarrow L_2^0(F)$ は

$$(\dot{\psi}^T b^*)(z) = b^*\{c(z) - \boldsymbol{E}_F c(Z)\}$$

で与えられる．ある $h \in \{L_2(P_F)\}^* = L_2(P_F)$ が存在して

$$c(z) - \boldsymbol{E}_F c(Z) = \int h(x)g(x-z)d\mu(x)$$

を満足すれば、定理 5.2 より、ν は経路ごとに微分可能であることがわかる．また、

$$c(z) = \int \tilde{h}(x)g(x-z)d\mu(x)$$

と書き直せる．このとき、Fubini の定理を使えば、

$$\begin{aligned}
\nu(P_F) &= \int c(z)dF(z) = \int\int \tilde{h}(x)g(x-z)d\mu(x)dF(z) \\
&= \int \tilde{h}(x)\int g(x-z)dF(z)d\mu(x) = \int \tilde{h}(x)dP_F(x)
\end{aligned}$$

となることよい、

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{(}X_i)$$

は有効推定量となることがわかる．

5.5　たたみこみ定理

$\boldsymbol{P}$ をモデルとする．つまり、ある確率空間 $(\boldsymbol{M}, \mathcal{M}_B)$ 上の確率測度の P の集まりである．簡単のために、ある測度 μ に対し、$P \ll \nu (P \in \boldsymbol{P})$ とする．また、写像

$$\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$$

を考える．ただし、$\boldsymbol{B}$ は norm $\| \cdot \|_B$ をもつ Banach 空間である．$\nu(P)$ を一般パラメータまたは $\boldsymbol{B}$ – 値パラメータとよぶことにする．ここで、$\boldsymbol{B}$ – 値パラメータを推定するときのたたみこみ定理を導出する．

5.5.1　影響作用素

$\boldsymbol{B}^*$ を $\boldsymbol{B}$ の双対空間[20] とし、任意の $b^* \in \boldsymbol{B}^*$ と $b \in \boldsymbol{B}$ に対し、$\langle b, b^* \rangle_B \equiv b^*(b)$ とする．

定義 5.7 ：$\boldsymbol{B}$ – 値パラメータ ν が $P_0 \in \boldsymbol{P}$ において道ごとに微分可能であるとは、有界線形作用素

$$\dot{\nu}(P_0) \equiv \dot{\nu} : \dot{\boldsymbol{P}} \longrightarrow \boldsymbol{B}$$

が存在して、任意の接ベクトル $h \in \dot{\boldsymbol{P}}$ をもつ[21] 曲線 $\{P_\eta\} \subset \boldsymbol{P}$ に対し、

$$\left\| \frac{\nu(P_\eta) - \nu(P_0)}{\eta} - \dot{\nu}(h) \right\|_B = o(\eta) \tag{5.55}$$

を満足することである．

ν が未知ごとに微分可能で微分 $\dot{\nu}$ をもてば、任意の $b^* \in \boldsymbol{B}^*$ に対し、

$$\langle \frac{\nu(P_\eta) - \nu(P_0)}{\eta} - \dot{\nu}(h), b^* \rangle_B = o(1) \tag{5.56}$$

が成立する．つまり、道ごとに微分可能ならば、$\boldsymbol{B}$ の弱位相の意味で微分可能であることがわかる．(5.56) が成立するとき、ν は**道ごとに弱微分可能**であるという．また、b^* の線形性より、(5.56) は、任意の b^* に対し、

$$\langle \nu(P_\eta), b^* \rangle_B = \langle \nu(P_0), b^* \rangle_B + \eta b^* \dot{\nu}(h) + o(1) \tag{5.57}$$

が成立することと同値である．また、$\dot{\nu}^T : \boldsymbol{B}^* \longrightarrow \dot{\boldsymbol{P}}$ の定義より

$$b^* \dot{\nu}(h) = \langle \dot{\nu}(h), b^* \rangle_B = \langle h, \dot{\nu}^T b^* \rangle_{L_2(P_0)} \tag{5.58}$$

[20] $\boldsymbol{B}$ 上の有界線形汎関数のなす空間

[21] $s_\eta = \sqrt{dP_\eta / d\mu}$ とおいてとき、

$$s_\eta = s_0 + \frac{1}{2}\eta h s_0 + o(\eta)$$

を満足する．

となる．$b^* \boldsymbol{B}^*$ を固定すれば、$b^*\nu(P) = \langle \nu(P), b^* \rangle_{B^*}$ は $\boldsymbol{P}$ 上の実数値パラメータになるので、道ごとに弱微分可能ならば、$b^*\nu(P)$ は定義 4.6 の意味で道ごとに微分可能で微分 $b^*\dot{\nu} : \dot{\boldsymbol{P}} \longrightarrow \boldsymbol{R}$ をもつことがわかる．

さらに、$b^*\dot{\nu}$ は Hilbert 空間[22] の部分空間 $\dot{\boldsymbol{P}}$ 上の有界線形作用素なので、Riesz の表現定理より、一意的に $\dot{\nu}_{P_0,b^*} \in \dot{\boldsymbol{P}}$ が存在して、

$$b^*\dot{\nu}(h) = \langle \dot{\nu}_{P_0,b^*}, h \rangle_{L_2(P_0)} \tag{5.59}$$

と書ける．(5.58) と (5.59) を比較すれば、

$$\dot{\nu}_{P_0,b^*} = \dot{\nu}^T b^* \tag{5.60}$$

を得る．$\dot{\nu}_{b^*}$ はしばしば P_0 における b^* の方向についての ν の **canonical gradient** とよぶ．

$\dot{\nu}$ は一意的に $\dot{\boldsymbol{P}}^0$ から $\dot{\boldsymbol{P}}$ へ拡張できる．また、ν はより大きなモデル $\boldsymbol{M}_0 \supset \boldsymbol{P}$ 上で定義されたパラメータ ν_e の $\boldsymbol{P}$ への制限とみなすことが自然なことがある．ν_e は $\boldsymbol{M}_0$ 上で道ごとに微分可能で微分 $\nu_e : \dot{\boldsymbol{M}}_0 \longrightarrow \boldsymbol{B}$ をもつと仮定する．このとき、

$$\dot{\nu}_e | \dot{\boldsymbol{P}} = \dot{\nu}$$

なので、

$$\dot{\nu}^T b^* = \Pi_0(\dot{\nu}_e^T b^* | \dot{\boldsymbol{P}})$$

である．

定義 5.8 ：ν を経路ごとに微分可能で微分 $\dot{\nu}$ をもつとき、ν を推定するときの $\boldsymbol{P}$ における**有効影響作用素** $\tilde{\boldsymbol{l}}(P_0|\nu, \boldsymbol{P}) \equiv \tilde{\boldsymbol{l}}_\nu : \boldsymbol{B}^* \longrightarrow \dot{\boldsymbol{P}}$ を

$$\tilde{\boldsymbol{l}}_\nu(b^*) = \dot{\nu}^T b^* = \dot{\nu}_{P_0,b^*} \tag{5.61}$$

で定義する．

このとき、$\tilde{\boldsymbol{l}}_\nu$ の値域はパラメータ $b^*\nu$ を推定するときの影響関数の集まりである．また、ν の任意の拡張 ν_e と $h \in \dot{\boldsymbol{P}}$ に対し、

$$\dot{\nu}_e^T b^* - \Pi_0(\dot{\nu}^T | \dot{\boldsymbol{P}}) \perp h \in \dot{\boldsymbol{P}}$$

であることに注意すれば、

$$\langle \dot{\nu}_e(h), b^* \rangle_B = \langle h, \dot{\nu}_e^T b^* \rangle_{L_2(P_0)} = \langle h, \Pi_0(\dot{\nu}_e^T | \dot{\boldsymbol{P}}) \rangle_{L_2(P_0)} = \langle h, \tilde{\boldsymbol{l}}_\nu \rangle_{L_2(P_0)} \tag{5.62}$$

となる．

定義 5.9 ：ν に対する**逆情報共分散汎関数** $I^{-1}(P_0|\nu, \boldsymbol{P}) \equiv I_\nu^{-1} : \boldsymbol{B}^* \times \boldsymbol{B}^* \longrightarrow \boldsymbol{R}$ を任意の $b_1^*, b_2^* \in \boldsymbol{B}^*$ に対し、

$$I_\nu^{-1}(b_1^*, b_2^*) = \boldsymbol{E}_{P_0}[\tilde{\boldsymbol{l}}_\nu(b_1^*) \tilde{\boldsymbol{l}}_\nu(b_2^*)] \tag{5.63}$$

で定義する．

[22] $(L_2(P_0), \langle \cdot, \cdot \rangle_{L_2(P_0)})$

I_ν^{-1} は確率過程 $b^* \longrightarrow \tilde{\boldsymbol{l}}_\nu(b^*)$ の分散共分散構造を与える．

さらに、

$$I_\nu^{-1}(b_1^*, b_2^*) = \langle \tilde{\boldsymbol{l}}_\nu(b_1^*), \tilde{\boldsymbol{l}}_\nu(b_2^*)\rangle_{L_2(P_0)} = \langle b_1^*, \tilde{\boldsymbol{l}}_\nu^T \tilde{\boldsymbol{l}}(b_2^*)\rangle_{B^*}$$

を得る．ただし、

$$\tilde{\boldsymbol{l}}_\nu^T : \dot{\boldsymbol{P}} \longrightarrow \boldsymbol{B}^*$$

である．よって、

$$\tilde{\boldsymbol{l}}_\nu^T \tilde{\boldsymbol{l}}_\nu : \dot{\boldsymbol{P}} \longrightarrow \boldsymbol{B}^*$$

となる．

定義 5.10 ：$\tilde{\boldsymbol{l}}_\nu^T \tilde{\boldsymbol{l}}_\nu : \boldsymbol{B}^* \longrightarrow \boldsymbol{B}^{**}$ をモデル $\boldsymbol{P}$ における ν に対する**情報限界作用素**とよぶ．

注意 5.4 ：$\boldsymbol{B} = \boldsymbol{R}^m$ もしくは Hilbert 空間である場合には、$\boldsymbol{B}, \boldsymbol{B}^*, \boldsymbol{B}^{**}$ が同一視でき、$\boldsymbol{B} = \boldsymbol{R}^m$ のときには、情報限界作用素は $I^{-1}(P_0|\nu, \boldsymbol{P})$ である．

定義 5.11 ：写像 $\tilde{\boldsymbol{l}} : \boldsymbol{X} \longrightarrow \boldsymbol{B}$ が存在して、任意の $b^* \in \boldsymbol{B}^*$ に対し、

$$b^* \tilde{\boldsymbol{l}} = \dot{\nu}_{P_0, b^*} = \dot{\nu}^T b^* = \tilde{\boldsymbol{l}}_\nu(b^*) \tag{5.64}$$

を満足するとき、$\tilde{\boldsymbol{l}}$ を**有効影響関数**とよぶ．

注意 5.5 ：もし、 $\dot{\boldsymbol{P}}$ が有限次元のとき、ν が経路ごとに微分可能ならば、有効影響関数は存在する．なぜならば、$\{h_1, \ldots, h_k\}$ を $\dot{\boldsymbol{P}}$ 直交基底とする．任意の $x \in \boldsymbol{X}$ に対し、$\tilde{\boldsymbol{l}}(x) \in \boldsymbol{B}$、$\dot{\nu}(h_j) \in \boldsymbol{B}$、$h_j(x) \in \boldsymbol{B}(j = 1, 2, \ldots, k)$ とおく．すると、

$$b^* \tilde{\boldsymbol{l}} = \sum_{j=1}^k h_j(x) b^* \dot{\nu}(h_j) = \sum_{j=1}^k h_j(x) \langle \dot{\nu}_{b^*}, h_j\rangle_{L_2(P_0)} = \dot{b^*}(x)$$

となり、$b^* \tilde{\boldsymbol{l}} = \dot{\nu}_{b^*}$ を満足する．

5.5.2　Banach - 値　パラメータにたいする正則推定量

$X_1, \ldots, X_n$ を $(\boldsymbol{X}, \mathcal{B})$ に値ととる確率変数とし、独立同一に分布 P に従うものとする．このとき、$\nu(P)$ の「推定量」[23] は $\boldsymbol{X}^n$ から $\boldsymbol{B}$ への写像 $T_n = t_n(X_1, \ldots, X_n)$ である．ここで、正則推定量の定義を Hoffmann-Jorgensen=Dudely の意味での弱収束を用いて定義を行う．

定義 5.12 ：$\nu(P)$ の推定量の列 $\{T_n\}$ が $P_0 \in \boldsymbol{P}$ において**正則**であるとは、$\boldsymbol{B}$ の値をとる tight な Borel 可測 random element Z が存在して、P_0 を通過する任意の曲線 $\{P_\eta\} \subset \boldsymbol{P}$ と $\eta_n = O(n^{-1/2})$ なるすべての数列 $\{\eta_n\}$ に対し、$P_n = P_{\eta_n}$ のもとで

$$\sqrt{n}(T_n - \nu(P_n)) \Rightarrow Z \qquad n \to \infty \tag{5.65}$$

が成立することである．

[23] 可測性についての仮定はおかない．

定義 5.13 ：推定量の列 $\{T_n\}$ が、ある線形作用素 $\psi_\nu(\cdot, P_0) \equiv \tilde{\psi}_{P_0,\nu} : \boldsymbol{B}^* \longrightarrow L_2^0(P_0)$ が存在して、すべての $b^* \in \boldsymbol{B}^*$ について、P_0 のもとで、$n \to \infty$ としたとき、

$$\sqrt{n}b^*(T_n - \nu(P_0)) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}_{P_0,\nu}(b^*)(X_i) \Rightarrow 0$$

を満足するならば、$\{T_n\}$ は、P_0 において**影響作用素** ψ_ν をもち、**弱漸近線形**であるという．

同様に、ある関数 $\psi_{P_0} : \boldsymbol{X} \longrightarrow \boldsymbol{B}$ が存在して、P_0 のもとで、$n \to \infty$ としたとき、

$$\sqrt{n}(T_n - \nu(P_0)) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{P_0}(X_i) \Rightarrow 0$$

を満足するならば、$\{T_n\}$ は影響関数 ψ をもち、P_0 において**漸近線形**であるという．

また、すべての $b^* \in \boldsymbol{B}^*$ に対し、P_0 のもとで、$n \to \infty$ としたとき、

$$\sqrt{n}b^*(T_n - \nu(P_0)) - \frac{1}{n}\sum_{i=1}^n b^*\psi_{P_0}(X_i) \Rightarrow 0$$

を満足するならば、$\{T_n\}$ は P_0 において影響関数 ψ をもち、**弱漸近線形**である．

5.5.3 たたみこみ定理

定理 5.3 （たたみこみ定理）：つぎを仮定する：
(i)．ν は $P_0 \in \boldsymbol{P}$ において経路ごとに微分可能で微分 $\dot{\nu}$ をもつ．
(ii)．$\{T_n\}$ は正則で極限分布 Z をもつ．
(iii)．$\dot{\boldsymbol{P}}^0$ は線形である．すなわち、$\dot{\boldsymbol{P}} = \overline{\dot{\boldsymbol{P}}^0}$ とする．
このとき、tight で Borel 可測な $\boldsymbol{B}$ – 値 random element Z_0 と Δ_0 が存在し、つぎを満足する．
A．$\boldsymbol{L}(Z) = \boldsymbol{L}(Z_0 + \Delta_0)$．
B．Z_0 と Δ_0 は独立．
C．Z_0 は平均が 0 で分散共分散関数が任意の $b_1^*, b_2^* \in \boldsymbol{B}^*$ に対し、

$$\boldsymbol{Cov}(b_1^* Z_0, b_2^* Z_0) = I_\nu^{-1}(b_1^*, b_2^*)$$

で与えられる Gauss 過程である．ただし、I_ν^{-1} は ν に対する逆情報分散共分散汎関数である．
D．有効影響関数 $\tilde{\boldsymbol{l}} : \boldsymbol{X} \longrightarrow \boldsymbol{B}$ が存在して、

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i) \Rightarrow Z_0 \tag{5.66}$$

を満足するとする．このとき。

$$\begin{bmatrix} \sqrt{n}(T_n - \nu(P_0)) - (1/\sqrt{n})\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i) \\ (1/\sqrt{n})\sum_{i=1}^n \tilde{\boldsymbol{l}}(X_i) \end{bmatrix} \Rightarrow \begin{bmatrix} \Delta_0 \\ Z_0 \end{bmatrix}$$

であり、Δ_0 と Z_0 は独立である．
E．(5.66) が成立するとき、Δ_0 が $0 \in \boldsymbol{B}$ に退化するための必要十分条件は $\{T_n\}$ が P_0 において影響関数が $\tilde{\boldsymbol{l}}$ で漸近線形であることである．

定義 5.14 ：$\{T_n\}$ が極限分布 $Z = Z_0$ で (5.65) を満たす正則推定量ならば、$\{T_n\}$ は P_0 において**有効**であるという．

系 5.5 ：定理 5.3 の仮定が成立し、$\{T_n\}$ は影響関数が ψ で、P_0 において漸近線形であるとする．このとき、$\{T_n\}$ が P_0 において有効であるための必要十分条件は、すべての $b^* \in \boldsymbol{B}^*$ に対し、$b^*\psi \in \dot{\boldsymbol{P}}$ が成立することである．この場合、有効影響関数 $\tilde{\boldsymbol{l}}$ が存在し、ψ と一致する．

5.5.4 たたみこみ定理の応用例

例 5.7 （$\boldsymbol{R}$ 上の分布関数の推定）: $(\boldsymbol{X}, \mathcal{A}) = (\boldsymbol{R}, \mathcal{A})$ とし、

$$\boldsymbol{P} = \{\boldsymbol{R}\ \text{上の確率測度}\ \boldsymbol{P} \ll \boldsymbol{\mu} \equiv \text{Lebesgue 測度}\}$$

とする．任意の $\boldsymbol{R}$ 上の測度に対し、$\nu(P)$ を対応する分布関数とする．すなわち、任意の $t \in \boldsymbol{R}$ に対し、

$$\nu(P)(t) \equiv P\{(-\infty, t]\} \equiv F(t)$$

とする．よって、$\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ で、$\boldsymbol{B} \equiv D[-\infty, \infty]$ は左極限が存在する右連続な関数の空間で、norm $\|\cdot\|_B$ は $\|\cdot\|_\infty$ で与えられる．このとき、例 5.4 より、ν は経路ごとに微分可能であり、微分 $\dot{\nu}(P_0) \equiv \dot{\nu} : \dot{\boldsymbol{P}} \longrightarrow \boldsymbol{B}$ は、$t \in \boldsymbol{R}$ に対し、

$$\dot{\nu}(h)(t) = \int [1_{(-\infty,t]}(x) - F_0(t)] h(x) dP_0(x)$$

で与えれることがわかる．また、$\boldsymbol{B}^*$ を $t \in \boldsymbol{R}$ における射影関数つまり t における関数の値とすれば、任意の $\pi_t \in \boldsymbol{B}^*$ と任意の $h \in \dot{\boldsymbol{P}}$ に対し、

$$\dot{\nu}(h)(t) = \langle \dot{\nu}(h), \pi_t \rangle_B = \langle h, \dot{\nu}^T \pi_t \rangle_{L_2(P_0)} = \int h(x)(\dot{\nu}^T \pi_t)(x) dP_0(x)$$

より、

$$(\dot{\nu}^T \pi_t)(x) = 1_{(-\infty,t]}(x) - F_0(t)$$

を得る．したがって、gradients $\dot{\nu}_t = \dot{\nu}^T \pi_t$ は任意の $t \in \boldsymbol{R}$ に対し、

$$\dot{\nu}_t(x) = 1_{(-\infty,t]}(x) - F_0(t)$$

で与えられる．例 4.2 より、$\dot{\boldsymbol{P}} = L_2^0(P_0)$ であることから、$\dot{\nu}_t \in \dot{\boldsymbol{P}}$ となる．したがって、

$$\tilde{\boldsymbol{l}} = \pi_t \tilde{\boldsymbol{l}} = \tilde{\boldsymbol{l}}_\nu(\pi_t) = \dot{\nu}_t$$

となる．更に、逆情報分散共分散関数は $\pi_s, \pi_t \in \boldsymbol{B}^*$ に対し、

$$\begin{aligned} I_\nu^{-1}(\pi_s, \pi_t) &= \boldsymbol{E}_{P_0}[\tilde{\boldsymbol{l}}_\nu(\pi_s)\tilde{\boldsymbol{l}}_\nu(\pi_t)] = \boldsymbol{E}[\dot{\nu}_s \dot{\nu}_t] \\ &= F_0(s) \wedge F_0(t) - F_0(s)F_0(t) \equiv I_\nu^{-1}(s,t) \end{aligned}$$

で与えれる．よって、定理 5.3 より Gauss 過程 $[0,1]$ 上の Brownian bridge と $\nu(P_0) = F_0$ で合成表現でき、

$$Z_0(t) = B_0(F_0(t))$$

で与えられる．よって、定理 5.3 は任意の正則推定量 $T = \{T_n\}$ の推定誤差

$$Z_n = \sqrt{n}(T_n - \nu(P_0))$$

の極限分布のばらつきは Z_0 のばらつきよりも小さくならない．

ここで、

$$T_n = F_n = (1/n)\sum_{i=1}^{n} 1\{X_i \leq \cdot\}$$

とおけば、$Z_n \equiv \sqrt{n}(F_n - F)$ は漸近線形で $\nu = F$ の正則推定量であり、$Z_n \Rightarrow Z = 0$ が知られている．よって、F_n は有効推定量である．

第6章　有効推定量の構成のための道具箱

6.1　距離空間上の分布の収束

6.1.1　距離とノルム空間

定義 6.1 ：空でない集合 $\boldsymbol{M}$ とつぎの性質 (i) $\sim$ (iii) をもつ写像 $d : \boldsymbol{M} \times \boldsymbol{M} \longrightarrow [0, \infty)$ があるとき、d を $\boldsymbol{M}$ 上の**距離**といい、$\boldsymbol{M}$ を**距離空間**という．

(i)．任意の $x, y \in \boldsymbol{M}$ に対し、$d(x, y) = 0 \Leftrightarrow x = y$ が成立する．

(ii)．任意の $x, y \in \boldsymbol{M}$ に対し、$d(x, y) = d(y, x)$ が成立する．

(iii)．任意の $x, y, z \in \boldsymbol{M}$ に対し、$d(x, z) \le d(x, y) + d(y, z)$ が成立する.

r を正の数とし、集合 $\{y : d(x, y) < r\}$ を**開球**という．A を $\boldsymbol{M}$ の部分集合とし、$x \in \boldsymbol{M}$ が A の**内点**であるとは、ある $r > 0$ に対して、$\{y : d(x, y) < r\} \subset A$ が成立することである．A の各点が A の内点であるとき、A は**開集合**であるという．A の補集合が開集合のとき、A を**閉集合**であるという．A に含まれる最大の開集合を A の**内部**といい、$\overset{\circ}{A}$ で記す．点 $x \in \boldsymbol{M}$ が A の**集積点**であるとは、任意の r に対しても $0 < d(z, x) < r$ を満たす点 $z \in A$ が存在することである．A とその集積点全体の集合の和集合を A の**閉包**とよび、$\overline{A}$ で記す．点列 $\{x_n\}_{n=1}^{\infty}$ に対して、$d(x_n, x) \to 0 (n \to \infty)$ なる点 $x \in \boldsymbol{M}$ が存在するとき、$\{x_n\}$ は x に**収束する**といい、$x_n \to x$ と記す．部分集合 A が $\boldsymbol{M}$ において**稠密**であるとは、$\overline{A} = \boldsymbol{M}$ が成立することである．$\boldsymbol{M}$ が**可分**であるとは、$\boldsymbol{M}$ のあるたかだか可算部分集合が $\boldsymbol{M}$ において稠密であることをいう．$\boldsymbol{M}$ のどんな Cauchy 列[1] も $\boldsymbol{M}$ のある点に収束するとき、$\boldsymbol{M}$ は**完備**であるという．$\boldsymbol{M}$ の部分集合 K が**コンパクト**であるとは、K の任意の開被覆から K の有限開被覆を選び出せることである．K が**全有界**であるとは、任意の $\epsilon > 0$ に対しても K の中から有限個の点 $x_1, \ldots, x_m$ を選び出して、$K \subset \bigcup_{i=1}^{m} \{y : d(y, x_i) < r\}$ とできることである．K が**点列コンパクト**であるとは、K の中のどんな無限点列 $\{x_n\}_{n=1}^{\infty}$ からも、$\boldsymbol{M}$ の中のある点に収束する部分列 $\{x_{n_j}\}_{j=1}^{\infty}$ が選びだせることである[2]．$(\boldsymbol{M}_1, d_1)$ と $(\boldsymbol{M}_2, d_2)$ を 2 つの距離空間とする．写像 $f : \boldsymbol{M}_1 \longrightarrow \boldsymbol{M}_2$ が点 $x_0 \in \boldsymbol{M}_1$ において**連続**であるとは、どんな正の数 ϵ に対しても適当に正の数 δ をとって、$d_1(x, x_0) < \delta$ ならば、$d_2(f(x), f(x_0)) < \epsilon$ となるようにできることである．

[1] 点列 $\{x_n\}$ が Cauchy 列であるとは、どんな正の数 ϵ に対しても、それに応じて適当な自然数 N をとって、$j, k > N$ ば、$d(x_j, x_k) < \epsilon$ となるようにできることである．

[2] これら 3 つの概念にはつぎのような同値関係がある．(i) $\sim$ (iii) は同値である．(i)．K は点列コンパクトである．(ii)．$\overline{K}$ はコンパクトである．(iii)．K は全有界かつ $\overline{K}$ は完備である．

定義 6.2 ：線形空間 $\boldsymbol{X}$ で定義された写像 $\|\cdot\| : \boldsymbol{X} \longrightarrow [0,\infty)$ がつぎの性質 (i) $\sim$ (iii) を満たすとき、$\|\cdot\|_X$ を $\boldsymbol{X}$ の**ノルム**とよび、ノルムが定義されている線形空間 $\boldsymbol{X}$ を**ノルム空間**いう．

(i) ．任意の $x \in \boldsymbol{X}$ に対して、$\|x\|_X \Leftrightarrow x = 0$ ．

(ii) ．任意の $x, y \in \boldsymbol{X}$ に対して、$\|x+y\|_X \leq \|x\|_X + \|y\|_X$ ．

(iii) ．任意の係数 α と任意の $x \in \boldsymbol{X}$ に対して、$\|\alpha x\|_X = |\alpha| \|x\|_X$ ．

6.1.2　分布の収束

$(\Omega, \mathcal{A}, P)$ と $(\Omega_n, \mathcal{A}_n, P_n)$　$(n = 1, 2, \ldots,)$ を確率空間とする．

定義 6.3 ：$\boldsymbol{M}$ 上の Borel σ－集合体とは、$\boldsymbol{M}$ 上の開集合全体を含む最小の σ – 加法族のことである．これを $\mathcal{M}_B$ と記す．確率空間 $(\Omega, \mathcal{A}, P)$ 上の写像 $X : \Omega \longrightarrow \boldsymbol{M}$ が Borel 可測[3] ならば、X を $\boldsymbol{M}$ – 値 random element とよぶ．

任意の写像 $U : \Omega \longrightarrow \overline{\boldsymbol{R}}$ に対して[4] 、**外期待値（outer expectation ）**を

$$\boldsymbol{E}^* U = \inf\{\boldsymbol{E}V : V \geq U, V : \Omega \longrightarrow \overline{\boldsymbol{R}} \text{ は可測で、} \boldsymbol{E}V \text{ が存在 }\}$$

で定義する[5] ．

定義 6.4 ：　$(\boldsymbol{M}, d)$ を距離空間とする．
（分布収束）：任意の写像の列 $X_n : \Omega_n \longrightarrow \boldsymbol{M}$ がある random element　X に**分布収束**するとは、任意の有界連続関数 f に対して、$n \to \infty$ のとき、

$$\boldsymbol{E}^* f(X_n) \to \boldsymbol{E} f(X)$$

が成立することである．ただし、X は Borel 可測とする．$X_n \rightsquigarrow X$ と記す．
（確率収束）：どんな正の数 ϵ に対しても、$n \to \infty$ のとき、

$$P^*(d(X_n, X) > \epsilon) \to 0$$

が成立するとき、X_n は X に**確率収束**するという．$X_n \xrightarrow{P} X$ と記す．
（概収束）：ある可測な確率変数の列 $\{\Delta_n\}$ が存在して、$n \to \infty$ のとき、

$$d(X_n, X) \leq \Delta_n \qquad \Delta_n \xrightarrow{as} 0$$

を満足するとき、X_n は X に**概収束**するという．$X_n \xrightarrow{as^*} X$ と記す．

---

[3] 任意の $C \in \mathcal{M}_B$ に対しても $X^{-1}(C) \in \mathcal{A}$ が成立する．
[4] $\overline{\boldsymbol{R}}$ は、実数 $\boldsymbol{R}$ に $\{\infty\}$ と $\{-\infty\}$ を加えたもの．
[5] $\boldsymbol{E}\max(V, 0)$ と $\boldsymbol{E}\max(0, -V)$ の少なくと一方は有限のとき、$\boldsymbol{E}V$ が存在するという．

補題 6.1 （Portmanteau の補題）: つぎの条件は同値である．
(i)．すべての有界連続関数に対して、$\boldsymbol{E}^* f(X_n) \to \boldsymbol{E} f(X)$．
(ii)．すべての有界 Lipshitz 関数に対して、$\boldsymbol{E}^* f(X_n) \to \boldsymbol{E} f(X)$．
(iii)．すべての開集合 G に対して、$\liminf_{n\to\infty} P_*(X_n \in G) \geq P(X \in G)$
(iv)．すべての閉集合 F に対して、$\limsup_{n\to\infty} P^*(X_n \in F) \leq P(X \in F)$．
(v)．すべての $P(\delta B = 0)$ なる Borel 可測集合 B に対して、$P^*(X_n \in B) \to P(X \in B)$．

定理 6.1 $X_n : \Omega_n \longrightarrow \boldsymbol{M}$ を任意の写像とし、X を $\boldsymbol{M}$ – 値 random element とする．
(i)．$X_n \xrightarrow{as^*} X$ ならば、$X_n \xrightarrow{P} X$．
(ii)．$X_n \xrightarrow{P} X$ ならば、$X_n \rightsquigarrow X$．
(iii)．c をある定数とする．このとき、$X_n \xrightarrow{P} c \Leftrightarrow X_n \rightsquigarrow c$．
(iv)．$X_n \rightsquigarrow X$ かつ $d(X_n, Y_n) \xrightarrow{P} 0$ ならば、$Y_n \rightsquigarrow X$．
(v)．c をある定数とする．$X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{P} c$ ならば、$(X_n, Y_n) \rightsquigarrow (X, c)$
(vi)．$X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y$ ならば、$(X_n, Y_n) \xrightarrow{P} (X, Y)$．

定理 6.2 （連続写像定理）: $\boldsymbol{M}_n \subset \boldsymbol{M}$ を任意の部分集合とし、$g_n : \boldsymbol{M} \longrightarrow \boldsymbol{D}$ を任意の写像で、ある点 $x \in \boldsymbol{M}_0$ に収束するどんなの点列 $x_n \in \boldsymbol{M}_n$ に対しも、$g_n(x_n) \to g_0(x)$ を満足するとする．このとき、任意の写像 $X_n : \Omega_n \longrightarrow \boldsymbol{M}_n$ と $\boldsymbol{M}$ – 値 random element X （したがって、$g_0(X)$ は $\boldsymbol{D}$ – 値 random element）に対し、
(i)．$X_n \rightsquigarrow X$ ならば、$g_n(X_n) \rightsquigarrow g_0(X)$
(ii)．$X_n \xrightarrow{P} X$ ならば、$g_n(X_n) \xrightarrow{P} g_0(X)$
(iii)．$X_n \xrightarrow{as^*} X$ ならば、$g_n(X_n) \xrightarrow{as^*} g_0(X)$
が成立する．

距離空間 $(\boldsymbol{M}, d)$ に値を取る Borel 可測な random element X が **tight** であるとは、どんな正の数 $\epsilon > 0$ に対しても、あるコンパクト集合 K が存在して、$P(X \notin K) < \epsilon$ を満足することである．また、任意の写像の列 $X_n : \Omega \longrightarrow \boldsymbol{M}$ が**漸近的に tight** であるとは、どんな正の数 $\epsilon > 0$ に対しても、あるコンパクトな集合 K が存在して、すべての $\delta > 0$ に対し、

$$\limsup_{n\to\infty} P^*(X \notin K^\delta) < \epsilon$$

を満足することである．ただし、$K^\delta = \{y : d(y, K) < \delta\}$ である．任意の写像の列 $X_n : \Omega \longrightarrow \boldsymbol{M}$ が**漸近可測**であるとは、すべての有界連続関数 $f : \boldsymbol{M} \longrightarrow \boldsymbol{R}$ に対し、$n \to \infty$ のとき、

$$\boldsymbol{E}^* f(X_n) - \boldsymbol{E}_* f(X_n) \to 0$$

を満足する． ただし、任意の写像 $U : \Omega \longrightarrow \boldsymbol{R}$ に対し、$\boldsymbol{E}_* U = \sup\{\boldsymbol{E} V : V : \Omega \longrightarrow \boldsymbol{R}$ は可測で $V \leq U$ を満足$\}$ とした．

定理 6.3 ：任意の写像 $X_n : \Omega_n \longrightarrow l^\infty(T)$ の列が tight な random element に弱収束するための必要十分条件はつぎのふたつである．
(i)　T のどんな有限個の点 $t_1, \ldots, t_k$ に対しても、列 $(X_{n,t_1}, \ldots, X_{n,t_k})$ は分布収束する．

(ii)　どんな $\epsilon > 0, \eta > 0$ に対しても、半径が $L_2(P)$ – norm に関して ϵ 以下の集合 $T_1, \ldots, T_k$ による T の分割 $T = \bigcup_{i=1}^k T_i$ が存在し、

$$\limsup_{n\to\infty} P^* \left(\sup_i \sup_{t,t' \in T_i} |X_{n,t} - X_{n,t'}| \geq \epsilon \right) \leq \eta$$

が成立する．

定理 6.4 （Prohorov の定理）：$X_n : \Omega \longrightarrow \boldsymbol{M}$ は任意の写像とする．
(i)．X をある tight な random element が存在し、$X_n \rightsquigarrow X$ ならば、$\{X_n\}_{n=1}^\infty$ は漸近 tight で漸近可測である．
(ii)．X_n が漸近 tight で漸近可測ならば、ある部分列 $\{X_{n_j}\}_{j=1}^\infty$ と tight な random element X が存在して、$j \to \infty$ のとき、$X_{n_j} \rightsquigarrow X$ が成立する．

6.2　経験過程

6.2.1　Glivenko-Cantelli および Donsker クラス

$X_1, \cdots, X_n$ を $(\boldsymbol{X}, \mathcal{A})$ 上の確率測度 P からのランダム標本とする．δ_x を x に退化したデイラック測度とし、$P_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ とする．可測関数 $f : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ に対して

$$P_n f = \frac{1}{n}\sum_{i=1}^n f(X_i) \qquad Pf = \int f dP$$

と記すことにする．

固定した f に対しては、大数の法則から、$Pf < \infty$ ならば、$P_n f \xrightarrow{as} Pf$ が成立する．可測関数のある族 $\mathcal{F}$ に関して一様にこの結果が成立する場合を考える．

定義 6.5 ：$\mathcal{F}$ を可測関数 $f : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ の任意の族とする．$\mathcal{F}$ が P - Glivenko – Cantelli であるとは、

$$\|P_n f - Pf\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |P_n f - Pf| \xrightarrow{as^*} 0$$

が成立することである．

関数 f における経験過程を

$$G_n \equiv \sqrt{n}(P_n f - Pf) = \sqrt{n}(P_n - P)f$$

で定義することにする．

固定した有限個の可測関数 $f_i (i = 1, 2, \ldots, k)$ で $Pf_i^2 < \infty$ なるものに対して、中心極限定理から

$$(G_n f_1, \ldots, G_n f_k) \rightsquigarrow (G_P f_1, \ldots, G_P f_k)$$

が成立する．ただし、右辺 G_P は平均ゼロで分散共分散が

$$\boldsymbol{E} G_P f G_P g = Pfg - Pf Pg \tag{6.1}$$

で与えられる多変量正規分布である．

定義 6.6 ：$\mathcal{F}$ を可測関数 $f : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ の任意の族とする．$\mathcal{F}$ が P **- Donsker** であるとは、確率過程 $\{G_n f : f \in \mathcal{F}\}$ が $l^\infty(\mathcal{F})$ において tight な極限確率過程に分布収束することである．ただし、極限確率過程は (6.1) なる分散共分散関数をもつ平均ゼロのガウス過程 G_P である．この確率過程を P – **Brownian bridge** とよぶ．

6.2.2 metric entorpy

ある関数族 $\mathcal{F}$ が Glivenko – Cantelli または Donsker であるかは $\mathcal{F}$ の「サイズ」に依存する．$\mathcal{F}$ のサイズを計る比較的簡単な測度として、metric entropy を利用するのが標準的な手段である．

いま、可測関数 f に対して、$L_r(P)$ – norm を

$$\|f\|_{P,r} = (P|f|^r)^{1/r}$$

で定義する．ふたつの関数 l, u に対して、bracket $[l, u]$ を $l \le f \le u$ を満足するすべての関数とする．$L_r(P)$ に関する ϵ – bracket を $P(u-l)^r < \epsilon^r$ を満足する bracket とする．bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ を $\mathcal{F}$ を覆うために必要な ϵ – bracket の最小の個数とする．このとき、l と u は有限の $L_r(P)$ – norm を持つが、$\mathcal{F}$ に属する必要はない．更に、**entropy with bracketing** を bracketing number の対数を取ったもので定義する．

通常 $\epsilon \downarrow 0$ のとき、$N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ は発散するが、$\mathcal{F}$ が Donsker であるための十分条件は $N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ が急には増大しないことである．増大の早さは **bracketing integral** で計ることができる：

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$$

補題 6.2 （Bernstein の不等式）：f を任意の有界可測関数とする．このとき、どんな正の数 $x > 0$ に対しても

$$P(|G_n f| > x) \le 2 \exp\left(-\frac{1}{4} \frac{x^2}{Pf^2 + x\|f\|_\infty/\sqrt{n}}\right)$$

が成立する．

証明：左側の裾を評価すれば十分であることに注意する．どんな $\lambda > 0$ に対しても、Markov の不等式より

$$P(G_n f > x) \le e^{-\lambda x} \boldsymbol{E} e^{\lambda G_n f}$$

が成立する[6]．更に、

$$\begin{aligned}
\boldsymbol{E} e^{\lambda G_n f} &= \Pi_{i=1}^n \boldsymbol{E} \exp\left(\frac{\lambda}{\sqrt{n}}(f(X_i) - Pf)\right) \\
&= \left\{1 + \boldsymbol{E} \sum_{k=1}^\infty \left(\frac{\lambda}{\sqrt{n}}\right)(f(X_1) - Pf)^k\right\}^n \\
&= \left\{1 + \sum_{k=1}^\infty \left(\frac{\lambda}{\sqrt{n}}\right) P(f(X_1) - Pf)^k\right\}^n
\end{aligned}$$

[6] $P(G_n f > x) = P(e^{\lambda G_n f} > e^{\lambda x}) \le e^{-\lambda x} \boldsymbol{E} e^{\lambda G_n f}$

となる[7]．

　ここで、

$$\lambda = \frac{1}{2}\frac{1}{Pf^2 + x\|f\|_\infty/\sqrt{n}} \le \min\left(\frac{x}{2Pf^2}, \frac{\sqrt{n}}{2\|f\|_\infty}\right)$$

とおく．すると、

$$\lambda^k \le \frac{x}{2Pf^2}\left(\frac{\sqrt{n}}{2\|f\|_\infty}\right)^{k-2}\lambda$$

と

$$|P(f - Pf)^k| \le Pf^2(2\|f\|_\infty)^{k-2}$$

に注意すれば、

$$\sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{\lambda}{\sqrt{n}}\right)^k P(f - PF)^k \le \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\lambda x$$

がわかる．更に、

$$\sum_{k=2}^{\infty}\frac{1}{k!} \le e - 2 \le 1 \quad \text{と} \quad (1 + a)^n \le e^{na}$$

を利用すれば、

$$\begin{aligned} P(G_n f > x) &\le e^{-\lambda x}\left(1 + \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\lambda x\right)^n \le e^{-\lambda x}\left(1 + \frac{1}{n}\frac{\lambda x}{2}\right)^n \\ &\le e^{-\lambda x}e^{(1/2)\lambda x} = e^{-(1/2)\lambda x} \end{aligned}$$

となり、補題は証明された．□

補題 6.3 ：$\mathcal{F}$ を有限個の有界な可測関数からなる族とする．その要素の数を $|\mathcal{F}|$ と記す．このとき、

$$\boldsymbol{E}_P\|G_n\|_{\mathcal{F}} \lesssim \max_f \frac{\|f\|_\infty}{\sqrt{n}}\log(1 + |\mathcal{F}|) + \max_f \|f\|_{P,2}\sqrt{\log(1 + |\mathcal{F}|)}$$

が成立する．

証明：$a = 24\|f\|_\infty/\sqrt{n}$ と $b = 24Pf^2$ とおく．$x > b/a$ の場合には

$$\frac{1}{4}\frac{x^2}{Pf^2 + x\|f\|_\infty/\sqrt{n}} \ge \frac{6x^2}{a(x + b/a)} \ge \frac{3x}{a} \tag{6.2}$$

となり、$x \le b/a$ の場合には

$$\frac{1}{4}\frac{x^2}{Pf^2 + x\|f\|_\infty/\sqrt{n}} \ge \frac{6x^2}{b\{1 + (a/b)x\}} \ge \frac{3x^2}{b} \tag{6.3}$$

が成立する．打ち切り型の確率変数

$$A_f = G_n 1\{|G_n f| > b/a\} \quad \text{と} \; B_f = G_n 1\{|G_n f| \le b/a\}$$

[7] f が有界であることに注意すれば、和と積分の交換が保障されることがわかる．

を考える．Bernstein の不等式と (6.2) と (6.3) を用いれば、

$$P\{|A_f| > x\} \leq 2\exp\left(\frac{-3x}{a}\right) \quad と \quad P\{|B_f| > x\} \leq 2\exp\left(\frac{-3x^2}{b}\right) \tag{6.4}$$

が成立する．$\psi_p = \exp x^p - 1$ として、Fubini の定理[8] と (6.4) の第一の不等式を利用すれば、

$$\boldsymbol{E}\psi_1\left(\frac{|A_f|}{a}\right) = \boldsymbol{E}\int_0^{|A_f|/a} e^x dx = \int_0^\infty P\{|A_f| > ax\}e^x dx \leq 1$$

がわかる．同様の議論[9] から

$$\boldsymbol{E}\psi_2\left(\frac{|B_f|}{\sqrt{b}}\right) \leq 1$$

を得る．

ψ_1 は非負凸関数なので、Jensen の不等式を用いれば、

$$\psi_1\left(\boldsymbol{E}\max_f \frac{|A_f|}{a}\right) \leq \boldsymbol{E}\psi_1\left(\max_f \frac{|A_f|}{a}\right) \leq \boldsymbol{E}\sum_f \psi_1\left(\max_f \frac{|A_f|}{a}\right) \leq |\mathcal{F}|$$

を得る．更に、$\psi_1^{-1}(u) = \log(1+u)$ は単調増加関数であることより、

$$\boldsymbol{E}\left(\max_f \frac{|A_f|}{a}\right) \leq \log(1+|\mathcal{F}|)$$

となり、

$$\boldsymbol{E}\max_f |A_f| \lesssim \frac{\|f\|_\infty}{\sqrt{n}}\log(1+|\mathcal{F}|)$$

を得る．同様な議論[10] より、

$$\boldsymbol{E}\max_f |B_f| \lesssim \max_f \|f\|_{P,2}\sqrt{\log(1+|\mathcal{F}|)}$$

がわかる．よって、補題は証明された． □

補題 6.4 ：δ をある正の数とする．$\mathcal{F}$ を可測関数 $f : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ で $Pf^2 < \delta^2$ を満たすものからなる族とする．このとき、

$$\boldsymbol{E}^*\|G_n\|_{\mathcal{F}} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(P)) + \sqrt{n}PF1\{F > \sqrt{n}a(\delta)\}$$

が成立する．ただし、$a(\delta) = \delta/\sqrt{1+\log N_{[]}(\delta, \mathcal{F}, L_2(P))}$ とし、F は $\mathcal{F}$ の envelope 関数とした．

---

[8] $\boldsymbol{E}\psi_i(u) = \int_0^\infty \int_0^u e^x dx dP = \int_0^\infty \int_x^\infty dP dx = \int_0^\infty P\{u > x\}dx$ に注意．

[9] $\boldsymbol{E}\psi_2\left(\frac{|B_f|}{\sqrt{b}}\right) = \boldsymbol{E}\int_0^{|B_f|^2/b} e^x dx = \int_0^\infty P\{|B_f| > \sqrt{bx}\}e^x dx \leq 1$

[10]

$$\psi_2\left(\boldsymbol{E}\max \frac{|B_f|}{\sqrt{b}}\right) \leq \boldsymbol{E}\psi_2\left(\max \frac{|B_f|}{\sqrt{b}}\right) \leq \boldsymbol{E}\sum_f \psi_2\left(\frac{|B_f|}{\sqrt{b}}\right) \leq |\mathcal{F}|$$

より、$\boldsymbol{E}\left(\max_f \frac{|B_f|}{\sqrt{b}}\right) \leq \sqrt{\log(1+|\mathcal{F})}$ となる． よって、$\boldsymbol{E}\max_f |B_f| \lesssim \max_f \sqrt{b}\sqrt{\log(1+|\mathcal{F}|)} \lesssim \max_f \|f\|_{P,2}\sqrt{\log(1+|\mathcal{F}|)}$ がわかる

証明：$|G_n f| \le \sqrt{n}(P_n + P)F \le 2\sqrt{n}F$ に注意すれば、

$$\boldsymbol{E}^*\|G_n f 1\{F > \sqrt{n}a(\delta)\}\|_{\mathcal{F}} \le 2\sqrt{n}PF1\{F > \sqrt{n}a(\delta)\}$$

となることがわかる．したがって、$\boldsymbol{E}^*\|G_n f 1\{F \le \sqrt{n}a(\delta)\}\|_{\mathcal{F}}$ を評価すればよい．関数族 $\{f1\{F \le \sqrt{n}a(\delta)\} : f \in \mathcal{F}\}$ の bracketing number は $\mathcal{F}$ よりも小さいので、一般性を失わず、すべての $f \in \mathcal{F}$ に対して $|f| \le \sqrt{n}a(\delta)$ が成り立つとしてよい．

$\delta < 2^{-q_0} < 2\delta$ を満足する整数 q_0 をひとつ選び、固定する．$q \ge q_0$ とし、いれこ状に細分化した排反な N_q 個の部分集合による $\mathcal{F}$ の分割、すなわち $\mathcal{F} = \bigcup_{i=1}^{N_q} \mathcal{F}_{qi}$ が存在し、

$$\begin{aligned} \sum_{q \ge q_0} 2^{-q} \log \sqrt{N_q} \;&\lesssim\; \int_0^\delta \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))d\epsilon} \\ \sup_{f,g \in \mathcal{F}_{qi}} |f-g| \;&\le\; \Delta_{qi}, \qquad P\Delta_{qi}^2 \le 2^{-2q} \end{aligned}$$

を満足するようにできる．なぜなら、$\mathcal{F}$ のサイズ $2^{-\delta}$ の $L_2(P)$ – bracket で覆い、二つの分割の積集合でさらに細分化すれば、$N_q \le 2^{-q} \times 2^{-q}$ となる．更に、$\bigcap_{j=q_0}^{q} F_{j,i_j}$ と細分化すれば、$\bar{N}_q = N_{q_0}N_{q_0+1} \times \cdots \times N_q$ が分割の個数となる．

それぞれの $q \ge q_0$ に対して、$\mathcal{F}_{qi}$ からひとつの元 f_{qi} を選び固定し、

$$f \in \mathcal{F}_{qi} \text{ ならば } \pi_q f_{qi} = f_{qi}, \qquad \Delta_q f = \Delta_{qi}$$

とおく．固定した n と $q \ge q_0$ に対し、

$$\begin{aligned} a_q &= 2^{-q}/\sqrt{\log N_{q+1}} \\ A_{q-1}f &= 1\{\Delta_{q_0}f \le \sqrt{n}a_{q_0}, \ldots, \Delta_{q-1}f \le \sqrt{n}a_{q-1}\} \\ B_q f &= 1\{\Delta_{q_0}f \le \sqrt{n}a_{q_0}, \ldots, \Delta_{q-1}f \le \sqrt{n}a_{q-1}, \Delta_q f > \sqrt{n}a_q\} \end{aligned}$$

とおく．すると、$A_q f$ と $B_q f$ は $\mathcal{F}_{qi}$（それぞれの q について）上で一定となる．また、定義より、$A_{q_0}f = 1$ となる[11]．

いま、

$$f - \pi_{q_0}f = \sum_{q=q_0+1}^{\infty}(f - \pi_q f)B_q f + \sum_{q=q_0+1}^{\infty}(\pi_q f - \pi_{q-1}f)A_{q-1}f$$

と分解する[12]．分割がいれこ状であることより、

$$\Delta_q f B_q f \le \Delta_{q-1} f B_q f \le \sqrt{n}a_{q-1} \tag{6.5}$$

となる[13]．また、$B_q f^2 = B_q f \le 1$ より

$$\|\Delta_q f B_q f\|_{P,2} = \sqrt{P(\Delta_q f)^2 B_q f} \le \sqrt{\max P\Delta_{qi}^2} \le 2^{-q} \tag{6.6}$$

---

[11] $\delta < 2^{-q_0}$ から、$N_{q_0} \ge N_{[\,]}(\delta, \mathcal{F}, L_2(P))$ よりわかる．

[12] すべての q に対して、 $B_q f = 1$ である（この場合には、$A_q f = 1$ となる）か、ある $q_1 = q_1(f, x)$ が存在して、$B_{q_1} f = 1$ となる（この場合には、$A_q f = 1 (g \ge q_1)$ で $A_q f = 0$（その他）となる）かのいずれかである．

[13] $Bqf = 1$ ならば、定義より $\Delta_{q-1} f \le \sqrt{n}a_{q-1}$ となる．その他の場合はゼロ．

となる．$|f| \le g$ ならば、

$$|G_n g| \le G_n g + 2\sqrt{n} Pg$$

であることと $|f - \pi_q f| \le \Delta_q f$ であることに注意すれば、

$$\boldsymbol{E}^* \| \sum_{q=q_0+1}^{\infty} G_n (f - \pi_q f) B_q f \|_{\mathcal{F}} \le \sum_{q=q_0+1}^{\infty} \boldsymbol{E}^* \| G_n \Delta_q f B_q f \|_{\mathcal{F}} + \sum_{q=q_0+1}^{\infty} 2\sqrt{n} \| P \Delta_q f B_q f \|_{\mathcal{F}}$$

となる．前の補題を用いて、(6.5) と (6.6) を使えば、

$$\begin{aligned} \boldsymbol{E}^* \| G_n \Delta_q f B_q f \|_{\mathcal{F}} &\le \max \frac{\| \Delta_q f B_q f \|_\infty}{\sqrt{n}} \log(1 + N_q) \\ &\quad + \max \| \Delta_q f B_q f \|_{P,2} \sqrt{\log(1 + N_q)} \\ &\lesssim a_{q-1} \log N_q + 2^{-q} \sqrt{\log(1 + N_q)} \end{aligned}$$

となる．また、$B_q f = 1$ のとき、$\Delta_q f > \sqrt{n} a_q$ から

$$\sqrt{n} \| P \Delta_q f B_q f \|_{\mathcal{F}} \le \frac{1}{a_q} \| P (\Delta_q f)^2 B_q f \|_{\mathcal{F}} \le \frac{1}{a_q} 2^{-2q}$$

となる．これらより、

$$\begin{aligned} \boldsymbol{E}^* \| \sum_{q=q_0+1}^{\infty} G_n (f - \pi_q f) B_q f \|_{\mathcal{F}} &\lesssim \sum_{q=q_0+1}^{\infty} \left[a_{q-1} \log N_q + 2^{-q} \sqrt{\log N_q} + \frac{1}{a^q} 2^{-2q} \right] \\ &\lesssim \sum_{q=q_0+1}^{\infty} \frac{1}{2^{-q}} \sqrt{\log N_q} \end{aligned} \tag{6.7}$$

を得る．

次に、$A_{q-1} f = 1$ 上で $\Delta_{q-1} f \le \sqrt{n} a_{q-1}$ に注意すれば、

$$|\pi_q f - \pi_{q-1} f| A_{q-1} f \le |\Delta_{q-1} f| A_{q-1} f \le \sqrt{n} a_{q-1}$$

となる．また、

$$\| \pi_q f - \pi_{q-1} f \|_{P,2} \le \| \Delta_{q-1} f \|_{P,2} \le 2^{-q+1}$$

となる．前の補題を用いて、さらに上の不等式に注意して同様な議論を行えば、

$$\begin{aligned} &\boldsymbol{E}^* \| \sum_{q=q_0+1}^{\infty} G_n (\pi q f - \pi_{q-1} f) A_{q-1} f \|_{\mathcal{F}} \\ &\lesssim \sum_{q=q_0+1}^{\infty} \left[a_{q-1} \log N_q + 2^{-q} \sqrt{\log N_q} + \frac{1}{a^q} 2^{-2q} \right] \\ &\lesssim \sum_{q=q_0+1}^{\infty} \frac{1}{2^{-q}} \sqrt{\log N_q} \end{aligned} \tag{6.8}$$

を得る．

最後に、$|\pi_{q_0} f| \leq F \leq a(\delta)\sqrt{n} \leq \sqrt{n} a_{q_0}$ と $P(\pi_{q_0} f)^2 \leq \delta^2$ に注意して、前の補題を再度利用すれば、

$$\begin{aligned} \boldsymbol{E}^* \|G_n \pi_{q_0} f\|_{\mathcal{F}} &\lesssim \max \frac{|\pi_{q_0} f|}{\sqrt{n}} \log(1 + N_{q_0}) + \max \|\pi_{q_0} f\|_{P,2} \sqrt{\log(1 + N_{q_0})} \\ &\lesssim \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_q} \end{aligned} \tag{6.9}$$

を得る．よって、(6.7) – (6.9) と三角不等式を用いれば、

$$\begin{aligned} \boldsymbol{E}^* \|G_n f 1\{F \leq \sqrt{n} a(\delta)\}\|_{\mathcal{F}} &\lesssim \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_q} \\ &\lesssim \int_0^{\infty} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon \end{aligned}$$

が成立することがわかる．よって、補題は証明された．

□

系 6.1 ：$\mathcal{F}$ を可測関数の任意の族とし、その envelope 関数を F とする．このとき、

$$\boldsymbol{E}_P^* \|G_n\|_{\mathcal{F}} \lesssim J_{[]}(\|F\|_{P,2}, \mathcal{F}, L_2(P))$$

が成立する．

証明：$\mathcal{F}$ はひとつの bracket $[-F, F]$ に含まれることから、$\delta = 2\|F\|_{P,2}$ とおけば、

$$N_{[]}(\delta, \mathcal{F}, L_2(P)) = 1$$

となることから、$J_{[]}(\delta, \mathcal{F}, L_2(P)) = 0$ となる．$a(\delta) = \eta \|F\|_{P,2}$（$\eta$ はある定数）となり、Markov の不等式を用いれば、

$$\begin{aligned} \sqrt{n} P F 1\{F > \sqrt{n} a(\delta)\} &\leq \sqrt{PF^2} \sqrt{n P 1\{F > \sqrt{n} a(\delta)\}} \\ &= \|F\|_{P,2} \sqrt{n \frac{1}{n a(\delta)^2} P F^2} \lesssim \|F\|_{P,2} \end{aligned}$$

となり、前の補題から系は証明された．□

補題 6.5 ：$\mathcal{F}$ を $\boldsymbol{X}$ 上の実数値可測関数の任意の族とする．ある正の数 δ と M に対して、、すべての $f \in \mathcal{F}$ は $Pf^2 < \delta^2$ と $\|f\|_\infty < M$ を満足するものとする．このとき、

$$\boldsymbol{E}^* \|G_n\|_{\mathcal{F}} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(P)) \left(1 + \frac{J_{[]}(\delta, \mathcal{F}, L_2(P))}{\delta^2 \sqrt{n}} M\right)$$

が成立する．

証明：VdVW の補題 3.4.2 および 2.14 節を参照されたい． □

$L_2(P)$ – norm の代わりに Bernstein norm

$$\|f\|_{P,B} = 2P(e^{|f|} - 1 - |f|)$$

を用いた最大不等式がある．

補題 6.6 ：$\mathcal{F}$ を $\boldsymbol{X}$ 上の実数値可測関数の任意の族とする．ある正の数 δ に対して、すべての $f \in \mathcal{F}$ が $\|f\|_{P,B} < \delta$ を満足するものとする．このとき、

$$\boldsymbol{E}_P^* \|G_n\|_{\mathcal{F}} \lesssim J_{[]}(\delta, \mathcal{F}, \|\cdot\|_{P,B}) \left(1 + \frac{J_{[]}(\delta, \mathcal{F}, \|\cdot\|_{P,B})}{\delta^2 \sqrt{n}}\right)$$

が成立する．

証明：VdVW の補題 3.4.3 および節 2.14 を参照されたい． □

bracketing entropy の代わりに uniform entropy を用いた最大不等式がある．$N(\epsilon, \mathcal{F}, L_2(Q))$ を $\mathcal{F}$ を覆うために必要な半径 ϵ の $L_2(Q)$ – ball の最小の個数とし、uniform entropy integral を

$$J(\delta, \mathcal{F}, L_2) = \int_0^\infty \sqrt{\log \sup_Q N(\epsilon \|F\|_Q, \mathcal{F}, L_2(Q))} d\epsilon$$

で定義する．ただし、δ は正の数とする．

補題 6.7 ：$\mathcal{F}$ を $\boldsymbol{X}$ 上の実数値可測関数の任意の族とし、その envelope 関数を F とする．このとき、

$$\boldsymbol{E}_P^* \|G_n\|_{\mathcal{F}} \lesssim J(1, \mathcal{F}, L_2) \|F\|_{P,2}$$

が成立する．

証明：VdVW の定理 2.14.1 を参照されたい． □

6.2.3 Glivenko – Cantelli および Donsker の定理

定理 6.5 （Glivenko – Cantelli の定理）：$\mathcal{F}$ を実数値可測関数の族とする．このとき、どんな正の数 $\epsilon > 0$ に対しても、$N_{[]}(\epsilon, \mathcal{F}, L_2(P)) < \infty$ ならば、$\mathcal{F}$ は P – Glivenko – Cantelli である．

定理 6.6 （Donsker の定理）：$\mathcal{F}$ を実数値可測関数の族とする．このとき、

$$J_{[]}(1, \mathcal{F}, L_2(P)) < \infty$$

ならば、$\mathcal{F}$ は P – Donsker である．

証明：$\mathcal{G} = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$ とする．$\mathcal{F}$ の ϵ − brackets $[u_i, l_i]$ を用いて、$[l_i - u_j, u_i - l_j]$ とすれば、これは $\mathcal{G}$ の $e\epsilon$ − brackets となることがわかる．したがって、$N_{[]}(\epsilon, \mathcal{G}, L_2(P)) \leq N_{[]}(\epsilon/2, \mathcal{F}, L_2(P))$ がわかる．

$\delta > 0$ を固定する．$\mathcal{F}$ の分割 $\mathcal{F} = \bigcup_{i=1} \mathcal{F}_i$ をとり、$\mathcal{F}_i$ の半径が $L_2(P)$ − norm に関して ϵ 以下になるようにする．$\mathcal{G}$ の部分集合として $\mathcal{G}_i = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}_i\}$ をとる．

$$a(\delta) = \frac{\delta}{\sqrt{\log N_{[]}(\delta, \mathcal{G}, L_2(P))}} \leq \frac{\delta}{\sqrt{\log N_{[]}(\frac{\delta}{2}, \mathcal{F}, L_2(P))}} < \infty$$

とすれば、補題より

$$\begin{aligned} &\boldsymbol{E}^* \sup_i \sup_{f_i, f_2 \in \mathcal{F}} |\mathbb{G}_n(f_1 - f_2)| \lesssim J_{[]}(\delta, \mathcal{G}, L_2(P)) + \sqrt{n} P 2F 1\{2F > a(\delta)\sqrt{n}\} \\ &\lesssim J_{[]}(\frac{\delta}{2}, \mathcal{F}, L_2(P)) + \sqrt{n} P 2F 1\{F > \frac{1}{2} a(\delta)\sqrt{n}\} \end{aligned} \tag{6.10}$$

が成立する．ただし、F は $\mathcal{F}$ の envelope 関数である．

Cauchy − Schwarz の不等式[14] を用いて

$$\begin{aligned} &\sqrt{n} P^* F 1\{F > (1/2)a(\delta)\sqrt{n}\} \\ &\leq \sqrt{n} \sqrt{P^* 1\{F > (1/2)a(\delta)\sqrt{n}\}} \sqrt{P^* F^2 1\{F > (1/2)a(\delta)\sqrt{n}\}} \\ &\leq \sqrt{n} \sqrt{\frac{4}{a(\delta)^2} \frac{1}{n} \|F\|^2_{P,2}} \sqrt{P^* F^2 1\{F > (1/2)a(\delta)\sqrt{n}\}} \\ &\lesssim \frac{1}{a(\delta)} \sqrt{P^* F^2 1\{F > (1/2)a(\delta)\sqrt{n}\}} \to 0 \qquad (n \to \infty \text{ のとき}) \end{aligned}$$

となる．したがって、$\delta \downarrow 0$ のとき、(6.10) の右辺はゼロの収束する．

最後に、Markov の不等式から、すべての $\epsilon > 0$ に対しても、$n \to \infty$ とすれば、

$$P^*(\sup_i \sup_{g \in \mathcal{G}_i} |\mathbb{G}_n g|) \leq \frac{1}{\epsilon} \boldsymbol{E}^* \sup_i \sup_{g \in \mathcal{G}_i} |\mathbb{G}_n g| \to 0$$

が成立するので、定理からこの定理は証明された．□

6.2.4 Random Functions

モデル P からの観測値 $X_1(\omega), \ldots, X_n(\omega)$ に依存する実数値関数

$$x \longrightarrow \hat{f}_n(x; \omega)$$

を **random functions** とよぶことにする．多くの例では、

$$\hat{f}_n(x) = \hat{f}_n(x; X_1, X_2, \ldots, X_n)$$

のように書ける．

[14] $\int ghdP \leq \int g^2 dP \int h^2 dP$ に対して、$g = 1\{F > (1/2)a(\delta)\sqrt{n}\}$ と $h = F1\{F > (1/2)a(\delta)\sqrt{n}\}$ して適用すればよい．

定理 6.7 ：$\mathcal{F}$ を可測関数の族で、P – Donsker class とする．$\hat{f}_n$ を random functions （$\hat{f}_n \in \mathcal{F}$ ）の列とし、ある $f_0 \in L_2(P)$ が存在して、

$$\int \{\hat{f}_n(x) - f_0(x)\}^2 dP(x) \xrightarrow{P} 0$$

を満足するとする．このとき、

$$G_n(\hat{f}_n - f_0) \xrightarrow{P} 0$$

が成立する．したがって、

$$G_n f_n \rightsquigarrow G_P f_0$$

である．

証明：一般性を失わずに $f_0 \in \mathcal{F}$ としてよい[15] ．関数 $g : l^\infty(\mathcal{F}) \times \mathcal{F} \longrightarrow \boldsymbol{R}$ を

$$g(z, f) = z(f) - z(f_0)$$

で定義する． すると、g は $f \mapsto z(f)$ が連続となる点 (z, f) において連続となる[16] ． 実際、$(z_n, \hat{f}_n) \to (z, f)$ とすれば、$\|z_n - z\|_\mathcal{F} \to 0$ から

$$z_n(\hat{f}_n) = z(\hat{f}_n) + o(1) \to z(f)$$

となることよりわかる．

仮定から $\hat{f}_n \xrightarrow{P} f$ と $G_n \rightsquigarrow G_P$ （$l^\infty(\mathcal{F})$ の中のおいて）が成立するので、$l^\infty(\mathcal{F}) \times \mathcal{F}$ の中において $(G_n, \hat{f}_n) \rightsquigarrow (G_P, f_0)$ がわかる．G_P は $\mathcal{F}$ 上でほとんど確実に連続であるので、g は (G_P, f_0) において連続である．これより、連続写像定理を用いれば、

$$G_n(\hat{f}_n - f_0) = g(G_n, \hat{f}_n) \rightsquigarrow g(G_P, f_0) = 0$$

がわかる．収束先が退化した分布の場合、分布収束と確率収束が同値なので、補題は証明された． □

定理 6.8 ：それぞれの正の整数 n に対して、$\mathcal{F}_n = \{f_{n,t} : t \in T\}$ を可測関数 $f : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ の族とし、その添字集合 T はある距離 ρ に関して全有界とする． また、F_n を $\mathcal{F}_n$ の envelope 関数とする．更に、

$$\begin{aligned} &PF_n^2 = O(1) \\ &PF_n^2 1\{F_n > \epsilon\sqrt{n}\} \qquad （すべての正の数\ \epsilon >\ ） \\ &\sup_{\rho(s,t)<\delta_n} P(f_{n,s} - f_{n,t})^2 \to 0 \qquad （\delta_n \downarrow 0\ のとき） \end{aligned} \tag{6.11}$$

を仮定する．このとき、すべての $\delta_n \downarrow 0$ に対して

$$J_{[]}(\delta_n, \mathcal{F}_n, L_2(P)) \to 0 \quad または \quad J(\delta_n, \mathcal{F}_n, L_2) \to 0$$

が成立し、$Pf_{n,s}f_{n,t} - Pf_{n,s}Pf_{n,t}$ が $T \times T$ 上の各点で収束するならば、$\{G_n f_{n,t} : t \in T\}$ はある tight な Gauss 過程に分布収束する．

[15] $f_0 \notin \mathcal{F}$ ならば、$\mathcal{F}' = \mathcal{F} \bigcup \{f_0\}$ とすればよい．しかし、$\mathcal{F}'$ の metric entropy は定数倍しか変化しない．

[16] 任意の $z \in l^\infty(\mathcal{F})$ に対して、norm を $\|z\|_\mathcal{F} = \sup_{f\in\mathcal{F}} |z(f)| < \infty$ で定義すれば、$l^\infty(\mathcal{F}) \times \mathcal{F}$ の norm を $\|\cdot\|_\mathcal{F} \wedge \|\cdot\|_{P,2}$ で定義すればよい．

証明：(T,ρ) が全有界であることから、どんな正の数 $\delta>0$ に対しても、有限個の $T_1,\dots,T_k$ が存在して、十分大きなすべての $\epsilon>0$ に対して

$$\sup_i \sup_{s,t\in T_i} P(f_{n,s}-f_{n,t})^2<\delta^2$$

とできる．よって、補題より

$$\boldsymbol{E}\sup_i \sup_{s,t\in T_i}|G_n(f_{n,s}-f_{n,t})|\lesssim J_{[]}(\delta,\mathcal{F}_n,L_2(P))+\frac{2PF_n^2 1\{F_n>a_n(\delta)\sqrt{n}\}}{a_n(\delta)} \tag{6.12}$$

が成立する．ただし、

$$a_n(\delta)=\frac{\delta}{\sqrt{\log N_{[]}(\delta,\mathcal{F}_n\times\mathcal{F}_n,L_2(P))}}$$

である．$\log N_{[]}(\delta,\mathcal{F}_n\times\mathcal{F}_n,L_2(P))$ は $\log N_{[]}(\delta,\mathcal{F}_n,L_2(P))$ の定数倍であることとどんな $\delta_n\downarrow 0$ にたいしても $\log N_{[]}(\delta_n,\mathcal{F}_n,L_2(P))\downarrow 0$ なので、どんな $\delta>0$ に対しても $\log N_{[]}(\delta,\mathcal{F}_n,L_2(P))=O(1)$ が成立するので、ある正の数 $\gamma>0$ が存在して、$a_n(\delta)>\gamma>0$ となる．したがって、固定したどんな $\delta>0$ に対しても、$n\to\infty$ のとき、

$$\frac{PF_n^2 1\{F_n>a(\delta)\sqrt{n}\}}{a(\delta)}\to 0$$

となる．また、仮定より (6.12) の右辺の第一項目もゼロに収束する．よって、定理は証明された． □

6.3　有効推定量と汎関数デルタ法

6.3.1　Hadamard 微分可能な汎関数

$\boldsymbol{D}$ と $\boldsymbol{M}$ をノルム空間とし、そのノルムを $\|\cdot\|_D$ と $\|\cdot\|_M$ と記すことにする．写像 $\phi:\boldsymbol{D}\longrightarrow\boldsymbol{M}$ に関する微分可能性には次の 3 つが考えられる．

Gateaux 微分可能性：ϕ が $\theta\in\boldsymbol{D}$ において **Gateaux 微分可能**であるとは、固定したどんな $h\in\boldsymbol{D}$ に対しても、連続線形写像 $\dot{\phi}_\theta:\boldsymbol{D}\longrightarrow\boldsymbol{M}$ が存在して、$t\downarrow 0$ のとき、

$$\left\|\frac{\phi(\theta+th)-\phi(\theta)}{t}-\dot{\phi}_\theta(h)\right\|_M\to 0 \tag{6.13}$$

を満足すること[17] である．

Hadamard 微分可能性：$\boldsymbol{D}_\phi$ をノルム空間 $\boldsymbol{D}$ の部分集合で θ を含むものとする．写像 $\phi:\boldsymbol{D}_\phi\longrightarrow\boldsymbol{M}$ が θ において **Hadamard 微分可能**であるとは、連続線形写像 $\dot{\phi}_\theta:\boldsymbol{D}\longrightarrow\boldsymbol{M}$ が存在して、$t\downarrow 0$ のとき、$h_t\to h\in\boldsymbol{D}_\phi$ となるような $\boldsymbol{D}_\phi$ のなかの列 $\{h_t\}$ で、十分小さな t に対して $\theta+th\in\boldsymbol{D}_\phi$ を満足するどんな $\{h_t\}$ に対しても、$t\downarrow 0$ のとき、

$$\left\|\frac{\phi(\theta+th_t)-\phi(\theta)}{t}-\dot{\phi}_\theta(h)\right\|_M\to 0 \tag{6.14}$$

を満足することである．

[17] これを $\phi(\theta+th)-\phi(\theta)=t\dot{\phi}_\theta(h)+o(t)$ と記すことにする．

注意 6.1 ：$\dot{\phi}_\theta$ は $\boldsymbol{D}$ 上のすべての領域で定義されていることを要請しているが、$\dot{\phi}_\theta$ が $\boldsymbol{D}$ の部分空間 $\boldsymbol{D}_0$ においてのみ定義されている場合には、$h \in \boldsymbol{D}_0$ に収束するものだけに限定した列 $\{h_t : h_t \in \boldsymbol{D}_\phi\}$ に対して (6.14) を満足すれば、ϕ は $\boldsymbol{D}_0$ **に接して** θ **において** Hadamard **微分可能** であるという．

注意 6.2 ：(6.14) の同値条件はどんなコンパクトな部分集合 $K \subset \boldsymbol{D}$ に対しても、$t \downarrow 0$ のとき、

$$\sup_{h \in K} \left\| \frac{\phi(\theta + th) - \phi(\theta)}{t} - \dot{\phi}_\theta(h) \right\|_M \to 0$$

が成立することである．このことより、Hadamard 微分は**コンパクト微分**とも呼ばれる．

Fréchet 微分可能性：$\boldsymbol{D}_\phi$ を $\boldsymbol{D}$ の部分空間とする．写像 $\phi : \boldsymbol{D}_\phi \longrightarrow \boldsymbol{M}$ が $\theta \in \boldsymbol{D}_\phi$ において **Fréchet 微分可能** であるとは、連続線形写像 $\dot{\phi}_\theta : \boldsymbol{D} \longrightarrow \boldsymbol{M}$ が存在して、$\|h\| \downarrow 0$（$h \in \boldsymbol{D}_\phi$）のとき、

$$\|\phi(\theta + h) - \phi(\theta) - \dot{\phi}_\theta(h)\|_M = o(\|h\|_D)$$

が成立することである．

6.3.2　デルタ法

定理 6.9 ：$\boldsymbol{D}$ と $\boldsymbol{M}$ をノルム空間とする．$\boldsymbol{D}_\phi$ と $\boldsymbol{D}_0$ を $\boldsymbol{D}$ の部分空間とする．写像 $\phi : \boldsymbol{D}_\phi \longrightarrow \boldsymbol{M}$ は $\boldsymbol{D}_0$ に接して θ において Hadamard 微分可能（その微分 $\dot{\phi}_\theta$ と記す）とする．T_n と T を $\boldsymbol{D}_\phi$ – 値ランダム要素とし、$n \to \infty$ としたとき、$r_n \to \infty$ となるような数列 $\{r_n\}$ に対して、$r_n(T_n - \theta) \rightsquigarrow T$ が成立するとする．このとき、

$$r_n(\phi(T_n) - \phi(\theta)) \rightsquigarrow \dot{\phi}_\theta(T) \tag{6.15}$$

が成立する．更に、$\dot{\phi}_\theta$ が $\boldsymbol{D}$ 上で定義された連続汎関数であれば、

$$r_n(\phi(T_n) - \phi(\theta)) = \dot{\phi}_\theta(r_n(T_n - \theta)) + o_P^*(1) \tag{6.16}$$

も成立する．

証明：(6.15) を示すために、自然数 n に対して、$\boldsymbol{D}_n = \{h : \theta + r_n^{-1}h \in \boldsymbol{D}_\phi\}$ 上の写像 $f_n(h) = r_n(\phi(\theta + r_n^{-1}h) - \phi(\theta))$ を導入する．Hadamard 微分可能性より、任意の部分列 $\{h_{n'}\}$ で $h_{n'} \to h \in \boldsymbol{D}_0$ となるものに対して、$f_{n'}(h_{n'}) \to \dot{\phi}_\theta(h)$ となる．したがって、連続写像定理を利用すれば、$f_{n'}(r_{n'}(T_{n'} - \theta)) \rightsquigarrow \dot{\phi}_\theta(T)$ が成立すること[18] がから、(6.15) がわかる．

---

[18] $f_{n'}(r_{n'}(T_{n'} - \theta)) = r'_n(\phi(T_{n'} - \phi(\theta))$ となることに注意．

(6.16) を示すために、写像 $\psi = (\phi, \dot{\phi}_\theta) : \boldsymbol{D}_\phi \longrightarrow \boldsymbol{M} \times \boldsymbol{M}$ を導入する．ψ は θ において Hadamard 微分可能で、微分 $\dot{\psi}_\theta = (\dot{\phi}_\theta, \dot{\phi}_\theta)$ をもつこと[19] がわかる．$\boldsymbol{M} \times \boldsymbol{M}$ の中において、$r_n(\psi(T_n) - \psi(\theta)) \rightsquigarrow \dot{\psi}_\theta(T) = (\dot{\phi}_\theta(T), \dot{\phi}_\theta(T))$ であることが (6.15) よりわかる．再度、連続写像定理を写像 $g(x,y) = x - y$ に適用すれば、

$$\begin{aligned} g\{r_n(\psi(T_n) - \psi(\theta))\} &= g\{r_n(\phi(T_n) - \phi(\theta)), r_n(\dot{\phi}_\theta(T_n) - \dot{\phi}_\theta(\theta))\} \\ &= r_n(\phi(T_n) - \phi(\theta)) - r_n(\dot{\phi}_\theta(T_n) - \dot{\phi}_\theta(\theta)) \\ &= r_n(\phi(T_n) - \phi(\theta)) - \dot{\phi}_\theta(r_n(T_n - \theta)) \\ &\rightsquigarrow g(\dot{\phi}_\theta(T), \dot{\phi}_\theta(T)) = \dot{\phi}_\theta(T) - \dot{\phi}_\theta(T) = 0 \end{aligned}$$

となることがわかる．極限が退化したときには、確率収束と分布収束が同値なので、(6.16) が証明された． □

補題 6.8 （連鎖微分）: $\boldsymbol{D}$、$\boldsymbol{F}$、$\boldsymbol{M}$ をノルム空間とし、$\boldsymbol{D}_\phi$ と $\boldsymbol{F}_\psi$ を $\boldsymbol{D}$ と $\boldsymbol{F}$ の部分集合とする．$\phi : \boldsymbol{D}_\phi \longrightarrow \boldsymbol{F}_\psi$ と $\psi : \boldsymbol{F}_\psi \longrightarrow \boldsymbol{M}$ を写像とする．ϕ は $\boldsymbol{D}_0$ に接して θ において Hadamard 微分可能（その微分を $\dot{\phi}_\theta$ と記す）で、ψ は $\dot{\phi}_\theta(\boldsymbol{D}_0)$ に接して $\phi(\theta)$ において Hadamard 微分可能（その微分を $\dot{\psi}_{\phi(\theta)}$ と記す）であるとする．合成写像 $\psi \circ \phi : \boldsymbol{D}_\phi \longrightarrow \boldsymbol{M}$ は $\boldsymbol{D}_0$ に接して θ において Hadamard 微分可能で微分 $\dot{\psi}_{\phi(\theta)} \circ \dot{\phi}_\theta$ を持つ．

証明: $h_t \to h \in \boldsymbol{D}_\phi$ に収束する $\boldsymbol{D}_\phi$ の中の任意の列 $\{h_t\}$ を取り、$g_t = t^{-1}(\phi(\theta + th_t) - \phi(\theta))$ とおく．すると、

$$\frac{\psi \circ \phi(\theta + th_t) - \psi \circ \phi(\theta)}{t} = \frac{\psi(\phi(\theta) + tg_t) - \psi(\theta)}{t}$$

と書け、ϕ の Hadamard 微分可能性から $g_t \to \dot{\phi}_\theta(h)$ が成立することに注意して、今度は ψ の Hadamard 微分可能性を用いれば、

$$\frac{\psi(\phi(\theta) + tg_t) - \psi(\theta)}{t} \to \dot{\psi}_{\phi(\theta)}(\dot{\phi}_\theta(h))$$

が成立することがわかる． □

---

[19] 実際、 $t \downarrow 0$ のとき、$h_t \to h \in \boldsymbol{D}$ なる $\boldsymbol{D}$ の中の列 $\{h_t\}$ に対しても、$\dot{\phi}_\theta$ の線形性に注意すれば、$t \downarrow 0$ のとき、

$$\begin{aligned} &\left\| \frac{\psi(\theta + th_t) - \psi(\theta)}{t} - \dot{\psi}_\theta(h) \right\|_{M \times M} \\ &= \left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \dot{\phi}_\theta(h) \right\|_M \vee \left\| \frac{\dot{\phi}_\theta(\theta + th_t) - \dot{\phi}_\theta(\theta)}{t} - \dot{\phi}_\theta(h) \right\|_M \\ &= \left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \dot{\phi}_\theta(h) \right\|_M \to 0 \end{aligned}$$

よりわかる．

6.3.3 有効性とデルタ法

$\boldsymbol{B}$ を Banach 空間とし、$\boldsymbol{B}^*$ をその双対空間[20] とする．モデルを $\boldsymbol{P}$ と記し、$\boldsymbol{B}$ – 値母数 $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ の推定を考える．T_n を母数 $\nu(P)$ の $\boldsymbol{B}$ – 値正則推定量とすれば、どんな $b^* \in \boldsymbol{B}^*$ に対しても、b^*T_n は実数値母数 $b^*\nu(P)$ の正則推定量となる．定義 5.14 にあるように、T_n が有効であるとは、真のモデル P のもとで、$\sqrt{n}(T_n - \nu(P))$ はある tight な極限に収束し、さらに、どんな $b^* \in \boldsymbol{B}^*$ に対しても、b^*T_n は P において $b^*\nu(P)$ の有効推定量であることである．また、ν が $P \in \boldsymbol{P}$ において微分可能であるとは、ある連続線形写像 $\dot{\nu}_P : L_2(P) \longrightarrow \boldsymbol{B}$ が存在して、任意の接ベクトル $h \in \dot{\boldsymbol{P}}$ をもつ[21] $\boldsymbol{P}$ の１次元母数部分モデル $\{P_\eta : eta$ は小さな正の実数 $\}$ に対して、$\eta \downarrow 0$ のとき、

$$\left\| \frac{\nu(P_\eta) - \nu(P)}{\eta} - \dot{\nu}_P(h) \right\|_B \to 0$$

を満足することである．上の式が成立すれば、どんな $b^* \in \boldsymbol{B}$ に対しも、$b^*\nu : \boldsymbol{P} \longrightarrow \boldsymbol{R}$ は P において微分可能となる．したがって、接空間 $\dot{\boldsymbol{P}}$ は Hilbert 空間であることから Riesz の表現定理を用いれば、どんな $b^* \in \boldsymbol{B}^*$ に対しても、$\dot{\boldsymbol{P}}$ の中にある関数 $\tilde{\nu}_{P,b^*} : \boldsymbol{X} \longrightarrow \boldsymbol{R}$ が存在して、すべての $h \in \dot{\boldsymbol{P}}$ に対して、

$$b^*\dot{\nu}_P(h) = \langle \tilde{\nu}_{P,b^*}, h \rangle_{L_2(P)} = P\tilde{\nu}_{P,b^*}h$$

を満足することがわかる．さらに、$b^*\nu$ の正則推定量 b^*T_n が有効であるための必要十分条件は

$$b^*\sqrt{n}(T_n - \nu(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\nu}_{P,b^*}(X_i) + o_P(1) \tag{6.17}$$

が成立することである．

定理 6.10 ：$\boldsymbol{B}$ と $\boldsymbol{M}$ を Banach 空間とし、その双対空間を $\boldsymbol{B}^*$ と $\boldsymbol{M}^*$ とする．写像 $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ は $\underline{P \in \boldsymbol{P}}$ において微分可能とし、$\boldsymbol{B}$ の部分集合 $\boldsymbol{B}_\phi$ の中に値を取るとし、$\phi : \boldsymbol{B}_\phi \longrightarrow \boldsymbol{M}$ は $\overline{[\dot{\nu}_P(\dot{\boldsymbol{P}})]}$ に接して $\nu(P)$ において Hadamard 微分可能であるとする．このとき、$\phi \circ \nu : \boldsymbol{P} \longrightarrow \boldsymbol{M}$ は P において微分可能である．さらに、T_n は $\boldsymbol{B}_\phi$ – 値推定量の列で、P において $\nu(P)$ の有効推定量ならば、$\phi(T_n)$ は P において $\phi \circ \nu(P)$ の有効推定量である．

証明：$\phi \circ \nu$ の可微分性は補題 6.8 よりわかる．また、その微分は $\dot{\phi}_{\nu(P)} \circ \dot{\nu}_P$ で与えられる．

つぎに、$\sqrt{n}(T_n - \nu(P))$ の極限分布 L は $\boldsymbol{B}$ の部分集合 $\overline{[\dot{\nu}_P(\dot{\boldsymbol{P}})]}$ に集中することを示す．Hahn – Banach の定理を利用すれば、どんな $S \subset \boldsymbol{B}$ に対しても

$$\overline{[\dot{\nu}_P(\dot{\boldsymbol{P}})]} \bigcap S = \bigcap_{b^* \in \boldsymbol{B}^* : b^*\dot{\nu}_P(\dot{\boldsymbol{P}}) = 0} \{b \in S : b^*b = 0\}$$

---

[20] すなわち、$\boldsymbol{B}$ 上の実数値線形写像の集まりの成す空間．

[21] すなわち、$h(x) = \frac{\partial}{\partial \eta} \log dP_\eta(x) \Big|_{\eta=0}$ が成り立つことである．

となる．S を可分集合とすれば、上の式の左辺の和は可算個の b^* に関する和とできる．L は tight なので、S に集中しており、すべて（可算個の）の b^* に対して、$L\{b \in \boldsymbol{B} : b^*b = 0\}$ ならば、

$$L\left\{\bigcap_{b^* \in \boldsymbol{B}^* : b^*\dot{\nu}(\dot{\boldsymbol{P}})=0} (b \in S : b^*b = 0)\right\} = \bigcap_{b^* \in \boldsymbol{B}^* : b^*\dot{\nu}(\dot{\boldsymbol{P}})=0} \{b \in S : b^*b = 0\} = 1$$

となる．極限分布は $N(0, \|\tilde{\nu}_{b^*,P}\|_{L_2(P)})$ である．しかし、どんな $h \in \dot{\boldsymbol{B}} \subset L_2(P)$ に対して、

$$0 = b^*\dot{\nu}_P(h) = \langle \tilde{\nu}_{b^*,P}, h\rangle_{L_2(P)}$$

が成立するので、$\tilde{\nu}_{b^*,P} = 0$ となり、L は退化した分布となるので、$L\{b \in S : b^*b = 0\} = 1$ となる．よって、ほとんど確実に $\sqrt{n}(T_n - \nu(P))$ は $\overline{[\dot{\nu}(\dot{\boldsymbol{P}})]}$ の中に値をとることがわかる．

ここで、汎関数デルタ法を使えば、$\sqrt{n}(\phi(T_n) - \phi \circ \nu(P))$ は tight な分布に収束することがわかる．さらに、どんな $m^* \in \boldsymbol{M}^*$ に対しても

$$\begin{aligned}
\sqrt{n}(m^*\phi(T_n) - m^*\nu(P)) &= m^*\{\sqrt{n}(\phi(T_n) - \phi \circ \nu(P))\} \\
&= m^*\{\dot{\phi}_{\nu(P)}(\sqrt{n}(T_n - \nu(P)) + o_P^*(1)\} \\
&= m^*\{\dot{\phi}_{\nu(P)}(\sqrt{n}(T_n - \nu(P))\} + o_P^*(1)
\end{aligned}$$

となる．さらに、$m^*\dot{\phi}_{\nu(P)} \in \boldsymbol{B}^*$ （$b^* = m^*\dot{\phi}_{\nu(P)}$ と記すことにする）であることと T_n が $\nu(P)$ の有効推定量であることから、

$$\sqrt{n}(m^*\phi(T_n) - m^*\phi \circ \nu(P)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{\nu}_{b^*,P}(X_i) + o_P^*(1) \tag{6.18}$$

となる．一方、$\phi \circ \nu$ の P における微分は $\dot{\phi}_{\nu(P)} \circ \dot{\nu}_P$ であることに注意すれば、任意の $h \in \dot{\boldsymbol{P}} \subset L_2(P)$ に対して、

$$m^*\dot{\phi}_{\nu(P)}\dot{\nu}_P(h) = b^*\dot{\nu}_P(h) = \langle \tilde{\nu}_{b^*,P}, h\rangle_{L_2(P)}$$

となることから、$m^*\phi\nu(P)$ を推定するときの有効影響関数は $\tilde{\nu}_{b^*,P} = \tilde{\nu}_{m^*\dot{\phi}_{\nu(P)},P}$ となる．このことと (6.18) から、$\phi(T_n)$ は $\phi \circ \nu(P)$ の有効推定量となる． □

以下、ある集合 S に対して、$\boldsymbol{B} = l^\infty(S)$ となる場合を考える．推定量の有効性を示すときにつぎの補題が有効である．

補題 6.9 ：写像 $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ は $P \in \boldsymbol{P}$ において微分可能であり、T_n を $\boldsymbol{B}$ – 値ランダム要素とする．ある部分集合 $\boldsymbol{B}_0^*(\subset \boldsymbol{B}^*)$ に含まれる b_o^* に対して、$b_0^*T_n$ は P において $b_0^*\nu(P)$ の有効推定量であり、ある定数 C が存在して、

$$\|b\|_B \leq C \sup_{b_0^* \in \boldsymbol{B}_0^* : \|b_0^*\|_{B^*} \leq 1} |b_0^*(b)| \tag{6.19}$$

が成立する．このとき、P のもとで、$\sqrt{n}(T_n - \nu(P))$ が漸近 tight ならば、T_n は P において $\nu(P)$ の有効推定量である．

証明：VdV の pages 389 – 390 を参照されたい． □

この補題を用いて、以下の場合の推定量の有効性が簡単にしめすことができる．

定理 6.11 （$l^\infty(S)$ における有効性）：写像 $\nu : \boldsymbol{P} \longrightarrow \boldsymbol{B}$ は $P \in \boldsymbol{P}$ において微分可能とし、T_n は $l^\infty(S)$ – 値ランダム要素とする．どんな $s \in S$ に対しても、$T_n(s)$ は P において $\nu(P)(s)$ の有効推定量とする．このとき、P のもとで、列 $\sqrt{n}(T_n - \nu(P))$ は $l^\infty(S)$ の中の tight な分布に分布収束するならば、T_n は P において $\nu(P)$ の有効推定量である．

証明：π_s を s における coordinate 射影とする．$\boldsymbol{B}_0^* = \{\pi_s : s \in S\}$ とおけば、どんな $b \in l^\infty(S)$ に対しても $\|b\|_\infty = \sup_{s\in S} |\pi_s b|$ となるので、(6.19) を満たすので、最前の補題より定理は証明される． □

$\boldsymbol{B}_1$ と $\boldsymbol{B}_2$ を Banach 空間とし、その双対空間を $\boldsymbol{B}_1^*$ と $\boldsymbol{B}_2^*$ とする．さらに、$\nu_i : \boldsymbol{P} \longrightarrow \boldsymbol{B}_i (i = 1, 2)$ を可微分写像とする．$\nu(P) = (\nu_1(P), \nu_2(P))$ とおけば、ν は $\boldsymbol{B}_1 \times \boldsymbol{B}_2$ の中に値と取る写像となる．

定理 6.12 （直積空間における有効性）：写像 $\nu_i : \boldsymbol{P} \longrightarrow \boldsymbol{B}_i (i = 1, 2)$ は $P \in \boldsymbol{P}$ において微分可能とし、$T_{n,i}$ は P において $\nu_i(P)$ の有効推定量とする．このとき、P のもとで、$\sqrt{n}(T_{n,i} - \nu_i(P))$ は漸近 tight ならば、P において $(T_{n,1}, T_{n,2})$ は $\nu(P) = (\nu_1(P), \nu_2(P))$ の有効推定量である．

証明：ある定数 a_1 と a_2 に対して、

$$\boldsymbol{B}_0^* = \{b_0^* : b_0^*(b_1, b_2) = a_1 b_1^*(b_1) + a_2 b_2^*(b_2), b_1^* \in \boldsymbol{B}_1^*, b_2^* \in \boldsymbol{B}_2^*\}$$

で定義する．$\|b_i\| = \sup_{b_i^* \in \boldsymbol{B}_i^*} |b_i^*(b_i)|$ となる[22] ので、

$$\sup_{b_0^* \in \boldsymbol{B}_0^* : \|b_0^*\|_{(B_1 \times B_2)^*}} |b_0^*(b_1, b_2)| \geq 2(a_1 \vee a_2)(\|b_1\|_{B_1} \vee \|b_2\|_{B_2}) \gtrsim \|(b_1, b_2)\|_{B_1 \times B_2}$$

となるので、(6.19) が満たされる[23] ．また、周辺 tightness から同時 tightness がわかるので、補題を用いれば、$(T_{n,1}, T_{n,2})$ の有効性がわかる． □

6.4 M – 推定量

定理 6.13 （一致性）：M_n を距離空間 (Θ, d) によって index される確率過程とし、$M : \Theta \mapsto \boldsymbol{R}$ を写像とし、以下を仮定する．

(i) Θ はコンパクト集合．

---

[22] $\rho_i^*(b_i) = \|b_i\|_{B_i}$ とおけば、$\rho_i^* \in \boldsymbol{B}_i^*$ となり、また、$\|\rho_i^*\|_{B_i^*} = \sup_{b_i \in \boldsymbol{B}_i} \frac{|\rho_i^*(b_i)|}{\|b_i\|_{B_i}} = \frac{\|b_i\|_{B_i}}{\|b_i\|_{B_i}} = 1$ より $\|\rho_i^*\|_{B_i^*} = 1$ となる．

[23] 直積空間 $\boldsymbol{B}_1 \times \boldsymbol{B}_2$ のノルム $\|\cdot\|_{B_1 \times B_2} = \|\cdot\|_{B_1} \vee \|\cdot\|_{B_2}$ で定義される．

(ii) コンパクトな部分集合 $K \subset \Theta$ に対しても

$$\limsup_{n\to\infty} \sup_{\theta\in K}(M_n(\theta) - M_n(\theta_0)) \le \sup_{\theta\in K}(M(\theta) - M(\theta_0))$$

がほとんどいたるところで成立する．

(iii) 写像 $M : \Theta \mapsto \boldsymbol{R}$ は上半連続とする．

(iv) θ_0 を含むどんな開集合 G に対しても

$$M(\theta_0) > \sup_{\theta\notin G} M(\theta)$$

が成立する．

このとき、$M_n(\hat\theta) \ge \sup_{\theta\in\Theta} M_n(\theta) - o_P(1)$ を満足するどんな列 $\{\hat\theta_n\}$ に対しても、

$$d(\hat\theta_n, \theta_0) \xrightarrow{P} 0$$

が成り立つ．

証明：任意の正の数 $\epsilon > 0$ を固定し、

$$G_\epsilon = \{\theta : d(\theta, \theta_0) < \epsilon\}, \qquad K = \Theta \setminus G_\epsilon$$

とする．(i) より K はコンパクトとなるので、(ii) を使えば、

$$P_{\theta_0}\left(\limsup_{n\to\infty} \sup_{\theta\in K}(M_n(\theta) - M_n(\theta_0)) \le \sup_{\theta\in K}(M(\theta) - M(\theta_0))\right) = 1$$

となり、$M(\theta)$ は上半連続より、コンパクト集合 K 上で最大値をとるので、$\delta = \sup_{\theta\in K}(M(\theta) - M(\theta_0))$ と記すことにする．更に、(iv) より $\delta < 0$ となる．よって、ある自然数 N があり、どんな $n > N$ なる自然数に対しても

$$\sup_{\theta\in K}(M_n(\theta) - M_n(\theta_0)) \le \frac{\delta}{2} < 0 \tag{6.20}$$

が成立する．また、(iv) より、

$$M_n(\hat\theta) - M_n(\theta_0) \ge \sup_{\theta\in\Theta}(M_n(\theta) - M_n(\theta_0)) - o_P(1) \ge -o_P(1) \tag{6.21}$$

が成立し、(6.20) と (6.21) を比較すれば、$\hat\theta \notin K$ となるので、確率 1 で、$d(\hat\theta, \theta_0) < \epsilon$ が成立することがわかる． □

定理 6.14 （argmax 連続写像定理）：M_n と M を距離空間 $\boldsymbol{H}$ で index された確率過程とし、任意のコンパクト集合 $K \subset \boldsymbol{H}$ に対して、$l^\infty(K)$ の中において $M_n \rightsquigarrow M$ が成立するとする．また、ほとんどすべての道 $h \mapsto M(h)$ は上半連続で、$\hat h$ で最大値ととり、$\hat h$ は tight であるとする．更に、$\hat h_n$ は一様に tight で $M_n(\hat h_n) \ge \sup M_n(h) - o_P(1)$ を満足するならば、$\hat h_n \rightsquigarrow \hat h$ が成立する．

証明：K をコンパクト集合とし、固定する．$G \subset \boldsymbol{H}$ を $\hat{h}$ を含む任意の開集合としたとき、

$$M(\hat{h}) > \sup_{h \notin G, h \in K} M(h)$$

が成立すること[24] に注意する．つぎに、連続写像定理と $M_n \rightsquigarrow M$ （in $l^\infty(K)$ ）から、任意の閉集合 $F \subset \boldsymbol{H}$ に対して、

$$\sup_{h \in F \cap K} M_n(h) - \sup_{h \in K} M_n(h) \rightsquigarrow \sup_{h \in F \cap K} M(h) - \sup_{h \in K} M(h)$$

が成立することがわかる．このことより、Stutsky の補題と Portmanteau の定理を使えば、

$$\begin{aligned}
\limsup_{n\to\infty} P^*(\hat{h}_n \in F \cap K) &\le \limsup_{n\to\infty} P^*(\sup_{h \in F \cap K} M_n(h) \ge \sup_{h \in K} M_n(h) - o_P(1)) \\
&\le P(\sup_{h \in F \cap K} M(h) \ge \sup_{h \in K} M(h)) \\
&\le P(\hat{h} \in F) + P(\hat{h} \notin K)
\end{aligned}$$

より[25] 、

$$P^*(\hat{h}_n \in F) \le P(\hat{h} \in F) + P(\hat{h} \notin K) + \limsup_{n\to\infty} P^*(\hat{h}_n \notin K)$$

が成立することがわかる．$\hat{h}$ が tight で $\hat{h}_n$ が一様に tight であることより、上の式の右辺の最後の二項はいくらでも小さくとれるので、

$$\limsup_{n\to\infty} P^*(\hat{h}_n \in F) \le P^*(\hat{h} \in F)$$

が成立することがわかる．最後に Portmantaeu の定理を再度使えば、この定理が証明される．
□

(Θ, d) をある距離空間とし、写像 $M : \Theta \mapsto \boldsymbol{R}$ を考える．$\theta_0 \in \Theta$ を写像 $\theta \mapsto M(\theta)$ が最大値を取る点[26] をする．

定理 6.15 （rate of convergence）：M_n を半距離空間 (Θ, d) によって添字付けられた確率過程とし、$M : \Theta \mapsto \boldsymbol{R}$ を写像とし、θ_0 のある近傍に含まれるどんな θ に対しても

$$M(\theta) - M(\theta_0) \lesssim -d^2(\theta, \theta_0) \tag{6.22}$$

---

[24] 背理法で証明する．もし、この式が成立しなければ、$h_m \in G^c \cup K$ の列で $M(h_m) \to M(\hat{h})$ を満足するものが存在する．$\{h_m\}$ の中から収束する部分列をとれることが K がコンパクトであることよりわかる．その収束先を $\bar{h}$ とすれば、$M(\hat{h}) = M(\bar{h})$ となり、数列の作り方から $\hat{h}\ \not\bar{h}$ となり、最大値をふたつのことなる点でとることになり、$h \mapsto M(h)$ が上半連続であることと矛盾する．

[25] $P^*(\hat{h}_n \in F \cap K) = 1 - P^*(\hat{h}_n \in F^c \cup K^c) \ge P^*(\hat{h}_n \in F) - P^*(\hat{h}_n \notin K)$ から $\limsup_{n\to\infty} P^*(\hat{h}_n \in F \cap K) \ge \limsup_{n\to\infty} P^*(\hat{h}_n \in F) - \limsup_{n\to\infty} P^*(\hat{h}_n \notin K)$ がわかる．

[26] $\theta \mapsto M(\theta)$ の一次微分は θ_0 でゼロとなり、2 次微分が θ_0 のまわりで負になるので、θ_0 のある近傍に含まれる θ に対して、$M(\theta) - M(\theta_0) \lesssim -d^2(\theta, \theta_0)$ が成立することは自然である．

を満足するとする．更に、すべての自然数 n と十分小さな正数 $\delta > 0$ に対しても、中心化された確率過程 $M_n - M$ は

$$\boldsymbol{E}^* \sup_{d(\theta,\theta_0)<\delta} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}} \tag{6.23}$$

を満足する．ただし、ϕ_n はある $\alpha > 2$ に対して、$\delta \mapsto \phi_n(\delta)/\delta^\alpha$ が減少関数[27] となるようなものである．すべての自然数 n に対しても

$$r_n^2 \phi\left(\frac{1}{r_n}\right) \le \sqrt{n} \tag{6.24}$$

とする．もし、 $\hat{\theta}_n$ が

$$M_n(\hat{\theta}_n) \ge M_n(\theta_0) - O_P^*(r_n^{-1})$$

を満たし、一致性

$$\hat{\theta}_n = \theta_0 + O_P^*(1)$$

が成立するならば、

$$r_n d(\hat{\theta}_n, \theta_0) = O_P^*(1) \tag{6.25}$$

が成立する．

証明：それぞれの n に対して、

$$S_{j,n} = \{\theta : 2^{j-1} < r_n d(\theta, \theta_0) \le 2^{\,\mathrm{j}}\,\}$$

とする[28]．(6.25) を示すためには、ある大きな正数 M が存在して、$n \to \infty$ のとき、

$$P^*(r_n d(\hat{\theta}, \theta_0) > 2^M) \to 0$$

を示せば十分である．

$\hat{\theta}$ は写像 $\theta \mapsto M_n(\theta)$ を最大化する点なので、$\hat{\theta} \in S_{j,n}$ ならば、

$$\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta)) \ge 0$$

となる．このことに注意すれば、どんな正数 $\eta > 0$ に対しも

$$\begin{aligned}
&P^*(r_n d(\hat{\theta}, \theta_0) > 2^M) \\
&\quad \le P^*(r_n d(\hat{\theta}, \theta_0) > 2^M, r_n d(\hat{\theta}, \theta_0) < \eta) + P^*(r_n d(\hat{\theta}, \theta_0) \ge \eta) \\
&\quad \le \sum_{j \ge M} P^*(\hat{\theta} \in S_{j,n}, r_n d(\hat{\theta}, \theta_0) < \eta) + P^*(r_n d(\hat{\theta}, \theta_0) \ge \eta) \\
&\quad \le \sum_{j \ge M, 2^j \le \eta r_n} P^*(\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \ge 0) + P^*(r_n d(\hat{\theta}, \theta_0) \ge \eta)
\end{aligned} \tag{6.26}$$

[27] すべての $c > 1$ に対して、$c\delta > \delta$ より、$\phi_n(c\delta)/(c\delta)^\alpha \le \phi(\delta)/\delta^\alpha$ から、$\phi_n(c\delta) \le c^\alpha \phi_n(\delta)$ を得る．

[28] したがって、$\Theta = \bigcup_{j=1}^\infty S_{j,n}$ である．

となる[29]．どんな正数 $\eta > 0$ に対しても、$n \to \infty$ のとき、(6.26) の最右辺の１項目はゼロに収束する．したがって、あとはその第二項目もゼロに収束することを示せばよい．

そのために、$\eta > 0$ を十分小さく取り、$d(\theta, \theta_0) < \eta$ を満足するどんな θ に対しても (6.22) が成立し、かつ $\delta < \eta$ なるどんな正数 $\delta > 0$ に対しても (6.23) が成立するようにする．すると、どんな $\theta \in S_{j,n}$ に対しても

$$M(\theta) - M(\theta_0) \lesssim -d^2(\theta, \theta_0) \leq -\frac{2^{2j-2}}{r_n^2}$$

が成立する．ここで、$W_n = M_n - M$ とおく．

$$\begin{aligned}
&\sup_{\theta \in S_{j,n}} (M_n(\theta) - M(\theta_0) + \frac{2^{2j-2}}{r_n^2} \\
&\quad \gtrsim \sup_{\theta \in S_{j,n}} \{(M_n(\theta) - M_n(\theta)) - (M(\theta) - M(\theta_0))\} \\
&\quad = \sup_{\theta \in S_{j,n}} |(M_n(\theta) - M_n(\theta)) - (M(\theta) - M(\theta_0))| \\
&\quad = \|W_n(\theta) - W_n(\theta_0)\|_{S_{j,n}} \gtrsim \frac{2^{2j-2}}{r_n^2}
\end{aligned}$$

より[30]、

$$\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq 0$$

となるので、

$$P^*(\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq 0) \leq P^*\left(\|W_n(\theta) - W_n(\theta_0\|_{S_{j,n}} \gtrsim \frac{2^{2j-2}}{r_n^2}\right)$$

がわかる．Markov の不等式と (6.24) を用いて、更に $c > 1$ に対して、$\phi_n(c\delta) \leq c^\alpha \phi_n(\delta)$ が成立することを利用すれば、

$$\begin{aligned}
&\sum_{j \geq M, 2^j \leq \eta r_n} P^*\left(\|W_n(\theta) - W_n(\theta_0)\|_{S_{j,n}} \gtrsim \frac{2^{2j-2}}{r_n^2}\right) \\
&\lesssim \sum_{j \geq M, 2^j \leq \eta r_n} \frac{r_n^2}{2^{2j-2}} \boldsymbol{E}\|W_n(\theta) - W_n(\theta_0)\|_{S_{j,n}} \lesssim \sum_{j > M} \frac{r_n^2}{2^{2j-2}} \frac{\phi_n(2^j/r_n)}{\sqrt{n}} \\
&\lesssim \sum_{j > M} \frac{r_n^2}{2^{2j}} \frac{\phi_n(2^j/r_n)}{\sqrt{n}} \leq \sum_{j > M} \frac{r_n^2}{2^{2j}} \frac{2^{j\alpha}\phi_n(1/r_n)}{\sqrt{n}} \\
&\leq \sum \frac{1}{2^{(2-\alpha)j}} \to 0 \qquad (M \to \infty)
\end{aligned}$$

となるので、定理は証明された． □

[29] 最後の不等式のところで、$d(\hat{\theta}_n, \theta_0) < \eta$ と $\hat{\theta} \in S_{j,n}$ から $d(\hat{\theta}_n, \theta_0) \leq 2^j/r_n$ から $2^j \leq \eta r_n$ の条件が出てくる．

[30] 最後からふたつ目の等号は $\hat{\theta}_n \in S_{j,n}$ のとき、$\sup_{\theta \in S_{j,n}}(M_n(\theta) - M_n(\theta_0)) \geq 0$ と $-(M(\theta) - M(\theta_0)) \geq 0$ よりわかる．

$X_1,\dots,X_n \sim P$（iid）とし、$P_n=(1/n)\sum_{i=1}^n \delta_{X_i}$ とおき、$M_n(\theta)=P_n m_\theta$ と $M(\theta)=Pm_\theta$ と書ける場合を考える．更に、$G_n=M_n-M$、$\mathcal{M}_\delta=\{m_\theta-m_{\theta_0}:d(\theta,\theta_0)<\delta\}$ とおく．

このとき、写像 $\theta\mapsto Pm_\theta$ が θ_0 において 2 回微分可能ならば、(6.22) は成立する．また、最大不等式から

$$\begin{aligned}\boldsymbol{E}_P^*\|G_n\|_{\mathcal{M}_\delta} &\lesssim J(1,\mathcal{M}_\delta)\sqrt{PM_\delta^2}\\ \boldsymbol{E}_P^*\|G_n\|_{\mathcal{M}_\delta} &\lesssim J_{[]}(1,\mathcal{M}_\delta,L_2(P))\sqrt{PM_\delta^2}\end{aligned}$$

が成立する．ただし、M_δ を $\mathcal{M}_\delta$ の envelope 関数とした．$\delta\downarrow 0$ としたときに、$J(1,\mathcal{M}_\delta)$ と $J_{[]}(1,\mathcal{M}_\delta,L_2(P))$ が有界である場合には、$\phi_n^2(\delta)=PM_\delta^2$ とすれば、(6.23) が成立する．したがって、収束のオーダーは

$$r_n^4\phi_n^2\left(\frac{1}{r_n}\right)=r_n^4PM_{\frac{1}{r_n}}^2\sim n \tag{6.27}$$

で与えられることがわかる．

以降、Θ を有限次元とし、ランダム標本のもとで議論を進める．(6.27) によって、収束のオーダーが与えられていると仮定し、その上で、$r_n(\hat\theta-\theta_0)$ の極限分布を求めることにする．そのために、$l^\infty(h:|h|\le K)$（ただし、K はある正の定数とした）上の確率過程

$$h\mapsto r_n^2P_n(m_{\theta_0+h/r_n}-m_{\theta_0})$$

についての極限分布を求める．中心化した確率過程

$$\begin{aligned}h &\mapsto \frac{r_n^2}{\sqrt{n}}G_n(m_{\theta_0+h/r_n}-m_{\theta_0})\\ &= r_n^2P_n(m_{\theta_0+h/r_n}-m_{\theta_0})-r_n^2(m_{\theta_0+h/r_n}-m_{\theta_0})\end{aligned}$$

を考える．すなわち、$(r_n^2/\sqrt{n})\mathcal{M}_{K/r_n}$ によって index された経験過程である．

定理 6.16 ：θ を Euclid 空間の開部分集合の元とし、それぞれの θ において m_θ は可測関数とする．以下の仮定が成立するとする．

(i) 写像 $\theta\mapsto Pm_\theta$ はその最大点 θ_0 において 2 回連続微分可能で、正則な 2 回微分 V をもつ．

(ii) エントロピー条件： ある $\delta_0>$ に対して、

$$\int_0^\infty \sup_{\delta<\delta_0}\sup_Q\sqrt{\log N(\epsilon\|\mathcal{M}_\delta\|_{Q,2},\mathcal{M}_\delta,L_2(Q))}d\epsilon<\infty$$

または

$$\int_0^\infty \sup_{\delta<\delta_0}\sqrt{\log N_{[]}(\epsilon\|\mathcal{M}_\delta\|_{P,2},\mathcal{M}_\delta,L_2(P))}d\epsilon<\infty$$

が成立する．

(iii) ある関数 ϕ が存在し、$\mathcal{M}_\delta$ の envelope 関数 M_δ に対して、$\phi^2(\delta) \geq P^* M_\delta^2$ を満たし、ある $\alpha < 2$ に対して、$\delta \mapsto \phi(\delta)/\delta^\alpha$ は減少関数となり、この ϕ とどんな $\eta > 0$ と $K > 0$ に対しても平均がゼロの Gauss 過程 G で $h = g$ のときのみに、ほとんど確実に $G(g) = G(h)$ が成立するようなものが存在し、

$$\begin{aligned}
\lim_{\delta \downarrow 0} \frac{P^* M_\delta^2 1\{M_\delta > \eta \delta^{-2} \phi^2(\delta)\}}{\phi^2(\delta)} &= 0 \\
\lim_{\epsilon \downarrow 0} \lim_{\delta \downarrow 0} \sup_{\|h-g\| \leq \epsilon, \|h\| < K, \|g\| < K} \frac{m_{\theta_0 + \delta g} - m_{\theta_0 + \delta h}}{\phi^2(\delta)} &= 0 \\
\lim_{\delta \downarrow 0} \frac{P(m_{\theta_0 + \delta g} - m_{\theta_0 + \delta h})^2}{\phi^2(\delta)} &= \boldsymbol{E}(G(g) - G(h))^2 \qquad (6.28)
\end{aligned}$$

を満足する．

このとき、コンパクトな集合上で有界連続な見本路をもつ Gauss 過程の version が存在し、r_n を $r_n^2 \phi(1/r_n) = \sqrt{n}$ の解としたとき、すべての自然数 n に対して、$\hat{\theta}_n$ で写像 $\theta \mapsto P_n m_\theta$ はほとんど確実に最大となり、$\hat{\theta}_n$ は θ_0 に確率収束するとき、$r_n(\hat{\theta}_n - \theta_0)$ は確率過程 $h \mapsto G(h) + (1/2)h'Vh$ を最大にする一意な点 h に分布収束する．

証明：確率過程

$$h \mapsto \frac{r_n^2}{\sqrt{n}} G_n(m_{\theta_0 + h/r_n} - m_{\theta_0}) \qquad (6.29)$$

が $l^\infty\{h : \|h\| < K\}$ において漸近的に tight であるための条件がエントロピー条件が成立とどんな $\eta > 0$ と $\delta_n \downarrow 0$ に対しても

$$\begin{aligned}
\frac{r_n^4}{n} P^* M_{K/r_n}^2 &= O(1) \\
\frac{r_n^4}{n} P^* M_{K/r_n}^2 1\{r_n^2 M_{K/r_n} > \eta n\} &= o(1) \\
\sup_{\|h-g\| \leq \delta_n} \frac{r_n^4}{n} P(m_{\theta_0 + g/r_n} - m_{\theta_0 + h/r_n})^2 &= o(1)
\end{aligned}$$

が成立こと[31] である．この条件は定理の仮定の下で成立する．また、写像 $\theta \mapsto Pm_\theta$ は 2 回連続微分可能であることから、有界関数 h に対して一様に

$$r_n^2 P(m_{\theta_0 + h/r_n} - m_{\theta_0}) \to \frac{1}{2} h'Vh$$

となる．したがって、確率過程

$$h \mapsto r_n^2 P_n(m_{\theta_0 + h/r_n} - m_{\theta_0})$$

は $l^\infty\{h : \|h\| < K\}$ において漸近的に tight であることがわかる．

[31] (6.11) において $F_n = (r_n^2/\sqrt{n}) M_{K/r_n}$ とすればよい．

θ_0 の近傍では確率過程 (6.29) は $l^\infty\{h : \|h\| < K\}$ において連続なので、$r_n(\hat\theta - \theta_0)$ も一様に tight であることがわかる．有限次元の収束は Linderg 条件から確認できる．また、平均と分散は $(1/2)h'Vh$ と $\boldsymbol{E}G(h)G(g)$ になることがわかる．したがって、$h \mapsto P_n(m_{\theta_0+h/r_n} - m_{\theta_0})$ は $l^\infty\{h : \|h\| < K\}$ において、Gauss 過程 $h \mapsto G(h) + (1/2)h'Vh$ に収束する．この過程の見本路はほとんど確実に一意的な点 $\hat h$ で最大となること[32] から連続写像定理を利用すれば、定理は証明される． □

6.5　Z – 推定量

母数空間 Θ を Banach 空間とし、$\boldsymbol{B}$ を他の Banach 空間とする．モデルを $\boldsymbol{P} = \{P_\theta : \theta \in \Theta\}$ とし、真のモデルを P_{θ_0} （これを簡略した、P とかく）とする．

$$\Phi_n : \Theta \longrightarrow \boldsymbol{B} \qquad \Phi : \Theta \longrightarrow \boldsymbol{B}$$

を $\boldsymbol{B}$ – 値のランダム写像と deterministic 写像とする．

ここでは、方程式

$$\Phi_n(\hat\theta_n) = 0 \tag{6.30}$$

の解として定義される「推定量」 $\hat\theta_n$ の漸近正規性を得るための十分条件を検討することにする．$\hat\theta_n$ を Z – 推定量と呼ぶことにする．

$\boldsymbol{B} = l^\infty(\mathcal{H})$ ならば、一般性を失わず、(6.30) は $h \in \mathcal{H}$ を動くときの方程式

$$\Phi_n(\hat\theta_n)h = 0$$

の集まりとなる．Φ を Φ_n の漸近版とし、θ_0 を方程式 $\Phi(\theta_0) = 0$ の解としたとき、$\hat\theta_n$ は θ_0 に確率収束することが期待できる．以下、一致性：$\hat\theta_n = \theta_0 + o_P(1)$ を仮定したうえで、$\sqrt{n}(\hat\theta_n - \theta_0)$ の分布収束の十分条件を求める．

つぎのような仮定をおくことにする．

<u>Φ の Fréchet 微分可能性</u>： Θ の接空間 $\dot\Theta$ から $\boldsymbol{B}$ への連続線形で 1 対 1 写像 $\dot\Phi_{\theta_0}$ が存在し、$\theta \to \theta_0$ のとき、

$$\|\Phi(\theta) - \Phi(\theta_0) - \dot\Phi_{\theta_0}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|) \tag{6.31}$$

を満足する．

<u>$\dot\Phi_{\theta_0}$ の逆写像の連続性</u>：Image($\dot\Phi_{\theta_0}$) から $\dot\Theta$ への逆写像（これを $\dot\Phi_{\theta_0}^{-1}$ と記すことにする）が存在し、その定義域上で連続とする．

定理 6.17 ：Φ_n と Φ を Θ からある Banach 空間へのランダム写像と deterministic 写像とし、つぎを満足すると仮定する．

(i)

$$\frac{\sqrt{n}(\Phi_n - \Phi)(\hat\theta_n) - \sqrt{n}(\Phi_n - \Phi)(\theta_0)}{1 + \sqrt{n}\|\hat\theta - \theta_0\|} = o_{P^*}(1)$$

[32] VdVW の Appendix A2.20 を参照されたい．

(ii) $\sqrt{n}(\Phi_n - \Phi)(\theta_0)$ は tight なあるランダム要素 Z に分布収束する．

(iii) 写像 $\theta \mapsto \Phi(\theta)$ は θ_0 において Fréchet 微分可能で、その微分 $\dot{\Phi}_{\theta_0}$ の連続な逆写像（$\dot{\Phi}_{\theta_0}^{-1}$ と記すことにする）が存在する．

このとき、$\Phi(\theta_0) = 0$ と $\Phi_n(\hat{\theta}) = o_{P^*}(1/\sqrt{n})$ が成立し、$\hat{\theta}_n$ が θ_0 に確率収束するならば、

$$\sqrt{n}\dot{\Phi}_{\theta_0}(\hat{\theta}_n - \theta) = -\sqrt{n}(\Phi_n - \Phi)(\theta_0) + o_{P^*}(1) \tag{6.32}$$

が成立する．したがって、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Phi}_{\theta_0}^{-1} Z$$

である．

証明：$\hat{\theta}_n$ の定義と定理の仮定 (i) から

$$\begin{aligned}\sqrt{n}(\Phi(\hat{\theta}_n) - \Phi(\theta_0)) &= \sqrt{n}(\Phi(\hat{\theta}_n) - \Phi_n(\hat{\theta}_n)) + o_{P^*}(1) \\ &= -\sqrt{n}(\Phi_n - \Phi)(\theta_0) + o_{P^*}(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|)\end{aligned} \tag{6.33}$$

となる．一方、定理の仮定 (iii) より、ある正数 $c > 0$ が存在して、どんな θ に対しても

$$\|\dot{\Phi}_{\theta_0}(\theta - \theta_0)\| \geq c\|\theta - \theta_0\|$$

とできる[33]．これと Φ の Fréchet 可微分性 (6.31) から

$$\begin{aligned}\|\Phi(\theta) - \Phi(\theta_0)\| &= \|\dot{\Phi}_{\theta_0}(\theta - \theta_0)\| + o(\|\theta - \theta_0\|) \\ &\geq c\|\theta - \theta_0\| + o(\|\theta - \theta_0\|)\end{aligned}$$

を得る．これを (6.33) の左辺に適用すれば、

$$\sqrt{n}\|\hat{\theta}_n - \theta_0\|(c + o_P(1)) \leq O_P(1) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta\|)$$

となり、

$$\sqrt{n}\|\hat{\theta}_n - \theta\| = O_{P^*}(1)$$

がわかる．この収束の rate と Φ の可微分性から

$$\begin{aligned}\|\Phi(\hat{\theta}_0) - \Phi(\theta_0)\| &= \|\dot{\Phi}_{\theta_0}(\hat{\theta}_n - \theta_0)\| + o_{P^*}(\sqrt{n}\|\hat{\theta}_n - \theta_0\|) \\ &= \|\dot{\Phi}_{\theta_0}(\hat{\theta}_n - \theta_0)\| + o_{P^*}(1)\end{aligned}$$

を得る．よって、(6.32) は証明された．更に、$\dot{\Phi}_{\theta_0}^{-1}$ の連続性から、$\sqrt{n}(\hat{\theta}_n - \theta_0)$ の分布収束はわかる． □

つぎに、Θ をあるノルム空間の部分空間とし、$\mathcal{H}$ を任意の集合としたとき、確率分布 P からの大きさ n のランダム標本を仮定し、ある実数値可測関数 $x \mapsto \phi_{\theta,h}(x)$ によって、

$$\Phi_n(\theta)(h) = P_n\phi_{\theta,h}, \qquad \Phi(\theta)h = P\phi_{\theta,h}$$

[33] Yoshida, K: Functional Analysis (Springer) の Corollary 3 （pages 43）を参照されたい．

と表現された場合を考えよう．ただし、どんな $h \in \mathcal{H}$ に対しても、$P\phi_{\theta_0,h} = 0$ である．すると、$\sqrt{n}(\Phi_n - \Phi)(\theta) = \{G_n\phi_{\theta,h} : h \in \mathcal{H}\}$ は関数族 $\{\phi_{\theta,h} : h \in \mathcal{H}\}$ によって添字付けるられた経験過程となり、前の定理の仮定 (i) は

$$\sup_{h\in\mathcal{H}} |G_n(\phi_{\hat{\theta}_n,h} - \phi_{\theta_0,h})| = o_P^*(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|) \tag{6.34}$$

となることがわかる．

定理 6.18 ：どんな $\theta \in \Theta$ と $h \in \mathcal{H}$ に対しても、$x \mapsto \phi_{\theta,h}$ は実数値可測関数とし、つぎの条件を満足すると仮定する．

(i) ある正数 $\delta > 0$ に対して、$\{\phi_{\theta,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\}$ は P – Donsker クラスで、有界な envelope 関数を持つ．

(ii) Θ から $l^\infty(\mathcal{H})$ への写像 $\theta \mapsto P\phi_{\theta,\cdot}$ は Fréchet 微分可能で、微分 $V : [\bar{\Theta}] \longrightarrow l^\infty(\mathcal{H})$ を持ち、V は Image(V) 上で定義された連続な逆写像を持つ．

(iii) $\theta \to \theta_0$ のとき、

$$\sup_{h\in\mathcal{H}} P(\phi_{\theta,h} - \phi_{\theta_0,h})^2 \to 0$$

が成立する．

このとき、

$$\sup_{h\in\mathcal{H}} |P_n\phi_{\hat{\theta}_n,h}| = o_P^*(1/\sqrt{n}), \qquad \hat{\theta}_n \xrightarrow{P} \theta_0$$

ならば、

$$V\sqrt{n}(\hat{\theta}_n - \theta_0) = -G_n\phi_{\theta_0} + o_P^*(1)$$

が成立[34] する．

証明：一般性を失わず、$\hat{\theta}_n$ は $\Theta_\delta = \{\theta \in \Theta : \|\theta - \theta_0\| < \delta\}$ の中に値を取るとしてよい．写像 $f : l^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \longrightarrow l^\infty(\mathcal{H})$ を任意の $z \in l^\infty(\Theta_\delta \times \mathcal{H})$ と $\theta \in \Theta_\delta$ に対して、

$$f(z,\theta)h = z(\theta,h) - z(\theta_0,h)$$

で定義する．(z_0,θ_0) を $\theta \to \theta_0$ のとき、

$$\sup_{h\in\mathcal{H}} |z_0(\theta,h) - z_0(\theta_0,h)| \to 0 \tag{6.35}$$

となる点とする．すると、f はこのような点 (z_0,θ_0) において連続[35] となる．

[34] $\sup_{h\in\mathcal{H}} |\{V\sqrt{n}(\hat{\theta}_n - \theta_0)\}(h) + G_n\phi_{\theta_0,h}| = o_P^*(1)$ が成立することである．

[35] 実際、$n \to \infty$ のときに、$z_n \to z_0$ （すなわち、$\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}} |z_n(\theta,h) - z_0(\theta,h)| \to 0$ ）なる列 $\{z_n\}_{i=1}^\infty$ を取る．$\theta_n \to \theta_0$ のとき、

$$\begin{aligned}\sup_{h\in\mathcal{H}} |f(z_n,\theta_n)h - f(z_0,\theta_0)h| &\le \sup_{h\in\mathcal{H}} |f(z_n,\theta_n)h - f(z_0,\theta_n)h| + \sup_{h\in\mathcal{H}} |f(z_0,\theta_n)h - f(z_0,\theta_0)h| \\ &\le \sup_{\theta\in\Theta_\delta,h\in\mathcal{H}} |z_n(\theta,h) - z_0(\theta,h)| + \sup_{h\in\mathcal{H}} |z_0(\theta_n,h) - z_0(\theta,h)| \to 0\end{aligned}$$

から f の連続性はわかる．

ここで、G を Brown 橋とすれば、仮定 (iii) より、$\theta \to \theta_0$ のとき、$\sup_{h\in\mathcal{H}} P(\phi_{\theta,h}-\phi_{\theta_0,h})^2 \to 0$ となるので、Brown 橋の一様連続性より、$\sup_{h\in\mathcal{H}} |G\phi_{\theta,h} - G\phi_{\theta_0,h}| \to 0$ となる．したがって、G はほとんど確実にすべての見本路に関して (6.35) [36] を満足するので、f はほとんど確実にすべての見本路に関して連続となることがわかる．

つぎに、定理の仮定 (i) と $\hat{\theta}_n = \theta_0 + o_P^*(1)$ から、Slutszky の定理を適用すれば、$l^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta$ の中において $(G_n\phi_{\theta,h}, \hat{\theta}_n) \rightsquigarrow (G\phi_{\theta,h}, \theta_0)$ となり、更に、連続写像定理を利用すれば、$l^\infty(\mathcal{H})$ の中において

$$f(G_n\phi, \hat{\theta}_n) \rightsquigarrow f(G\phi, \theta_0) = 0$$

となることがわかる．極限が退化したときには、分布収束から確率収束がわかるので、

$$o_P^*(1) = f(G_n\phi, \hat{\theta}_n)h = G_n\phi_{\hat{\theta}_n,h} - G_n\phi_{\theta_0,h}$$

となり、(6.34) が成立することがわかるので、最後に定理 6.12 を適用すれば、この定理は証明された． □

定理 6.19 ：$\phi_{\theta,h}$ に対応する Φ と Φ_n は定理 6.16 の条件を満足するとする．$\hat{\theta}_n$ が正則推定量であるための必要十分条件は、どんな $h \in \mathcal{H}$ に対しても

$$\dot{\Phi}_{\theta_0}(\dot{\theta}_0)h = -P_0\dot{\boldsymbol{l}}(\dot{\theta}_0)\phi_{\theta_0,h} \tag{6.36}$$

が成立することである．ただし、θ_0 を通る Θ の中の曲線

$$\{\theta_t : t \text{ は小さな非負の実数}, \theta_t|_{t=0} = \theta_0\}$$

は、$t \downarrow 0$ のとき、

$$\left\| \frac{\theta_t - \theta_0}{t} - \dot{\theta}_0 \right\| \to 0$$

を満足するものとし、そのような曲線に対する $\dot{\theta}_0$ の集まりの linear span の閉包、すなわち接空間を $\dot{\Theta}$ と記す．また、$\dot{\boldsymbol{l}}_0 : \dot{\Theta} \longrightarrow \dot{\boldsymbol{P}} \subset L_2(P_0)$ は、どんな $\{P_{\theta_t}\}$ に対しても、$t \downarrow 0$ のとき、

$$\int \left[\frac{\sqrt{dP_{\theta_t}} - \sqrt{dP_0}}{t} - \frac{1}{2}\dot{\boldsymbol{l}}_0(\dot{\theta}_0)\sqrt{dP_0} \right]^2 \to 0$$

が成立するものである．

証明：$t_n = O(n^{-1/2})$ なる数列 $\{t_n\}$ を取る．するとモデルの正則性より

$$\log \Pi_{i=1}^n \frac{dP_{\theta_{t_n}}}{dP_0}(X_i) = t_n \sum_{i=1}^n \dot{\boldsymbol{l}}_0(X_i) - \frac{1}{2}(\sqrt{n}t_n)^2 P_0 \dot{\boldsymbol{l}}_0^2 + o_P(1)$$

が成立する．ただし、$o_{P_0}(1) = o_P(1)$ とした．また、仮定より、P_0 のもとで、$l^\infty(\mathcal{H})$ – 値の tight なある random element Z_0 が存在して、

$$\sqrt{n}(\Phi_n - \Phi)(\theta_0) \rightsquigarrow Z_0$$

[36] この式において、$z_0(\theta,h) = G\phi_{\theta,h}$ と観ればよい．

が成立する．よって、漸近的 tight であるので、$(\sqrt{n}(\Phi_n-\Phi)(\theta_0), n^{-1/2}\sum_{i=1}^n \dot{\boldsymbol{l}}_0(X_i))$ は $l^\infty(\mathcal{H})\times\boldsymbol{R}$ の中で同時に漸近 tight になる．したがって、$\{n\}$ の部分列（それも n とかく）をとれば、ある Gauss 過程に同時に弱収束する．Le Cam の第三補題を適用すれば、$P_{\theta_{t_n}}$ のもとで、

$$\sqrt{n}(\Phi_n-\Phi)(\theta_0) \rightsquigarrow Z_0 + P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot}$$

が成立すること[37] がわかる．一方、定理 6.16 より、

$$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_0)) = -\sqrt{n}(\Phi_n-\Phi)(\theta_0) + o_P(1)$$

である．さらに、LAN であることより、$P_{t_n} \lhd P_0$ がわかるので、

$$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) = -\sqrt{n}(\Phi_n-\Phi)(\theta_{t_n}) + o_{P_n}(1)$$

が成立する．ただし、$o_{P_n}(1) = o_{P_n}(1)$ とした．このふたつの式に Slutzky の補題を利用すれば、

$$(\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_0)), \dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) + \sqrt{n}(\Phi_n-\Phi)(\theta_{t_n})) \rightsquigarrow (Z_0 + P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot}, 0)$$

を得る．写像 $(a,b)\mapsto b-a$ に対して連続写像定理を上の式に適用[38] すれば、

$$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_0)) \rightsquigarrow -Z_0 - P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot} - \dot{\Phi}_0(\dot{\theta}_0)$$

が成立することがわかり、定理は証明された．

□

[37] $P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot}$ はふたつの過程の共分散関数である．

[38]
$$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) - \dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_0)) + \dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) \rightsquigarrow -Z_0 - P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot}$$

となり、$\theta_{t_n}\to\theta_0$ から

$$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) \rightsquigarrow -Z_0 - P_0\dot{\boldsymbol{l}}_0\phi_{\theta,\cdot}$$

を得る．さらに、$\sqrt{n}(\theta_{t_n}-\theta_0) = \dot{\theta}_0 + o(1)$ に注意すれば、$\dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_{t_n})) = \dot{\Phi}_0(\sqrt{n}(\hat{\theta}_n-\theta_0)) + \dot{\Phi}_0(\dot{\theta}_0)$ に注意すればよい．

6.6 ノンパラメトリック最尤推定量

$\boldsymbol{B}$ を Banach 空間とし、$\boldsymbol{A}_0 \subset \boldsymbol{B}$ とする．正則モデル $\boldsymbol{P}$ に対して、$\nu : \boldsymbol{P} \longrightarrow \boldsymbol{A}_0$ の推定をを考えよう．そのために、$\boldsymbol{P}$ の拡張

$$\boldsymbol{M}_c = \boldsymbol{P} \cup \{\text{コンパクトな台を持つ確率分布}\}$$

と $\boldsymbol{A}_0$ の拡張 $\boldsymbol{A}$ を考え、$\boldsymbol{A} \times \boldsymbol{M}_c$ 上の実数値写像 D で、どんな $P \in \boldsymbol{P}$ に対しても $D(\cdot, P) : \boldsymbol{A} \longrightarrow \boldsymbol{R}$ は $\nu = \nu(P)$ で最小値をとるものとする．さらに、D_n も $\boldsymbol{A} \times \boldsymbol{M}_c$ 上の実数値写像とし、どんな $\nu \in \boldsymbol{A}$ と $P \in \boldsymbol{P}$ に対しても、$n \uparrow \infty$ のとき、$D_n(\nu, P) \to D(\nu, P)$ を満足するものとする．Minimum Contrast 推定量 $\hat{\nu}_n$ を

$$D_n(\nu, P_n) = \min\{D_n(\nu, P_n) : \nu \in \boldsymbol{A}\}$$

で定義する．

$\nu(P) = P$ のとき、コントラストとして、$D(P, Q) = \log(dP/d\mu)dQ$ を考える．ただし、μ は Lebeague 測度とする．経験分布 P_n も議論の対象に含むようにしたいので、$\boldsymbol{A} = \boldsymbol{M}_c$ とすると、D は定義されなくなる．そこで、$\boldsymbol{P}$ を μ に支配されていない測度を含む族 $\bar{\boldsymbol{P}}$ を考え、P の推定量を $\hat{P} \in \bar{\boldsymbol{P}}$ をつぎのように定める．どんな $P \in \boldsymbol{P}$ に対しても

$$\int \log \left[\frac{d\hat{P}}{d\mu_n}\right] dP_n \geq \int \log \left[\frac{dP}{d\mu_n}\right] dP_n$$

を満足するものとする．ただし、ν_n は P と $\hat{P}$ の優測度とする．これを **Nonparametric Maximum Likelihood Estimator（NMPLE）**とよぶ．

罰則つき NPMLE：

sivesd NPMLE：

Regularized NPMLE：

6.7　有効推定量の構成：セミパラメトリックの観点から

ここでは、セミパラメトリックモデル

$$\boldsymbol{P} = \{P_{\theta,g} : \theta \in \Theta \subset \boldsymbol{R}^k, g \in \boldsymbol{G}\}$$

を考え、Euclid 母数 $\nu(P_{\theta,g}) = \theta$ の推定を考える．ただし、$\boldsymbol{G}$ はある Hilbert 空間の（無限次元）部分空間とする．

6.7.1　有効影響関数

$\boldsymbol{P}$ の部分モデルとして、

$$\begin{aligned}\boldsymbol{P}_1 &= \{P_{\theta,g} : 任意の \theta \in \Theta, 固定した\ g \in \boldsymbol{G}\}\\ \boldsymbol{P}_2 &= \{P_{\theta,g} : 固定した \theta \in \Theta, 任意の\ g \in \boldsymbol{G}\}\end{aligned}$$

とする．$\boldsymbol{G}$ の中のなめらかな曲線 $t \mapsto g_t$（t は正の実数）を取る．これとベクトル $a \in \Theta$ に対して、$\boldsymbol{P}$ の中の曲線 $t \mapsto P_{\theta+ta,g_t}$ を考える．このとき、

$$\frac{\partial}{\partial t} \log dP_{\theta+ta,g_t}\bigg|_{t=0} = a^T \dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2$$

のように書けた仮定する．ただし、$\dot{\boldsymbol{l}}_1$ は θ に関する通常のスコア関数[39] とし、$\dot{\boldsymbol{l}}_2$ は θ を固定したときの g に対するスコア関数[40] と考えることができる．さらに、これの集まりを

$$\dot{\boldsymbol{P}}_2^0 = \{\dot{\boldsymbol{l}}_2 : \{g_t\}\ は\ \boldsymbol{G}\ の中の任意のなめらかな曲線\ \}$$

とし、$\dot{\boldsymbol{P}}_2 = \overline{[\dot{\boldsymbol{P}}_2^0]}$ である．

ある関数 $\tilde{\nu}_{\theta,g} \in L_2(P_{\theta,g})$ が存在して、

$$a = \frac{\partial}{\partial t}\nu(P_{\theta+ta,g_t}) = \dot{\nu}(a^T\dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2) = \langle \tilde{\nu}_{\theta,g}, a^T\dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2 \rangle_{L_2(P_{\theta,g})} \tag{6.37}$$

を満足するとする．$a = 0$ とおけば、$\tilde{\nu}_{\theta,g}$ は $\dot{\boldsymbol{P}}_2$ に直交しなければならないことがわかる．$L_2(P_{\theta,g})$ の中における $\dot{\boldsymbol{P}}_2$ への直交射影を $\Pi_0(\cdot|\dot{\boldsymbol{P}}_2)$ と書くことにする．

定義 4.9 のように ν の有効スコア関数を

$$\boldsymbol{l}_1^* = \dot{\boldsymbol{l}}_1 - \Pi_0(\dot{\boldsymbol{l}}_1|\dot{\boldsymbol{P}}_2)$$

で定義し、有効情報行列を

$$I(P_{\theta,g}|\nu, \boldsymbol{P}) : k \times k = P_{\theta,g}\{\dot{\boldsymbol{l}}_1^*(\dot{\boldsymbol{l}}_1^*)^T\}$$

---

39 すなわち、θ における $\dot{\boldsymbol{l}}_1 = (\partial/\partial\theta) \log dP_\theta$ である．

40 正確には $\boldsymbol{G}$ の接空間から $\dot{\boldsymbol{G}}$ から $L_2(P_{\theta,g})$ への線形写像であるスコア作用素の像である．

とかくことにする．どんな $a \in \boldsymbol{R}^k$ と $\dot{\boldsymbol{l}}_2 \in \dot{\boldsymbol{P}}_2$ に対しても $\boldsymbol{G}$ の中の曲線 $\{g_t\}$ が存在して、$t \downarrow 0$ のとき、

$$\int \left[\frac{\sqrt{dP_{\theta+ta,g_t}} - \sqrt{dP_{\theta,g}}}{t} - \frac{1}{2}(a^T \dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2)\sqrt{dP_{\theta,g}} \right]^2 \to 0 \tag{6.38}$$

を満足するとする[41]．このとき、$I(P_{\theta,g}|\nu, \boldsymbol{P})$ が正則であれば、系 4.3 から、$\nu(P_{\theta,G}) = \theta$ は $P_{\theta,g}$ において経路毎に微分可能[42] であることがわかる[43]．さらに、有効影響関数は

$$\tilde{\boldsymbol{l}}_1 = I^{-1}(P_{\theta,g}|\nu, \boldsymbol{P})\boldsymbol{l}_1^*$$

で与えられることがわかる．したがって、$P_{\theta,g}$ からのランダム標本 $X_1, \ldots, X_n$ に基づく θ の正則推定量を $\{T_n\}$ としたとき、

$$\sqrt{n}(T_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I^{-1}(P_{\theta,g}|\nu, \boldsymbol{P})\boldsymbol{l}_1^*(X_i) + o_P(1) \tag{6.39}$$

ならば、$\{T_n\}$ は有効推定量となる．

6.7.2 有効スコアー関数方程式

母数モデル $\boldsymbol{Q} = \{Q_\theta : \theta \in \Theta \subset \boldsymbol{R}^k\}$ において Q_θ からのランダム標本 $X_1, \ldots, X_n$ に基づいて θ を推定するとき、θ の最尤推定量は方程式

$$\sum_{i=1}^n \dot{\boldsymbol{l}}_\theta(X_i) = 0$$

の解として得られる．ただし、$\dot{\boldsymbol{l}}_\theta = (\partial/\partial\theta)\log dQ_\theta$ である．

セミパラメトリックモデル $\boldsymbol{P} = \{P_{\theta,g} : \theta \in \Theta \subset \boldsymbol{R}^k, g \in \boldsymbol{G}\}$ において θ を推定する場合には、有効スコア関数[44] による方程式

$$\sum_{i=1}^n \boldsymbol{l}_1^*(X_i; \theta, \hat{g}_n) = 0$$

[41] すなわち、$\dot{\boldsymbol{P}}$ を $\boldsymbol{P}$ の接空間とすれば、$\dot{\boldsymbol{P}} = \dot{\boldsymbol{P}}_1 + \dot{\boldsymbol{P}}_2$ が成立することである．

[42] (6.37) を満足する $a \in \boldsymbol{R}^k$ と $\dot{\boldsymbol{l}}_1 \in \dot{\boldsymbol{P}}_1$ に対して、$\dot{\boldsymbol{P}}$ 上の実数値関数 $dotnu_{\theta,g}$ が存在して、$t \downarrow 0$ のとき、

$$\left| \frac{\nu(P_{\theta+ta,g_t}) - \nu(P_{\theta,g})}{t} - \dot{\nu}_{\theta,g}(a^T \dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2) \right| \to 0$$

を満足することである．

[43] $\dot{\boldsymbol{l}}_1 = (l_{11}, \ldots, l_{1k})^T$ 、$\dot{\boldsymbol{l}}_1^* = (l_{11}^*, \ldots, l_{1k}^*)^T$ 、$a = (a_1, \ldots, a_k)^T$ 、$I^{-1}(P_{\theta,g}|\nu, \boldsymbol{P}) = (I^{ij})$ とおけば、$i = 1, \ldots, k$ に対して、

$$\langle \sum_j I^{ij} l_{1j}^*, a^T \dot{\boldsymbol{l}}_1 + \dot{\boldsymbol{l}}_2 \rangle_{L_2(P_{\theta,g})} = \sum_j \sum_k I^{ij} \langle l_{1i}, l_{1k} \rangle_{L_2(P_{\theta,g})} a_k = a_i$$

から (6.37) が成立することがわかる．

[44] これを今後のこの節では、$\boldsymbol{l}_1^*(\cdot; \theta, g) = \boldsymbol{l}_1^*(\cdot, P_{\theta,g}|\theta, \boldsymbol{P})$ と記すことにする．

の解を用いて θ の推定を行うことができる．ただし、$\hat{g}_n$ は g の適当な推定量である．

また、与えられた θ の値に対して、$\hat{g}_n(\theta)$ を求めて、つぎに、方程式

$$\sum_{i=1}^{n} \boldsymbol{l}_1^*(X_i;\theta,\hat{g}_n(\theta)) = 0$$

の解から θ の推定をすることもできる．有効スコア関数を陽にもとめることができる場合には、g の推定量の収束のオーダーが通常の $O(n^{-1/2})$ でなくともよい．

定理 6.20 ：モデル $\boldsymbol{P}$ は $P_{\theta,g}$ において正則で、正則な有効情報行列 $I(P_{\theta,g}|\theta,\boldsymbol{P})$ を持つとする．$\hat{\theta}_n$ を $\sqrt{n}P_n\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n) = o_P(1)$ を満足し、さらに、θ の一致推定量とする．局外母数 g の推定量 $\hat{g}_n$ は

$$\begin{aligned}
\sqrt{n}P_{\hat{\theta},g}\boldsymbol{l}^*(\cdot;\hat{\theta}_n,\hat{g}_n) &= o_P(1+\sqrt{n}\|\hat{\theta}_n-\theta\|) &(6.40)\\
P_{\theta,g}\|\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)-\boldsymbol{l}_1^*(\cdot;\theta,g)\| &\xrightarrow{P} 0 &(6.41)\\
P_{\hat{\theta}_n,g}\|\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)\|^2 &= O_P(1) &(6.42)
\end{aligned}$$

を満足するとする．さらに、$\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)$ は 2 乗可積分な envelope 関数をもつある Donsker クラスに確率 1 で含まれるとする．このとき、$\hat{\theta}_n$ は θ の有効推定量となる．

証明：$H_n(\theta',g') = \sqrt{n}(P_n-P_{\theta,g})\boldsymbol{l}_1^*(\cdot;\theta',g')$ を関数 $\boldsymbol{l}_1^*(\cdot;\theta',g')$ で添字付けられた経験過程とする．$\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)$ は確率 1 である Donsker クラスに含まれるので、

$$H_n(\hat{\theta}_n,\hat{g}_n) = H_n(\theta,g)+o_P(1) \tag{6.43}$$

がわかる[45]．$\hat{\theta}_n$ の定義と (6.40) を用いれば、

$$\sqrt{n}(P_{\hat{\theta}_n,g}-P_{\theta,g})\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n) = H_n(\theta,g)+o_P(1+\sqrt{n}\|\hat{\theta}_n-\theta\|) \tag{6.44}$$

がわかる．

つぎに上の式の左辺と $I^{-1}(P_{\theta,g};\theta,\boldsymbol{P})\sqrt{n}(\hat{\theta}_n-\theta)$ の差が $o_P(\sqrt{n}\|\hat{\theta}_n-\theta\|)$ で評価することを示す．$I(P_{\theta,g};\theta,\boldsymbol{P}) = P_{\theta,g}\{\boldsymbol{l}_1^*(\cdot;\theta,g)(\boldsymbol{l}_1^*(\cdot;\theta,g))^T\}$ であることに注意して書き換えれば、

$$\begin{aligned}
&\sqrt{n}(P_{\hat{\theta}_n,g}-P_{\theta,g})\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n) - I^{-1}(P_{\theta,g};\theta,\boldsymbol{P})\sqrt{n}(\hat{\theta}_n-\theta)\\
&= \sqrt{n}\int \boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})\\
&\times[(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})-\frac{1}{2}(\hat{\theta}_n-\theta)^T\dot{\boldsymbol{l}}_1\sqrt{p_{\theta,g}}]d\mu\\
&+\int \boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})\frac{1}{2}\sqrt{p_{\theta,g}}\dot{\boldsymbol{l}}_1^T d\mu\sqrt{n}(\hat{\theta}_n-\theta)\\
&-\int(\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)-\boldsymbol{l}_1^*(\cdot;\theta,g))\dot{\boldsymbol{l}}_1^T\sqrt{p_{\theta,g}}d\mu\sqrt{n}(\hat{\theta}_n-\theta)
\end{aligned} \tag{6.45}$$

---

[45] 実際、定理 6.2 と (6.41) から、$A_n = \{P_{\theta,g}\|\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)-\boldsymbol{l}_1^*(\cdot;\theta,g)\|^2 \le \eta\}$ とおいたとき、どんな正数 $\epsilon > 0$ と $\eta > 0$ に対しても、

$$\begin{aligned}
&P^*(\|\sqrt{n}(P_n-P_{\theta,g})(\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)-\boldsymbol{l}_1^*(\cdot;\theta,g))\| > \epsilon)\\
&\le P^*(\sup_{A_n}\|\sqrt{n}(P_n-P_{\theta,g})(\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)-\boldsymbol{l}_1^*(\cdot;\theta,g))\| > \epsilon)+P(A_n^c) \to 0
\end{aligned}$$

となる[46]．一項目[47]と三項目は Cauchy – Schwarz の不等式、モデルの正則性、(6.41)、(6.42) を用いれば、$o_P(\sqrt{n}\|\hat{\theta}_n-\theta\|)$ と評価できることがわかる．Cauchy – Schwarz の不等式を利用すれば、$n\to\infty$ のとき、$m_n\uparrow\infty$ となるどんな数列に対しても、

$$
\begin{aligned}
&\left\|\int \boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})\frac{1}{2}\sqrt{p_{\theta,g}}\dot{\boldsymbol{l}}_1^T d\mu\right\|^2\\
&\quad\lesssim \int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|\le m_n}\left\|\cdots\right\|^2 d\mu+\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|> m_n}\left\|\cdots\right\|^2 d\mu\\
&\quad\le \int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|\le m_n}\|\boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|^2 p_{\theta,g}d\mu\\
&\quad\times\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|\le m_n}(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})^2 d\mu\\
&\quad+\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|> m_n}\|\boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)\|^2(p_{\hat{\theta}_n,g}+p_{\theta,g})d\mu\\
&\quad\times\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|> m_n}\|\boldsymbol{l}_1^*(\,\cdot\,;\theta,g)\|^2 p_{\theta,g}d\mu
\end{aligned}
$$

となる．ここで、$m_n\|\hat{\theta}_n-\theta\|\xrightarrow{P}0$ となるようにゆっくりとしたスピードで $m_n\uparrow\infty$ となるように $\{m_n\}$ をとれば、

$$
\begin{aligned}
&\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|\le m_n}\|\boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)\boldsymbol{l}_1^*(\,\cdot\,;\theta,g)\|^2 p_{\theta,g}d\mu\int_{\|\dot{\boldsymbol{l}}_1(\,\cdot\,;\theta,g)\|\le m_n}(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})^2 d\mu\\
&\quad\le m_n^2 O_P(1)o_P(\|\hat{\theta}_n-\theta\|^2)=O_P(1)o_P(m_n^2\|\hat{\theta}_n-\theta\|^2)=o_P(1)
\end{aligned}
$$

となる．また、(6.42) から、

$$
\int_{\|\boldsymbol{l}_1^*(\,\cdot\,;\theta,g)\|> m_n}\|\boldsymbol{l}_1^*(\,\cdot\,;\theta,g)\|^2 p_{\theta,g}d\mu=o_P(1)
$$

[46] モデル $\boldsymbol{P}$ のある優測度 μ に対して、$p_{\theta,g}=dP_{\theta,g}/d\mu$ とした．

[47] 実際、Cauchy – Schwarz の不等式をも用いれば、

$$
\begin{aligned}
&\left|\sqrt{n}\int \boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)(\sqrt{p_{\hat{\theta}_n,g}}+\sqrt{p_{\theta,g}})[(\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}})-\frac{1}{2}(\hat{\theta}_n-\theta)^T\dot{\boldsymbol{l}}_1\sqrt{p_{\theta,g}}]d\mu\right|^2\\
&\quad\le\int\{\boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)\}^2(\sqrt{p_{\hat{\theta}_n,g}}+\sqrt{p_{\theta,g}})^2 d\mu\\
&\quad\times\int\left[\frac{\sqrt{p_{\hat{\theta}_n,g}}-\sqrt{p_{\theta,g}}}{1/\sqrt{n}}-\frac{1}{2}(\hat{\theta}_n-\theta)^T\dot{\boldsymbol{l}}_1\sqrt{p_{\theta,g}}\right]^2 d\mu\\
&\quad=O_P(1)o_p(n\|\hat{\theta}_n-\theta\|^2)
\end{aligned}
$$

となる．なぜなら、$\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n)$ は 2 乗可積分な envelope 関数を持つ Donsker クラスの確率 1 でふくまれるので、(6.41) と (6.42) から

$$
\int\{\boldsymbol{l}_1^*(\,\cdot\,;\hat{\theta}_n,\hat{g}_n)\}^2(\sqrt{p_{\hat{\theta}_n,g}}+\sqrt{p_{\theta,g}})^2 d\mu=O_P(1)
$$

がわかり、モデルの正則性より

$$
\left\{\int\left[\frac{\sqrt{p_{\theta+(1/\sqrt{n})\sqrt{n}(\hat{\theta}_n-\theta),g}}-\sqrt{p_{\theta,g}}}{1/\sqrt{n}}-\frac{1}{2}\sqrt{n}(\hat{\theta}_n-\theta)^T\dot{\boldsymbol{l}}_1\sqrt{p_{\theta,g}}\right]^2 d\mu\right\}^{1/2}=o_P(\sqrt{n}\|\hat{\theta}_n-\theta\|)
$$

となることがわかる．

がわかる．これらから、(6.45) の右辺の第二項目は $o_P(1)$ と評価ができる．したがって、

$$\sqrt{n}(P_{\hat{\theta}_n,g} - P_{\theta,g})\boldsymbol{l}_1^*(\cdot;\hat{\theta}_n,\hat{g}_n) - I^{-1}(P_{\theta,g};\theta,\boldsymbol{P})\sqrt{n}(\hat{\theta}_n - \theta) = o_P(\sqrt{n}\|\hat{\theta}_n - \theta\|) \tag{6.46}$$

最後に、(6.43) – (6.46) を合わせれば、

$$\begin{aligned} I^{-1}(P_{\theta,g}|\theta,\boldsymbol{P})\sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}(P_n - P)\boldsymbol{l}_1^*(\cdot;\theta,g) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta\|) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{l}_1^*(X_i;\theta,g) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta\|) \end{aligned}$$

が成立することがわかる．よって、定理は証明された． □

定理における条件 (6.40) 以外が成立していれば、同様な議論により

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n I^{-1}(P_{\theta,g}|\theta,\boldsymbol{P})\boldsymbol{l}_1^*(X_i|\theta,g) \\ &\quad + I^{-1}(P_{\theta,g}|\theta,\boldsymbol{P})\sqrt{n}P_{\hat{\theta}_n,g}\boldsymbol{l}_1^*(\cdot|\hat{\theta}_n,\hat{g}_n) + o_P(1) \end{aligned}$$

が成立することがわかる．したがって、(6.39) より、$\hat{\theta}_n$ が有効であるためには、条件 (6.40) が必要になることがわかる．

無バイアス条件 (6.40) はモデルの性質に依存することが以下の議論からわかる． まず、$\hat{\theta}_n = \theta + o_P(1)$ であれば、条件 (6.40) は

$$\sqrt{n}P_{\theta,g}\boldsymbol{l}_1^*(\cdot|\theta,\hat{g}_n) \xrightarrow{P} 0 \tag{6.47}$$

となることがわかる．さらに、$\boldsymbol{G}$ の中の曲線 $\{g_t = g + th\}$ に対して、$\dot{\boldsymbol{l}}_2 : \dot{\boldsymbol{G}} \longrightarrow L_2(P_{\theta,g})$ が存在し、

$$\|\sqrt{dP_{\theta,g_t}} - \sqrt{dP_{\theta,g}} - \frac{1}{2}\dot{\boldsymbol{l}}_2(h)\sqrt{dP_{\theta,g}}\|_{L_2(\mu)} = o(t)$$

となるので、

$$P_{\theta,g}\left\|\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}} - \frac{1}{2}\dot{\boldsymbol{l}}_2(\hat{g}_n - g)\right\| = o(\|\hat{g}_n - g\|)$$

となること[48] に注意する．さらに、$\boldsymbol{l}_1^*(\cdot|\theta,g)$ と $\dot{\boldsymbol{P}}_2$ の直交性を用いれば、

$$\begin{aligned} P_{\theta,g}\boldsymbol{l}_1^*(\cdot|\theta,\hat{g}_n) &= (P_{\theta,g} - P_{\theta,\hat{g}_n})(\boldsymbol{l}_1^*(\cdot|\theta,\hat{g}_n) - \boldsymbol{l}_1^*(\cdot|\theta,g)) \\ &\quad - P_{\theta,g}\boldsymbol{l}_1^*(\cdot|\theta,g)\left[\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}} - \dot{\boldsymbol{l}}_2(\hat{g}_n - g)\right] \end{aligned} \tag{6.48}$$

48

$$\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}} = \frac{2\sqrt{p_{\theta,g}}(\sqrt{p_{\theta,\hat{g}_n}} - \sqrt{p_{\theta,g}} - (1/2)\dot{\boldsymbol{l}}_2\sqrt{p_{\theta,g}})}{p_{\theta,g}} + \frac{(\sqrt{p_{\theta,\hat{g}_n}} - \sqrt{p_{\theta,g}})^2}{\sqrt{p_{\theta,g}}}$$

より、

$$\begin{aligned} &\left\{P_{\theta,g}\left\|\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}} - \dot{\boldsymbol{l}}_2(\hat{g}_n - g)\right\|\right\}^2 \\ &\le 4\int p_{\theta,g}^2 d\mu \int \{\sqrt{p_{\theta,\hat{g}_n}} - \sqrt{p_{\theta,g}} - (1/2)\dot{\boldsymbol{l}}_2\sqrt{p_{\theta,g}}\}^2 d\mu + \int (\sqrt{p_{\theta,\hat{g}_n}} - \sqrt{p_{\theta,g}})^2 d\mu \to 0 \end{aligned}$$

となることからわかる．

がわかる[49]．さらに、(6.41) とモデルの正則性から (6.48) の右辺の 1 項目と 2 項目は $o_P(\|\hat{g}_n - g\|)$ となること[50] がわかる．したがって、$\|\hat{g}_n - g\| = O_P(n^{-1/2})$ ならば、(6.47) が成立することがわかる．また、(6.48) の右辺のふたつの項のオーダーが $O_P(\|\hat{g}_n - g\|^2)$ ならば、$\|\hat{g}_n - g\| = o_P(n^{-1/4})$ であれば、(6.47) が成立する．

---

[49]

$$
\begin{aligned}
P_{\theta,g}\boldsymbol{l}_1^*(\,\cdot\,|\theta,g) &= (P_{\theta,g} - P_{\theta,\hat{g}_n})\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) \qquad \text{(なぜならば、} P_{\theta,\hat{g}_n}\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) = 0 \quad \text{)} \\
&= (P_{\theta,g} - P_{\theta,\hat{g}_n})(\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) - \boldsymbol{l}_1^*(\,\cdot\,|\theta,g)) - (P_{\theta,\hat{g}_n} - P_{\theta,g})\boldsymbol{l}_1^*(\,\cdot\,|\theta,g) \\
&= (P_{\theta,g} - P_{\theta,\hat{g}_n})(\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) - \boldsymbol{l}_1^*(\,\cdot\,|\theta,g)) - P_{\theta,g}\left\{\left(\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}}\right)\boldsymbol{l}_1^*(\,\cdot\,|\theta,g)\right\}
\end{aligned}
$$

となる．さらに、$\boldsymbol{l}_1^*(\,\cdot\,|\theta,g)$ の定義から、$P_{\theta,g}\boldsymbol{l}_1^*(\,\cdot\,|\theta,g)\dot{\boldsymbol{l}}_2 = 0$ であることに注意すれば、(6.48) はわかる．

[50] 実際、(6.41) を使えば、

$$
\begin{aligned}
&|(P_{\theta,g} - P_{\theta,\hat{g}_n})(\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) - \boldsymbol{l}_1^*(\,\cdot\,|\theta,g))| \\
&\quad \le \sqrt{\int\left(\frac{p_{\theta,g} - p_{\theta,\hat{g}_n}}{p_{\theta,g}}\right)^2 d\mu}\sqrt{\int\{\boldsymbol{l}_1^*(\,\cdot\,|\theta,\hat{g}_n) - \boldsymbol{l}_1^*(\,\cdot\,|\theta,g)\}^2 d\mu} \\
&\quad = O(1)o_P(\|\hat{g}_n - g\|)
\end{aligned}
$$

となる．また、モデルの正則性から、

$$
\begin{aligned}
&\left|P_{\theta,g}\boldsymbol{l}_1^*(\,\cdot\,|\theta,g)\left[\frac{p_{\theta,\hat{g}_n} - p_{\theta,g}}{p_{\theta,g}} - \dot{\boldsymbol{l}}_2(\hat{g}_n - g)\right]\right| \\
&\quad \le \sqrt{\int\{\boldsymbol{l}_1^*(\,\cdot\,|\theta,g)\}^2 p_{\theta,g}d\mu}\sqrt{\int\{(p_{\theta,\hat{g}_n} - p_{\theta,g} - \dot{\boldsymbol{l}}_2(\hat{g}_n - g)\}^2 d\mu} \\
&\quad = O(1)o_P(\|\hat{g}_n - g\|)
\end{aligned}
$$

となることがわかる．

6.7.3　一般の推定方程式

6.7.4　最尤推定量

6.7.5　漸近的に最も不利な部分モデル

6.7.6　同時最尤方程式

第II部

いろいろなモデルにおける有効推定量

第7章　Random Censoring モデル

7.1　Censor モデルとマルチンゲール

7.1.1　マルチンゲールと計数過程

$(\Omega, \mathcal{F}, P)$ を固定された完備[1] な確率空間とする．σ – 加法族の集まり $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t \subset \mathcal{F} : t \in [0,\infty]\}$ がつぎを満たすときに**情報増大系**（filtration）という：(i) $s < t$ に対し、$\mathcal{F}_s \subset \mathcal{F}_t$、(ii) $\{\mathcal{F}_t\}_{t\geq 0}$ は右連続：$\cap_{s>t}\mathcal{F}_s = \mathcal{F}_t$、(iii) $\{\mathcal{F}_t\}$ は完備[2] である．

$\mathcal{B}$ を $[0,\infty)$ のボレロ σ – 加法族とする．確率過程 $X = \{X_t : t \in [0,\infty\}$ が**可測**であるとは、$(t,w) \mapsto X_t(\omega)$ で与えられる $\boldsymbol{R}^+ \times \Omega$ 上の写像が、直積 σ – 加法族 $\mathcal{B} \times \mathcal{F}$ に関して可測であることである．また、$X = \{X_t\}$ が $(\boldsymbol{\mathcal{F}})$ – 適合（adapted ）であるとは、すべての $t \geq 0$ に対して、X_t は $\mathcal{F}_t$ 可測であることである．

$\overline{\boldsymbol{R}^+} = [0,\infty]$ の中の値を取る確率変数 T が $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t\}_{t\geq 0}$ に関する**停止時間**（stopping time）とは、任意の $t \geq 0$ に対し、$\{T \leq t\} \equiv \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t$ が成立するときにいう．

確率過程 $X = \{X_t\}_{t\geq 0}$ が**可積分**（integrable）とは、$\sup_{0\leq t<\infty} \boldsymbol{E}|X_t| < \infty$ が成立するときをいう．また、$\{X_t\}$ が**一様可積分**（uniformly integrable）とは、$\{X_t\}$ は可積分で、さらに、$\lim_{a\to\infty}\sup_{0\leq t<\infty} \boldsymbol{E}|X_t|1\{|X_t| \geq a\} = 0$ が成立するときをいう． また、$\{X_t\}$ が **２乗可積分**（squared integrable）とは、$\sup_{0\leq t<0} \boldsymbol{E}X_t^2 < \infty$ が成立するときをいう．確率過程 $\{X_t\}$ が局所的にある性質をもつとは、停止時間の増大列 $\{T_n\}$ で確率１で $T_n < 1$ で、かつ $T_n \uparrow \infty (n \to \infty)$ なるものが存在し、局所化された確率過程 $X(\cdot \wedge T_n), n = 1,2,\ldots$ がその性質をもつことである．

右連続な確率過程 $X = \{X_t\}_{t\geq 0}$ が $(\mathcal{F}_t)$ に関して**マルチンゲール**であるとは、

(i) $\{X_t\}$ は可積分であり、

(ii) $\{X_t\}$ は $(\mathcal{F}_t)$ – 適合、

(iii) $0 \leq s \leq t$ に対して、ほとんど確実に $\boldsymbol{E}[X_t|\mathcal{F}_s] = X_s$

が成立するときをいう． また、(iii) において、$\boldsymbol{E}[X_t|\mathcal{F}_s] \geq X_s$ が成立するとき、X は**劣マルチンゲール**（submartingale）といい、$\boldsymbol{E}[X_t|\mathcal{F}_s] \leq X_s$ が成立するとき、**優マルチンゲール**（supermartingale）という．

---

[1] $N' \subset N \in \mathcal{F}$ で $P(N) = 0$ ならば、$N' \in \mathcal{F}$ であること．したがって、$P(N') = 0$ である．確率空間 $(\Omega, \mathcal{F}, P)$ は $\tilde{\mathcal{F}} = \{A' = A \cup N' : A \in \mathcal{F}, N' \subset N \in \mathcal{F}, P(N) = 0\}$ と $\tilde{P}(A') = P(A)$ とすることにより、つねに完備な確率空間 $(\tilde{\Omega}, \tilde{\mathcal{F}}, P)$ に拡張可能である．このことを**完備化**という．

[2] $\mathcal{F}_0$ は $\mathcal{F}$ のすべての確率ゼロの集合を含む．

確率過程 $X=\{X_t\}$ が predictable とは、$(0,\infty)$ 上の左連続な適合過程の集まりによって生成さらた $(0,\infty)\times\Omega$ 上の σ – 加法族に関して X が可測であるときにいう．これは、$(s,t]\times A(A\in\mathcal{F}_s$ の集まりによって生成された σ – 加法族に関して X が可測であることと同値である．

$A=\{A_t\}_{t\geq 0}$ が $(\mathcal{F}_t)$ に関して可積分な**増分過程**であるとは、(i) $\{A_t\}$ は右連続、かつ $(\mathcal{F}_t)$ – 適合であり、(ii) 確率１で $A_0=0$ 、(iii) $t\mapsto A_t$ は非減少であり、(iv) 各 $t\geq 0$ に対し、$\boldsymbol{E}|A_t|<\infty$ である条件を満足するときにいく．

定理 7.1 （Doob – Meyer 分解）: $\{\mathcal{F}_t\}$ は右連続な情報増大系で、$X=\{X_t\}_{t\geq 0}$ は $(\Omega,\mathcal{F},P,\{\mathcal{F}_t\})$ 上の非負右連続 $(\mathcal{F}_t)$ – 適合な局所劣マルチンゲールであるとする．このとき、$\{X_t\}$ が局所一様可積分ならば、*predictable* な可積分増分過程 A_t で、各 $t\geq 0$ に対し、$P(A_t<\infty)=1$ で、X_t-A_t が右連続局所マルチンゲールとなるようなものが一意的に存在する．T_n を局所化停止時間とすれば、各 $t\geq 0$ に対し、$A_t=\lim_{n\to\infty}A_n(t)$ とすることができる．ただし、$A_n(t)$ は $X_{t\wedge T_n}$ に対する compensator である：すなわち、$X_{t\wedge T_n}-A_n(t)$ がマルチンゲールになるような増分過程である．

しばらく、$T>0$ と固定し、$t\in[0,T]$ とする．$\{X_t\}$ が２乗可積分マルチンゲールならば、Jensen の不等式から X_t^2 は劣マルチンゲールになり、しかも局所一様可積分になるので、Doob – Meyer 分解から、predictable で右連続増分過程 A_t が一意的に存在し、$X_t^2-A_t$ がマルチンゲールになることがわかる．この A_t を $\langle M\rangle_t$ と記し、X の**２次変分**（quadratic variateion）という．また、$X=\{X_t\}$ と $Y=\{Y_t\}$ が２乗可積分マルチンゲールとしたとき、X と Y に対し、

$$\begin{aligned}\langle X,Y\rangle &= (1/4)\{\langle X+Y\rangle-\langle X-Y\rangle\}\\ &= (1/2)\{\langle X+Y\rangle-\langle X\rangle-\langle Y\rangle\}\end{aligned}$$

とおけば、

$$XY-\langle X,Y\rangle=\{X_tY_t-\langle X,Y\rangle_t\}$$

はマルチンゲールとなる．$\langle X,Y\rangle$ を X と Y の**２次共分散過程**または**２次変分**という．

7.1.2 R – 作用素と L – 作用素

F を $\boldsymbol{R}$ 上の連続型分布関数とし、その確率密度関数を f とすれば、ハザード関数は

$$\lambda(t)=\frac{f(t)}{1-F(t)} \tag{7.1}$$

で定義される．これを確率密度関数からハザード関数への写像

$$f\mapsto\lambda=\frac{f}{1-F}$$

とみなすことができる．逆に、ハザード関数から確率密度関数への写像

$$\lambda\mapsto f=\lambda\exp(-\int_{-\infty}^{\cdot}\lambda(s)ds) \tag{7.2}$$

を考える．$\{f_r : r \in \boldsymbol{R}\}$ を f を通る曲線とし、それに対応したハザード関数と分布関数の曲線を $\{\lambda_r\}$ と $\{F_r\}$ と記し、(7.1) の両辺に対数をとり、t に関して微分すれば、

$$\frac{\partial}{\partial r}\log\lambda_r = \frac{\partial}{\partial r}\log f_r - \frac{1}{1-F_r}\int_{\cdot}^{\infty}\frac{\partial}{\partial r}\log f_r(s)ds$$

を得る．このことから、$a \in L_2^0(F)$ に対して、写像

$$(Ra)(t) = a(t) - \frac{1}{1-F(t)}\int_t^{\infty} a(s)ds \tag{7.3}$$

を定義する．これは

$$(Ra)(t) = -\boldsymbol{E}[a(X)-a(t)|X>t] \tag{7.4}$$

とも表現できる．(7.2) に関して同様な議論を行えば、

$$\frac{\partial}{\partial t}\log f_t = \frac{\partial}{\partial t}\log\lambda_t - \int_{-\infty}^{\cdot}\left(\frac{\partial}{\partial t}\log\lambda_t\right)\log\lambda_t(s)ds$$

となるので、$b \in L_2^0(F)$ に対して、

$$(Lb)(t) = b(t) - \int_{-\infty}^{t} b(s)\lambda(s)ds$$

を導入する．累積ハザード関数を

$$\Lambda(t) = \int_{-\infty}^{t}\lambda(s)ds$$

と書けば、

$$\begin{aligned}(Lb)(t) &= b(t) - \int_{-\infty}^{t} b(s)d\Lambda(s)\\ &= b(t) - \int_{-\infty}^{\infty} 1\{t \geq s\}b(s)d\Lambda(s)\end{aligned}$$

と表現できる．L は「マルチンゲール作用素」とみなすことができる：すなわち、X の分布関数を F とし、

$$M(t) = 1\{X \leq t\} - \int_{-\infty}^{t} 1\{X \geq s\}b(s)d\Lambda(s) \tag{7.5}$$

を計数過程マールチンゲールとすれば、

$$(Lb)(X) = \int_{-\infty}^{\infty} bdM$$

と表現できる．

命題 7.1 ：F を連続型分布関数とし、R と L をそれぞれ (7.1) と (7.2) によって定義される $L_2(F)$ から $L_2(F)$ への作用素とする．このとき、R と L は以下を満足する．
(i)．R と L は有界で、$\|R\| = \|L\| = 1$ である．

(ii) ．$L \circ Ra = a - \boldsymbol{E}[a(X)]$ と $R \circ La = a$ が成立する．したがって、$L_2^0(F)$ 上では、$R^{-1} = L$ である．
(iii) ．$L^T = R$ が成立し、$L^T L = R^T R = id$ である．
(iv) ．$\boldsymbol{var}[a(X)] = \boldsymbol{E}[a(X) - \tilde{A}(X)]^2$ が成立する．ただし、Y を X のコピーとして、$\tilde{A}(t) = \boldsymbol{E}[a(Y)|Y > t]$ である．で定義する．

証明：まず、(i) を証明する．任意に $a \in L_2(F)$ に対して、

$$Z(t) = \int_{-\infty}^{t} a dM$$

と定義する．ただし、M は (7.5) で定義された計数過程マルチンゲールで、変分過程 $\langle M \rangle(t) = \int_{-\infty}^{t} 1\{X \geq t\} d\Lambda(s)$ である．このとき、Z は 2 乗可積分なマルチンゲールで、$t \uparrow \infty$ のとき、

$$Z(t) \xrightarrow{a.s.} La(X)$$

で変分過程が

$$\langle Z \rangle(t) = \int_{-\infty}^{t} a^2 d\langle M \rangle$$

である．したがって、

$$\begin{aligned} \boldsymbol{E}[Z^2(t)] &= \boldsymbol{E}[\langle Z \rangle(t)] = \boldsymbol{E}[\int_{-\infty}^{t} a^2 d\langle M \rangle] \\ &= \boldsymbol{E}[\int_{-\infty}^{t} a^2(s) 1\{t \geq s\} d\Lambda(s) = \int_{-\infty}^{t} a^2 dF(s) \end{aligned}$$

となる．さらに、Fatou の補題を用いれば、

$$\boldsymbol{E}[(La(X))^2] = \boldsymbol{E}[\lim_{t \to \infty} Z^2(t)] \leq \liminf_{t \to \infty} \int_{-\infty}^{t} a^2 dF = \boldsymbol{E}[a^2(X)] < \infty$$

から、$L : L_2(F) \mapsto L_2(F)$ は有界であることがわかる．
　つぎに、(ii) を証明する．

$$\begin{aligned} (L \circ Ra)(t) &= L\left(a(t) - \frac{\int_t^\infty a(u) dF(u)}{1 - F(t)}\right) \\ &= a(t) - \frac{\int_t^\infty a(u) dF(u)}{1 - F(t)} - \int_{-\infty}^{t} \left\{a(s) - \frac{\int_s^\infty a(u) dF(u)}{1 - F(s)}\right\} d\Lambda(s) \end{aligned}$$

となることより、$L \circ Ra = a - \boldsymbol{E}[a(X)]$ を示すために、

$$\begin{aligned} \int_{-\infty}^{t} \left\{a(s) - \frac{\int_s^\infty a(u) dF(u)}{1 - F(s)}\right\} d\Lambda(s) &= \int_{-\infty}^{t} a(u) d\Lambda(u) + \frac{\int_t^\infty a(u) dF(u)}{1 - F(t)} \\ &\quad - \int_{-\infty}^{\infty} a(u) dF(u) \end{aligned}$$

を示せばよいことがわかる．実際、

$$\begin{aligned}
&\int_{-\infty}^{t}\int_{s}^{\infty}\frac{a(u)}{1-F(s)}dF(u)d\Lambda(s)=\int_{-\infty}^{t}\int_{s}^{\infty}\frac{a(u)}{(1-F(s))^2}dF(u)dF(s)\\
&=\int_{-\infty}^{t}a(u)\left[\int_{-\infty}^{u}\frac{dF(u)}{(1-F(s))^2}\right]dF(u)+\int_{f}^{\infty}a(u)\left[\int_{-\infty}^{t}\frac{dF(u)}{(1-F(s))^2}\right]dF(u)\\
&=\int_{-\infty}^{t}a(u)\left(\frac{1}{1-F(u)}-1\right)dF(u)+\int_{t}^{\infty}a(u)\left[\frac{1}{1-F(t)}-1\right]dF(u)
\end{aligned}$$

よりわかる．

また、

$$\begin{aligned}
&R\circ La=R\left(b(t)-\int_{-\infty}^{t}b(u)d\Lambda(u)\right)\\
&=b(t)-\int_{-\infty}^{t}b(u)d\Lambda(u)-\frac{1}{1-F(t)}\int_{t}^{\infty}\left\{a(s)-\int_{-\infty}^{s}b(u)d\Lambda(u)\right\}dF(s)
\end{aligned}$$

となることより、$R\circ La=a$ を示すために、

$$\int_{t}^{\infty}\int_{-\infty}^{s}b(u)d\Lambda(u)dF(s)=(1-F(t))\int_{-\infty}^{t}b(u)d\Lambda(u)+\int_{t}^{\infty}b(s)dF(s)$$

を示せばよい．実際、上の式の左辺は

$$\int_{-\infty}^{t}b(u)\left[\int_{t}^{\infty}dF(s)\right]d\Lambda(u)+\int_{t}^{\infty}b(u)\left[\int_{u}^{\infty}dF(s)\right]d\Lambda(u)$$

と変形すればわかる．

つぎに、(iii) を示す．Fubini の定理を用いれば、任意の $a,b\in L_2(F)$ に対して、

$$\begin{aligned}
\langle La,b\rangle_{L_2(F)}&=\int_{-\infty}^{\infty}b(t)\left\{a(t)-\int_{-\infty}^{t}a(u)d\Lambda(u)\right\}dF(t)\\
&=\int_{-\infty}^{\infty}a(t)\left\{b(t)-\frac{\int_{t}^{\infty}b(u)dF(u)}{1-F(t)}\right\}dF(t)=\langle a,Rb\rangle_{L_2(F)}
\end{aligned}$$

からわかる[3]．さらに、このことより、$R:L_2^0(F)\mapsto L_2^0(F)$ と $L:L_2^0(F)\mapsto L_2^0(F)$ は isometry であることがわかるは

$$\begin{aligned}
\|La\|_{L_2(F)}^2&=\langle La,La\rangle_{L_2(F)}=\langle a,L^TLa\rangle_{L_2(F)}=\langle a,RLa\rangle_{L_2(F)}\\
&=\langle a,a\rangle_{L_2(F)}=\|a\|_{L_2(F)}
\end{aligned}$$

[3] 実際、

$$\begin{aligned}
\int_{-\infty}^{\infty}b(t)\left\{\int_{-\infty}^{t}a(u)d\Lambda(u)\right\}dF(t)&=\int_{\infty}^{\infty}b(t)\left\{\int_{-\infty}^{t}\frac{a(u)}{1-F(u)}dF(u)\right\}dF(t)\\
&=\int_{-\infty}^{\infty}\frac{a(u)}{1-F(u)}\left\{\int_{u}^{\infty}b(t)dF(t)\right\}dF(u)
\end{aligned}$$

に注意すればよい．

よりわかる．

(iv) はつぎの計算からわかる．$\langle \boldsymbol{E}a, a - \boldsymbol{E}a\rangle_{L_2(F)} = 0$ と (ii) から

$$\begin{aligned}\boldsymbol{var}[a(X)] &= \langle a, a - \boldsymbol{E}a\rangle_{L_2(F)} = \langle a, LRa\rangle_{L_2(F)} = \langle a, R^T Ra\rangle_{L_2(F)} \\ &= \langle Ra, Ra\rangle_{L_2(F)} = \|Ra\|_{L_2(F)} = \boldsymbol{E}[a(X) - \tilde{A}(X)]^2\end{aligned}$$

からわかる． □

7.1.3 $R-$ 作用素とマルチンゲールとの関係

X を確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数とし、その分布関数を F と記す．すべての $t \geq 0$ に対し、

$$\mathcal{F}_t = \sigma\{1\{X \leq s\} : s \leq t\} = \sigma\{X \wedge t, 1\{X \leq t\}\} \tag{7.6}$$

で定義された情報増大系 $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t\}$ を考える．

固定した $a \in L_2(F)$ と $t \geq 0$ に対し、

$$Y(t) \equiv \boldsymbol{E}[a(X)|\mathcal{F}_t]$$

とおく．モデル $\boldsymbol{Q} = \{q_\theta : \theta \in \Theta \subset \boldsymbol{R}\}$ における θ のスコア関数を $(\partial/\partial\theta)\log q_\theta = a$ とかけば、$Y(t)$ は欠損値データ $X \wedge t$ に基づくモデル $\boldsymbol{P} = \{p_\theta : \theta \in \Theta \subset \boldsymbol{R}\}$ における θ のスコア関数となることが命題 5.1 よりわかる．また、(7.6) より

$$\begin{aligned}Y(t) &= 1\{X \leq t\}a(X) + 1\{X > t\}\frac{\boldsymbol{E}[a(X)1\{X > t\}}{1 - F(t)} \\ &= 1\{X \leq t\}a(X) + 1\{X > t\}\tilde{A}(t)\end{aligned} \tag{7.7}$$

と表現できることがわかる．

つぎに命題はマルチンゲールと計数過程マルチンゲール

$$M(t) \equiv N(t) - A(t) = 1\{X \leq t\} - \int_{-\infty}^{t} 1\{X \geq s\}d\Lambda(s) \tag{7.8}$$

との関係を示すものである．

命題 7.2 ：$a \in L_2^0(F)$ とする[4]．$Y(t)$ と $M(t)$ を (7.7) と (7.8) で定義されたマルチンゲール確率過程とする．このとき、

$$Y(t) = \int_{-\infty}^{t} Ra(s)dM(s) \tag{7.9}$$

が成立する．ただし、

$$Ra(t) = a(t) - \frac{\int_t^\infty adF}{1 - F(t)} = a(t) - \tilde{A}(t)$$

である．

[4] すなわち、$\int a^2 dF < \infty$ かつ $\int adF = 0$ である．

証明：$Ra = a - \tilde{A}$ に注意すれば、

$$\int_{-\infty}^{t} RadN = a(X)1\{X \le t\} - \tilde{A}(X)1\{X \le t\} \tag{7.10}$$

となる[5]．さらに、Fubini の定理を用いれば、

$$\begin{aligned}
&\int_{-\infty}^{t}\left\{\frac{1}{1-F(s)}\int_{s}^{\infty} a(u)dF(u)\right\}d\Lambda(s) \\
&= \int_{-\infty}^{t} a(u)\left\{\int_{-\infty}^{u}\frac{d\Lambda(s)}{1-F(s)}\right\}dF(u) + \int_{t}^{\infty} a(u)\left\{\int_{-\infty}^{t}\frac{d\Lambda(s)}{1-F(s)}\right\}dF(u) \\
&= \int_{-\infty}^{t} a(u)d\Lambda(u) + \frac{1}{1-F(t)}\int_{t}^{\infty} a(u)dF(u) \\
&= \int_{-\infty}^{t} a(u)dF(u) + \boldsymbol{E}[a(X)|X > t]
\end{aligned}$$

となること[6]から、

$$\int_{-\infty}^{u} Rad\Lambda = \int_{-\infty}^{u} ad\Lambda - \int_{-\infty}^{t}\left\{\frac{1}{1-F(s)}\int_{s}^{\infty} adF\right\}d\Lambda(s) = -\tilde{A}(u)$$

のなることがわかる．これを用いれば、

$$\begin{aligned}
\int_{-\infty}^{t} RadA &= \int_{-\infty}^{t\wedge X} Rad\Lambda \\
&= \int_{-\infty}^{t} Rad\Lambda 1\{X \le t\} + \int_{-\infty}^{t} Rad\Lambda 1\{X > t\} \\
&= -\tilde{A}(X)1\{X \le t\} - \tilde{A}(t)1\{X > t\}
\end{aligned} \tag{7.11}$$

を得る．$M = N - A$ に注意すれば、(7.10) と (7.11) から (7.9) が導かれることがわかる． □

7.2 Random censoring モデル：ノンパラメトリックの観点から

確率変数 Y と C は独立で $\boldsymbol{R}^{+} = [0, \infty)$ 上の分布関数 F と G を持つ[7]とする．Y を生存時間とし、C を打ち切り時間（censoring time）とする． $X^0 = (Y, C)$ を未観測な確率変数、$X = (T, \Delta)$ を観測できる確率変数とする．ここで、

$$T \equiv Y \wedge C, \qquad \Delta = 1\{Y \le C\}$$

[5] $N = 1\{X \le t\}$ より、$\int f(t)dN(t) = f(X)1\{X \le t\}$ なることに注意すればよい．

[6] 最後からふたつ目の等式は

$$\int_{-\infty}^{u}\frac{d\Lambda(s)}{1-F(s)} = \int_{-\infty}^{u}\frac{dF(s)}{(1-F(s))^2} = \frac{1}{1-F(u)}$$

に注意すればよい．

[7] ただし、どちらか一方は ∞ 上に mass をもつことも許すとする．

とする．F と G が Lebeague 測度に関する確率密度関数 f と g をもつと仮定し、$\boldsymbol{R}^* \times \{0,1\}$ の中に値を取る確率変数 $X = (T, \Delta)$ の確率密度関数[8] $p(\cdot : F, G)$ は $t \in \boldsymbol{R}^*$ と $\delta \in \{0,1\}$ に対して、

$$p(t,\delta) = \{(1-G(t))f(t)\}^{\delta}\{(1-F(t))g(t)\}^{1-\delta} \tag{7.12}$$

で与えられる．

ここで、

$$\begin{aligned} H(t) &\equiv P(T \le t) = H_0(t) + H_1(t) \\ &\equiv \int_0^t p(s,0)ds + \int_0^t p(s,1)ds \\ &= \int_0^t (1-F(s))ds + \int_0^t (1-G(s))ds \end{aligned}$$

とおく．これから、

$$1 - H(t) = P(T > t) = P(Y > t, C > t) = (1-F(t))(1-G(t))$$

であるので、

$$\frac{f(t)}{1-F(t)} = \frac{p(t,1)}{1-H(t)}, \qquad \frac{g(t)}{1-G(t)} = \frac{p(t,0)}{1-H(t)} \tag{7.13}$$

となる．また、上のそれぞれの式の両辺を積分すれば、

$$1 - F(t) = \exp\left(-\int_0^t \frac{p(s,1)}{1-H(s)}ds\right), \qquad 1 - G(t) = \exp\left(-\int_0^t \frac{p(s,0)}{1-H(s)}ds\right) \tag{7.14}$$

を得る．このことより、F と G は $[0, H^{-1}(1))$ 上で identifiable である．ただし、

$$H^{-1}(1) = \inf\{s : H(s) = 1\} = F^{-1}(1) \wedge G^{-1}(1)$$

である． $\boldsymbol{M}_\mu$ を $\boldsymbol{R}^+\{0,1\}$ 上の確率測度で μ に関して絶対連続で proper [9] なものの集まりとし、

$$\boldsymbol{P} = \{P \text{ は } \boldsymbol{R}^+ \times \{0,1\} \text{ 上の確率測度で } p = \tfrac{dP}{d\mu} \text{ は (7.13) の形}\}$$

とすれば、(7.13) と (7.14) から

$$\boldsymbol{P} = \boldsymbol{M}_\mu$$

が成立すること[10] がわかる．

---

[8] これは、$\mu \equiv$ Lebeague 測度 $\times$ counting measure に関するものである．

[9] すなわち、$\lim_{t\to-\infty} P(T \le t, \Delta = \cdot) = 0$ かつ $\lim_{t\to\infty} P(T \le t, \Delta = \cdot) = 1$ を満足するもの．

[10] P が与えられたとき、F か G のどちらか一方が proper であることをあとは示せばよい．実際、p を proper とすれば、$t \uparrow \infty$ のとき、

$$-\log\{(1-F(t))(1-G(t))\} = \int_0^t \frac{p(s,0)+p(s,1)}{1-H(s)}ds = \int_0^t \frac{dH(s)}{1-H(s)} = -\log(1-H(s)) \to \infty$$

となることより、すくなくともどちらか一方は proper である．すなわち、$\lim_{t\to\infty} F(t) \wedge G(t) = 1$ が成立する．

したがって、固定した $P \in \boldsymbol{P}$ に対して、

$$\dot{\boldsymbol{P}} = \{h \in L_2(P) : \int ddP = 0\} = L_2^0(P)$$

を定義する．しかし、この接空間は未知の分布関数 F と G に依存する．ここで、

$$\nu(P) \equiv \Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s-)} \qquad (0 \le t \le \tau 0)$$

と

$$\kappa(P) \equiv 1 - F(t) = \Pi_{0 \le s \le t}(1 - d\Lambda(s)) = \exp(-\Lambda^c(t))\Pi_{0 \le s \le t}(1 - \Delta\Lambda(s))$$

の推定問題を考えよう．ただし、τ は $1 - H(\tau) > 0$ なる点とし、

$$\Delta\Lambda = \Lambda(t) - \Lambda(t-), \qquad \Lambda^c(t) = \Lambda(t) - \sum_{0 \le s \le t} \Delta\Lambda(s)$$

とする．

7.2.1 有効影響作用素と逆情報作用素の計算

まず、形式的に $\dot{\nu}_P$ を計算する．そのために、$h \in \dot{\boldsymbol{P}}$ に対して、

$$p_\eta(s, \delta) = p(s, \delta) \exp(\eta h(s, \delta)) \qquad (\eta \in \boldsymbol{R})$$

とし、p_η に対応する分布関数を P_η と書くことにする．すると、

$$\begin{aligned}\frac{\partial}{\partial \eta} \nu(P_\eta)(t)\Big|_{\eta=0} &= \int_0^\infty \left\{ \frac{\mathbf{1}_{[0,t]}(s)}{\bar{H}(s)} - C(s \wedge t) \right\} h(s,1)p(s,1)ds \\ &\quad - \int_0^\infty C(s \wedge t)h(s,0)p(s,0)ds\end{aligned}$$

となること[11]がわかる．ただし、

$$C(t) = \int_0^t \frac{1}{\bar{H}^2} dH_1 = \int_0^t \frac{1}{\bar{H}} d\Lambda \tag{7.15}$$

[11] これは以下の計算からわかる．(7.13) と (7.14) において、p_η に対応して得られるものを添字 η を付けて表現することにする．まず、$H_{1\eta}(t) = \int_0^t p_\eta(s,1)ds = \int_0^s p(s,1)\exp(\eta h(s,1))ds$ と

$$\begin{aligned}\bar{H}_\eta(t) &= \int_0^t p_\eta(s,0)ds + \int_0^t P_\eta(s,1)ds \\ &= \int_0^t p(s,0)\exp(\eta h(s,0))ds + \int_0^t p(s,1)\exp(\eta h(s,1))ds\end{aligned}$$

に注意すれば、

$$\begin{aligned}\frac{\partial}{\partial \eta} \int_0^t \frac{dH_{1\eta}(s)}{\bar{H}_\eta(s)}\Big|_{\eta=0} &= \int_0^t \frac{1}{\bar{H}(s)} \frac{\partial}{\partial \eta} dH_{1\eta}(s) + \int_0^t \frac{1}{\bar{H}^2(s)} \frac{\partial}{\partial \eta} \bar{H}_\eta(s)\Big|_{\eta=0} dH_1(s) \\ &= \int_0^t \frac{p(s,1)h(s,1)}{\bar{H}(s)} ds - \int_0^t \frac{1}{\bar{H}^2(s)} \left[\int_0^s \{p(u,0)h(u,1) + p(u,1)h(u,1)\} ds \right] dH_1(s) \\ &= \int_0^t \left\{ \frac{\mathbf{1}_{[0,t]}(s)}{\bar{H}(s)} - C(s \wedge t) \right\} h(s,1)p(s,1)ds - \int_0^\infty C(s \wedge t)h(s,0)p(s,0)ds\end{aligned}$$

となることからわかる．

である．厳密には、

$$\sum_{\delta=0}^{1}\int\{\sqrt{p_\eta}-\sqrt{p}-\frac{1}{2}\eta\dot{\nu}_P\sqrt{p}\}^2 ds=o(\eta)$$

を示せば、$\dot{\nu}_P$ は ν の微分であることがわかる．直接計算するのではなく、Van der Vaart の可微分性定理を用いてそれを後で証明することにする．ここで、微分 $\dot{\nu}_P$ が与えられたと仮定して議論を進める．

(7.12) の形から

$$\begin{aligned}\dot{\nu}_P(h)(t) &= \sum_{\delta=0}^{1}\int h(s,\delta)\left\{\frac{\delta 1_{[0,t]}(s)}{\bar{H}(s)}-C(s\wedge t)\right\}dP(s,\delta)\\ &= \langle h,\frac{\delta 1_{[0,t]}(s)}{\bar{H}(s)}-C(s\wedge t)\rangle_{L_2(P)}\end{aligned}$$

なので、canonical gradient は、$0<s<\tau_H$ と $\delta\in\{0,1\}$ に対して、

$$\tilde{\nu}_{P,t}(s,\delta)=\frac{\delta 1_{[0,t]}(s)}{\bar{H}(s)}-C(s\wedge t)$$

で与えられる．したがって、(5.61) から、有効影響作用素は

$$\tilde{\boldsymbol{l}}_\nu(\pi_t)(s,\delta)=\tilde{\nu}_{P,t}(s,\delta)$$

となることがわかる．さらに、(5.63) から、逆情報共分散汎関数 $I^{-1}(P|\nu,\boldsymbol{P})$ は

$$I^{-1}(P|\nu,\boldsymbol{P})(s,t)=I_\nu^{-1}(s,t)=\boldsymbol{E}[\tilde{\nu}_{P,s}(T,\Delta)\tilde{\nu}_{P,t}(T,\Delta)]$$

で与えられることがわかる．上の式の右辺をさらに計算する．そのために、

$$\begin{aligned}\int_0^t C^2 dH &= -\int_0^t C^2\bar{H}=\int_0^t \bar{H}d(C^2)-C^2(t)\bar{H}(t)\\ &= 2\int_0^t\frac{C}{\bar{H}}dH_1-C^2(t)\bar{H}(t) \qquad (7.16)\end{aligned}$$

と $s\le t$ としたとき、

$$\int_s^t CdH=-\int_s^t Cd\bar{H}=\int_s^t\frac{1}{\bar{H}}dH_1-\bar{H}(t)C(t)+\bar{H}(s)C(s) \qquad (7.17)$$

が部分積分から用いれば得られることに注意する．つぎに、

$$\begin{aligned}\boldsymbol{E}[C(s\wedge T)C(t\wedge T)] &= C(s)C(t)\int_{s\vee t}^{\infty}dH+C(s\wedge t)\int_{s\wedge t}^{s\vee t}C(u)dH(u)\\ &\quad+\int_0^{s\wedge t}C^2(u)dH(u)dH(u)\end{aligned}$$

となることに注意する．(7.16) と (7.17) を上の式の右辺の二項目と三項目に対応すれば、

$$\begin{aligned}&C(s\wedge t)\int_{s\wedge t}^{s\vee t}C(u)dH(u)\\ &=C(s\wedge t)\left\{\int_{s\wedge t}^{s\vee t}\frac{dH_1}{\bar{H}}-\bar{H}(s\vee t)C(s\vee t)+\bar{H}(s\wedge t)C(s\wedge t)\right\} \qquad (7.18)\end{aligned}$$

と

$$\int_0^{s\wedge t} C^2(u)dH(u)dH(u) = 2\int_0^{s\wedge t}\frac{C}{\bar{H}}dH_1 - C^2(s\wedge t)\bar{H}(s\wedge t) \tag{7.19}$$

となることわかる．また、

$$\begin{aligned}
\boldsymbol{E}\left[\frac{\Delta 1_{[0,t]}(T)C(s\wedge T)}{\bar{H}^2(T)}\right] &= \int_0^t \frac{C(s\wedge u)}{\bar{H}(u)}dH_1(u) \\
&= \int_0^{s\wedge t}\frac{C(u)}{\bar{H}(u)}dH_1(u) + C(s\wedge t)\int_{s\wedge t}^t \frac{1}{\bar{H}(u)}dH_1(u) \qquad (7.20) \\
\boldsymbol{E}\left[\frac{\Delta 1_{[0,s]}(T)C(t\wedge T)}{\bar{H}^2(T)}\right] &= \int_0^s \frac{C(t\wedge u)}{\bar{H}(u)}dH_1(u) \\
&= \int_0^{s\wedge t}\frac{C(u)}{\bar{H}(u)}dH_1(u) + C(s\wedge t)\int_{s\wedge t}^s \frac{1}{\bar{H}(u)}dH_1(u) \qquad (7.21)
\end{aligned}$$

となる．さらに、

$$\boldsymbol{E}\left[\frac{\Delta 1_{[0,s]}(T)\Delta 1_{[0,t]}(T)}{\bar{H}^2(T)}\right] = \int_0^{s\wedge t}\frac{dH_1(u)}{\bar{H}^2(u)} = C(s\wedge t) \tag{7.22}$$

となることがわかる．したがって、(7.18) – (7.22) から

$$\boldsymbol{E}[\tilde{\nu}_{P,s}(T,\Delta)\tilde{\nu}_{P,t}(T,\Delta)] = C(s\wedge t)$$

が得られる．これから、定理 5.3 における Gauss 過程 Z_0 は

$$Z_0(t) = B(C(t)) \qquad (0\le t\le\tau)$$

で与えられる．ただし、B は Brown 過程である．

つぎに、F は連続と仮定（したがって、Λ も連続）して、$0\le t\le\tau$ に対して、

$$\kappa(P)(t) \equiv 1 - F(t) = \Pi_{0\le s\le t}(1-d\Lambda) = \exp(-\Lambda(t)) = \alpha(\nu(P))(t)$$

の推定を考える．ただし，$b\in l^\infty([0,\tau])$ に対して、$\alpha : l^\infty([0,\tau]) \mapsto l^\infty([0,\tau])$ を $\alpha(b) = \exp\circ(-b)$ で定義した．すると、$\dot{\alpha}_b(\cdot) : l^\infty([0,\tau]) \mapsto l^\infty([0,\tau])$ は、任意の $h\in l^\infty([0,\tau])$ に対して、

$$\dot{\alpha}_b(h) = -\alpha(b)h$$

で与えられること[12] がわかる．したがって、$h\in L_2(P)$ に対して、

$$\dot{\kappa}_P(h) = \dot{\alpha}_\Lambda(\dot{\nu}_P(h)) = -\alpha(\Lambda)\dot{\nu}_P(h)$$

となることが chain rule（補題 6.8）よりわかる．さらに、Reisz の表現定理より、κ の canonical gradient $\tilde{\kappa}_{P,t}$ は、$0\le t\le\tau$ と $(s,\delta)\in[0,\infty)\times\{0,1\}$ に対して、

$$\tilde{\kappa}_{P,t}(s,\delta) = -\alpha(\Lambda)\tilde{\nu}_{P,t}(s,\delta) = -(1-F(t))\tilde{\nu}_{P,t}(s,\delta)$$

[12] 実際、$\frac{\alpha(b+\eta h)-\alpha(b)}{\eta} = \frac{1}{\eta}\exp(-\eta h)\alpha(b)$ と $\lim_{\eta\to 0}\frac{1}{\eta}\exp(-\eta h) = -h$ となることからわかる．

で与えられることがわかる．したがって、κ に対する逆情報共分散汎関数 $I^{-1}(P|\kappa,\boldsymbol{P})$ は

$$\begin{aligned} I^{-1}(P|\kappa,\boldsymbol{P})(s,t) &= I_\kappa^{-1}(s,t) = (1-F(s))(1-F(t))C(s\wedge t) \\ &= \frac{\bar{F}}{\bar{K}}(s)\frac{\bar{F}}{\bar{H}}(t)\{K(s\wedge t)-K(s)K(t)\} \end{aligned}$$

で与えられることがわかる．ここで、

$$K = \frac{C}{1+C}, \qquad \bar{K} = 1-K = \frac{1}{1+C}, \qquad \bar{F} = 1-F$$

とした．さらに、定理 5.3 における Gauss 過程 Z_0 は

$$\boldsymbol{L}(Z_0) = \boldsymbol{L}(\bar{F}B(C)) = \boldsymbol{L}\left(\frac{\bar{F}}{\bar{K}}B_0(K)\right)$$

で与えられることがわかる．ただし、B と B_0 は Brown 過程 と Brown 橋である．

7.2.2 スコアー作用素とマルチンゲール

．いま、$\dot{\boldsymbol{G}}_1 = L_2^0(F)$ と $\dot{\boldsymbol{G}}_2 = L_2^0(G)$ から $\dot{\boldsymbol{P}}$ へのスコア作用素 $\dot{\boldsymbol{l}}_1 : \dot{\boldsymbol{G}}_1 \mapsto \dot{\boldsymbol{P}}$ と $\dot{\boldsymbol{l}}_2 : \dot{\boldsymbol{G}}_2 \mapsto \dot{\boldsymbol{P}}$ は命題 5.1 から、任意の $a \in \dot{\boldsymbol{G}}_1$ と $b \in \dot{\boldsymbol{G}}_2$ に対して、

$$\begin{aligned} (\dot{\boldsymbol{l}}_1 a)(T,\Delta) &= \boldsymbol{E}[a(X)|(T,\Delta)] \\ &= \Delta a(T) + (1-\Delta)\frac{\int_T^\infty a dF}{1-F(T)} \\ &= \Delta a(T) + (1-\Delta)\boldsymbol{E}[a(Y)|Y>T] \end{aligned} \tag{7.23}$$

と

$$\begin{aligned} (\dot{\boldsymbol{l}}_2 b)(T,\Delta) &= \boldsymbol{E}[b(C)|(T,\Delta)] \\ &= (1-\Delta)b(T) + \Delta\frac{\int_T^\infty b dG}{1-G(T)} \\ &= (1-\Delta)b(T) + \Delta\boldsymbol{E}[b(C)|C>T] \end{aligned}$$

で与えられる．

さらに、

$$M_{uc}(t) \equiv 1\{T\le t, \Delta=1\} - \int_0^T 1\{T\ge s\}d\Lambda(s)$$

と

$$M_c(t) \equiv 1\{T\le t, \Delta=0\} - \int_0^T 1\{T\ge s\}\frac{dG(s)}{1-G(s)}$$

とおけば、

$$(\dot{\boldsymbol{l}}_1 a)(T,\Delta) = \int R_F a dM_{uc}$$

と

$$(\dot{\boldsymbol{l}}_2 b)(T,\Delta) = \int R_G b dM_c$$

と表現できること[13] がわかる．ただし、R_F と R_G は (7.4) で定義された F と G に関する R – 作用素である．

マルチンゲールの性質より、$\langle M_{uc}, M_c\rangle = 0$ がわかる．これより、

$$\langle \dot{\boldsymbol{l}}_1 a, \dot{\boldsymbol{l}}_2 b\rangle_{L_2(P)} = \boldsymbol{E}\left[\int_0^\infty (R_F a R_G b) d\langle M_{uc}, M_c\rangle\right] = 0$$

となることがわかる．したがって、

$$\mathrm{Image}(\dot{\boldsymbol{l}}_1) \perp \mathrm{Image}(\dot{\boldsymbol{l}}_2) \tag{7.24}$$

で

$$\dot{\boldsymbol{l}}_1^T \dot{\boldsymbol{l}}_2 = 0, \qquad \dot{\boldsymbol{l}}_2^T \dot{\boldsymbol{l}}_1 = 0$$

が成立する．

さらに、この直交性を $\dot{\boldsymbol{P}}_1 \supset \overline{\mathrm{Image}(\dot{\boldsymbol{l}}_1)}$ と $\dot{\boldsymbol{P}}_2 \supset \overline{\mathrm{Image}(\dot{\boldsymbol{l}}_2)}$ の直交性：$\dot{\boldsymbol{P}}_1 \perp \dot{\boldsymbol{P}}_2$ まで拡張できることを示す．

まずはじめに、どんな $h \in \dot{\boldsymbol{P}}_1^0$ に対しても

$$h(t,0) = -\frac{1}{\bar{F}(t)}\int_0^t h(u,1)dF(u), \qquad 0 \le t \le \tau_H \tag{7.25}$$

が成立することを示す．ただし、$\bar{F} = 1 - F$ である．これを示すために、接ベクトル $h \in \dot{\boldsymbol{P}}_1^0$ をもつ $\boldsymbol{P}$ の中の曲線[14] $\{P_\eta : \eta \ge 0\}$ を取る．この曲線に対応する確率密度関数の曲線[15] を $\{p_\eta\}$ と記し、これから得られる $\boldsymbol{G}_1$ の中の曲線（確率密度関数で表現されたもの）を f_η と書

---

[13]

$$\begin{aligned}
\int R_F a dM_{uc} &= \int \left\{a(t) - \frac{\int_t^\infty a(s)dF(s)}{1-F(t)}\right\} dM_{uc} \\
&= \Delta\left\{a(T) - \frac{\int_T^\infty a(s)dF(s)}{1-F(T)}\right\} - \int_0^T a(t)d\Lambda(t) + \int_0^T \frac{\int_t^\infty a(s)dF(s)}{1-F(t)} d\Lambda(t)
\end{aligned}$$

となることに注意する．さらに、Fubini の定理から

$$\begin{aligned}
\int_0^T \frac{\int_t^\infty a(s)dF(s)}{1-F(t)} d\Lambda(t) &= \int_0^T a(s)\left\{\int_0^s \frac{d\Lambda(t)}{1-F(t)}\right\} dF(s) + \int_T^\infty a(s)\left\{\int_0^T \frac{d\Lambda(t)}{1-F(t)}\right\} dF(s) \\
&= \int_0^T \frac{a(s)}{1-F(s)} dF(s) + \frac{\int_T^\infty a(s)dF(s)}{1-F(T)}
\end{aligned}$$

となる．このふたつの式を合わせると求めたい式の一番目のものを得る．

[14] 真のモデルを $P = P_\eta|_{\eta=0}$ とする．

[15] 真のモデルの確率密度関数を $p = p_\eta|_{\eta=0}$ とする．

く．モデルの可微分性の定義と (7.13) から、

$$\begin{aligned} o(\eta^2) &= \sum_{\delta=0}^{1}\int_0^\infty \left\{\sqrt{p_\eta(t,\eta)}-\sqrt{p(t,\eta)}-\frac{1}{2}\eta h(t,\delta)\sqrt{p(t,\eta)}\right\}^2 dt \\ &\geq \int_0^s \left\{\sqrt{\bar{F}_\eta}-\sqrt{\bar{F}}-\frac{1}{2}h(t,0)\sqrt{\bar{F}}\right\}^2 dG(t) \\ &\quad +\int_0^s \left\{\sqrt{f_\eta}-\sqrt{f}-\frac{1}{2}h(t,1)\sqrt{f}\right\}^2 \bar{G}(t)dt \end{aligned} \tag{7.26}$$

を得る．ただし、F_η、f_η、G を P_η に対応する Y の分布関数と確率密度関数、および C の分布関数とし、$\bar{F}_\eta = 1-F_\eta$ と $\bar{G}=1-G$ とした．さらに、補題 5.3（$Z=0$ として）を利用すれば、

$$\begin{aligned} &\int_0^s \left\{\sqrt{\bar{F}_\eta}-\sqrt{\bar{F}}-\frac{1}{2}\eta h(t,0)\sqrt{\bar{F}}\right\}^2 dG(t) \\ &\quad = \int_0^s \left\{\sqrt{\bar{F}_\eta}-\sqrt{\left(\bar{F}+\frac{1}{2}\eta h\sqrt{\bar{F}}\right)^2}\right\}^2 dG(t) \\ &\quad \geq \left\{\sqrt{\int_0^s \bar{F}_\eta dG}-\sqrt{\int_0^s \left(\sqrt{\bar{F}}+\frac{1}{2}\eta h\sqrt{\bar{F}}\right)dG}\right\}^2 \end{aligned} \tag{7.27}$$

がわかる．これより

$$\int_0^s (\bar{F}_\eta(t)-\bar{F}(t))dG(t) = \eta\int_0^s h(t,0)\bar{F}(t)dG(t)dG(t)+o(\eta) \tag{7.28}$$

がわかる．一方、

$$\begin{aligned} &\int_0^s \left\{\sqrt{f_\eta}-\sqrt{f}-\frac{1}{2}\eta h(t,1)\sqrt{f}\right\}^2 \bar{G}(t)dt \\ &\quad \geq \bar{G}(s)\int_0^s \left\{\sqrt{f_\eta}-\sqrt{\left(\sqrt{f}+\frac{1}{2}\eta h(t,1)\sqrt{f}\right)^2}\right\}^2 dt \\ &\quad \geq \bar{G}(s)\left\{\sqrt{\int_0^s f_\eta dt}-\sqrt{\int_0^s \left(\sqrt{f}+\frac{1}{2}\eta h(t,1)\sqrt{f}\right)^2 dt}\right\}^2 \end{aligned}$$

から

$$\int_0^s f_\eta dt = \int_0^s \left(\sqrt{f}+\frac{1}{2}\eta h(t,1)\sqrt{f}\right)^2 dt \tag{7.29}$$

を得る．(7.29) を (7.28) に代入すれば、s に関して一様に

$$\begin{aligned} &-\int_0^s \left\{\eta\int_0^t h(u,1)f(u)du+\frac{1}{4}\eta^2\int_0^t h^2(u,1)f(u)du+[\bar{G}(s)]^{1/2}0(\eta)\right\}dG(t) \\ &\quad = \eta\int_0^s h(t,0)\bar{F}(t)dG(t)+o(\eta) \end{aligned}$$

が成立することがわかる．したがって、

$$\eta \int_0^s \left\{ h(t,0) + \frac{1}{\bar{F}(t)} \int_0^t h(u,1) dF(u) \right\} \bar{F}(t) dG(t) = o(\eta) \tag{7.30}$$

が s に関して一様にわかる．このこと[16] から、(7.25) が示せた．

つぎに、$F^*(t) \equiv \int_0^t \bar{G} dF / \int_0^\infty \bar{G} dF$ とおき、任意の $a \in L_2^0(F^*)$ に対して、

$$(\boldsymbol{l}_1 a)(T, \Delta) = \Delta a(T) - \frac{1-\Delta}{\bar{F}(T)} \int_0^T a(u) dF$$

とする．$\int a^2 dF^* \geq \int a^2 dF / \int \bar{G} dF$ から $L_2^0(F^*) \supset L_2^0(F)$ がわかる．さらに、$h \in \dot{\boldsymbol{P}}_1$ に対して、$a = h(t,1)$ とおけば、(7.25) から $\dot{\boldsymbol{l}}_1 a = h(t,\delta)$ となるので、$\dot{\boldsymbol{P}}_1 \subset L_2^0(F^*)$ が成立することがわかる．したがって、

$$\overline{\{\dot{\boldsymbol{l}}_1 a : a \in L_2^0(F)\}} \subset \dot{\boldsymbol{P}}_1 \subset \overline{\{\dot{\boldsymbol{l}}_1 a : a \in L_2^0(F^*)\}} \tag{7.31}$$

が成立する．$G^*(t) = \int_0^t \bar{F} dG / \int_0^\infty \bar{F} dG$ とおけば、同様な議論から

$$\overline{\{\dot{\boldsymbol{l}}_2 b : b \in L_2^0(G)\}} \subset \dot{\boldsymbol{P}}_2 \subset \overline{\{\dot{\boldsymbol{l}}_2 b : b \in L_2^0(G^*)\}}$$

が成立することがわかる．F と G を F^* と G^* に置き換えて (7.24) を得るための議論と同様のものを再度行えば、

$$\{\dot{\boldsymbol{l}}_1 a : a \in L_2^0(F^*)\} \perp \{\dot{\boldsymbol{l}}_2 b : b \in L_2^0(G^*)\}$$

が成立するので、$\dot{\boldsymbol{P}}_1 \perp \dot{\boldsymbol{P}}_2$ が示せた．

また、$\int (1/\bar{F}) dG < \infty$ ならば、(7.31) において等号が成立する．これを示すために、どんな $a \in L_2^0(F^*)$ と正数 $\epsilon > 0$ に対しも、$a_M \in L_2^0(F)$ が存在して、

$$\boldsymbol{E}[\{\dot{\boldsymbol{l}}_1 (a - a_M)\}^2] < \epsilon \tag{7.32}$$

を満足することを示せばよい．$l^\infty([0,\infty))$ の中の可測関数の集合は $L_2^0(F^*)$ に稠密であるので、一般性を失わず、a を有界としてよい．ある $0 < M < \infty$ に対して、

$$a_M(t) = a(t) 1\{t \leq M\} - \frac{1}{\bar{F}(M)} \int_0^M a dF 1\{t > M\}$$

おく．明らかに、$a_M \in L_2^0(F)$ である．また、(7.13) から、(7.32) の左辺は $M \uparrow \tau_H$ とすれば、

$$\begin{aligned}
&\sum_{\delta=0}^{1} \int \{\dot{\boldsymbol{l}}_1 (a - a_M)(t,\delta)\}^2 p(t,\delta) dt \\
&= \int_M^\infty \left\{ a + \frac{1}{\bar{F}} \int_0^M a dF \right\}^2 \bar{G} dF + \int_M^\infty \frac{1}{\bar{F}} \left\{ \int_M^t \left[a + \frac{1}{\bar{F}(M)} \int_0^M a dF \right] dF \right\}^2 dG(t) \\
&= o(1) + O\left[\frac{1}{\bar{F}^2(M)} \int_M^\infty \bar{G} dF \right] + O(1) \left[\int_M^\infty \frac{1}{\bar{F}} dG \right] = o(1)
\end{aligned}$$

[16] 任意の s について、(7.30) が成立し、さらにその被積分関数が η に依存しないことに注意すればよい．

となる．最後の等式は、$M \uparrow \tau_H$ のとき、

$$\frac{1}{\bar{F}^2(M)}\int_M^\infty \bar{G}dF \le \frac{\bar{G}(M)}{\bar{F}^2(M)}\int_M^\infty dF \le \frac{\bar{G}(M)}{\bar{F}(M)} = \frac{1}{\bar{F}(M)}\int_M^\infty dG \le \int_M^\infty \frac{dG}{\bar{F}} = o(1)$$

となることよりわかる．したがって、(7.31) の最右辺は最左辺と等しくなるので、

$$\mathrm{Image}(\dot{\boldsymbol{l}}_1) = \dot{\boldsymbol{P}}_1$$

がわかる．

任意の $a \in L_2(F)$ に対して、写像 $L_{uc} : L_2(F) \mapsto L_2(P)$ を

$$(L_{uc}a)(T,\Delta) = \int adM_{uc}$$

で定義する．すると、(7.23) は

$$(\dot{\boldsymbol{l}}_1 a)(T,\Delta) = (L_{uc} \circ R_F)a$$

と書き換えることができる．これより、F の確率密度関数 f に対する情報作用素 $\dot{\boldsymbol{l}}_1^T\dot{\boldsymbol{l}}_1 : L_2(F) \mapsto L_2(F)$ は

$$(\dot{\boldsymbol{l}}_1^T\dot{\boldsymbol{l}}_1 a)(s) = \{R_F^T(L_{uc}^T L_{uc})R_F a\}(s)$$

で与えられる．$\dot{\boldsymbol{l}}_1^T\dot{\boldsymbol{l}}_1$ を求めるためには、$L_{uc}^T L_{uc}$ を計算すれば十分である．そのために、$a, b \in L_2(F)$ を取る．このとき、マルチンゲールの性質と Fubini の定理を用いれば、

$$\begin{aligned}
\langle L_{uc}^T L_{uc}a, b\rangle_{L_2(F)} &= \langle L_{uc}a, L_{uc}b\rangle_{L_2(F)} = \boldsymbol{E}[\int_0^\infty a(s)b(s)1\{T \ge s\}d\Lambda(s)] \\
&= \int_0^\infty \boldsymbol{E}[a(s)b(s)1\{T \ge s\}]d\Lambda(s) \\
&= \int_0^\infty a(s)b(s)(1-G(s))d\Lambda(s) = \langle (1-G)a, b\rangle_{L_2(F)}
\end{aligned}$$

となる．これは、すべての $b \in L_2(F)$ に対して成立するので、$L_{uc}^T L_{uc}$ は対角作用素で、

$$(L_{uc}^T L_{uc}a)(s) = (1-G(s))a(s) \equiv Da(s)$$

と表現できる．したがって、

$$\dot{\boldsymbol{l}}_1^T\dot{\boldsymbol{l}}_1 = R_F^T D R_F$$

となる．

$t \uparrow \tau_G$ のとき、$1-G(s) \downarrow 0$ となるので、$\tau_G = \tau_G = \tau_F \wedge \tau_G$ のとき、Iamge$(\dot{\boldsymbol{l}}_1^T\dot{\boldsymbol{l}}_1)$ は閉にはならない．

$0 \le t \le \tau < \tau_H$ において、

$$\kappa(P)(t) \equiv 1 - F(t)$$

の推定を考える．ここで、写像 $\psi(f) = 1 - F$ を考える．これは、任意の $h \in L_2^0(F)$ に対して、

$$\dot{\psi}_F(h)(t) = -\int_0^\infty \{1_{[0,t]}(z) - F(t)\}h(z)dF(z)$$

で定義される微分 $\dot{\psi}_F$ を持つ[17]．したがって、

$$\dot{\psi}_F(h)(t) = \langle -(1_{[0,t]}(\cdot) - F(t)), h \rangle_{L_2(F)}$$

より、canonical gradient は

$$\tilde{\psi}_{F,t}(z) = -(1_{[0,t]}(z) - F(t))$$

となる．

これを用いて、$\kappa(P)(t) = 1 - F(t)$ が可微分性を調べる．系 5.2.C と定理 5.1 から

$$\tilde{\psi}_{F,t} \in \mathrm{Image}(\dot{\boldsymbol{l}}_1^T \dot{\boldsymbol{l}}_1) = \mathrm{Image}(R_F^{-1} D R_F) \tag{7.33}$$

を示せばよい．しかし、

$$(R_F \tilde{\psi}_{F,t})(z) = -1_{[0,t]}(z) \frac{1 - F(t)}{1 - F(z)}$$

となること[18]に注意して、

$$a_t^*(z) = \left[R_F^{-1} \left(-1_{[0,t]}(\cdot) \frac{1 - F(t)}{1 - F(\cdot)} \right) \right](z) \in L_2(F)$$

とおけば、

$$\begin{aligned}
\dot{\boldsymbol{l}}_1^T \dot{\boldsymbol{l}}_1 a_t^* &= \dot{\boldsymbol{l}}_1^T \dot{\boldsymbol{l}}_1 \left[R_F^{-1} \left(-1_{[0,t]}(\cdot) \frac{1 - F(t)}{1 - F(\cdot)} \right) \right] \\
&= R_F^{-1} D R_F \left[R_F^{-1} \left(-1_{[0,t]}(\cdot) \frac{1 - F(t)}{1 - F(\cdot)} \right) \right] \\
&= R_F^{-1} \left(-1_{[0,t]}(\cdot) \frac{1 - F(t)}{1 - F(\cdot)} \right) = \tilde{\psi}_{F,t}
\end{aligned} \tag{7.34}$$

となるので、(7.33) が成立することがわかる．したがって、$\kappa(P) = 1 - F$ が定義 5.7 の意味において経路ごとに微分可能であることがわかる．

また、定理 5.2.B より、有効影響作用素 $\tilde{\boldsymbol{l}}_\kappa$ は方程式 $\dot{\psi}_F^T = \dot{\boldsymbol{l}}_1 \tilde{\boldsymbol{l}}_\kappa$ の解より、

$$\begin{aligned}
\tilde{\boldsymbol{l}}_\kappa(\pi_t)(T, \Delta) &= (\dot{\boldsymbol{l}}_1 a_t^*)(T, \Delta) = \int R_F a_t^* dM_{uc} \\
&= -(1 - F(t)) \int \frac{1_{[0,t]}(\cdot)}{(1 - F(\cdot))(1 - G(\cdot))} dM_{uc} \\
&= -(1 - F(t)) \left\{ \frac{\Delta 1_{[0,t]}(T)}{1 - F(T)} - C(T \wedge t) \right\}
\end{aligned}$$

---

[17] 例 5.5 を参照のこと．

[18] R_F の定義より、

$$(R_F \tilde{\psi}_{F,t})(z) = -(1_{[0,t]}(z) - F(t)) - \frac{\int_z^\infty \{1_{[0,t]}(u) - F(t)\} dF(u)}{1 - F(z)}$$

となる．さらに、

$$\int_z^\infty 1_{[0,t]}(u) dF(u) = 1_{[0,t]}(z) \int_z^\infty dF(u) = 1_{[0,t]}(z)(F(t) - F(z))$$

になることよりわかる．

となる[19]．ただし、

$$C(t) = \int_0^t \frac{d\Lambda(s)}{(1-F(s))(1-G(s))}$$

である．したがって、$\bar{H} = (1-F)(1-G)$ とおき計算すれば、

$$\begin{aligned}
I_\kappa^{-1}(s,t) &= \boldsymbol{E}[\tilde{\boldsymbol{l}}_\kappa(\pi_t)\tilde{\boldsymbol{l}}_\kappa(\pi_s)] \\
&= \boldsymbol{E}\left[(1-F(t))\int \frac{1_{[0,t]}}{(1-F)(1-G)} dM_{uc}(1-F(s))\int \frac{1_{[0,s]}}{(1-F)(1-G)} dM_{uc}\right] \\
&= (1-F(s))(1-F(t))\boldsymbol{E}\left[\int \frac{1_{[0,t]}(u)1_{[0,s]}(u)}{\bar{H}(u)^2} 1\{T \geq u\} d\Lambda(u)\right] \\
&= (1-F(s))(1-F(t))\int_0^{s\wedge t} \frac{\boldsymbol{E}1\{T \geq u\}}{\bar{H}^2(u)} d\Lambda(u) \\
&= (1-F(s))(1-G(t))\int_0^{s\wedge t} \frac{d\Lambda(u)}{\bar{H}(u)} \\
&= (1-F(s))(1-F(t))C(s\wedge t)
\end{aligned}$$

となる．

7.2.3 ハザードレートに対する NPMLE

．(7.12) で与えられるモデル $\boldsymbol{P}$ を F と G を任意とする（∞ に確率をおくことも許す）モデル $\bar{\boldsymbol{P}}$ に拡張する．すると、

$$dP(t,\delta) = \delta(1-G(t-))dF(t) + (1-\delta)(1-F(t))dG(t) \tag{7.35}$$

と書くことができる．

はじめに、$\bar{\boldsymbol{P}}$ は $[0,\infty]\times\{0,1\}$ 上で定義される確率測度の全体と一致することを示す．そのためには、$[0,\infty]\times\{0,1\}$ 上の任意の確率測度 P が (7.35) の表現を満足することを示す．この P に対し、

$$\bar{H}(t-) = P(T \geq t) = 1 - (H_1(t-) + H_0(t-))$$

[19] $M_{uc}(s) = 1\{s \leq T, \Delta = 1\} - \int_0^s 1\{T \geq u\} d\Lambda(u)$ に注意すれば、

$$\begin{aligned}
&\int \frac{1_{[0,t]}(s)}{(1-F(s))(1-G(s))} dM_{uc}(u) \\
&\quad = \frac{\Delta 1_{[0,t]}(T)}{(1-F(T))(1-G(T))} - \int \frac{1_{[0,t]}(s)}{(1-F(s))(1-G(s))} 1\{T \geq s\} d\Lambda(s)
\end{aligned}$$

となる．さらに、

$$\int \frac{1_{[0,t]}(s)1\{T \geq s\}}{(1-F(s))(1-G(s))} d\Lambda(s) = \int_0^{t\wedge T} \frac{d\Lambda(s)}{(1-F(s))(1-G(s))} = C(t\wedge T)$$

となることよりわかる．

とおく．ただし、

$$H_1(t) = P(T \le t, \Delta = 1), \qquad H_0(t) = P(T \le t, \Delta = 0)$$

である．$[0,\infty]$ 上の分布関数 F が与えられば、ハザード関数は

$$\Lambda(t) = \int_{[0,t]} \frac{dF(t)}{1-F(t-)}$$

で与えられる． また、$1 - H(t-) = P(T \ge t) = (1-F(t-))(1-G(t-))$ と (7.35) より $dH_1(t) = dP(T \le t, \Delta = 1)$ であることに留意して、

$$\Lambda(t) = \int_{[0,\infty]} \frac{1\{\bar{H}(s-) > 0\}}{1-H(s-)} dH_1(s) \tag{7.36}$$

とおく．ここで、$\tau = \sup\{s : \bar{H}(s) > 0\}$ とおけば、

$$1 - F(t) = \begin{cases} \Pi_{0 \le s \le t}(1 - d\Lambda(s)) & 0 \le t < \tau \\ dF(\infty) & t = \tau \end{cases} \tag{7.37}$$

と定める．さらに、

$$1 - G(t-) = \begin{cases} \frac{\bar{H}(t-)}{1-F(t-)} & 0 \le t < \tau \\ dG(t) & t = \tau \end{cases} \tag{7.38}$$

と定義する．(7.35) と (7.37) より、

$$dH_1(t) = \frac{\bar{H}(t-)dF(t)}{1-F(t-)} = (1-G(t-))dF(t)$$

を得る．また、(7.38) より、

$$(1-G(t-))dF(t) + (1-F(t))dG(t) = d(H_1(t) + H_0(t)) \tag{7.39}$$

を得る[20] ．したがって、$\bar{\boldsymbol{P}}$ は $[0,\infty] \times \{0,1\}$ 上の確率測度の全体と一致する．

いま、$(T_1, \Delta_1), \ldots, (T_n, \Delta)$ を P からのランダム標本とすれば、このことから、P の NPMLE $\hat{P}_n$ は

$$\hat{P}_n(t,\delta) = \frac{1}{n}\sum_{i=1}^{n} 1\{t = T_i, \delta = \Delta_i\}$$

[20] (7.38) より、$\bar{H}(t-) = (1-G(t-))(1-F(t-))$ となる．したがって、

$$\begin{aligned} &(1-G(t))(1-F(t)) - \{(1-G(t-))(1-F(t-))\} \\ &\quad = -(1-F(t))(G(t)-G(t-)) - (1-G(t-))(F(t)-F(t-)) \\ &\quad = -(1-F(t))dG(t) - (1-G(t-))dF(t) \end{aligned}$$

となる．一方、

$$\bar{H}(t) - \bar{H}(t-) = P(T > t) - P(T \ge t) = -\{P(T \le t) - P(T < t)\} = -d(H_1(t) + H_0(t))$$

となることよりわかる．

で与えられることがわかる．したがって、(7.36) と (7.37) より、F の NPMLE $\hat{F}_n$ は $t \le T_{(n)}$ に対して、

$$1 - \hat{F}_n(t) = \Pi_{0 \le s \le t}\left(1 - \frac{d\hat{H}_{n1}(s)}{1 - \hat{H}_n(s-)}\right) \tag{7.40}$$

を満足することがわかる．ただし、$T_{(1)} < \ldots < T_{(n)}$ は $T_1, \ldots, T_n$ の順序統計量であり、

$$\begin{aligned} \hat{H}_n(t-) &= \hat{H}_{n1}(t-) + \hat{H}_{n0}(t-) \\ d\hat{H}_{n1}(t) &= d\hat{P}_n(t,1), \quad d\hat{H}_{n0}(t) = d\hat{P}_n(t,0) \end{aligned}$$

である．

もちろん、(7.40) は Kaplan – Meier 推定量である．$\hat{P}_n\{\infty, \cdot\} = 0$ なので、$T_{(n)}$ が打ち切られないならば、$\hat{F}_n$ の台は打ち切られていない $T_i (i = 1, \ldots, n)$ を台にもつ確率測度として唯一定まる．また、$T_{(n)}$ が打ち切られた場合には、$T_{(n)}$ を越えたどこかに確率をもつ確率測度になる．

7.2.4 Nelson – Aalen 推定量と Kaplan – Meier 推定量の漸近線形性

P_0 を真のモデルとして、τ_0 を $P(T > \tau_0) > 0$ となる点とする．$\mu(P) = \Lambda$ を $[0, \tau_0]$ 上で定義される累積ハザード関数とする．$\mathcal{F} = \{1_{[0,s]} : s \le \tau_0\}$ とし、Λ を $\boldsymbol{B} \equiv l^\infty(\mathcal{F})$ の中に値を取る関数とする．すなわち、$\Lambda(1_{[0,s]}) = \Lambda(s)$ とする．このとき、Nelson – Aalen 推定量を $\hat{\Lambda} = q(\hat{P}_n)$ と書く．ただし、

$$q(P) = \int_0^{\cdot} \frac{1 - H(t-)}{dH_1(u)}$$

で、$H(t-) = P(T \ge t)$、$H_1(t) = P(T \le t, \Delta = 1)$ で、$\hat{P}_n(t, \delta) = (1/n)\sum_{i=1}^n 1\{t = T_i, \delta = \Delta_i\}$ である[21]．

定理 7.2 ：推定量 $\hat{\Lambda}$ は影響関数

$$\tilde{\nu}_{P,t}(s, \delta) = \delta \frac{1_{[0,t]}(s)}{1 - H(s)} - C(s \wedge t) \tag{7.41}$$

をもつ[22] Gauss 正則な推定量である．

証明：事象 $\{1 - \hat{H}_n(y) > 0\}$ 上で、

$$\sqrt{n}(\hat{\Lambda}(y) - \Lambda(y)) = \sqrt{n}\left\{\int_0^y \frac{\hat{H}_{n1}(t)}{1 - \hat{H}_n(t-)} - \int_0^y d\Lambda_0\right\} \tag{7.42}$$

となる．マルチンゲール中心極限定理を用いれば、Gauss 過程に弱収束することがわかるが、ここでは別証明を与える．

[21] 観測を $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ とする．

[22] ただし、$C(t) = \int_0^t \frac{dH_1}{(1-H(s))^2} = \int_0^t \frac{d\Lambda(s)}{1-H(s)}$ である．

まず、(7.42) の右辺を書き換えると[23]

$$\sqrt{n}\left\{\int_0^y \frac{d\hat{H}_{n1}(t)}{1-\hat{H}_n(t-)} - \int_0^y \frac{dH_{01}}{1-H_0(t-)}\right\} = \sqrt{n}\int_0^y \frac{d(\hat{H}_{n1}-H_{01})(t)}{1-H_0(t-)}$$
$$+\int_0^y \frac{\sqrt{n}(\hat{H}_n-H_0)(t-)}{(1-H_0(t-))^2}dH_{01}(t) + \sqrt{n}\int_0^y \frac{(\hat{H}_n(t-)-H_0(t-))^2}{(1-H_n(t-))(1-H_0(t-))^2}dH_{01}$$
$$+\sqrt{n}\int_0^y \frac{H_n(t-)-H_0(t-)}{(1-\hat{H}_n(t-))(1-H_0(t-))}d(\hat{H}_{n1}-H_{01})(t) \tag{7.43}$$

を得る．Donsker 定理より、$\sqrt{n}(\hat{H}_{n1}-H_{01})$ と $\sqrt{n}(\hat{H}_n-H_0)$ は $l^\infty(\mathcal{F})$ の中に値をとる Gauss 過程に弱収束するので、(7.43) の右辺の１項目と２項目はそれぞれの経験過程の連続関数となるので、ある Gauss 過程に弱収束することがわかる．また、$1-H_n(t-)=1-H_0-(1/\sqrt{n})(H_n(t-)-H_0(t-))$ から、

$$\int_0^y \frac{dH_{01}}{(1-H_n(t-))(1-H_0(t-))} = O_P(1)$$

がわかる．したがって、

$$\left|\sqrt{n}\int_0^y \frac{(\hat{H}_n(t-)-H_0(t-))^2}{(1-\hat{H}_n(t-))(1-H_0(t-))}dH_{01}\right| \le \sqrt{n}\|\hat{H}_n-H_0\|_\infty O_P(1) = O_P(n^{-1/2})$$

となることがわかる．つぎに、(7.43) の右辺の最後の項を書き直せば、

$$\sqrt{n}\int_0^y \frac{\hat{H}_n(t-)-H_0(t-)}{(1-\hat{H}_n(t-))(1-H_0(t-))}d(\hat{H}_{n1}(t)-H_{01}(t))$$
$$=\int_0^{\hat{H}_{n1}(y)} W_n(\hat{H}_{n1}^{-1}(s))ds - \int_0^{H_{01}(y)} W_n(H_{01}^{-1}(s))ds$$
$$=\int_0^{\hat{H}_{n1}^{-1}(y)} \{W_n(\hat{H}_{n1}^{-1}(s)-W_n(H_{01}^{-1}(s))\}ds - \int_{\hat{H}_{n1}(y)}^{H_{01}(y)} W_n(H_{01}^{-1}(s))ds$$

となる. ただし、

$$W_n(t) = \sqrt{n}\frac{\hat{H}_n(t-)-H_0(t-)}{(1-H_n(t-))(1-H_0(t-))}$$

---

[23] これは以下の議論からわかる．

$$\frac{d\hat{H}_{n1}}{1-\hat{H}_n} - \frac{d\hat{H}_{01}}{1-H_0} = \frac{d\hat{H}_{n1}}{1-\hat{H}_n} - \frac{d\hat{H}_{n1}}{1-H_0} - \frac{d(\hat{H}_{n1}-H_0)}{1-H_0}$$
$$= \frac{(\hat{H}_n-H_0)d\hat{H}_{n1}}{(1-\hat{H}_n)(1-H_0)} + \frac{d(\hat{H}_{n1}-H_{01})}{1-H_0} + \frac{(\hat{H}_n-H_0)dH_{01}}{(1-\hat{H}_n)(1-H_0)} - \frac{(\hat{H}_n-H_0)dH_{01}}{(1-H_0)^2}$$
$$+\frac{(\hat{H}_n-H_0)dH_{01}}{(1-H_0)^2}$$
$$= \frac{(\hat{H}_n-H_0)d\hat{H}_{n1}}{(1-\hat{H}_n)(1-H_0)} + \frac{d(\hat{H}_{n1}-H_{01})}{1-H_0} + \frac{(\hat{H}_n-H_0)d(\hat{H}_{n1}-H_{01})}{(1-\hat{H}_n)(1-H_0)^2} + \frac{(\hat{H}_n-H_0)dH_{01}}{(1-H_0)^2}$$

とした．W_n はほとんど確実に連続な経路をもつ確率過程に収束するので、

$$\sup\{|H_{n1}^{-1}(s) - H_{01}^{-1}(s)| : 0 \le s \le \tau_0\} = O_P(n^{-1/2})$$

であることより、

$$\left|\int_0^{\hat{H}_{n1}^{-1}(y)} \{W_n(\hat{H}_{n1}^{-1}(s) - W_n(H_{01}^{-1}(s))\} ds\right|$$
$$\le \sup\{|W_n(\hat{H}_{n1}^{-1}(s)) - W_n(H_{01}^{-1}(s)| : |s| \le H_{n1}(\tau_0)\} O_P(1) = o_P(1)$$

と

$$\left|\int_{\hat{H}_{n1}(y)}^{H_{01}(y)} W_n(H_{01}^{-1}(s)) ds\right| \le \sup|W_n(t)| \sup|\hat{H}_{n1}(y) - H_{01}(y)| = O_P(1) o_P(1) = o_P(1)$$

となることより、

$$\left|\sqrt{n}\int_0^y \frac{\hat{H}_n(t-) - H_0(t-)}{(1-\hat{H}_n(t-))(1-H_0(t-))} d(\hat{H}_{n1}(t) - H_{01}(t))\right| = o_P(1)$$

がわかる．

あとは、推定量 Λ_n が影響関数 $\nu_{P,t}$ をもつことを示せばよい． □

定理 7.3 ：推定量 $\hat{F}_n$ は影響関数

$$\tilde{\kappa}_{P,t}(s,\delta) = -(1-F(t))\tilde{\nu}_{P,t}(s,\delta)$$

をもつ Gauss 正則な推定量である．ただし、$\tilde{\nu}_{P,t}$ は (7.41) で与えられるものである．

証明： □

7.3 Cox モデル

生存データ解析で重要な Cox モデル

$$\lambda(t|z) = e^{\theta^T z}\lambda(t)$$

を考える．ただし、λ は基準ハザード関数（baseline hazard function）で、基準確率密度関数 g と分布関数 G によって、$\lambda = g/(1-G)$ と表現できる．また、$Z \in \boldsymbol{R}^k$ は共変量で、分布 H [24] に従い、$\theta \in \boldsymbol{R}^k$ は未知の回帰係数である．

[24] その確率密度関数を h と記すことにする．

7.3.1 情報量計算

以下、簡単のために、$k=1$ として、議論を進める．(Z,T) を観測とすれば、その同時確率密度関数は

$$f(z,t)=rg(t)(1-G(t))^{r-1}h(z) \tag{7.44}$$

で与えられる[25]．ただし、$r=r(\theta z)=e^{\theta z}$ である．

簡単な計算より、θ に対するスコア関数は

$$\dot{\boldsymbol{l}}(z,t)=z[1-\Lambda_r(t)]$$

となる[26]．ただし、

$$\Lambda_r(t)\equiv r(\theta z)\Lambda(t)=e^{\theta z}\int_0^t\frac{dG}{1-G}$$

である．つぎに、g に対するスコア作用素を求める．そのために、g を通る[27] $\{g_\eta:\eta\in\boldsymbol{R}\}$ を $\boldsymbol{G}\equiv\{G:G$ は絶対連続な分布関数$\}$ の中のなめらかな曲線を取る．$a=(\partial/\partial\eta)g_\eta|_{\eta=0}$ とおけば、

$$(\dot{\boldsymbol{l}}_2a)(z,t)=a(t)+(r-1)\frac{\int_t^\infty adG}{1-G(t)} \tag{7.45}$$

で与えられる[28]．

いま、2組の L – 作用素と R – 作用素を考える．すなわち、基準生存関数 $1-G$ と基準ハザード関数 λ に対応する L – 作用素と R – 作用素で、これらを L と R と記す．もう一方は、$Z=z$ という条件が与えられたときに、生存関数 $(1-G)^r$ とハザード関数 $\lambda_r=r\lambda$ に対応する L – 作用素と R – 作用素で、これらを L_r と R_r と記すことにする．したがって、任意の $a\in\dot{\boldsymbol{G}}$ [29] に対し、

$$(L_ra)(t) \;=\; a(t)-\int_0^t ad\Lambda_r=a(t)-r\int_0^t ad\Lambda$$

---

[25] 実際、$1-G=\exp(-\int_0^\cdot\lambda(u)dU)$ に注意すれば、

$$\begin{aligned}f(z,t) &= h(z)\frac{\partial}{\partial t}\left[1-\exp\left(-\int_0^t\lambda(u|z)du\right)\right]=h(z)\lambda(t|z)\exp\left(-\int_0^t\lambda(u|z)du\right)\\ &= h(z)e^{\theta z}\frac{g}{1-G}\left[\exp\left(-\int_0^t\lambda(u|z)du\right)\right]^{\theta z}=h(z)\theta^{\theta z}\frac{g}{1-G}(1-G)^r\end{aligned}$$

となることよりわかる．

[26] なぜならば、$\frac{\partial}{\partial\theta}\log f(z,t)=\frac{\partial}{\partial\theta}\left[\log e^{\theta z}+(e^{\theta z}-1)\log(1-G)\right]=z+ze^{\theta z}\log(1-G)$ と $\log(1-G)=-\int_0^t\frac{dG}{1-G}$ よりわかる．

[27] $g=g_\eta|_{\eta=0}$ とした．

[28] g_η に対応して、$f_\eta(z,t)$ と G_η などと書けば、

$$\frac{\partial}{\partial\eta}\log f_\eta(z,t)\bigg|_{\eta=0}=\frac{\partial}{\partial\eta}\left[\log g_\eta(t)+(e^{\theta z}-1)\log(1-G_\eta)\right]\bigg|_{\eta=0}=a-(e^{\theta z}-1)(1-G)^{-1}g_\eta\big|_{\eta=0}$$

と $\int_t^\infty adG=\int_t^\infty\frac{\partial}{\partial\eta}\log g_\eta\Big|_{\eta=0}dG=-g(t)$ よりわかる.

[29] $\dot{\boldsymbol{G}}$ はモデル $\boldsymbol{G}=\{G:G$ は絶対連続な分布関数$\}$ の接空間である．

$$(R_r a)(t) \;=\; a(t) - \frac{\int_t^\infty a dF(\cdot|z)}{1-F(t|z)} = -\boldsymbol{E}[a(T)-a(t)|Z=z, T>t]$$

である．ただし、$F(\cdot|z)$ は $f(\cdot|z)$ の分布関数である．このふたつの作用素を用いれば、θ と g のスコア作用素は

$$\begin{aligned} \dot{\boldsymbol{l}}_1(z,t) &= z(L_r 1)(t) \\ (\dot{\boldsymbol{l}}_2 a)(z,t) &= (L_r R a)(t) \end{aligned}$$

と表現できること[30] がわかる．

いま、

$$M_r(t) = 1\{T \le t\} - \int_0^t 1\{T \ge s\} d\Lambda_r(s)$$

とおけば、命題 7.2 から、

$$\dot{\boldsymbol{l}}_1(Z,T) \;=\; Z\int_0^\infty dM_r(s) \tag{7.46}$$

$$\dot{\boldsymbol{l}}_2(Z,T) \;=\; \int_0^\infty (Ra)(s) dM_r(s) \tag{7.47}$$

と表現できることがわかる．

θ に対する有効スコア関数 $\boldsymbol{l}_1^*$ を求めるために、つぎの関係を満足する $a^* \in L_2^0(G)$ を見つける必要がある：どんな $a \in L_2^0(G)$ に対しても、

$$\boldsymbol{l}_1^* \equiv \dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^* \perp \dot{\boldsymbol{l}}_2 a \qquad （L_2(F) \text{ の中において}）$$

---

[30] (7.3) を使えば、

$$\int_t^\infty Ra d\Lambda = \int_t^\infty \left\{ a(s) - \frac{\int_s^\infty a(u)du}{1-G(s)} \right\} d\Lambda(s) = \int_t^\infty \frac{adG}{1-G} - \int_t^\infty \frac{\int_s^\infty a(u)du}{(1-G(s))^2} dG(s)$$

と

$$\begin{aligned} \int_t^\infty \frac{\int_s^\infty a(u)du}{(1-G(s))^2} dG(s) &= \int_t^\infty a(u) \int_t^u \frac{dG(s)}{(1-G(s))^2} du = \int_t^\infty \left[\frac{1}{1-G(s)}\right]_t^u du \\ &= \int_t^\infty \frac{a(u)}{1-G(u)} du - \frac{1}{1-G(t)} \int_t^\infty a(u) du \end{aligned}$$

から

$$\int_t^\infty Ra d\Lambda = \frac{1}{1-G(t)} \int_t^\infty a(s) ds$$

になることに注意する．再度、(7.3) を使えば、

$$\begin{aligned} a(t) + (r-1)\frac{\int_t^\infty adG}{1-G(t)} &= (Ra)(t) + r\frac{\int_t^\infty adG}{1-G(t)} = (Ra)(t) - r\int_t^\infty Ra d\Lambda \\ &= (Ra)(t) - \int_0^t Ra d\Lambda_r = (L_r Ra)(t) \end{aligned}$$

となることがわかる．

を満足する．すなわち、

$$\boldsymbol{E}[\{\dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^*\}\dot{\boldsymbol{l}}_2 a] = 0 \tag{7.48}$$

である．

Z を条件付きにすれば、(7.48) の左辺は (7.46) と (7.47) によって、与えられるマルチンゲール変換の 2 次変分の期待値として計算できる．したがって、(7.48) の左辺は以下のように変形できる．

$$\begin{aligned}
\boldsymbol{E}[\boldsymbol{E}[\{\dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^*\}\dot{\boldsymbol{l}}_2 a]|Z] &= \boldsymbol{E}[\boldsymbol{E}[\int_0^\infty (Z-(Ra^*))dM_r \int_0^\infty (Ra)dM_r]|Z] \\
&= \boldsymbol{E}[\boldsymbol{E}[\int (Z-(Ra^*))(Ra)1\{T \geq s\}e^{\theta z}d\Lambda]|Z] \\
&= \int \left\{\boldsymbol{E}[Ze^{\theta Z}1\{T \geq s\}] - \boldsymbol{E}[e^{\theta Z}1\{T \geq s\}](Ra^*)(s)\right\}(Ra)(s)d\Lambda(s) \\
&= \int \{S_1(s) - S_0(s)(Ra^*)\}(Ra)(s)d\Lambda(s)
\end{aligned} \tag{7.49}$$

となる．ただし、

$$S_i(s) \equiv \boldsymbol{E}[Z^i e^{\theta Z}1\{T \geq s\}] \qquad i = 0, 1$$

である．(7.49) より、

$$a^* \equiv L\left(\frac{S_1}{S_0}\right) \qquad \text{または} \qquad Ra^* = \frac{S_1}{S_0}$$

とすればよいことがわかる．したがって、θ の有効スコア関数は

$$\begin{aligned}
\boldsymbol{l}_1^*(Z,T) &= \dot{\boldsymbol{l}}_1(Z,T) - (\dot{\boldsymbol{l}}_2 a^*)(Z,T) = \int_0^\infty \{Z-(Ra^*)(t)\}dM_r(t) \\
&= \int_0^\infty \{Z - \frac{S_1}{S_0}\}dM_r(t) = \int_0^\infty \{Z - \boldsymbol{E}[Z|T=t]\}dM_r(t)
\end{aligned}$$

となることがわかる．上の式の最後の等式は

$$\frac{S_1}{S_0}(t) = \boldsymbol{E}[Z|T=t]$$

より[31] わかる．したがって、θ に対する情報量はマルチンゲール計算と (7.44) から

$$I(\theta) \;=\; \boldsymbol{E}[(\boldsymbol{l}_1^*(Z,T))^2] = \boldsymbol{E}[\boldsymbol{E}[\int_0^\infty \{Z-\boldsymbol{E}(Z|T=t)\}^2 1\{T \geq t\}e^{\theta z}d\Lambda(t)]|Z]$$

[31] $T = t$ が与えられたときの Z の条件付き確率密度関数は

$$\frac{f(z,t)}{\int f(z,t)dz} = \frac{e^{\theta z}(1-G(t))^{r-1}h(z)}{\int e^{\theta z}(1-G(t))^{r-1}h(z)dz}$$

である．また、

$$\boldsymbol{E}[1\{T \geq t\}] = \exp\left(-\int_0^t e^{\theta z}\lambda(s)ds\right) = \left\{-\int_0^t \lambda(s)ds\right\}^{e^{\theta z}} = (1-G(t))^r$$

から、$i = 0, 1$ に対して、

$$\boldsymbol{E}[Z^i e^{\theta Z}1\{T \geq t\}] \;=\; \int_0 \infty \int_t^\infty z^i e^{\theta z} f(z,s)dzds = \int_0^\infty \int_t^\infty z^i r^2 (1-G(s))^{r-1}h(z)dG(s)dz$$

$$
\begin{aligned}
&= \boldsymbol{E}\left[\int_0^\infty \{Z - \boldsymbol{E}(Z|T=t)\}^2 e^{\theta z}(1-G(t))^{r-1} dG(t)\right] \\
&= \int_0^\infty \int_0^\infty \{Z - \boldsymbol{E}(Z|T=t)\}^2 e^{\theta z} h(z) dG(t) dz \\
&= \boldsymbol{E}\{Z - \boldsymbol{E}(Z|T=t)\}^2
\end{aligned}
$$

となることがわかる．

7.4 打ち切り型線形回帰モデル

$X^0 \equiv (Y, Z, C) \in \boldsymbol{R}^3$ は線形回帰モデル

$$Y = \theta Z + \epsilon$$

を想定する. ただし、ϵ は (Z, C) と独立とし同時確率密度関数 $h(z, c)$ を持ち、ϵ の分布関数 F は確率密度関数 f を持つとする. したがって、モデル $\boldsymbol{Q}$ は

$$\boldsymbol{Q} = \{Q : \frac{dQ}{d\mu}(y, z, c) = f(y - \theta z)h(z, c),\ f \text{ と } h \text{ は確率密度関数}\}$$

で与えられる.

$V \equiv Y \wedge C, \Delta \equiv 1\{Y \leq C\}$ とおき、観測できる確率変数を $X = (Z, V, \Delta)$ とかく. 写像 $T : X^0 \mapsto X$ によって誘導されるモデルを $\boldsymbol{P} \equiv \boldsymbol{Q}T^{-1}$ と記すことにする.

7.4.1 情報量計算

まず、θ と f に対するモデル $\boldsymbol{Q}$ におけるスコア関数を求める. θ のスコア関数は

$$\dot{\boldsymbol{l}}_1(y, z; \boldsymbol{Q}) = -z\frac{\dot{f}}{f}(y - \theta z) \equiv z\psi(y - \theta z)$$

であることが簡単にわかる. f に対するスコア関数を求めるために、１次元正規部分モデル[32] $\{f_\eta : \eta \in \boldsymbol{R}\}$ を考え、$a = (\partial/\partial\eta)\log f_\eta|_{\eta=0}$ とする. このとき、

$$\dot{\boldsymbol{l}}_2(y, z; \boldsymbol{Q}) = a(y - \theta z)$$

$$
\begin{aligned}
&= \int_0^\infty r[1 - G(s)]_t^\infty h(z) dz = \int_0^\infty r(1 - G(s))^r h(z) dz \\
&= \int_0^\infty r\boldsymbol{E}[1\{T \geq t\} h(z) dz
\end{aligned}
$$

となる．これから、に注意すれば

$$\boldsymbol{E}[Z|T=t] = \frac{\int z e^{\theta z}(1-G(t))^{r-1} h(z) dz}{\int e^{\theta z}(1-G(t))^{r-1} h(z) dz} = \frac{\boldsymbol{E}[Ze^{\theta Z} 1\{T \geq t\}]}{\boldsymbol{E}[Ze^{\theta Z} 1\{T \geq t\}]}$$

となることがわかる．

[32] $f_0 = f$ とする.

と書ける．つぎに、誘導されたモデル $\boldsymbol{P} = \boldsymbol{Q}T^{-1}$ におけるスコア関数を計算する．誘導されたモデルにおける確率密度関数は

$$p(z, v, \delta) = \{f(v-\theta z)\bar{H}(v|z)\}^{\delta}\{h(v|z)\bar{F}(v-\theta z)\}^{1-\delta}h(z) \tag{7.50}$$

で与えられること[33] がわかる．ただし、$h(z)$ は Z の周辺確率密度関数とし、$h(v|z)$ は $Z = z$ が与えられたときの V の条件付き確率密度関数とし、$H(v|z)$ はその分布関数とする．また、$\bar{F} = 1 - F$ と $\bar{H} = 1 - H$ とした．

確率密度関数 (7.50) から直接スコア関数を導出する．

$$\begin{aligned}
\dot{\boldsymbol{l}}_1(x; \boldsymbol{P}) &= \frac{\partial}{\partial\theta}\log p(z, v; \delta)\Big|_{\theta=0} \\
&= -\delta z\frac{\dot{f}(v-\theta z)}{f(v-\theta z)} + (1-\delta)z\frac{f(v-\theta z)}{1-F(v-\theta z)}
\end{aligned}$$

となる．f に対するスコア関数を求めるために、1 次元正規母数モデル $\{f_\eta : \eta \in \boldsymbol{R}\}$ を考え、$a = (\partial/\partial\eta)\log f_\eta$ とおけば、

$$\dot{\boldsymbol{l}}_2(x; \boldsymbol{P}) = \delta a(v-\theta z) + (1-\delta)\frac{\int_{v-\theta z}^{\infty}}{1-F(v-\theta z)}$$

となることがわかる．

また、命題 5.1 から T が与えられたときの $\dot{\boldsymbol{l}}_1(y, z; \boldsymbol{Q})$ と $\dot{\boldsymbol{l}}_2(y, z; \boldsymbol{Q})$ の条件付き期待値を計算することにより、モデル $\boldsymbol{P}$ のもとでのスコア関数を求めることもできる．そのために、

$$\epsilon = Y - \theta Z \qquad \text{と} \qquad W = C - \theta Z$$

---

[33] $\Delta = 0$ のとき、$V = Y$ から

$$\begin{aligned}
&P(Z \le z, V \le v, \Delta = 0) = \int_{-\infty}^{z}\int_{-\infty}^{v}\int_{v}^{\infty} f(u-\theta w)h(w, r)dudrdw \\
&= \int_{-\infty}^{z} h(w)\left\{\int_{-\infty}^{v}(1-F(v-\theta w))h(w|r)dr\right\}dw = \int_{-\infty}^{z}\int_{-\infty}^{v} h(w)\bar{F}(v-\theta w)h(w|r)drdw
\end{aligned}$$

から

$$\frac{\partial^2}{\partial v\partial z}P(Z \le z, V \le v, \Delta = 0) = h(z)\bar{F}(v-\theta w)h(w|r)$$

がわかる．一方、$\Delta = 1$ のとき、$V = C$ から

$$\begin{aligned}
&P(Z \le z, V \le v, \Delta = 10 = \int_{-\infty}^{z}\int_{-\infty}^{v}\int_{v}^{\infty} f(u-\theta w)h(w, r)drdudw \\
&= \int_{-\infty}^{z} h(w)\int_{-\infty}^{v}\left\{f(u-\theta w)\int_{v}^{\infty} h(r|w)dr\right\}dwdz = \int_{-\infty}^{z}\int_{-\infty}^{v} h(w)f(u-\theta w)(1-H(v|w))dwdz
\end{aligned}$$

から

$$\frac{\partial^2}{\partial v\partial z}P(Z \le z, V \le v, \Delta = 1) = h(z)f(v-\theta w)\bar{H}(w|r)$$

となることよりわかる．

とおけば、

$$Y \wedge C - \theta Z = \epsilon \wedge W \qquad と \qquad 1\{Y \le C\} = 1\{\epsilon \le W\}$$

となる．また、$\tilde{T} \equiv (Z, \epsilon \wedge W, 1\{\epsilon \wedge W\})$ を与えることと T を与えることは同値にある．$W = w$ として、

$$\mathcal{F}_w \equiv \sigma\{1\{\epsilon \le s\}, s \le w\} = \sigma\{\epsilon \wedge w, 1\{\epsilon \le w\}\}$$

とおき、命題 5.1 に注意すれば、

$$\dot{\boldsymbol{l}}_i(T; \boldsymbol{P}) = \boldsymbol{E}[\dot{\boldsymbol{l}}_i(T; \boldsymbol{Q})|T] = \boldsymbol{E}[\dot{\boldsymbol{l}}_i(T; \boldsymbol{P})|Z, \mathcal{F}_w], (i = 1, 2)$$

となる．よって、(7.7) と命題 7.2 から

$$\begin{aligned}
\boldsymbol{E}[\dot{\boldsymbol{l}}_i(T; \boldsymbol{P})|Z, \mathcal{F}_w] &= \boldsymbol{E}[Z\psi(w)|Z, \mathcal{F}_w] \\
&= Z[1\{\epsilon \le w\}\psi(w) + 1\{\epsilon > w\}\frac{\psi(\epsilon)1\{\epsilon > w\}}{1 - F(w)}] \\
&= Z[1\{\epsilon \le w\}\psi(w) + 1\{\epsilon > w\}\frac{f}{1 - F}(w)] = Z\int_{-\infty}^{2} R\psi dM
\end{aligned}$$

となる[34]．

さらに、

$$M_{uc}(u) = 1\{\epsilon \wedge w \le u, \Delta = 1\} - \int_{-\infty}^{u} 1\{\epsilon \wedge w \ge s\}\Lambda(s)$$

とおけば、

$$\int_{-\infty}^{w} R\psi dM = \int_{-\infty}^{\infty} 1\{w \ge s\}R\psi(s)dM(s) = \int_{-\infty}^{\infty} R\psi(s)dM_{uc}(s) \tag{7.51}$$

となる[35]．また、

$$\begin{aligned}
\dot{\boldsymbol{l}}_2 a(X|\boldsymbol{P}) &= \boldsymbol{E}[\dot{\boldsymbol{l}}_2 a(X|\boldsymbol{Q})|Z, \mathcal{F}_w] \\
&= \boldsymbol{E}[a(\epsilon|Z, \mathcal{F}_w] \\
&= 1\{\epsilon \le w\}a(\epsilon) + \frac{\boldsymbol{E}[a(\epsilon)1\{\epsilon > w\}]}{1 - F(w)}
\end{aligned}$$

---

[34]

$$\begin{aligned}
R\psi(s) &= \psi(s) - \boldsymbol{E}[\psi(\epsilon)|\epsilon > s] = \psi(s) - \lambda(s) = -\frac{\dot{f}}{f}(s) - \frac{f}{1 - F}(s) \\
&= -\frac{d}{ds}\log \lambda(s) = -\frac{\dot{\lambda}}{\lambda}(s)
\end{aligned}$$

である．

[35] これは、$M(s) = 1\{\epsilon \le s\} - \int_{-\infty}^{s} 1\{\epsilon \ge u\}d\Lambda(u)$ から

$$\begin{aligned}
\int_{-\infty}^{\infty} R\psi(s)dM_{uc}(s) &= \Delta R\psi(\epsilon \wedge w) - \int_{-\infty}^{i} nfty1\{\epsilon \wedge w \ge s\}R(s)d\Lambda(s) \\
&= \int_{-\infty}^{i} nfty1\{w \ge s\}R\psi(s)dM(s)
\end{aligned}$$

からわかる．上の式の最後の等号は $1\{w \ge s\}R\psi(\epsilon) = \Delta R\psi(\epsilon \wedge w)$ に注意すればよい．

$$
\begin{aligned}
&= \mathbf{1}\{\epsilon \le w\}a(\epsilon) + \frac{1}{1-F(w)}\int w^{\infty} a(s)dF(s) \\
&= \int_{-\infty}^{\infty} RadM_{uc} \qquad (7.52)
\end{aligned}
$$

を得る[36]．

つぎに、θ に対する有効スコア関数を求める．そのために、任意の $a \in L_2^0(F)$ に対して、$a^* \in L_2^0(F)$ で

$$
\boldsymbol{l}_1^* \equiv \dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^* \perp \dot{\boldsymbol{l}}_2 a^2
$$

を $L_2(P)$ において満足するもの[37] を見つければよい．これは

$$
\boldsymbol{E}[(\dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^*)\dot{\boldsymbol{l}}_2 a] = 0 \qquad (7.53)
$$

と書ける．Z と W を与えれば、(7.53) は predictable covariation 過程の期待値になることに

---

[36] 上の式の最後の等号は、

$$
M_{uc}(u) = \mathbf{1}\{\epsilon \wedge w \le u, \Delta = 1\} - \int_{-\infty}^{u} \mathbf{1}\{\epsilon \wedge w \ge s\} d\Lambda(s)
$$

に注意すれば、

$$
\begin{aligned}
\int_{-\infty}^{\infty} RadM_{uc} &= \int_{-\infty}^{\infty} \left\{ a(u) - \frac{\int_u^{\infty} a(v)dF(v)}{1-F(u)} \right\} dM_{uc}(u) \\
&= \Delta \left\{ a(\epsilon \wedge w) - \frac{\int_{\epsilon \wedge w}^{\infty} a(v)dF(v)}{1-F(u)} \right\} \\
&\quad - \int_{-\infty}^{\epsilon \wedge w} a(u)d\Lambda(u) + \int_{-\infty}^{\epsilon \wedge w} \frac{\int_u^{\infty} a(v)dF(v)}{1-F(u)} d\Lambda(u)
\end{aligned}
$$

を得る．ここで、

$$
\{-\infty < u \le \epsilon \wedge w, u \le v < \infty\} = \{-\infty < v \le \epsilon \wedge w, -\infty < U \le v\} \cup \{\epsilon \wedge w < v < \infty, -\infty < u \le \epsilon \wedge w\}
$$

に注意して、Fubini の定理を用いれば、

$$
\begin{aligned}
\int_{-\infty}^{\epsilon \wedge w} \frac{\int_u^{\infty} a(v)dF(v)}{1-F(u)} d\Lambda(u) &= \int_{-\infty}^{\epsilon \wedge w} \int_{-\infty}^{v} \frac{a(v)}{(1-F(u))^2} dF(u)dF(v) + \int_{\epsilon \wedge w}^{\infty} \int_{-\infty}^{\epsilon \wedge w} \frac{a(v)}{(1-F(u))^2} dF(u)dF(v) \\
&= \int_{-\infty}^{\epsilon \wedge w} a(v)d\Lambda(v) + \frac{1}{1-F(\epsilon \wedge w)} \int_{\epsilon \wedge w}^{\infty} a(v)dF(v)
\end{aligned}
$$

となるので

$$
\int_{-\infty}^{\infty} RadM_{uc} = \Delta a(\epsilon \wedge w) + \frac{1-\Delta}{1-F(\epsilon \wedge w)} \int_{\epsilon \wedge w}^{\infty} a(v)dF(v)
$$

を得る．

[37] すなわち、

$$
\langle \dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^*, \dot{\boldsymbol{l}}_2 a \rangle_{L_2(P)} = \sum_{\delta=0,1} \int \{\dot{\boldsymbol{l}}_1(x,\delta) - \dot{\boldsymbol{l}}_2 a^*(x,\delta\}\dot{\boldsymbol{l}}_2(x,\delta)p(x,\delta)dx = 0
$$

を満たすことである．ただし、$x = (z,\delta)$ で $p(z,\delta)$ は (7.50) で与えられるものである．

注意すれば、

$$
\begin{aligned}
\boldsymbol{E}\boldsymbol{E}[(\dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^*)\dot{\boldsymbol{l}}_2 a|Z,W] &= \boldsymbol{E}\boldsymbol{E}[\{Z\int_{-\infty}^{\infty} R\psi dM_{uc} - \int_{-\infty}^{\infty} Ra^* dM_{uc}\}\{\int_{-\infty}^{\infty} Ra^* dM_{uc}\}|Z,W] \\
&= \boldsymbol{E}\boldsymbol{E}[\int_{-\infty}^{\infty} 1\{\epsilon \wedge W \geq s\}[ZR\psi(s) - Ra^*(s)]Ra(s)d\Lambda(s)|Z,W] \\
&= \boldsymbol{E}[\int_{-\infty}^{\infty} 1\{W \geq s\}[ZR\psi(s) - Ra^*(s)]1\{\epsilon \geq s\}Ra(s)d\Lambda(s)] \\
&= \boldsymbol{E}[\int_{-\infty}^{\infty} 1\{W \geq s\}[ZR\psi(s) - Ra^*(s)]Ra(s)dF(s)] \\
&= \int_{-\infty}^{\infty} \{\boldsymbol{E}[Z1\{W \geq s\}R\psi(s)] - \boldsymbol{E}[1\{W \geq s\}Ra^*(s)\}Ra(s)dF(s) \\
&= \langle \boldsymbol{E}[Z1\{W \geq s\}R\psi(s)] - \boldsymbol{E}[1\{W \geq s\}Ra^*(s), Ra(s)dF(s)\rangle_{L_2(F)} \\
&= \langle L\{\boldsymbol{E}[Z1\{W \geq s\}R\psi(s)] - \boldsymbol{E}[1\{W \geq s\}Ra^*(s)\}, a(s)\rangle_{L_2(F)} \\
&= \int_{-\infty}^{\infty} L\{\boldsymbol{E}[Z1\{W \geq s\}R\psi(s)] - \bar{G}(s)Ra^*(s)\}a(s)dF(s)
\end{aligned}
$$

を得る．ただし、最後から 2 番目の等号は $R^T = L$ よりわかり、

$$\bar{G}(s) = 1 - G(s) = P\{W \geq s\} = \boldsymbol{E}1\{W \geq s\}\boldsymbol{E}\{W \geq s\}$$

と記した．ここで、

$$K(s) \equiv \boldsymbol{E}[Z|W \geq s]$$

とし、

$$a^* \equiv L(KR\psi) = L(K(s)R\psi(s)) \tag{7.54}$$

とおけば、$R \circ L = \mathrm{id}$ から

$$Ra^*(s) = K(s)R\psi(s)$$

と

$$\bar{G}(s)Ra^*(s) = \boldsymbol{E}[Z1\{W \geq s\}]R\psi(s)$$

となることより、(7.53) で与えられる a^* は任意の $a \in L_2^0(F)$ に対して、(7.53) を満足することがわかる．(7.52) – (7.54) から

$$
\begin{aligned}
\dot{\boldsymbol{l}}_1^* &= \dot{\boldsymbol{l}}_1 - \dot{\boldsymbol{l}}_2 a^* \\
&= \int_{-\infty}^{\infty} \{ZR\psi - RLKR\psi\}dM_{uc}dM_{uc} \\
&= \int_{-\infty}^{\infty} \{Z - \boldsymbol{E}[Z|W \geq s]\}R\psi(s)dM_{uc} \\
&= \int_{-\infty}^{\infty} \{Z - \boldsymbol{E}[Z|C - \theta Z \geq s]\}R\psi(s)dM_{uc}
\end{aligned}
$$

を得る．

参考文献

関連図書

[BKRW] Bickel, P.J., Klaassen, C.A., Ritov, Y. and Wellner, J. *Efficient and adaptive estimation for semiparametric models*, Springer, 1998.

[G] Groeneboom, P. *Lectures on inverse problems* in " Lectures on Probability Theory and Statistics", Lecture notes in Mathematics 1648, 67–164, 1996.

[GW] Groeneboom, P. and Wellner, J. *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser, 1992.

[horowitz] Horowitz, J.L. *Semiparametric methods in Econometrics* Lecture notes in Statistics 131, Springer, 1998.

[huang] Huang, J. and Wellner, J. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1, *Statistica Nddrlandica* **49**, 153–163, 1995. Ibragimov, I.A. and Has'minskii, R.Z. *Statistical estimation*, Applications in Mathematics 16, Springer, 1981.

[stein] Stein, C. Efficient nonparametric testing and estimation, " *Proc. Third Berkeley Symp. Math. Statist. Prob.*" **1**, 187–195, Univ. California Press, Berkley, 1956.

[vdL] van der Laan, M.J. Efficient and inefficient estimation in semiparametric models, CWI Tracts 114, Center for Mathematics and Computer Science, 1995.

[VdVa] Van der Vaart, A. *Statistical estimation in large parameter spaces*, CWI Tract 44, Center for Mathematics and Computer Science, 1988.

[VdV] Van der Vaart, A. *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilisitc Mathematics, 1998.

[VdVW] van der Vaart, A. and Wellner, J. *Weak convergence and empirical preocess*, Springer, 1996.

Adaptive Estimation、the Hájek のたたみこみ定理、局所漸近ミニマックス定理にかんする代表的な論文

1. Adaptive Estimation.

a. Stein, C (1956): Efficient nonparametric estimation and testing. In *Proc. Third. Berkeley Symp. Math. Statist. Prob.* **1**, 185–195.

b. Hájek, J. and Zidak, Z. (1967): *Theory of Rank Tests.* Academic Press, New York.

c. van Eeden, C. (1970): Efficiency-robust estimation of location. *Ann. Math. Statist.* **41**, 172–181.

d. Beran, R. (1974): Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2**, 63–74.

e. Sacks, J. (1975): An asymptotically efficient sequence of estimators of location parameter. *Ann. Statist.* **3**, 285–298.

f. Stone, C.J. (1975): Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 297–284.

g. Bickel, P.J. (1982): On adaptive estimation. *Ann. Statist.* **10**, 647–671.

2. The convolution theorem.

a. Hájek, J. (1970): A characterization of limiting distribution of regular estimates. Z. Wahrsheinlichkeitstheorie **14**, 323–330.

b. Inagai, N. (1970): On the limiting distribution of a sequence of regular estimates. *Ann. Inst. Statist. Math.* **22**, 1–13.

c. Roussas, G. (1972): *Contiguity of Probability Measures; some application to statistics* Cambridge Univ. Press.

d. Hájek, J. (1972): Local asymptotic minimax and admissibility in estimation. In *Proc. of Sixth Berkeley Symp.* **1**, 175–194.

e. Le Cam, L. (1972): Limits of experiments. In *Proc. of Sixth Berkeley Symp.* **1**, 245–261.

f. Beran, R. (1977): Robust location estimates. *Ann. Statist.* **5**, 431–444.

g. Beran, R. (1977): Estimating a distribution function. *Ann. Statist.* **5**, 445–463.

h. Ibragimov, I.A. and Hasminskii, R.Z. (1981): *Statistical Estimation: Asymptotic Theory*, Springer-Verlag.

i. Fabian, V. and Hannan, J. (1982): On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrschenlichkeitstheorie v. Geb.* **56**, 459–478.

j. Pfanzagl, J. (1982): *Contributions to a General Asymptotic Statistical Theory* (with the assistanc of Wefelmeyer, W.). Lecture Notes in Statistics **13**, Springer-Verlag.

k. Millar. P.W. (1983): The minimax principle in asymptotic statistical theory. Ecole d'Eé Probabilités de St Flour XI – 1981. Lecture Notes in Math. **876**, 76–262. Springer-Verlag.

l. Begun, J. Hall. W.J., Huang, W.M., and Wellner, J. (1983): Informaiton and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432–452.

m. Millar, P.W. (1985): Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahrscheinlichkeitstheorei.* **68**, 545–556.

3. Asymptotic Minimax Theorems.

a. Le Cam, L. (1953): On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. Univ. California Publ. Statist. **1**, 277–330.

b. Chernoff, H. (1956): Large sample theory: Parmetric case. *Ann. Math. Statist.* **27**, 1–22.

c. Dvoretzky, A., Kiefer,J. and Wolfowitz, J. (1956): Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642–669.

d. Kiefer, J. and Wolfowitz, J. (1959): Asymptotic minimax character of the sample distribution for vecter chance variables. *Ann. Math. Statist.* **30**, 463–489.

e. Kiefer, J. and Wolfowitz, J. (1976): Asymptotic minimax estimation of concave and convex distributions. *Z Wahrsheinlichkeitstheorie* **34**, 73–85.

f. Le Cam, L. (1979): On a theorem of J. Hájek. In *Contribution to Staitistics; jaroslav Hájek Memorial Volumun.* Jureckova, J. (ed.), 119–135. Reidel.

h. Millar, P.W. (1979): Asumptotic minimax theorems for the sample distribution function. *Z. Wahrscheinlichkeitstheorie* **48**, 233–252.

i. Millar. P.W. (1983): The minimax principle in asymptotic statistical theory. Ecole d'Eé Probabilités de St Flour XI – 1981. Lecture Notes in Math. **876**, 76–262. Springer-Verlag.

セミパラメトリック推測理論に関係する最近の論文

1. Profile Likelihoood.

a. Severini, T.A. and Wong, W.H. (1992): Profile likelihood and conditionally parametric models. *Ann. Statist.* **20**, 1768–1862

b. Qin, J. and Wong, A. (1996): Empirical likelihood in a semiparametric model. *Scand. J. Statist.* **23**, 209–220.

c. Murphy, S. and Van der Vaart, A.W. (1997): On profile likelihood. Preprint, Penssylvania State University.

d. Huang, J. and Wellner, J.A. (1996): Regression models with interverl censoring. In *Probability theory and mathematical statistics*, 269–296. Ibragimov, I.A. and Zaitsev, Q.Y. eds, Gordon and Breach, Amsterdam.

2. Testing in and between semiparametric models.

a. Murphy, S.A. and Van der Vaart, A.W. (1997): Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–1509.

b. Choi, S., Hall, W.J., and Schick, A. (1996): Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, **24**, 841–861.

c. Cox, D. Kohn, E. Wahba, G., and Yandell, B.S. (1988)* Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16**, 113–119.

d. Rotnitzky, A. and Jewell, N.P. (1990): Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.

3. Missing Data models.

a. Robbins, J.M., Rotintzky, A., and Zhao, L.P. (1994): Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 866–866.

b. Rotinitzky, A. and Bobins, J. (1995): Semi-parametric estimation of models for means and covariance in the presence of missing data. *Scand. J. Statist.* **22**, 323–333.

c. Robins, J.M., Hsieh, F., and Newwey, W. (1995): Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *J. Royal Statist. Soc.* B **57**, 409–424.

4. Regression models for survival data: alternative to the Cox model.

a. Lin, D.Y. and Ying, Z. (1994): Semiparametric analyssi of the additive risk model. *Biometrika* **81**, 61–71.

b. Lin, D.Y. and Ying, Z. (1996): Semiparametric analysis of general addtive-multiplicative intensity models for counting processes. *Ann. Statist.* **23**, 1712–1734.

c. Sasieni, P. (1992): Non-orthogonal projections and their application to calculation the information in a partly linear Cox model. *Scand. J. Statist.* **19**, 215–234.

d. Huang, J. (1997): Efficient estimation in the partly linear additive Cox model. Technical Report, University of Iowa.

5. Attainment and nonattainment of information bounds.

a. Van der Vaart, A.W. (1995): Efficiency of infinite dimensional M-estimators. *Statistica Neerl.* **49**, 9–30.

b. Ritov, Y. and Bickel, P.J. (1990): Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18**, 925–938.

c. Ibragimov, I.A. and Khas:mininskii, R.Z. (1991): Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* **18**, 1681–1724.

d. Birgé, L. and Maasart, P. (1995): Estimation of integral functionals of a density. *Ann. Statist.* **23**, 11–29.

e. Laurent, B. (1996): Efficient estimation of integral functionals of a desnsity. *Ann. Statist.* **24** 669–681.

f. Efromovich, S. and Low, M.G.@(1996): On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Ann. Statist.* **24**, 682–686.

g. Laurent, B. (1997): Estimation of integral functionals of a density and its derivatives. *Bernoulli* **3**, 181–211.

6. Lower bounds for "inverse" problems and nonstandard asymptotics.

a. Groeneboom, P. and Jongbloed, G. (1995): Isotonic estimation and rates of convergence in Wicksll's problem. *Ann. Statist.* **23**, 1518–1542.

b. Donoho, D.L. and Low, M.G. (1990): Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20**, 944–970.

c. Huber,C. (1997): Lower bounds for function estimation. In *Festchrift for Lucien Le Cam*, 245–258. Pollard, D., Torgersen, E., and Yang, G.L. (eds.), Springer-Verlag, New York.

d. Yu, B. (1997): Assouad, Fano, and Le Cam. In *Festchrift for Lucien Le Cam*, 424–435. Pollard, D., Torgersen, E., and Yang, G.L. (eds.), Springer-Verlag, New York.

7. Estimation for binary choice and interval censoring models.

a. Cosslett, S.R. (1983): Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **55**, 559–585.

b. Huang, J. and Wellner, J.A. (1997): Interval censored suvrival data; a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Anslysis*, Lecture notes in Statitics **123**, 123–169.

c. Lin, D.Y., Oakes, D., and Ying, Z. (1997): Additive harards regression with current status data. *Biometrika* **85**, 289–299.

8. Information bounds for multiphase and case-cohort designs.

a. Breslow, N.E. and Holubkov, R. (1997): Mzximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *J. Royal Statist. Soc.* B **59**, 447–461.

b. Scott, A.J. and Wild, C.J. (1991): Fitting logistic models in stratified case-control studies. *Biometrics* **47**, 497–510.

c. Scott, A.J. and Wild, C.J. (1997): Fitting regression models to case-control data by maximum likelihood. *Biometrika* **84**, 57–84.

d. Prentice, R.L. (1986): A case-cohort design for epidemmiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

e. Self, S.G. and Prentice, R.L. (1988): Asymptotic distribution theory and efficient results for case-cohort studies. *Ann. Statist.* **16**, 64–81.

f. Goldstein, L. and Langholz, B (1992): Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann. Statist.* **20**, 1903–1928.

9. Information bounds for models with dependence.

a. Greenwood, P.E. and Wefelmeyer, W. (1995): Efficiency of empirical estimation for Markov chains. *Ann. Statist.* **23**, 132–143.

b. Ibragimov, I.A. and Khas'miniskii (1975): Local asymptotic normality of non-identically distributed observations. *Theor. Prob. Applic.* **20**, 246–260.

c. Jeganathan, P. (1981): On a decomposition of the limit distribution of a sequence estimation. *Sankhyā* **43**, 26–36.

d. Jeganathan, P. (1982): On the asymptotic theory of estiamtion when the limit of the log-likelihood ratios is mixed normal. *Sankhyā* **44**, 173–212.

e. Jeganathan, P. (1995): Some aspects of asymptotic theory with application to time series models. *Econometric Theory* **11**, 818–887.

f. Schick, A. (1988): On estimation in LAMN families when there are nuisance parameters present. *Sankhyā* **50**, 249–268.

10. Semiparametric generalized linear models.

a. Chen, H. (1995): Asymptotically efficient estimaiton in semiparametric generalized linear models. *Ann. Statist.* **23**, 1102–1129.

b. Mammen, E. and Van der Geer, S. (1997): Penalized quasi-likelihood estimation i partial linear models. *Ann. Statist.* **25**, 1014–1035.

c. Emond, M.J. and Self, S. (1993): An efficient estimation for the generalized semilinear model. *J. Amer. Statist. Assoc.* **92**, 1033–1040.

11. General M- and Z- theorems.

a. Huber, P. (1967): The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**, 221–233, Univeristy of Califorinia Press, Berkeley.

b. Pollard, D. (1985): New ways to prove central limit theorems. *Econometric Theory* **1**, 295–314.

c. Pakes, A. and Pollard, D. (1989); Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

d. Gill, R.D. (1989): Non- and semi-parametric maximum likelihood estimators and the von-Mises method - I, *Scand. J. Statist.* **16**, 97–128.

e. Gill, R.D. and Van der Vaart, A.W. (1993): Non- and semi-parametric maximum likelihood estimators and the von-Mises method - II, *Scand. J. Statist.* **20**, 271–288.

f. Wong, W.H. and Severini, T.A. (1991): On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19**, 603–632

g. Van der Vaart, A.W. (1995); Efficiency of infinite dimensional M-estimators. *Staistica Neerl.* **49**, 9–30.

h. Huang, J. (1996): Efficient estimation for the proposional harards model with interval censoring. *Ann. Statist.* **24**, 540–568.

i. Newey, W. (1994): The asymptotic variance of semiparametric estimatoirs. *Econometrica* **62**, 1349–1382.

12. Empirical process methods in estimation of general parameters.

a. Birgé, L. and Massart, P. (1993): Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields* **97**, 113–150.

b. Birgé, L. and Massart, P. (1997): From model seleciton to adaptive estimation. In *Festschrift for Lucien Le Cam*, 55–87, Pollard, D., Torgersen, E., and Yang, G.L. (eds.), Springer-Verlag, New York.

c. Shen, X. and Wong, W.H. (1994): Covnergence rate of sieve estiamtes. *Ann. Statist.* **22**, 580–615.

d. Wong, W.H. and Shen, X. (1995): Probability inequalities for likelihood ratios and convergence rates of sieve MLE's. *Ann. Statist.* **23**, 339–363.

e. Van der Geer, S. (1993): Helinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14–44.

f. Van der Geer, S. (1996): Rates of convergence for maximum likelihood estimators in mixture models. *Nonparametric Statistics* **6**, 293–310.

13. Mixture models.

a. Van der Vaart, A.W. (1995): Efficient maximum likelihood estimation in semi-parametric mixture models. *Ann. Statist.* **24**, 862–878.

b. Roeder, K., Caroll, R.J., and Lindsay, B.G. (1996): A semiparametric mixture approach to case-control studies with errors in covariables. *J. Amer. Statist. Assoc.* **91**, 722–732.

14. Information geometry approach.

a. Amari, S. and Kumon, M. (1988): Estimation in the presence of infinitely many nuisance parameters-geometry of estimating functions. *Ann. Statist.* **16**, 1044–1068.

b. Amari, S. and Kawanabe, M. (1997): Information geometry of estimating functions in semi-parametric statistical models. *Bernoulli* **3** 29–54.

15. Copula models.

a. Klaassen, C.A.J. and Wellner, J.A. (1997): Efficient estimation in the bibariate normal copula model: normal margins are least favourable. *Bernoulli* **3** 55-77.

モダンな経験過程に関する論文

1. Dudley, R. M. (1978): Central limit theorems for empirical measures. *Ann. of Prob.* **6**, 899 929. Correction: *Ann. Prob.* **7**, 909 – 911.

2. Dudley, R. M. (1984): A course on empirical processes. In *École d'Étd́e Probabilités de Saint-Flour XII-1982* Lecture Notes in Mathematics 1097, 2 – 141. Springer-Verlag, New York.

3. Gine E. and Zinn, J. (1984): Some limit theorems for empirical processes. *Ann. of Prob.* **12**, 929 – 989.

4. Dudley, R. M. (1987): Universal Donsker classes and metric entropy. *Ann. Prob.* **15**, 1306–1326.

5. Pollard, D. (1984): *Convergence of Stochastic Processes.* Springer-Verlag, New York.

6. Pollard, D. (1990): *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2, Society for Industrial and Applied Mathematics, Philadelphia. Institute of Mathematical Statistics and American Statistical Association, Hayward.

7. Gine E., and Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. Lecture Notes in Mathematics **1221**, 50–113. Springer-Verlag, New York.

8. Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics* Wiley, New York.

9. Van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

経験過程理論の統計学への応用

1. Basic M- and Z- Estimator Theorems.

a. Huber, P. J. (1967): The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**, 221–233. Univ. California Press.

b. Pollard, D. (1985): New ways to prove central limit theorems. *Econometric Theory* **1**, 295–314.

c. Pales, A. and Pollard, D. (1989): Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

d. Kim, J. and Pollard, D. (1990): Cube root asymptotics. *Ann. Statist.* **18**, 191–219.

d. Nolan, D. (1991): The excess-mass ellipsoid. *J. Mult. Analysis* **39**, 348–371.

e. Huang, J. (1993): Central limit theorems for M-estimates. Technical Report 251, Department of Statistics, University of Washington.

2. Infinite-Dimensional M- and Z- Estimator Theorems: .

a. Gill, R. D. (1989): Non- and semi-parametric maximum likelihood estimators and the von-Mises method - I. *Scand. J. Statist.*.**16**, 97–128.

b. Gill, R. D. and Van der Vaart, A. W. (1993): Non- and semi-parametric maximum likelihood estimators and the von-Mises method - II. *Scand. J. Statist.* **20**, 271 - 288.

c. Van der Vaart, A. W. (1995): Efficiency of infinite dimensional M-estimators. *Statistica Neerl.* **49**, 9–30.

3. Bootstrapping M- and Z- Estimators.

a. Arcones, M. A. and Gine E. (1992): On the bootstrap of M - estimators and other statistical functionals. In *Exploring the Limits of Bootstrap*, LePage R. and L. Billard, (eds.). 14–47. Wiley, New York. R.

b. Wellner, J. A. and Zhan, Y. (1996): Bootstrapping infinite-dimensional Z- estimators. Submitted to Bernoulli.

c. Zhan, Y. (1996). Bootstrapping Functional M-estimators. Unpublished Ph.D. dissertation, University of Washington.

4. Rates of Convergence in Nonparametric Estimation via Empirical Processes.

a. Birgé L. and Massart, P. (1993): Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields* **97**, 113–150.

b. Shen, X., and Wong, W. H. (1994): Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580–615.

c. Van de Geer, S. (1993): Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14–44.

d. Wong, W. H., and Shen, X. (1995): Probability inequalities for likelihood ratios and convergence rates of sieve MLE's. *Ann. Statist.* **23**, 339–363.

5. Recent Literature Using Empirical Process Methods in Semiparametric and NonparametricModels.

a. Emond, M. J. and Self, S. (1993): An efficient estimator for the generalized semilinear model. *J. Amer. Statist. Assoc.* **92**, 1033–1040.

b. Gilbert, P. (1996): Semiparametric biased sampling models. Submitted to Ann. Statist.

c. Geskus, R. and Groeneboom, P. (1997): Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. Technical Report, Delft University. Submitted to Ann. Statist.

d. Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, 540–568.

e. Mammen, E. and Van de Geer, S. (1997): Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25**, 1014–1035.

f. Murphy, S. A. (1995): Asymptotic theory for the frailty model. *Ann. Statist. 23*, 182–198.

g. Murphy, S. A. and Van der Vaart, A.W. (1995): Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–1509.

h. Parner, E. (1998): Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

i. Van der Vaart, A. W. (1994): Maximum likelihood estimation with partially censored data. *Ann. Statist.* **22**, 1896–1916.

j. Van der Vaart, A. W. (1996): Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24**, 862–878.

k. Wellner, J. A. and Zhang, Y. (1998). Large sample theory for an estimator of the mean of a counting process with panel count data. Technical Report No. 327, Department of Statistics, University of Washington. Submitted to Ann. Statist.