統計的推測理論の漸近理論

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

参考文献 217

1

確率論からの準備

·····1.1·····

可測空間

Ω を空でない集合とする．

定義 1.1　：Ω の部分集合族 $\mathcal{F}$ がつぎを満たすとき，σ – **加法族**という．

(i)　$\Omega \in \mathcal{F}$
(ii)　$A \in \mathcal{F}$ ならば，$A^c \in \mathcal{F}$
(iii)　$A_i \in \mathcal{F}, i = 1, 2, \ldots$ ならば，$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

定義 1.2　：Ω の部分集合族 $\mathcal{M}$ が増加列と減少列に対して閉じている[1)]とき，**単調族**という．

注意 1.1　：(i) $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ ならば，$\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$ である．
(ii) $\emptyset \in \mathcal{F}$ である．

命題 1.1　：(i) $\mathcal{F}_s, s \in S(\neq \emptyset)$ を Ω 上の σ-加法族とする．このとき，$\bigcap_{s \in S} \mathcal{F}_s$ も σ-加法族である．
(ii) $\mathcal{C}$ を Ω の部分集合族とする．$\mathcal{C}$ を含む最小の σ-加法族が唯一存在する．これを $\mathcal{C}$ のよって生成される σ-加法族とよび，$\sigma(\mathcal{C})$ と書く．
(iii) $\mathcal{F}$ が σ-加法族となる必要十分条件は，$\mathcal{F}$ が単調族となることである．

証明：(i) と (ii) は西尾 28 ページを参照．
(iii) 西尾 31 ページを参照.　□

定義 1.3　：一般にの位相空間 S に対し，S の開集合全体を含む最小の σ-加法族を**ボレロ集合族**[2)]といい，$\mathcal{B}(S)$ と書く．

例 1.1　：定義より，$\mathbb{R}$ のボレロ集合族 $\mathcal{B}(\mathbb{R})$ は $\mathbb{R}$ の開集合全体 $\mathcal{C}$ により生成される．すなわち，$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ である．しかし，$\mathcal{B}(\mathbb{R})$ を生成する集合族は他にもいくつかある．たとえば，$\mathcal{E} = \{(-\infty, r], r \in \mathbb{R}\}$ は $\mathcal{B}(\mathbb{R})$ の生成集合族である．これを示すためには，$\sigma(\mathcal{E}) \subset \sigma(\mathcal{C})$ と $\sigma(\mathcal{E}) \supset \sigma(\mathcal{C})$ を示せばよい．

[1)] すなわち，$A_i \in \mathcal{M}$ に対し

(i)　$A_i \subset A_{i+1}, i = 1, 2, \ldots,$ ならば，$\displaystyle\bigcup_{i=1}^{n} A_i \in \mathcal{M}$

(ii)　$A_i \supset A_{i+1}, i = 1, 2, \ldots,$ ならば，$\displaystyle\bigcup_{i=1}^{n} A_i \in \mathcal{M}.$

[2)] あるいは，位相的 σ-加法族ともいう．

まず, $\sigma(\mathcal{E}) \subset \sigma(\mathcal{C})$ を示そう. そのために, $\mathcal{E} \subset \sigma(\mathcal{C})$ を示せば十分[3)]である. これは, $(-\infty, r] = \cap_{n=1}^{\infty}(-\infty, r+n^{-1})$ と表現できることからわかる.

つぎに, $\sigma(\mathcal{E}) \supset \sigma(\mathcal{C})$ を示すために $\mathcal{C} \subset \sigma(\mathcal{E})$ を示そう. これは, $\mathbb{R}$ 上の任意の開集合は開区間の可算個の和として表現できることと任意の開区間は $(a, b) = (-\infty, b) \cap (-\infty, a]^c$ と $(-\infty, b) = \cup_{n=1}^{\infty}(-\infty, b-n^{-1}]$ から構成できることからわかる. よって, $\mathcal{C} \subset \sigma(\mathcal{E})$ は示された.

定義 1.4　: μ が $(\Omega, \mathcal{F})$ 上の測度とは
(i) $\mu : \mathcal{F} \to [0, \infty]$ かつ $\mu(\emptyset) = 0$,
(ii) $A_n \in \mathcal{F}, n \in \boldsymbol{N}$ が互いに素ならば,

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$

のときをいう. $(\Omega, \mathcal{F}, \mu)$ を**測度空間**という.

定義 1.5　: (i) $\mu(\Omega) = 1$ ならば, μ を**確率測度**という. 確率測度をとくに P と記すことにする.
(ii) 測度 μ が σ-**有限**であるとは, $\cup_{n=1}^{\infty} A_n = \Omega$ かつ すべての $n \geq 1$ に対して $\mu(A_n) < \infty$ なる $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ が存在することである.

命題 1.2　: $(\Omega, \mathcal{F}, \mu)$ を測度空間とする.
(i) 増大列 $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ に対して

$$\mu(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$$

が成立する.
(ii) 減少列 $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ に対して

$$\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$$

が成立する.

証明: (i) $A_0 = \emptyset$ と記す.

$$\begin{aligned}
\mu(\bigcup_{n=1}^{\infty} A_n) &= \mu\{\cup_{n=1}^{\infty}(A_n \setminus A_{n-1})\} \\
&= \sum_{n=1}^{\infty} \mu(A_n \setminus A_{n-1}) \qquad (\sigma\text{-加法性}) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i \setminus A_{i-1} \\
&= \lim_{n\to\infty} \mu\{\bigcup_{i=1}^{n}(A_i \setminus A_{ni-1})\} \qquad (\text{有限加法性}) \\
&= \lim_{n\to\infty} \mu(A_n)
\end{aligned}$$

3) なぜならば, $\sigma(\mathcal{E})$ と $\sigma(\mathcal{C})$ はともに $\mathcal{E}$ を含む σ-集合族となり, $\sigma(\mathcal{E})$ の最小性より, $\sigma(\mathcal{E}) \subset \sigma(\mathcal{C})$ がわかる.

(ii) $B_n = A_1 \setminus A_n = A_1 \cap A_n^c$ とおけば, B_n は増大列となる.

$$\begin{aligned}\lim_{n\to\infty}\mu(B_n) &= \mu(\bigcup_{n=1}^{\infty} B_n)\\ &= \mu\{\bigcup_{n=1}^{\infty}(A_1 \cap A_n^c)\}\\ &= \mu\{A_1 \bigcap (\bigcup_{n=1}^{\infty} A_n^c)\}\\ &= \mu\{A_1 \bigcap \{\bigcap_{n=1}^{\infty} A_n\}^c\}\\ &= \mu(A_1) - \mu\{\bigcap_{n=1}^{\infty} A_n\} \qquad \text{（有限加法性）}\end{aligned}$$

一方, 有限加法性から $\mu(A_1) = \mu(A_1 \setminus A_n) + \mu(A_n)$ なので,

$$\lim_{n\to\infty}\mu(B_n) = \mu(A_1) - \lim_{n\to\infty}\mu(A_n).$$

このふたつの式より (ii) は証明された. □

定義 1.6 ：(i)

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k\\ &= \{\omega \in \Omega : \omega\text{は有限個の } A_k \text{ を除いたすべての } A_k \text{ に含まれる}\\ &= \{A_n, a.a\}\end{aligned}$$

(ii)

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k\\ &= \{\omega \in \Omega : \omega\text{は可算個の } A_k \text{ に含まれる}\\ &= \{A_n, i.o\}\end{aligned}$$

(iii) $\limsup A_n = \liminf A_n = A$ のとき[4], $\lim_{n\to\infty} A_n = A$ と記す.

定義 1.7 ：$A \in 2^{\Omega}$ に対し, 指示関数を

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{その他} \end{cases}$$

で定義する.

[4] 一般には, $\liminf A_n \subset \limsup A_n$ である. また, $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ ならば, $\liminf A_n \in \mathcal{F}$ かつ $\limsup A_n \in \mathcal{F}$ である.

定義 1.8　：$(\Omega, \mathcal{F}, \mu)$ を測度空間とする. Ω の部分集合 A に対し, $A \subset E$ かつ $\mu(E) = 0$ を満たす $E \in \mathcal{F}$ が存在するとき, A を μ-**零集合**という. 測度空間 $(\Omega, \mathcal{F}, \mu)$ に対し $\mathcal{F}$ がすべての μ-零集合を含むとき, $(\Omega, \mathcal{F}, \mu)$ は**完備測度空間**[5)]という.

ここで., $\Omega = \mathbb{R}$ とする. $F : \mathbb{R} \to \mathbb{R}$ を右連続かつ単調非減少とする. この F に対しつぎの条件を満たす $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の測度 μ_F を F に対応する **Lebeague-Stieltjes** 測度[6)]という.

$$\mu_F((a, b]) = F(b) - F(a), \qquad (-\infty < a < b < \infty) \tag{1.1}$$

[5)] 測度空間はつねに完備測度空間に一意的に拡張できる. これについては西尾 *43* ページを参照.

[6)] 任意の右連続かつ単調非減少関数 F に対し, (1.1) を満たす測度 μ が一意的に存在することが保障される. 証明については西尾 45 ページを参照.

·····1.2·····

可測関数と積分

定義 1.9 ：$(\Omega, \mathcal{F})$ を測度空間とする．Ω 上の実数値関数 X が任意の $B \in \mathcal{B}(\mathbb{R})$ に対し

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

のとき，$\mathcal{F}$-可測という．

命題 1.3 ：X が $\mathcal{F}$-可測関数であるための必要十分条件は $\mathbb{R}$ の部分集合族 $\mathcal{C}$ で $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ なるものに対し

$$X^{-1}(\mathcal{C}) = \{X^{-1}(C) : C \in \mathcal{C}\} \subset \mathcal{F}$$

が成立することである．

証明：（$\Leftarrow$） は明らか．
（$\Rightarrow$）逆像 X^{-1} は集合の演算を保存するので

$$X^{-1}(\mathcal{B}(\mathbb{R})) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C}))$$

となる．さらに，X が可測性なので，その定義から $X^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{F}$ となる．よって，題意は示せた． □

注意 1.2 ：上の命題において，たとえば$\mathcal{C} = \{(-\infty, r] : r$ は有理数$\}$ とすればよいことがわかる．

定義 1.10 ：確率空間 $(\Omega, \mathcal{F}, P)$ 上で定義された可測関数を**確率変数**[1)] という．

定義 1.11 ：確率変数 X の**分布関数**を

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$$

で定義する．ただし，$x \in \boldsymbol{R}$ である．

<u>分布関数の性質</u>：

---

1) したがって，$X : \Omega \to \boldsymbol{R}$ で，どんな実数 r に対しても

$$X^{-1}(-\infty, r] = \{\omega \in \Omega, X(\omega) \leq r\} \in \mathcal{F}$$

を満足するものとする．X を**確率変数**と定義することもある．

(i)　どんな $x \in \boldsymbol{R}$ に対しても，$0 \le F_X(x) \le 1$
(ii)　$F_X(\,\cdot\,)$ は非減少関数
(iii)　$\lim_{x\to-\infty} F_X(x) = 0, \lim_{x\to\infty} F_X(x) = 1$
(iv)　F_X は右連続[2)]で左極限[3)]が存在する
(v)　$F_X(\,\cdot\,)$ の不連続点 x における jump は

$$P(X = x) = F_X(x) - F_X(x-)$$

で表現される .

記法：$F_X^{-1}(y) = \inf\{x : F_X(x) \ge y\}$

命題 1.4　：X, Y を可測関数とする. このとき，$X \pm Y$，XY，X/Y，$X^+ \equiv X1\{X \ge 0\}$，$X^- \equiv -X1\{X \le 0\}$，$|X|$，$g(X)$ も可測関数である. ただし，g は可測である.

命題 1.5　：$\{X_n\}_{n=1}^{\infty}$ は可測関数列とする. このとき，

$$\text{(i)}\ \sup_n X_n, \qquad \text{(ii)}\ \inf_n X_n \qquad \text{(iii)}\ \limsup_{n\to\infty} X_n \qquad \text{(iv)}\ \liminf_{n\to\infty} X_n \qquad \text{(v)}\ \lim_{n\to\infty} X_n$$

も可測である.

証明：(i) $\{\sup_n X_n \le x\} = \cup_{n=1}^{\infty}\{X_n \le x\}$
(ii) $\inf X_n = -\sup(-X_n)$
(iii) $\limsup_{n\to\infty} X_n = \inf_n(\sup_{k\ge n} X_k)$
(iv) $\liminf_{n\to\infty} X_n = -\limsup_{n\to\infty}(-X_n)$
(v) $\lim_{n\to\infty} X_n$ が存在するならば，$\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$ □

ここで，X を非負可測関数とする. このとき

$$\begin{aligned} X_n &\equiv \frac{1}{2^n}\sum_{i=1}^{4^n} 1\{X \ge \frac{1}{2^n}\} \\ &= 2^n 1\{X \ge 2^n\} + \sum_{i=1}^{4^n} \frac{i-1}{2^n} 1\{\frac{i-1}{2^n} \le X < \frac{i}{2^n}\} \end{aligned} \tag{1.2}$$

とする. これより，X_n は各点で X に収束すること[4)]がわかる. また，X_n は単調増加である[5)]. この X_n を**単関数**とよぶ.

---

2) すなわち，$F_X(x+) = \lim_{y\to x+0} F_X(y)$
3) これを $F_X(x-) = \lim_{y\to x-0} F_X(y)$ とかく .
4) すなわち

もし $X(\omega) = \infty$ ならば，すべての n に対し $X_n(\omega) = 2^n$
もし $X(\omega) < \infty$ ならば，十分お大きな n に対し $|X_n(\omega) - X(\omega)| \le 2^{-n}$

5) なぜならば,

$$\begin{aligned} X_n &= \frac{1}{2^{n+1}}\sum_{i=1}^{4^n} 2 1\{X \ge \frac{2i}{2^{n+1}}\} \\ &\le \frac{1}{2^{n+1}}\sum_{i=1}^{4^n}\left(1\{X \ge \frac{2i}{2^{n+1}}\} + 1\{X \ge \frac{2i-1}{2^{n+1}}\}\right) \\ &\le \frac{1}{2^{n+1}}\sum_{i=1}^{4^{n+1}} 1\{X \ge \frac{i}{2^{n+1}}\} \end{aligned}$$

命題 1.6　：非負実数値関数 X が可測であるための必要十分条件は X が (1.2) で定義された関数列 $\{X_n\}$ の極限であることである．

証明： □

定義 1.12　：(i) $X = \sum_{i=1}^n x_i 1_{A_i}$ $(x_i \geq 0, \cup_{i=1}^n A_i = \Omega)$ に対し，

$$\int X \, d\mu = \sum_{i=1}^n x_i \mu(A_i)$$

(ii) $X \geq 0$ に対し

$$\int X \, d\mu = \lim_{n\to\infty} \int X_n \, d\mu$$

ただし，X_n は非負単関数の任意の増加列で $X_n \to X,\ \mu - \text{a.e.}$ なるものである．
(iii) 可測関数 X に対し，$\int X^* \, d\mu$ か $\int X^- \, d\mu$ のどちらか一方が有限ならば,

$$\int X \, d\mu = \int X^* \, d\mu - \int X^- \, d\mu$$

$\int X \, d\mu$ は存在する．
(iv) $\int X \, d\mu$ が有限ならば，X は μ-可積分．

定義 1.13　：X を確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数とし，X が P-可積分のとき

$$\mathbb{E}(X) = \int_\Omega X(\omega) \, dP(\omega)$$

とおき，$\mathbb{E}(X)$ を X の**期待値**という．

定義 1.14　：可測関数列 $\{X_n\}$ が X に**ほとんどいたるところで収束**するとは，ある零集合[6)] N が存在し，すべての $\omega \in \Omega \setminus N$ に対し，$X_n(\omega) \to X(\omega)$ が成立すること[7)]である．これを

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \qquad \mu - \text{a.e.}\omega \qquad \text{または} \qquad X_n \to_{\text{a.e.}} X$$

などと記す．特に，$\mu = P$ のとき，$X \xrightarrow{a.s.} X$ と記し，X_n は X に**ほとんど確実に収束する**という[8)]．

命題 1.7　：$\int X \, d\mu$，$\int Y \, d\mu$，$\int (X+Y) \, d\mu$ は存在すると仮定する．
(i) $\int (X+Y) \, d\mu = \int X \, d\mu + \int Y \, d\mu, \qquad \int cX \, d\mu = c \int X \, d\mu$
(ii) $X \geq 0, \mu - a.e.$ ならば[9)]，$\int X \, d\mu \geq 0$; $X \geq Y, \mu - \text{a.e.}$ ならば，$\int X \, d\mu \geq \int Y \, d\mu$; $X = Y, \mu - \text{a.e.}$ ならば，$\int X \, d\mu = \int Y \, d\mu$
(iii) X は可積分 $\Leftrightarrow$ $|X|$ は可積分
Y は可積分とする．このとき，$|X| \leq Y,\ \mu - \text{a.e.}$ ならば，$|X|$ も可積分.

---

6) $\mu(N) = 0$
7) X_n と X は同じ可測空間 $(\Omega, \mathcal{F})$ 上で定義されている．
8) ある確率的な命題が確率 *1* で成立するとき，その命題はほとんど確実に成立するという．これを $P - a.s.$ の成立すると記す．
9) $P(X \geq 0) = P\{\omega \in \Omega : X(\omega) \geq 0\}$ のとき，$X \geq 0, \mu - \text{a.e.}$ と書く．

証明：　□

·····1.3·····

積分の収束定理

定理 1.1　（単調収束定理）：$\{X_n\}$ を非負確率変数の非減少列とする．このとき

$$\lim_{n\to\infty}\int X_n\,d\mu=\int\lim_{n\to\infty}X_n\,d\mu.$$

証明： □

定理 1.2　（Fatou の lemma）：$\{X_n\}$ を非負確率変数列とする．このとき

$$\liminf_{n\to\infty}\int X_n\,d\mu\geq\int\liminf_{n\to\infty}X_n\,d\mu$$

である．

証明： □

定理 1.3　（有界収束定理）：$|X_n|\leq Y,\mu-a.e.$ かつ Y は可積分とし，$X_n\to_{a.e.}X$ とする．このとき，

$$\lim_{n\to\infty}\int X_n\,d\mu=\int X\,d\mu.$$

証明：$Z_n=|X_n-X|$ と $Z=2Y$ とおく．仮定より，$Z_n\to_{a.e.}0$ と $Z_n\leq|X_n|+|X|\leq Z$ が成立する．よって，$Z-Z_n\geq0$ に対し Fatou の補題を適用すれば，

$$\begin{aligned}\int Z\,d\mu&=\int\liminf_{n\to\infty}(Z-Z_n)\,d\mu\\&\leq\liminf_{n\to\infty}\int(Z-Z_n)\,d\mu\\&=\int Z\,d\mu-\limsup\int Z_n\,d\mu\end{aligned}$$

となり，

$$\limsup_{n\to\infty}\int Z_n\,d\mu=\limsup_{n\to\infty}\int|X-X_n|\,d\mu\leq0$$

したがって，

$$|\int X_n\,d\mu-\int X\,d\mu|\leq\int|X_n-X|\,d\mu\to0$$

□

命題 1.8　：各 $t \in (a, b)$ に対し，$X(\omega, t) : \Omega \to \mathbb{R}$ は可測とする．
(i) $t = t_0$ において $X(\omega, t)$ はほとんどいたるところ連続とし，t_0 の近傍において

$$|X(\omega, t)| \leq Y(\omega) \qquad a.e$$

とする．ただし，$Y(\omega)$ は可積分とする．このとき，

$$\int X(\,\cdot\,, t)\,d\mu$$

は $t = t_0$ において連続である．
(ii) $t \in (a, b)$ に対し，

$$\frac{\partial}{\partial t} X(\omega, t)$$

はほとんどいたるところ存在し，可積分関数 $Y(\omega)$ が存在し，すべての $t \in (a, b)$ に対し，

$$\left|\frac{\partial}{\partial t} X(\omega, t)\right| \leq Y(\omega), \qquad a.e.$$

が成立する．このとき，

$$\frac{\partial}{\partial t} \int X(\omega, t)\,d\mu(\omega) = \int \frac{\partial}{\partial t} X(\omega, t)\,d\mu(\omega)$$

が成立する．

証明：(i) 省略．
(ii) 中間値の定理から，t と $t + h$ の間にある s が存在し

$$\frac{X(\omega, t+h) - X(\omega, t)}{h} = \frac{\partial}{\partial u} X(\omega, u)\bigg|_{t=s}$$

と書ける．また，

$$\lim_{h \to 0} \frac{X(\omega, t+h) - X(\omega, t)}{h} = \frac{\partial}{\partial t} X(\omega, t) \qquad a.e.$$

であることと仮定から

$$\left|\frac{X(\omega, t+h) - X(\omega, t)}{h}\right| \leq Y(\omega)$$

となる．したがって，有界収束定理から

$$\int \frac{\partial}{\partial t} X(\omega, t)\,d\mu(\omega) = \lim_{h \to 0} \left\{\frac{X(\omega, t+h) - X(\omega, t)}{h}\right\} = \int \frac{\partial}{\partial t} X(\omega, t)\,d\mu$$

より、命題は示せた． □

定理 1.4　：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． □

定義 1.15　（一様可積分）：確率変数の列 $\{X_n\}_{n=1}^{\infty}$ 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 1.5　：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 1.16　：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 1.6　：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 1.7　：$X_n\xrightarrow{d}X$ 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 1.8　：$r>0$ とする．$X_n\xrightarrow{r-\mathrm{th}}X$ 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（$\Leftarrow$）．c_r- 不等式[2) より、

$$|X_n-X|^r\le c_r|X_n|^r+c_r|X|^r\equiv Y_n\ge 0$$

となる．(i) と (ii) より、

$$Y_n\xrightarrow{P}2c_r|X|^r\qquad かつ\qquad \boldsymbol{E}Y_n\longrightarrow 2c_r\boldsymbol{E}|X|^r \tag{1.3}$$

となる．$Y=2c_r|X|^r$ とおき、(1.3) と有界収束定理をもちいれば

$$\begin{aligned}\boldsymbol{E}[Y_n1\{|Y_n|\ge a\}] &= \boldsymbol{E}Y_n-\boldsymbol{E}[Y_n1\{|Y_n|<a\}]\\ &\longrightarrow \boldsymbol{E}Y-\boldsymbol{E}[Y1\{|Y_n|<a\}]\\ &= \boldsymbol{E}[Y1\{|Y_n|\ge a\}]\downarrow 0,\qquad (a\to\infty\ のとき)\end{aligned}$$

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

---

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

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

となることである．

2) Loéve の page 153

·····1.4·····

ヒルベルト空間

ヒルベルト空間は通常のユークリット空間を無限次元の空間に自然に拡張したものである．

1.4.1 ヒルベルト空間の定義

$\mathcal{H}$ を実数体 $\mathbb{R}$ 上のベクトル空間とする．$f,\, g \in \mathcal{H}$ に対して実数 $\langle f,\, g\rangle$ が対応し
(i) $\langle f,\, f\rangle \geq 0, \quad \langle f,\, f\rangle = 0 \Longleftrightarrow f = 0$
(ii) $\langle f,\, g\rangle = \langle g,\, f\rangle$
(iii) $\langle \alpha f + \beta g,\, h\rangle = \alpha\langle f,\, g\rangle + \beta\langle g,\, h\rangle, \qquad f,\, g \in \mathcal{H}, \alpha,\, \beta \in \mathbb{R}$
をみたすとき，写像

$$\langle\, \cdot\,,\, \cdot\,\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$$

を**内積**といい，内積の定義されている空間を**内積空間**または**前ヒルベルト空間**という．

定義 1.17　：ノルム $\|h\| = \sqrt{\langle h,\, h\rangle}$ について完備[1)]な内積空間を**ヒルベルト**空間という．

ヒルベルト空間の基本的不等式と等式
三角不等式：$\|g + h\| \leq \|g\| + \|h\|$
シュワルツの不等式：$\langle g,\, h\rangle \leq \|g\|\|h\|$
平行四辺形等式：$\|g + h\|^2 + \|g - h\|^2 = 2\{\|g\|^2 + \|h\|^2\}$

$(\Omega,\, \mathcal{F},\, \mu)$ を測度空間とする．正数 $p \geq 1$ に対し，

$$\int_\Omega |f(\omega)|^p \mu(\,d\omega) < \infty$$

をみたす Ω 上の実数値 $\mathcal{F}$ - 可測関数の全体を $\mathcal{L}^p(\Omega,\, \mathcal{F},\, \mu)$ または $\mathcal{L}^p(\mu)$ と記す．$\mathcal{L}^p(\mu)$ に同値関係を

$$f \sim g \Longleftrightarrow f = g, \quad \mu - a.e.$$

によって導入する．これによって分割される同値類の全体を $L(\Omega,\, \mathcal{F},\, \mu)$ または $L^p(\mu)$ と書く．特に，$p = 2$ のとき内積を

$$\langle g,\, h\rangle = \int_\Omega g(\omega)h(\omega)\mu(\,d\omega)$$

---

1) $\{h_n\}_{n=1}^\infty$ が $\mathcal{H}$ の中のコーシー列，すなわち，$\|h_n - h_m\| \to 0(\min(m,\, n) \to \infty)$ ならば，$\|h_n - h\| \to 0(n \to \infty)$ をみたす $h \in \mathcal{H}$ が存在することである．

で定義すれば，$L^2(\mu)$ はヒルベルト空間になること[2)]がわかる．

1.4.2 直交射影

ヒルベルト空間 $\mathcal{H}$ の部分集合 $\mathcal{H}_0$ が線形[3)] のとき，$\mathcal{H}_0$ を**部分空間**という．$\mathcal{H}_0$ が閉部分空間であるとは，$\|h - h_n\| \to 0, (n \to \infty)$ なる $\mathcal{H}_0$ の任意の列 $\{h_n\}$ に対して $h \in \mathcal{H}_0$ が成立することである．

$\mathcal{H}$ の元 g, h が $\langle g, h\rangle = 0$ であるとき，g と h は**直交**するという．集合 $\mathcal{H}_0 \subset \mathcal{H}$ に対し，

$$\mathcal{H}_0^{\perp} = \{h: \text{ すべての } g \in \mathcal{H}_0 \text{ に対して，} \langle g, h\rangle = 0\ \}$$

を $\mathcal{H}_0$ の**直交補空間**という．

定理 1.9　：$\mathcal{H}_0$ を $\mathcal{H}$ の閉部分空間とするとき，

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^{\perp}$$

すなわち，任意の $f \in \mathcal{H}$ は

$$f = g + h, \qquad g \in \mathcal{H}_0, h \in \mathcal{H}_0^{\perp}$$

と一意に表される．このとき，$\mathcal{H}$ から $\mathcal{H}_0$ への対応 $f \mapsto g$ は $\mathcal{H}_0$ の上への線形写像である．さらに，g は $\tilde{g} \in \mathcal{H}_0$ に対し $\|\tilde{g} - f\|$ を最小にする点である．

証明：　□

[2)] 完備性を示せば十分である．$\{h_n\}$ を $L^2(\mu)$ のコーシー列とすると，任意の $\epsilon > 0$ に対してある番号 $n(\epsilon)$ が存在し

$$\|h_m - h_n\| < \epsilon, \qquad m, n \geq n(\epsilon)$$

とできる．ここで，$\{h_n\}$ からつぎの条件をみたす部分列 $\{h_{n_k}\}$ をとる：

$$\|h_{n_{k+1}} - h_{n_k}\| \leq 2^{-k}, \qquad k \geq 1$$

このとき，$h \in L(\mu)$ が存在し，$\{h_{n_k}\}_{k=1}^{\infty}$ は h に $\mu - a.e.$ で収束することを示す．三角不等式と単調収束定理を用いれば

$$\begin{aligned}\int \left(\sum_{k=1}^{\infty} |h_{n_{k+1}}(\omega) - h_{n_k}(\omega)|\right)^2 \mu(\,d\omega) &= \lim_{N\to\infty} (\|\sum_{k=1}^{N} |h_{n_{k+1}} - h_{n_k}|\,\|)^2 \\ &\leq \lim_{N\to\infty} (\sum_{k=1}^{\infty} \|h_{n_{k+1}} - h_{n_k}\|)^2 \\ &\leq \left(\sum_{k=1}^{\infty} 2^{-k}\right)^2 = 1\end{aligned}$$

であるから，

$$\sum_{k=1}^{\infty} |h_{n_{k+1}} - h_{n_k}| < \infty, \qquad \mu - a.e.$$

が成立する．そこで，

$$h = h_{n_1} + \sum_{k=1}^{\infty} (h_{n_{k+1}} - h_{n_k})$$

とおくと，右辺の和は絶対収束するので，h は $\mu - a.e.$ で定義され

$$\lim_{k\to\infty} h_{n_k}(\omega) = h(\omega), \qquad \mu - a.e.\ \omega$$

である．Fatou の補題から

$$\|h - h_n\| \leq \liminf_{k\to\infty} \|h_{n_k} - h_n\| \leq \epsilon$$

となるので，$\{h_n\}$ は f にノルムの意味で収束することがわかる．

[3)] $g, h \in \mathcal{H}_0 \Longrightarrow \alpha g + \beta h \in \mathcal{H}_0, (\alpha, \beta \in \mathbb{R})$

$h_0 \in \mathcal{H}$ を固定する．各 $h \in \mathcal{H}$ に対して，

$$T(h) = \langle h,\, h_0 \rangle$$

とおくと，T は $\mathcal{H}$ 上の（実数値）**線形汎関数**である．すなわち，

$$T(\alpha g + \beta h) = \alpha T(g) + \beta T(h), \qquad g,\, h \in \mathcal{H}, \alpha,\, \beta \in \mathbb{R} \tag{1.4}$$

をみたす．さらに，Schwarz の不等式から $\|h_0\| = C$ とおいて

$$|T(h)| \leq C\|h\| \tag{1.5}$$

となる．

一般に，条件 (1.4) と (1.5) をみたす $\mathcal{H}$ 上の汎関数を**有界線形汎関数**という．

命題 1.9　：T を $\mathcal{H}$ 上の線形汎関数とする．このとき，

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

証明：　□

定理 1.10　（Riesz - Fréchet の定理）：T を $\mathcal{H}$ 上の有界線形汎関数とする．このとき，つぎの関係式をみたす $h_0 \in \mathcal{H}_0$ が存在する：

$$T(h) = \langle h,\, h_0 \rangle, \qquad h \in \mathcal{H}.$$

さらに，このような $h_0 \in \mathcal{H}$ は一意的に定まる．

証明：　□

·····1.5·····

条件付き期待値

定義 1.18　：$X \in L^2(\Omega, \mathcal{F}, P)$ とし，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ-集合族とする．このとき，$\mathcal{G}$ を与えたときの X の**条件付き期待値**は $L^2(\Omega, \mathcal{G}, P)$ の元 $\mathbb{E}\{X|\mathcal{G}\}$ でつぎを満足する唯一のものである：すべての $Y \in L^2(\Omega, \mathcal{G}, P)$ に対して

$$\mathbb{E}\{XY\} = \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\}. \tag{1.6}$$

前節の議論から $L^2(\Omega, \mathcal{F}, P)$ は内積 $\langle X, Y\rangle = \mathbb{E}[XY]\,(X, Y \in L^2(\Omega, \mathcal{F}, P))$ をもつヒルベルト空間となる．また，$\mathcal{G} \subset \mathcal{F}$ から $L^2(\Omega, \mathcal{G}, P)$ は $L^2(\Omega, \mathcal{F}, P)$ の部分空間となるので，定理 1.9 から $L^2(\Omega, \mathcal{G}, P)$ 上への $X(\in L^2(\Omega, \mathcal{F}, P))$ の射影が一意的に存在することが保障される．それを $\mathbb{E}\{X|\mathcal{G}\}$ としたわけである．

定理 1.11　：$X \in L^2(\Omega, \mathcal{F}, P)$ とし，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ 集合族とする．このとき，以下が成立する：
(i) $X \geq 0, a.s.$ ならば，$\mathbb{E}\{X|\mathcal{G}\} \geq 0, a.s.$.
(ii) 確率変数 Y に対し，$\mathcal{G} = \sigma(Y)$ とおいたとき，ボレロ可測関数 f が存在し

$$\mathbb{E}\{X|\mathcal{G}\} = f(Y)$$

とできる．
(iii) $\mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}\} = \mathbb{E}\{X\}$.
(iv) 写像 $X \to \mathbb{E}\{X|\mathcal{G}\}$ は線形.

証明：
(i) $X \geq 0, a.s.$ と仮定する．$Y = 1\{\mathbb{E}\{X|\mathcal{G}\} < 0\}$ とおく．このとき，$P\{\mathbb{E}\{X|\mathcal{G}\} < 0\} > 0$ ならば,

$$\mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\} = \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}1\{\mathbb{E}\{X|\mathcal{G}\} < 0\}\} < 0$$

となる．しかし，これは (1.6) から

$$\mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\} = \mathbb{E}\{XY\} \geq 0$$

となるので矛盾．したがって，$P\{\mathbb{E}\{X|\mathcal{G}\} < 0\} = 0$ となる．
(ii) は略.
(iii) は定義 1.18 から明らか.

(iv) $U, V \in L^2(\Omega, \mathcal{F}, P)$, $a \in \mathbb{R}$ とする．このとき，すべての $Y \in L^2(\Omega, \mathcal{G}, P)$ に対して

$$\begin{aligned}\mathbb{E}\{(U+aV)Y\} &= \mathbb{E}\{UY\} + a\mathbb{E}\{VY\} \\ &= \mathbb{E}\{\mathbb{E}\{U|\mathcal{G}\}Y\} + a\mathbb{E}\{\mathbb{E}\{V|\mathcal{G}\}Y\} \\ &= \mathbb{E}\{(\mathbb{E}\{U|\mathcal{G}\} + a\{\mathbb{E}\{V|\mathcal{G}\})Y\}\end{aligned}$$

と一意性から

$$\mathbb{E}\{U|\mathcal{G}\} + a\mathbb{E}\{V|\mathcal{G}\} = \mathbb{E}\{U+aV|\mathcal{G}\}$$

を得る． □

定義 1.18 を可積分確率変数に拡張する．

$L^+(\Omega, \mathcal{F}, P)$ を非負確率変数の集合[1)]とする．

補題 1.1 ：$X \in L^+(\Omega, \mathcal{F}, P)$ とし，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ 集合族とする．このとき，$L^+(\Omega, \mathcal{G}, P)$ の元 $\mathbb{E}\{X|\mathcal{G}\}$ でつぎを満足する唯一のものである：すべての $Y \in L^+(\Omega, \mathcal{G}, P)$ に対して

$$\mathbb{E}\{XY\} = \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\}. \tag{1.7}$$

また，$X \in L^2(\Omega, \mathcal{F}, P)$ ならば，定義 1.18 で与えられたものと同じになる．さらに，$0 \le X \le \tilde{X}$ ならば，

$$\mathbb{E}\{X|\mathcal{G}\} \le \mathbb{E}\{\tilde{X}|\mathcal{G}\}. \tag{1.8}$$

証明：もし，$X \in L^2(\Omega, \mathcal{F}, P)$ ならば，定義 1.18 で $\mathbb{E}\{X|\mathcal{G}\}$ を定義する．

$Y \in L^+(\Omega, \mathcal{G}, P)$ のとき，$Y_n = Y \wedge n$ とおけば，$Y_n \in L^2(\Omega, \mathcal{G}, P)$ となる．$X \in L^2(\Omega, \mathcal{F}, P)$ に対して

$$\begin{aligned}\mathbb{E}\{XY\} &= \lim_{n\to\infty} \mathbb{E}\{XY_n\} && \text{（単調収束定理）} \\ &= \lim_{n\to\infty} \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y_n\} && \text{(1.6) より} \\ &= \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\} && \text{（単調収束定理）}\end{aligned}$$

となり，$X \in L^2(\Omega, \mathcal{G}, P)$ のとき，すべての非負確率変数 Y に対して

$$\mathbb{E}\{XY\} = \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\} \tag{1.9}$$

が成立する．

つぎに，$X \in L^+(\Omega, \mathcal{F}, P)$ とする．$X_m = X \wedge m$ とおけば，$X_m \in L^2(\Omega, \mathcal{F}, P)$ となる．したがって，

$$\begin{aligned}\mathbb{E}\{XY\} &= \lim_{m\to\infty}\lim_{n\to\infty} \mathbb{E}\{X_mY_n\} && \text{（単調収束定理）} \\ &= \lim_{m\to\infty}\lim_{n\to\infty} \mathbb{E}\{\mathbb{E}\{X_m|\mathcal{G}\}Y_n\} && \text{((1.9) より)} \\ &= \lim_{m\to\infty} \mathbb{E}\{\mathbb{E}\{X_m|\mathcal{G}\}Y\} && \text{（単調収束定理）} \\ &= \mathbb{E}\{\lim_{m\to\infty} \mathbb{E}\{X_m|\mathcal{G}\}Y\}. && \text{（単調収束定理）}\end{aligned}$$

[1)] a.s. に等しいものは同値とみる．また，値 $+\infty$ も　　　許すとする．

したがって，

$$\mathbb{E}\{X|\mathcal{G}\} = \lim_{m\to\infty} \mathbb{E}\{X_m|\mathcal{G}\}$$

と定義すれば (1.7) を満足する．

$X \le \tilde{X}$ ならば，$X \wedge m \le \tilde{X} \wedge m$ となる．したがって，

$$\mathbb{E}\{X \wedge m\} \le \mathbb{E}\{\tilde{X} \wedge m\}$$

と定理 1.11 よりなる．よって，(1.9) は成立．

最後に一意性を示す．U, V は $\mathcal{G}$ 可測とし，すべての $Y \in L^+(\Omega, \mathcal{G}, P)$ に対して

$$\mathbb{E}\{XY\} = \mathbb{E}\{UY\} = \mathbb{E}\{VY\}$$

を満足するとする．

$$\Lambda_n = \{U < V \le n\}$$

とし，$P\{\Lambda_n\} > 0$ と仮定する．$\Lambda_n \in \mathcal{G}$ より，

$$\mathbb{E}\{X1_{\Lambda_n}\} = \mathbb{E}\{U1_{\Lambda_n}\} = \mathbb{E}\{V1_{\Lambda_n}\}$$

となる．さらに，$0 \ge U1_{\Lambda_n} \le V1_{\Lambda_n} \le n$ と $P\{1_{\Lambda_n}\} > 0$ より，$U1_{\Lambda_n}$ と $V1_{\Lambda_n}$ は a.s. に一致はしない．したがって，$\mathbb{E}\{U1_{\Lambda_n}\} < \mathbb{E}\{V1_{\Lambda_n}\}$ となり矛盾．よって，$P\{1_{\Lambda_n}\} = 0$ となる．さらに，$\{U < V\} = \cup_{n\ge1}\Lambda_n$ より

$$P\{U < V\} = P\{\bigcup_{n\ge1}\Lambda_n\} = \lim_{n\to\infty} P\{\Lambda_n\}$$

となる．同様に，$P\{U > V\} = 0$ も示すことができる．これらから $P\{U = V\} = 0$ となる．□

定理 1.12　：$X \in L(\Omega, \mathcal{F}, P)$ とし，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ 集合族とする．このとき，$L(\Omega, \mathcal{G}, P)$ の元 $\mathbb{E}\{X|\mathcal{G}\}$ でつぎを満足する唯一のものである：すべての有界 $\mathcal{G}$ 可測な Y に対して

$$\mathbb{E}\{XY\} = \mathbb{E}\{\mathbb{E}\{X|\mathcal{G}\}Y\}. \tag{1.10}$$

さらに，$X \in L^2(\Omega, \mathcal{F}, P)$ ならば，定義 1.18 のものと一致する．また，$X \ge 0$ ならば，補題 1.1 のものと一致する．さらに，

(i) $X \ge 0, a.s.$ ならば，$\mathbb{E}\{X|\mathcal{G}\} \ge 0, a.s.$

(ii) 写像 $X \to \mathbb{E}\{X|\mathcal{G}\}$ は線形．

証明：$X \in L(\Omega, \mathcal{F}, P)$ なので，$X^{=} \max(X, 0)$ と $X^- = \max(-X, 0)$ とおく．

$$\mathbb{E}\{X|\mathcal{G}\} = \mathbb{E}\{X^+|\mathcal{G}\} - \mathbb{E}\{X^-|\mathcal{G}\}$$

と定義すれば，補題 1.1 より $\mathbb{E}\{X|\mathcal{G}\}$ は (1.10) を満足することがわかる．のこりは補題 1.1 と同じように証明できる．□

定理 1.13　：$\{X_n\}_{n=1}^{\infty}$ を $(\Omega,\, \mathcal{F},\, P)$ 上の確率変数列とし，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ 集合族とする．
(i)（単調収束定理）$X_n \geq 0$ とし，$X_n \nearrow X,\, a.s.$ とする．このとき，

$$\lim_{n\to\infty} \mathbb{E}\{X_n|\mathcal{G}\} = \mathbb{E}\{X|\mathcal{G}\}, a.s..$$

(ii)（Fatou の補題）$X_n \geq 0$ とする．このとき，

$$\mathbb{E}\{\liminf_{n\to\infty} X_n|\mathcal{G}\} \leq \liminf_{n\to\infty} \mathbb{E}\{X_n|\mathcal{G}\}, a.s..$$

(iii)（有界収束定理）$\lim_{n\to\infty} X_n = X, a.s.$ とし，$|X_n| \leq Z (n \geq 1)$ なる $Z \in L(\Omega,\, \mathcal{F},\, P)$ が存在するとする．このとき，

$$\lim_{n\to\infty} \mathbb{E}\{X_n|\mathcal{G}\} = \mathbb{E}\{X|\mathcal{G}\}, a.s..$$

証明：(1.8) から $\mathbb{E}\{X_n|\mathcal{G}\} \leq \mathbb{E}\{X_{n+1}|\mathcal{G}\}, a.s.$ となり，

$$U = \lim_{n\to\infty} \mathbb{E}\{X_n|\mathcal{G}\}, a.s.$$

が存在する．すべての $\mathcal{G}$ 可測 Y に対して

$$\begin{aligned}
\mathbb{E}\{UY\} &= \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{E}\{\mathbb{E}\{X_n|\mathcal{G}\}Y \wedge m\} \qquad \text{（単調収束定理）} \\
&= \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{E}\{X_n Y \wedge m\} \\
&= \mathbb{E}\{XY\} \qquad \text{（単調収束定理）}
\end{aligned}$$

となり，$U = \mathbb{E}\{X|Y\}$ がわかる．
(ii) と (iii) については単調収束定理から Fatou の補題や有界収束定理を導出するときと同じようにすればよい．　□

定理 1.14　（Jensen の不等式）：$g : \mathbb{R} \to \mathbb{R}$ を凸関数とし，X と $g(X)$ は可積分とする．このとき，任意の部分 σ 集合族 $\mathcal{G}$ を $\mathcal{F}$ に対して，

$$g(\mathbb{E}\{X|\mathcal{G}\}) \leq \mathbb{E}\{g(X)|\mathcal{G}\}, a.s.$$

が成立する．

証明：　□

補題 1.2　：$(\Omega,\, \mathcal{F},\, P)$ を確率空間，$\mathcal{G}$ を $\mathcal{F}$ の部分 σ-集合族とし，X を $\mathcal{G}$-可測確率関数とする．このとき，任意の $B \in \mathcal{G}$ に対して，

$$\mathbb{E}\{X 1_B(X)\} = 0$$

が成立するならば，$X = 0, a.s.$ である．

証明：どんな $\epsilon > 0$ に対しても，$\{X \geq \epsilon\} \in \mathcal{G}$ であることに注意すれば，

$$0 \leq \epsilon P\{X \geq \epsilon\} = \mathbb{E}[\epsilon \mathbf{1}\{X \geq \epsilon\}] \leq \mathbb{E}[X \mathbf{1}\{X \geq \epsilon\}] = 0$$

となるので，$P\{X \geq \epsilon\} = 0$ が成立する．同様にすれば，$P\{X \leq -\epsilon\} = 0$ も成り立つことがわかる．したがって，どんな $\epsilon > 0$ に対しても

$$P\{-\epsilon < X < \epsilon\} = 1$$

となる．

いま，$A_n = \{-1/n < X < 1/n\}$ とおけば，

$$P(A_n) = 1 \quad \text{かつ} \quad \{X = 0\} = \cap_{n=1}^{\infty} A_n$$

となる．$\{A_n\}$ は減少列であることに注意すれば，

$$P\{X = 0\} = \lim_{n \to \infty} P(A_n) = 1$$

となり，補題は証明された． □

·····1.6·····

絶対連続，Radon = Nikodim 定理，Fubini 定理

$(\Omega, \mathcal{F}, \mu)$ を測度空間とし，X を Ω 上の非負可測関数とする．任意の $A \in \mathcal{F}$ に対し

$$\nu(A) = \int_A X(\omega)\, d\mu(\omega) = \int X(\omega) 1_A(\omega)\, d\mu \tag{1.11}$$

とおく．X が μ- 可積分ならば，ν は $(\Omega, \mathcal{F})$ 上の別の測度となる．このとき，(1.11) によって定義される測度 ν は μ に関する**密度** X をもつという．

定義 1.19　：μ, ν を $(\Omega, \mathcal{F})$ 上の測度とする．任意の $A \in \mathcal{F}$ に対し，

$$\mu(A) = 0 \Longrightarrow \nu(A) = 0$$

が成立するとき，ν は μ に関して**絶対連続**であるといい，$\nu \ll \mu$ と記す．また，ν は μ に**優越**されるという．

定理 1.15　：$(\Omega, \mathcal{F}, \mu)$ を σ- 有限測度空間とし，ν は $(\Omega, \mathcal{F})$ 上の測度で $\nu \ll \mu$ を満足する．このとき，非負可測関数 X が存在し，任意の $A \in \mathcal{F}$ に対し

$$\nu(A) = \int X(\omega) 1_A(\omega) d\mu(\omega)$$

とできる．さらに，

$$X \equiv \frac{d\nu}{d\mu}$$

は一意的[1)]に定まる．X を μ に関する **ラドン=ニコディムの微分**という．

証明：　□

系 1.1　：ν と μ は $(\Omega, \mathcal{F}$ 上の σ- 有限測度で $\nu \ll \mu$ とする．さらに，Z は可測関数とし，$\int Z\, d\mu$ が存在するとする．このとき，$A \in \mathcal{F}$ に対し

$$\int_A Z(\omega) d\nu(\omega) = \int_A Z(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega)$$

が成立する．

[1)] Y を別のものとすれば，$\mu\{\omega \in \Omega : Y(\omega) \neq X(\omega)\} = 0$ となることである．これを $X = Y, \mu - a.e.$ と書く．

証明：<u>第一段階</u>　$Z = 1_B, B \in \mathcal{F}$ と仮定する．このとき，ラドン=ニコディムの定理から

$$\int_A 1_B(\omega)\, d\nu(\omega) = \nu(A \cap B) = \int_{A\cap B} \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega)$$

が成立する．

<u>第二段階</u>　$A_1, A_2, \ldots, A_m \in \mathcal{F}$ は互いに排反とし，$Z = \sum_{i=1}^m z_i 1_{A_i} z_i \in \mathbb{R}$ とおく．このとき

$$\begin{aligned}\int_A Z(\omega)\, d\nu(\omega) &= \sum_{i=1}^m \int_A 1_{A_i}\, d\nu(\omega) \\ &= \sum_{i=1}^m \int_A 1_{A_i} \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega) \qquad \text{（第一段階から）} \\ &= \int_A Z(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega).\end{aligned}$$

<u>第三段階</u>　$Z \geq 0$ と仮定する．Z_n を非負単関数列で $Z_n \uparrow Z$ とする[2)]．このとき

$$\begin{aligned}\int_A Z(\omega)\, d\nu(\omega) &= \lim_{n\to\infty} \int_A Z_n(\omega)\, d\nu(\omega) \qquad \text{（単調収束定理）} \\ &= \lim_{n\to\infty} \int_A Z_n(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega) \qquad \text{（第二段階）} \\ &= \lim_{n\to\infty} \int_A Z(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega) \qquad \text{（単調収束定理）}.\end{aligned}$$

<u>第四段階</u>　Z は可測関数とし，Z^+ と Z^- のどちらか一方は ν- 可積分とする．このとき

$$\begin{aligned}\int_A Z(\omega)\, d\nu(\omega) &= \int_A Z^+(\omega)\nu(\omega) - \int_A Z^-(\omega)\, d\nu(\omega) \\ &= \int_A Z^+(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega) - \int_A Z^-(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega) \qquad \text{（第三段階）} \\ &= \int_A Z(\omega) \frac{d\nu}{d\mu}(\omega)\, d\mu(\omega).\end{aligned}$$

□

例 1.2　：$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$ を確率空間とする．P は $\mathbb{R}^n$ 上の測度 μ に関して密度 f をもつとすれば，

$$P(A) = \int_A f(\omega) d\mu(\omega), \qquad A \in \mathcal{F}$$

と書ける．μ が $\mathbb{R}^n$ 上のルベーグ測度ならば，f を**密度関数**とよび，μ が $\mathbb{R}^n$ 上の計数測度ならば，f を**頻度関数**または **mass function** とよぶ．

定理 1.16　（Scheffé の定理）：ν と $\nu_n, n = 1, 2, \ldots$ を $(\Omega, \mathcal{F})$ 上の測度とし，f と f_n をそれぞれ ν と ν_n に関する密度とし，$\nu(\Omega) = \nu_n(\Omega) = 1$ とし，

$$f_n(\omega) \to f_(\omega), \qquad \nu - a.e.$$

---

2) このような単関数列がとれることは (1.11) からわかる．

とする．このとき，

$$\sup_{A\in\mathcal{F}}|\nu_n(A)-\nu(A)| = \frac{1}{2}\int_\Omega |f_n(\omega)-f(\omega)|\,d\nu(\omega)$$

が成立する．

証明：$A\in\mathcal{F}$ に対し

$$|\nu_n(A)-\nu(A)| = |\int_A (f_n(\omega)-f(\omega))\,d\nu| \le \int_A |f_n(\omega)-f(\omega)|\,d\nu \le \int_\Omega |f_n(\omega)-f(\omega)|\,d\nu$$

が成立する．ここで，$g_n = f_n - f$ とおく．仮定より，$g_n \to 0, \nu - a.e.$ で $g_n^+ \le f$ となる[3)]．f は可積分なので，有界収束定理を用いれば

$$\int g_n^+\,d\nu \to 0$$

となる．しかし，

$$0 = \int (f_n - f)\,d\nu = \int g_n\,d\nu = \int (g_n^+ - g_n^1\,d\nu$$

より

$$\int g_n^+\,d\nu = \int g_n^-\,d\nu$$

を得る．よって，

$$\int |g_n|\,d\nu = \int g_n^+\,d\nu + \int g_n^-\,d\nu = 2\int g_n^+\,d\nu$$

となる．あとは

$$\int g_n^+\,d\nu \to 0$$

を示せば，定理は証明される．そのために，$B_n = 1\{\omega\in\Omega : f_n(\omega)-f(\omega)\ge 0\}$ とおく．このとき，

$$\begin{aligned}
\sup_{A\in\mathcal{F}}|\nu_n(A)-\nu(A)| &\ge |\nu_n(B)-\nu(B)| \\
&= \int_{\{\omega:f_n(\omega)-f(\omega)\ge 0\}} (f_n(\omega)-f(\omega))\,d\nu(\omega) \\
&= \int_{\{\omega:g_n^+(\omega)\ge 0\}} g_n^+(\omega)\,d\nu \\
&= \frac{1}{2}\int |f_n(\omega)-f(\omega)|\,d\nu(\omega)
\end{aligned}$$

---

3) $g_n^+ = \max(g_n, 0)$ と $g_n^- = \max(-g_n, 0)$ と定義.

となる．しかし

$$\begin{aligned}
|\nu_n(A) - \nu(A)| &= |\int_A (f_n(\omega) - f(\omega))\,d\nu(\omega)| \\
&= \int_{A\cap B} (f_n(\omega) - f(\omega))\,d\nu(\omega) + \int_{A\cap B^c} (f_n(\omega) - f(\omega))\,d\nu(\omega) \\
&\le \int g_n^*(\omega)\,d\nu(\omega) \\
&= \frac{1}{2}\int |f_n(\omega) - f(\omega)|\,d\nu(\omega) \to 0
\end{aligned}$$

を仮定より得る．よって，定理は証明された． □

補題 1.3　：X と Y は可積分な実数値確率関数とし，同時密度関数 $f_{X,Y}(x,\,y)$ をもつとする．このとき，

$$E\{X|Y\} = \frac{\int x f_{X,Y}(x,\,y)\,dx}{\int f_{X,Y}(x,\,y)\,dx}, \quad a.s. \tag{1.12}$$

で与えられる．

証明：(1.12) が成立することを示すには任意のボレロ可測集合 B に対して

$$\mathbb{E}\{X1_B(Y)\} = \mathbb{E}\{g(Y)1_B(Y)\} \tag{1.13}$$

を満足する $\sigma(Y)$-可測関数 $g(y)$ を求めれば，条件付き期待値の定義から

$$\mathbb{E}[\mathbb{E}\{X|Y\}1_B(Y)] = \mathbb{E}\{g(Y)1_B(Y)\}, \quad a.s.$$

となるので，補題 1.2 を用いれば

$$\mathbb{E}\{X|Y\} = g(Y), \quad a.s.$$

となることがわかる．

同時密度関数と Fubini の定理を用いれば，(1.13) の両辺を計算する：

$$\mathbb{E}\{X1_B(Y)\} = \int\int x1_B(y) f_{X,Y}(x,\,y)\,dx\,dx = \int_B \left(\int_{\mathbb{R}} x f_{X,Y}(x,\,y)\,dx\right) dy$$

となる．一方，

$$\mathbb{E}\{g(Y)1_B(Y)\} = \int\int g(y)1_B(y) f_{X,Y}(x,\,y)\,dx\,dy = \int Bg(y) \left(\int_{\mathbb{R}} f_{X,Y}(x,\,y)\,dy\right) dy$$

となる．このふたつの式と補題 1.2 から

$$\int_{\mathbb{R}} x f_{X,Y}(x,\,y)\,dy = g(y)\int_{\mathbb{R}} f_{X,Y}(x,\,y)\,dx, a.s.$$

が成立することがわかる．よって，補題は証明された． □

注意 1.3　：(1.12) において

$$f_{X|Y}(x,\,y) = \frac{f_{X,Y}(x,\,y)}{f_Y(y)}$$

を Y が与えられたときの X の条件付き密度関数と呼ぶことにする．ただし，

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,\,y)\,dx$$

である．

·····1.7·····

直積空間と独立性

定義 1.20　：$\mathcal{F}$ の部分 σ-加法族 $\mathcal{F}_i, i \in I$ が独立であるとは，I のすべての有限部分集合 J と任意の $A_i \in \mathcal{F}_i, i \in J$ に対して

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$$

が成立することである．

$(\Omega, \mathcal{F})$ と $(W, \mathcal{W})$ を可測空間とし，確率変数 $X : (\Omega, \mathcal{F}) \to (W, \mathcal{W})$ に対し，

$$X^{-1}(\mathcal{W}) = \{\{\omega \in \Omega : X(\omega) \in B\} \text{ for any } B \in \mathcal{W}\}$$

で定義された集合族は $\mathcal{F}$ の部分 σ-集合族となることがわかる．この部分 σ - 集合族を X から**生成された σ - 集合族**という．可測空間 $(W_i, \mathcal{W}_i)$, $i \in I$ に対し，$X_i : (\Omega, \mathcal{F}) \to (W_i, \mathcal{W}_i)$ を W_i 値の確率変数とする．このとき，X_i, I が独立であるとは，$X_i^{-1}(\mathcal{W}_i)$ が独立であることをいう．

$\mathcal{Y}$ と $\mathcal{Z}$ をそれぞれ空間 Z と W の σ - 加法族とする．Y と Z の直積集合

$$\mathcal{Y} \times \mathcal{Z} = \{A \times B : A \in \mathcal{Y},\, B \in \mathcal{Z}\}$$

を含む $Y \times Z$ 上の σ - 加法族の中で最小のものを $\sigma(\mathcal{Y} \times \mathcal{Z})$ を $\mathcal{Y} \times \mathcal{Z}$ の**直積 σ - 加法族**といい，これを

$$\mathcal{Y} \otimes \mathcal{Z} = \sigma(\mathcal{Y} \times \mathcal{Z})$$

と記すことにする．

$(Y, \mathcal{Y}, \mu)$ と $(Z, \mathcal{Z}, \nu)$ をふたつの σ - 有限測度空間とする．可測空間 $(Y \times Z, \mathcal{Y} \otimes \mathcal{Z})$ 上の測度 π を

$$\pi(A \times B) = \mu(A)\nu(B), \qquad A \in \mathcal{Y},\, B \in \mathcal{Z} \tag{1.14}$$

で定義する．これを**直積測度**という．

定理 1.17　：可測関数 $f : (Y \times Z, \mathcal{Y} \otimes \mathcal{Z}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ に対して，$y \in Y$ の切り口 $z \to f(y, z)$ は $\mathcal{Z}$ - 可測となる．

証明：(第一段階)　$C \in \mathcal{Y} \otimes \mathcal{Z}$ に対し，$f^y(z) = 1_C(y, z)$ と書けると仮定する．さらに，固定した $y \in Y$ に対し，

$$\mathcal{H} = \{C \in \mathcal{Y} \otimes \mathcal{Z} : z \to 1_C(y, z) \text{ は } \mathcal{F} \text{ - 可測 }\}$$

を定義する．すると，$\mathcal{H}$ は σ - 加法族になること[1)]がわかる．また，$\mathcal{H}$ は $\mathcal{Y} \times \mathcal{Z}$ を含んでいる．よって，$\mathcal{Y} \otimes \mathcal{Z} \subset \mathcal{H}$ となる．しかし，作り形から $\mathcal{H} \subset \mathcal{Y} \otimes \mathcal{Z}$ となるので，$\mathcal{H} = \mathcal{Y} \otimes \mathcal{Z}$ となり，$f^y(z) = 1_C(y,\, z), C \in \mathcal{H} \subset \mathcal{Y} \otimes \mathcal{Z}$ は $\mathcal{Z}$ - 可測．
（第二段階）　$f^y(z) = \sum_{i=1}^n a_i 1_{C_i}(y,\, z)$ と仮定する．第一段階から $f^y(z)$ は $\mathcal{Z}$ - 可測．
（第三段階）　非負関数 $f(y,\, z)$ に対し，$\{f_n(y,\, z)\}_{n=1}^\infty$ は単関数列とし，$f_n(y,\, z) \nearrow f(y,\, z)$ を満足するものとする．このとき，$f_n^y(z)$ は $\mathcal{Z}$ - 可測となり，

$$f^y(z) = \lim_{n \to \infty} f_n^y(z) = f(y,\, z)$$

も $\mathcal{F}$ - 可測．
（第四段階）　$f(y,\, z)$ を任意の可測関数とする．$f^y(z) = (f^+)^y(z) - (f^-)^y(z)$ とすれば，切り口 $g(z) = f(y,\, z)$ も $\mathcal{Z}$ - 可測となることがわかる．このとき，

□

定理 1.18　（Tonelli-Fubini の定理）：$(Y,\, \mathcal{Y},\, \mu)$ と $(Z,\, \mathcal{Z},\, \nu)$ を σ - 有限測度空間とする．
(i) $C \in \mathcal{Y} \otimes \mathcal{Z}$ に対し，

$$\pi(C) = \int_A \nu(C^y)\, d\mu = \int B\mu(C^y)\, d\nu \tag{1.15}$$

とおくと π は $(Y \times Z,\, \mathcal{Y} \otimes \mathcal{Z})$ 上の σ - 有限測度で (1.14) を満たす．さらに，(1.15) を満たす $(Y \times Z,\, \mathcal{Y} \otimes \mathcal{Z})$ 上の測度は一意的である．これを　$\mu \otimes \nu$ と書くことにする．
(ii) $f(y,\, z)$ を $\mathcal{Y} \otimes \mathcal{Z}$ - 可測関数とする．f が非負または $\mu \otimes \nu$ に関して可積分であるならば，$y \to \int f(y,\, z)\, d\nu(y)$ は $\mathcal{Y}$ - 可測である．そして，

$$\int f(y,\, z)\, d(\mu \otimes \nu) = \int \int \{f(y,\, z)\, d\nu(z)\}\, d\mu(y) = \int \int \{f(y,\, z)\, d\mu(y)\}\, d\nu(z)$$

が成立する．

証明：(i) まず，π の σ - 加法性を示す．そのために，$\{C_n\}_{n=1}^\infty$ を排反な $\mathcal{Y} \otimes \mathcal{Z}$ - 可測集合の列を取る．このとき，$\{(C_n)^y\}$ も互いに排反な $\mathcal{Z}$ - 可測集合の列となるので，

$$\nu(\bigcup_{n=1}^\infty (C_n)^y) = \sum_{n=1}^\infty \nu\{(C_n)^y\}$$

[1)] (i) $1_\Omega(y,\, z)$ は $\mathcal{F}$ - 可測．
(ii) $1_{C_1}(y,\, z)$ と $1_{C_2}(y,\, z)$ を $\mathcal{F}$ - 可測とすれば，$1_{C_1 \cap C_2}(y,\, z) = 1_{C_1}(y,\, z) 1_{C_2}(y,\, z)$ も $\mathcal{F}$ - 可測．したがって，$1_{C_1 \cup C_2}(y,\, z) = 1_{C_1}(y,\, z) + 1_{C_2}(y,\, z) - 1_{C_1 \cap C_2}(y,\, z)$ も $\mathcal{F}$ - 可測．$1_{\cup_{i=1}^\infty C_i}(y,\, z) = \lim_{i \to \infty} 1_{C_i}(y,\, z)$ から $C_1, C_2, \ldots \in \mathcal{H}$ ならば，$\cup_{i=1}^\infty C_i \in \mathcal{H}$ がわかる．

となる．さらに，$\nu\{(C_n)^y\}$ は非負であるので，(1.14) と単調収束定理から

$$\begin{aligned}\pi(\bigcup_{n=1}^{\infty} C_n) &= \int_Y \nu\{\bigcup_{n=1}^{\infty} C_n\}\, d\mu(y) \\ &= \int_Y \sum_{n=1}^{\infty} \nu\{(C_n)^y\}\, d\mu(y) \\ &= \sum_{n=1}^{\infty} \int_Y \nu\{(C_n)^y\}\, d\mu(y) \\ &= \sum_{n=1}^{\infty} \pi(C_n)\end{aligned}$$

より示せた．つぎに，π の一意性を示す．

(ii) 第一段階　(i) において $f(y,\, z) = 1_C(y,\, z), C in \mathcal{Y} \otimes \mathcal{Z}$ の場合についてはすでに示した．

第二段階　f を単関数とする．積分の線形性より結果は成立する．

第三段階　f は非負の $\mathcal{Y} \otimes \mathcal{Z}$ - 可測関数とする．f_n を単関数とし，$f_n \nearrow f$ とする．このとき，

$$\inf f(y,\, z)\, d(\mu \otimes \nu)(y,\, z) = \lim_{n\to\infty} \int f_n(y,\, z)\, d(\mu \otimes \nu)(y,\, z) = \lim_{n\to\infty} \int \{\int f_n(y, z)\, d\nu(y)\}\, d\mu(y)$$

となる．さらに，$y \to \int f_n(y,\, z)\, d\nu(z)$ は単調増加（n に関して）なので，$y \to \int f(y,\, z)\, d\nu(z)$ に収束する[2)]ので，再度単調収束定理を用いれば，

$$\begin{aligned}\lim_{n\to\infty} \int \{\int f_n(y,\, z)\, d\nu(z)\}\mu(y) &= \int \{\lim_{n\to\infty} \int f_n(y,\, z)\, d\nu(z)\}\, d\mu(y) \\ &= \int \{\int \lim_{n\to\infty} f_n(y,\, z)\, d\nu(y)\}\, d\mu(z) \\ &= \int \{\int f(y,\, z)\, d\nu(z)\}\, d\mu(z)\end{aligned}$$

となる．残りの部分は同様に示せる．

第四段階

□

例 1.3　：Fubini の定理を用いて

$$\lim_{n\to\infty} \int_0^N \frac{\sin x}{x}\, dx$$

を示そう．

任意の正の正数 n に対して

$$I_n \equiv \int_0^n \frac{\sin x}{x}\, dx = \int_0^n \frac{\sin x}{x}\, dx \int_0^{\infty} e^{-ux}\, du$$

---

2) $f_n^y(z)$ は単調増加で $f^y(z)$ に収束するので，単調収束定理から

$$\lim_{n\to\infty} \int f_n^y(z)\, d\nu(z) = \int f^y(z)\, d\nu(z)$$

がいえる．

であることに注意する．μ を **Lebeague** 測度とし，直積測度空間

$$((0,\, n] \times (0,\, \infty), \mathcal{B}((0,\, n]) \times \mathcal{B}((0,\, \infty)), \mu \times \mu)$$

を考える．このとき，$f(x,\, u) = e^{-ux} \sin x$ は $(0,\, n] \times (0,\, \infty)$ 上の連続関数なので，２次元 Borel 可測関数である．また，$\sup_{x \in \mathbb{R}} |sinx/x| \leq 1$ から

$$\int_0^n dx \int_0^\infty \left| e^{-ux} \frac{\sin x}{x} \right| du \leq= \int_0^n \left| \frac{\sin x}{x} \right| dx \leq n < \infty$$

から $f(x,\, u)$ は可積分であることがわかる．したがって，Fubini の定理から

$$I_n = \int_0^\infty du \int_0^n e^{-ux} \frac{\sin x}{x}\, dx$$

が得られる．簡単な部分積分の計算から

$$\int_0^n e^{-ux} \frac{\sin x}{x}\, dx = \frac{1}{1+u^2}\{1 - e^{-un}(u \sin x + \cos n)\}$$

が得られる．これから

$$I_n = \int_0^\infty \frac{1}{1+u^2}\, du - \int_0^\infty e^{-un} \frac{u \sin x + \cos n}{1+u^2}\, du$$

となる．しかし，

$$\int_0^\infty \left| e^{-un} \frac{u \sin x + \cos n}{1+u^2} \right| du \leq 2 \int_0^\infty e^{-un}\, du = \frac{2}{n}$$

となり，

$$\lim_{n \to \infty} \int_0^\infty e^{-un} \frac{u \sin x + \cos n}{1+u^2}\, du = 0$$

となる．したがって，

$$\lim_{n \to \infty} I_n = \int_0^\infty \frac{1}{1+u^2}\, du = \frac{\pi}{2}$$

がわかる．

問い 1.1　：X を確率変数とし, ϕ を正値でかつ $(0, \infty)$ 上で増加関数とし, $\phi(-x) = \phi(x)$ を満足するとする．このとき, 任意の $x > 0$ に対して,

$$P(|X| \geq x) \leq \frac{\mathbb{E}\phi(X)}{\phi(x)}$$

を示せ.

問い 1.2　：X を離散型確率変数とし, $\{0, 1, 2, \ldots\}$ をその値域とする．このとき, $\mathbb{E}X < \infty$ ならば,

$$\mathbb{E}X = \sum_{n=1}^\infty P(X \geq n)$$

が成立することを示せ.

問い 1.3　：X を確率変数とし，分布関数 F_X を持つとする．このとき，$X \geq 0$ ならば，

$$\mathbb{E}X = \int [1 - F_X(x)]\,dx$$

が成立することを示せ．一般に，$\mathbb{E}X$ が存在するならば，

$$\mathbb{E}X = \int_0^\infty [1 - F_X(x)]\,dx - \int_{-\infty}^0 F_X(x)\,dx$$

が成立することを示せ．

·····1.8·····

確率変数のいろいろな収束

1.8.1 確率収束

定義 1.21　：確率変数の列 $\{X_n\}_{n=1}^{\infty}$ が定数 c に**確率収束**するとは，どんな $\epsilon > 0$ に対しても，$n \uparrow \infty$ のとき，

$$P(|X_n - c| \geq \epsilon) \to 0$$

を満足することである．これを $X_n \xrightarrow{P} c$ と記す．

つぎは確率収束を示すときに有用な補題である．

補題 1.4　：どんな確率変数 X と定数 $a > 0$ に対しても，

$$P(|X| \geq a) \leq \frac{\boldsymbol{E}(X^2)}{a^2}$$

が成立する．

証明：簡単のために，X が確率密度関数 $f_X(x)$ を持つ場合のみを示す．

$$\begin{aligned}
\boldsymbol{E}(X^2) &= \int x^2 f_X(x)dx \\
&= \int_{|x|\geq a} x^2 f_X(x)dx + \int_{|x|<a} x^2 f_X(x)dx \\
&\geq \int_{|x|\geq a} x^2 f_X(x)dx \geq a^2 \int_{|x|\geq a} f_X(x)dx \\
&= a^2 P(|X| \geq a)
\end{aligned}$$

より補題は示された．□

補題 1.5　：g を $\boldsymbol{R}$ 上の非負偶関数で $(0, \infty)$ 上で正かつ増大とする．$a > 0$ に対し，

$$P(|X| \geq a) \leq \frac{\boldsymbol{E}[g(X)]}{g(a)}$$

が成立する．

証明：$|X| \geq a$ と $\sqrt{g(X)} \geq \sqrt{g(a)}$ が同値であることに注意して，前の補題を用いる． □

定理 1.19　：$X_n \xrightarrow{P} c$ となるための十分条件は

$$\boldsymbol{E}(X_n - c)^2 \to 0, \qquad (n \to \infty) \tag{1.16}$$

である．

証明：与えられたどんな正の数 $\epsilon > 0$ に対しても

$$P(|X_n - c| \geq \epsilon) \leq \frac{1}{\epsilon^2}\boldsymbol{E}(X_n - c)^2$$

となり，$n \uparrow \infty$ とすれば,

$$P(|X_n - c| \geq \epsilon) \to 0$$

を得る． □

注意 1.4　：(1.16) が成立するとき，X_n は c に**2 次平均収束**するという．

例 1.4　：$\epsilon_1, \ldots, \epsilon_n$ を成功する確率が p の独立な n 回のベルヌーイ試行とし，$S_n = \sum_{i=1}^n \epsilon_i$ とおく．$n \uparrow \infty$ のとき,

$$\boldsymbol{E}\left(\frac{S_n}{n} - p\right)^2 = \boldsymbol{var}\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{n} \to 0$$

となるので,

$$\frac{S_n}{n} \xrightarrow{P} p$$

となることがわかる．

定理 1.20　：$X_1, \ldots, X_n$ は独立同一の分布に従い，その平均と分散は $\boldsymbol{E}(X_1) = \xi$ と $\boldsymbol{var}(X_1) = \sigma^2 < \infty$ で与えられるとする．このとき,

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

とおけば,

$$\bar{X}_n \xrightarrow{P} \xi$$

が成立する．

証明：$n \uparrow \infty$ のとき,

$$\boldsymbol{E}(\bar{X}_n - \xi)^2 = \boldsymbol{var}(\bar{X}_n) = \frac{\sigma^2}{n} \to 0$$

となることから，補題 1.1 を使えば，定理は証明される． □

補題 1.6　：$\{A_n\}_{n=1}^{\infty}$ と $\{B_n\}_{n=1}^{\infty}$ を事象の列とする．$n \uparrow \infty$ のとき，

$$P(A_n) \to 1, \qquad P(B_n) \to 1$$

ならば，

$$P(A_n \cap B_n) \to 1$$

が成立する．

証明：$P(A_n^c) = 1 - P(A_n) \to 0 (n \to \infty)$ に注意すれば，

$$P\{(A_n \cap B_n)^c\} = P(A_n^c \cup B_n^c) \leq P(A_n^c) + P(B_n^c) \to 0, \quad (n \to \infty)$$

よりわかる．　□

定理 1.21　：$\{X_n\}_{n=1}^{\infty}$ と $\{Y_n\}_{n=1}^{\infty}$ を確率変数列で

$$X_n \xrightarrow{P} a, \qquad Y_n \xrightarrow{P} b, \qquad (n \to \infty)$$

を満足するとする．ただし，a と b は定数とする．このとき，

(i)　$X_n \pm Y_n \xrightarrow{P} a \pm b$
(ii)　$X_n Y_n \xrightarrow{P} ab$
(iii)　$b \neq 0$ ならば，$X_n / Y_n \xrightarrow{P} a/b$

が成立する．

証明：(i) の証明．$|(X_n + Y_n) - (a+b)| \leq |X_n - a| + |Y_n - b|$ から，どんな正の数 $\epsilon > 0$ に対しても

$$|X_n - a| < \frac{\epsilon}{2} \quad \text{かつ} \quad |Y_n - b| < \frac{\epsilon}{2} \tag{1.17}$$

ならば，

$$|(X_n + Y_n) - (a+b)| < \epsilon$$

であるので

$$\{|X_n - a| < \frac{\epsilon}{2}\} \cap \{|Y_n - b| < \frac{\epsilon}{2}\} \subset \{|(X_n + Y_n) - (a+b)| < \epsilon\}$$

より，$n \uparrow \infty$ のとき，

$$P\{|(X_n + Y_n) - (a+b)| < \epsilon\} \geq P[\{|X_n - a| < \frac{\epsilon}{2}\} \cap \{|Y_n - b| < \frac{\epsilon}{2}\}] \to 1$$

となる[1)]．なぜならば，$P\{|X_n - a| < \frac{\epsilon}{2}\} \to 1$ と $P\{|Y_n - b| < \frac{\epsilon}{2}\} \to 1$ から補題 1.3 を用いればわかる．

(ii) の証明．$X_nY_n - ab = (X_n - a)(Y_n - b) + b(X_n - a) + a(Y_n - b)$ に注意する．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned}P\{|X_nY_n - ab| \geq \epsilon\} &\leq P\{|(X_n - a)(Y_n - b)| \geq \frac{\epsilon}{3}\} + P\{|(X_n - a)| \geq \frac{\epsilon}{3|b|}\} \\ &\quad + P\{|(Y_n - b)| \geq \frac{\epsilon}{3|a|}\}\end{aligned}$$

となる．どんな正の数 $\delta > 0$ に対しても

$$\begin{aligned}P\{|(X_n - a)(Y_n - b)| \geq \frac{\epsilon}{3}\} &= P\{|(X_n - a)(Y_n - b)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - b| \geq \delta\} \\ &\quad + P\{|(X_n - a)(Y_n - b)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - b| < \delta\} \\ &\leq P\{|Y_n - b| \geq \delta\} + P\{|(X_n - a)(Y_n - b)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - b| < \delta\} \\ &\leq P\{|Y_n - b| \geq \delta\} + P\{|X_n - a| \geq \frac{\epsilon}{3\delta}\} \to 0, \qquad (n \to \infty)\end{aligned}$$

となることからわかる．

(iii) の証明．$1/Y_n \xrightarrow{P} 1/b$ を示せば (ii) よりわかる．十分小さな正の数 $\delta > 0$ に対して，$|Y_n - b| \leq \delta$ ならば，$|Y_n| \geq (1/2)|b|$ より

$$P\{|Y_n| \geq \frac{1}{2}|b|\} \geq P\{|Y_n - b| \leq \delta\} \to 1, \qquad (n \to \infty)$$

となる．また，$|Y_n| \geq (1/2)|b|$ のとき，

$$\left|\frac{1}{Y_n} - \frac{1}{b}\right| \leq \frac{|Y_n - b|}{|Y_n||b|} \leq \frac{2}{|b|^2}|Y_n - b|$$

が成立する．これらを用いれば，

$$\begin{aligned}P\left\{\left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \epsilon\right\} &= P\left\{\left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \epsilon, |Y_n| \geq \frac{1}{2}|b|\right\} + P\left\{\left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \epsilon, |Y_n| < \frac{1}{2}|b|\right\} \\ &\leq P\left\{|Y_n - b| \geq \frac{|b|^2}{2}\epsilon\right\} + P\left\{|Y_n| < \frac{1}{2}|b|\right\} \\ &\to 0, \qquad (n \to \infty) \end{aligned} \tag{1.18}$$

より示せた．□

---

1) (1.17) の対偶をとれば，

$$|(X_n+Y_n)-(a+b)| \geq \epsilon \quad \text{ならば} \quad |X_n-a| \geq \frac{\epsilon}{2} \text{ また } |Y_n-b| \geq \frac{\epsilon}{2}$$

より

$$\begin{aligned}P\{|(X_n + Y_n) - (a + b)| \geq \epsilon\} &\leq P[\{|X_n - a| \geq \frac{\epsilon}{2}\} \cup \{|Y_n - b| \geq \frac{\epsilon}{2} \\ &\leq P[\{|X_n - a| \geq \frac{\epsilon}{2}\}] + P[\{|Y_n - b| \geq\end{aligned}$$

としても示せる．

定理 1.22　（連続写像定理）：g を実数値連続関数とする．このとき，$Y_n \xrightarrow{P} b$（b は定数）ならば，$n \uparrow \infty$ のとき，

$$g(Y_n) \xrightarrow{P} g(b)$$

が成立する．

証明：どんな正の数 $\epsilon > 0$ に対してもある正の数 $\delta > 0$ が存在して，

$$|Y_n - b| < \delta \quad \text{ならば} \quad |g(Y_n) - g(b)| < \epsilon$$

を満足するので，

$$P\{|Y_n - b| < \delta\} \leq P\{|g(Y_n) - g(b)| < \epsilon\}$$

より，$n \uparrow \infty$ のとき，

$$P\{|g(Y_n) - g(b)| \geq \epsilon\} \leq P\{|Y_n - b| \geq \delta\} \to 0$$

より定理は示せた．□

定義 1.22　：g をある関数とする．T_n をモデル P_θ からの大きさ n のランダム標本に基づく推定量とする．推定量の列 $\{T_n\}_{n=1}^{\infty}$ が母数 $g(\theta)$ の**一致推定量**であるとは，$n \uparrow \infty$ のとき，

$$T_n \xrightarrow{P} g(\theta)$$

を満足することである．

例 1.5　：$X_1, \ldots, X_n$ は独立同一にある分布に従い，$\boldsymbol{E}(X_1) = \xi$ と $\boldsymbol{var}(X_1) = \sigma^2 < \infty$ とする．定理 1.2 より，標本平均 $\bar{X}_n$ は ξ の一致推定量であることがわかる．さらに，定理 1.2 から，$\bar{X}_n^2$ は ξ^2 の一致推定量であることがわかる．

定理 1.23　：推定量 T_n が $g(\theta)$ の一致推定量であるための十分条件は，$n \uparrow \infty$ のとき，T_n のバイアスと分散がゼロに収束することである．

証明：補題 1.1 を用いれば，どんな $\epsilon > 0$ にたいしても

$$\begin{aligned} P\{|T_n - g(\theta)| > \epsilon\} &\leq \frac{1}{\epsilon^2}\boldsymbol{E}|T_n - g(\theta)|^2 \\ &= \frac{1}{\epsilon^2}\left(\boldsymbol{E}|T_n - \boldsymbol{E}(T_n)|^2 + \{\boldsymbol{E}(T_n) - g(\theta)\}^2\right) \end{aligned}$$

が成立することよりわかる．□

注意 1.5　：$Y_n \xrightarrow{P} c$（c は定数）であっても，必ずしも $\boldsymbol{E}(Y_n) \to c$ や $\boldsymbol{E}(Y_n - c)^2 \to 0$ は成立しないことに注意．
反例：$c = 1$ とし，

$$P(Y_n = 1) = 1 - \frac{1}{n} \qquad P(Y_n = n) = \frac{1}{n}$$

とすれば，$Y_n \xrightarrow{P} 1$ となるが，すべての n に対して，$\boldsymbol{E}(Y_n) \to 2 (n \to \infty)$ となる．また，

$$P(Y_n = 1) = 1 - \frac{1}{\sqrt{n}} \qquad P(Y_n = n) = \frac{1}{\sqrt{n}}$$

とすれば，$Y_n \xrightarrow{P} 1$ となるが，

$$\begin{aligned}\boldsymbol{E}(Y_n - 1)^2 &= P(Y_n = 1)(1-1)^2 + P(Y_n = n)(n-1)^2 \\ &= 0 + \frac{1}{\sqrt{n}}(n-1)^2 \to \infty \qquad (n \to \infty)\end{aligned}$$

となる．

記号：(i) $\{X_n\}_{n=1}^{\infty}$ と $\{\epsilon_n\}_{n=1}^{\infty}$ を確率変数列とする．

$$\frac{X_n}{\epsilon_n} \xrightarrow{P} 0 \qquad n \to \infty$$

のとき，$X_n = o_P(\epsilon_n)$ と書く．特に，$X_n \xrightarrow{P} 0$ のとき，

$$X_n = o_P(1)$$

と記す．
(ii) どんな $\delta > 0$ に対してもある $M = M(\delta) > 0$ と正数 $n_0 = n_0(\delta) > 0$ が存在して，どんな $n > n_0$ に対しても

$$P(|X_n| \leq M|\epsilon_n|) \geq 1 - \delta$$

が成立するとき，

$$X_n = O_P(\epsilon_n)$$

と書く．特に，$P(|X_n| \leq M) \geq 1 - \delta$ ならば，$X_n = O_P(1)$ と書く．
(iii) どんな $\delta > 0$ に対しても，ある $0 < m(\delta) = m < M(\delta) = M < \infty$ と $n_0 = n_0(\delta)$ が存在し，どんな $n > n_0$ に対しても

$$P\left(m < \left|\frac{X_n}{\epsilon_n}\right| < M\right) \geq 1 - \delta$$

が成立するとき，$X_n \asymp_P \epsilon_n$（同じオーダー）と記す．
o_P と O_P の性質をまとめる：
(i) $X_n = o_P(\epsilon_n)$ かつ $Y_n = o_P(\epsilon_n)$ ならば，$X_n \pm Y_n = o_P(\epsilon_n)$ である．
(ii) $X_n = o_P(\epsilon_n)$ かつ $Y_n = O_P(\delta_n)$ ならば，$X_nY_n = o_P(\delta_n\epsilon_n)$ である．

(iii) 特に,

$$\begin{aligned} o_P(1) + o_P(1) &= o_P(1) \\ o_P(1) + O_P(1) &= O_P(1) \\ O_P(1)o_P(1) &= o_P(1) \\ (1 + o_P(1))^{-1} &= O_P(1) \\ o_P(\epsilon_n) &= \epsilon_n o_P(1) \\ o_P(1)(O_P(1)) &= o_P(1) \end{aligned}$$

である．

問い 1.4　：$X_1, X_2, \ldots, X_n$ を確率密度関数 $f(x) = 1_{(\xi,\xi+1)}(x)$ からの実数値ランダム標本とする．このとき, $X_{(1)}$ が ξ の弱一致推定量であることを示せ.

1.8.2 分布収束

$\{H_n\}_{n=1}^{\infty}$ を分布関数列とし,

$$H(x) \equiv \lim_{n\to\infty} H_n(x)$$

とおく．

明らかに,

(i)　$0 \le H(x) \le 1$
(ii)　$H(x)$ は非減少関数
(iii)　$H(x)$ は右連続関数

となるが,

$$\lim_{x\to-\infty} H(x) = 0 \quad と \quad \lim_{x\to\infty} H(x) = 1 \tag{1.19}$$

は保障されない．

反例：

$$H_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n}\int_{-\infty}^{x} \exp\left(-\frac{t^2}{2\sigma_n^2}\right) dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x/\sigma_n} \exp\left(-\frac{t^2}{2}\right) dt$$

とする．$\sigma_n \uparrow \infty$ とすれば，どんな x に対しても

$$H_n(x) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{2}$$

となる．したがって，$H(x) = 1/2$ となる．

定義 1.23　：確率変数列 $\{X_n\}_{n=1}^{\infty}$ が**確率有界**であるとは，どんな $\epsilon > 0$ に対しても，ある定数 K と自然数 n_0 が存在して，どんな $n > n_0$ に対しても

$$P(|X_n| \le K) \ge 1 - \epsilon$$

が成立することである．すなわち，$X_n = O_P(1)$ である．

定理 1.24　：$\{X_n\}_{n=1}^{\infty}$ を確率変数列とする．X_n の分布関数を H_n とし，$H(x) = \lim_{n\to\infty} H_n(x)$ とおく．このとき，$X_n = O_P(1)$ ならば，(1.19) が成立する．

証明：どんな正の数 $\epsilon > 0$ に対しても，ある x_0 と正の整数 n_1 が存在して，どんな $n \ge n_1$ にたいしても

$$1 - H_n(x) = P\{X_n > x_0\} \le \frac{\epsilon}{2}$$

が条件より成立する．また，ある正の整数 n_2 が存在して，どんな $n \ge n_2$ に対しても

$$|H_n(x_0) - H(x_0)| \le \frac{\epsilon}{2}$$

が成立する．したがって，どんな $x > x_0$ と $n > \max(n_1, n_2)$ に対しても

$$1 - H(x) \le 1 - H(x_0) \le 1 - H_n(x_0) + |H_n(x_0) - H(x_0)| \le \epsilon$$

が成立することがら定理は示された．□

定義 1.24　：分布関数列 $\{H_n(\cdot)\}_{n=1}^{\infty}$ が分布関数 $H(\cdot)$ に**分布収束**するとは，H のすべての連続点 x において，

$$H_n(x) \to H(x), \qquad (n \to \infty) \tag{1.20}$$

が成立することである．

記号：X_n の分布関数を $H_n(\cdot)$，X の分布関数を $H(\cdot)$ とする．H_n が H に分布収束するとき，

$$X_n \xrightarrow{L} X \quad \text{または} \quad H_n \xrightarrow{L} H$$

と書く．

問い 1.5　：$X_n \xrightarrow{P} c$（c は定数）かつ $P(X = c) = 1$ とする．
(i) X の分布関数 $H(x)$ を求めよ．H の不連続点は $x = c$ のみになることを確認せよ．
(ii) X_n の分布関数を $H_n(x)$ とする．

$$\begin{aligned} x < c \quad &\text{のとき，}\quad H_n(x) \to H(x) \\ x > c \quad &\text{のとき，}\quad H_n(x) \to H(x) \end{aligned}$$

が成立することを確認して，$X_n \xrightarrow{L} X$ をしめせ．

定理 1.25　(Slutsky の定理)：$\{X_n\}_{n=1}^{\infty}$，$\{A_n\}_{n=1}^{\infty}$，$\{B_n\}_{n=1}^{\infty}$ を確率変数列とし $X_n \xrightarrow{L} X$ ，$A_n \xrightarrow{P} a$，$B_n \xrightarrow{P} b$ を満足するとする．ただし，X は確率変数，a と b は定数とする．このとき，

$$A_n + B_n X_n \xrightarrow{L} a + bX$$

が成立する．

証明：まず，

$$A_n + X_n \xrightarrow{L} a + X$$

を示す．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned} F_{(A_n+X_n)}(x) &= P\{A_n + X_n \le x\} \\ &= P\{A_n + X_n \le x, A_n \ge a - \epsilon\} + P\{A_n + X_n \le x, A_n < a - \epsilon\} \end{aligned} \tag{1.21}$$

となることに注意する．十分小さな ϵ ととり，$x \pm \epsilon$ が $F_{(a+X)}(\cdot) = P\{a + X \le \cdot)$ の連続点になるようにとる[2)]．(1.21) から

$$F_{(A_n+X_n)}(x) \le P\{X_n \le x - a + \epsilon\} + P\{|A_n - a| > \epsilon\}$$

となる．また，$F_{X+a}(\cdot)$ の連続点 x に対して，

$$\begin{aligned} F_{(X_n+a)}(x) &= P\{X_n + a \le x) = F_{X_n}(x - a) \\ &\to F_X(x - a) = P(X \le x - a) = F_{X+a}(x) \end{aligned}$$

となる ことより，$X_n + a \xrightarrow{L} X + a$ が成り立つ．したがって，

$$\begin{aligned} \limsup_{n\to\infty} F_{(X_n+A_n)}(x) &\le \lim_{n\to\infty} \{P(X_n \le x - a + \epsilon) + P(|A_n - a| > \epsilon)\} \\ &= \lim_{n\to\infty} F_{(X_n+a)}(x + \epsilon) + \lim_{n\to\infty} P(|A_n - a| > \epsilon) \\ &= F_{(X+a)}(x + \epsilon) \end{aligned} \tag{1.22}$$

となる．また，

$$\begin{aligned} 1 - F_{(X_n+A_n)}(x) &= P\{X_n + A_n > x\} \\ &\le P\{X_n > x - a - \epsilon\} + P\{|A_n - a| \ge \epsilon\} \end{aligned}$$

から

$$\liminf_{n\to\infty} F_{(X_n+A_n)}(x) \ge \lim_{n\to\infty} \{F_{(X_n+a)}(x - \epsilon) + P(|A_n - a| > \epsilon)\} = F_{X+a}(x - \epsilon) \tag{1.23}$$

となる．(1.22) と (1.23) から

$$F_{(X+a)}(x - \epsilon) \le \liminf_{n\to\infty} F_{(X_n+A_n)}(x) \le \limsup_{n\to\infty} F_{(X_n+A_n)}(x) \le F_{(X+a)}(x + \epsilon)$$

[2)] 不連続点が高々可算個しかないことからこのような　点がとれることが保障される．

を得る．したがって，

$$\lim_{n\to\infty} F_{(X_n+A_n)}(x) = F_{(X+a)}(x)$$

が成り立つ．

つぎに，一般性を失わずに $b=1$ として，$B_n X_n \xrightarrow{L} X$ を示せば，定理は示される．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned} F_{B_nX_n}(x) &= P\{B_nX_n \le x\} \\ &= P\left\{B_nX_n \le x, \left|\frac{1}{B_n}-1\right| \le \frac{\epsilon}{|x|}\right\} + P\left\{B_nX_n \le x, \left|\frac{1}{B_n}-1\right| > \frac{\epsilon}{|x|}\right\} \\ &\le P\{X_n \le x+\epsilon\} + P\left\{\left|\frac{1}{B_n}-1\right| > \frac{\epsilon}{|x|}\right\} \end{aligned}$$

より，$n\uparrow\infty$ のとき，

$$\limsup_{n\to\infty} F_{B_nX_n}(x) \le \lim_{n\to\infty}\left[P\{X_n \le x+\epsilon\} + P\left\{\left|\frac{1}{B_n}-1\right| > \frac{\epsilon}{|x|}\right\}\right] \to F_X(x+\epsilon)$$

を得る．同様な議論により，

$$\liminf_{n\to\infty} F_{B_nX_n}(x) \ge F_X(x-\epsilon)$$

を得る．したがって，

$$F_X(x-\epsilon) \le \liminf_{n\to\infty} F_{B_nX_n}(x) \le \limsup_{n\to\infty} F_{B_nX_n}(x) \le F_X(x+\epsilon)$$

が成り立つので，

$$F_{B_nX_n}(x) \to F_X(x)$$

を得る．よって，定理は示された． □

定義 1.25　：確率変数列 $\{X_n\}_{n=1}^{\infty}$ が確率変数 X に確率収束するとは，$X_n - X \xrightarrow{P} 0$ が成立することをいう．記号で，$X_n \xrightarrow{P} X$ と記す．

定理 1.26　：$X_n \xrightarrow{P} X$ ならば，$X_n \xrightarrow{L} X$ が成立する．

証明：$R_n = X_n - X$ とおく．条件より，$R_n \xrightarrow{P} 0$ となり，$X_n = X + R_n$ に対して，Slutsky の定理を用いれば，この定理は示される． □

注意 1.6　：上の定理の逆は一般には成立しないが，c をある定数とすると，

$$X_n \xrightarrow{L} c \quad \Leftrightarrow \quad X_n \xrightarrow{P} c$$

である．実際，X_n と X の分布関数を H_n と H かけば，すべての $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} H_n(c-\epsilon-) = H(c-\epsilon-) = 0$$

と

$$\lim_{n\to\infty} H_n(c+\epsilon) = H(c+\epsilon) = 1$$

から

$$\begin{aligned} P(|X_n - c| \geq c) &= P(X_n \geq c+\epsilon) + P(X_n \leq c-\epsilon) \\ &= 1 - H_n(c+\epsilon-) + H_n(c-\epsilon) \to 0 \end{aligned}$$

となる．

定理 1.27 ：$X_n \xrightarrow{L} X$ となるための必要十分条件は，すべての有界連続な関数 f に対し，

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

が成立することである．

証明：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 $B > 0$ とある正の整数 n_0 が存在して，どんな $n \geq n_0$ に対しても $P(|X_n| \geq B) \leq \epsilon/(3A)$ ともできる．h を実数値関数とし，$0 \leq h(x) \leq 1$ でかつ

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

を満足するものとする．このとき，

$$\begin{aligned} |\boldsymbol{E}[f(X_n)] - \boldsymbol{E}[f(X)]| &\leq |\boldsymbol{E}[f(X_n)] - \boldsymbol{E}[f(X_n)h(X_n)]| \\ &\quad + |\boldsymbol{E}[f(X_n)h(X_n)] - \boldsymbol{E}[f(X)h(X)]| + |\boldsymbol{E}[f(X)h(X)] - \boldsymbol{E}[f(X)]| \\ &\leq \frac{\epsilon}{3} + |\boldsymbol{E}[f(X_n)h(X_n)] - \boldsymbol{E}[f(X)h(X)]| \end{aligned}$$

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，

$$|\boldsymbol{E}[g(X_n)] - \boldsymbol{E}[g(X)]| < \frac{\epsilon}{3}$$

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，

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

となる[3)]．さらに，n を十分おおきくとれば，

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

とできることからわかる． □

[3)]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

問い 1.6　：$\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ を確率変数列とし,

$$X_n \xrightarrow{L} X \qquad (\text{確率変数}), \qquad Y_n \xrightarrow{L} c \qquad (\text{定数})$$

を満足する．このとき,

$$X_n Y_n \xrightarrow{L} cX$$

が成立することを示せ.

1.8.3 **概収束**

定義 1.26　：確率変数列 $\{X_n\}_{n=1}^{\infty}$ が確率変数 X に**概収束**するとは,

$$P\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$$

が成立することである．これを $X_n \xrightarrow{as} X$ と記すことにする．

注意 1.7　：任意の $\epsilon > 0$ に対し,

$$\lim_{n\to\infty} P\left\{\omega \in \Omega : \text{すべての } k, l \geq n \text{ にたいし}, |X_k(\omega) - X_l(\omega)| < \epsilon\right\} = 1$$

となることが，$X_n \xrightarrow{as} X$ であるための必要十分条件である[4)].

注意 1.8　：$X_n \xrightarrow{as} X$ ならば，$X_n \xrightarrow{P} X$ であるが，逆は必ずしも成立しない．
反例：$Z \sim U(0,1)$ とし,

$$X_1 = 1, X_2 = \mathbf{1}_{[0,1/2)}(Z), X_3 = \mathbf{1}_{[1/2,1)}(Z), X_4 = \mathbf{1}_{[0,1/4)}(Z), X_5 = \mathbf{1}_{[1/4,1/2)}(Z), \ldots$$

とおく．一般に，$n = 2^k + m (0 \leq m \leq 2^k)$ に対して,

$$X_n = \mathbf{1}_{[m/2^k,(m+1)/2^k)}(Z)$$

とする．このとき，どんな $Z \in (0,1)$ に対しても $\{X_n(Z)\}_{n=1}^{\infty}$ は収束しない．しかし,

$$P(|X_n| > \epsilon) < \frac{1}{2^k} \to 0, \qquad n \to \infty$$

より，$X_n \xrightarrow{P} X$ となる．

定理 1.28　：$\{X_n\}_{n=1}^{\infty}$ を確率変数列とし, X を確率変数とする．このとき,

$$X_n \xrightarrow{a.s} X$$

であるために必要十分条件は任意の $\epsilon > 0$ に対して,

$$\lim_{n\to\infty} P\{\sup_{k\geq n} |X_k - X| \geq \epsilon\} = 0$$

が成立することである.

---

4) 西尾の確率論 *71* ページを参照のこと．

証明： 略. □

補題 1.7　（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，

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

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

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

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． □

問い 1.7　：$\{X_i\}_{n=1}^{\infty}$ を独立同一分布に従う確率変数列で $\mathbb{E}X_1^4 < \infty$ を仮定する．さらに，$S_n = X_1 + X_2 + \cdots + X_n, \mathbb{E}X_1 = \xi, \mathbb{VAR}X_1 = \sigma^2$ と記すことにする．
(i)

$$\mathbb{E}|S_n - n\xi|^4 \le 3n(n-1)\{\mathbb{VAR}(X_1)\}^2 + n\mathbb{E}[(X_1 - \xi)^4]$$

を示せ．
(ii) Chebyshev の不等式を利用して，任意の $\epsilon > 0$ に対して，

$$P(|\frac{S_n}{n} - \xi| \ge \epsilon) = O\left(\frac{1}{n^2}\right)$$

を示せ．
(iii) Borel-Cantelli の補題と上の定理を用いて，

$$\frac{S_n}{n} \xrightarrow{a.s} \xi$$

を示せ．（この結果は大数の強法則から保障されるが，有限な４次積率をもつことを仮定するとこの法則は比較的簡単に証明できるわけである.）

·····1.9·····

種々の収束の関係

定理 1.29　：$\{X_n\}_{n=1}^{\infty}$ 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_n 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_n 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_n は X に法則収束する．

注意 1.9　：上の定理の逆は一般には成立しない．

定義 1.27　（r 次平均収束）：確率変数の列 $\{X_n\}$ が確率変数 X に**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．

定理 1.30　（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[1)]関数に対し，

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

が成立する．
(iv)．すべての非負連続関数 f に対し，

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

[1)] 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．
(v)．すべての開集合 G に対し，

$$\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$$

が成立する．
(vi)．すべての閉集合 F に対し，

$$\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$$

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

定理 1.31 （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． □

·····1.10·····

特性関数

定義 1.28　：確率変数 X の**特性関数**とは，変数 $t \in \boldsymbol{R}$ の複素値関数

$$\phi_X(t) = \boldsymbol{E}[e^{\sqrt{-1}tX}] = \boldsymbol{E}[\cos(tX)] + \boldsymbol{E}[\sin(tX)]$$

である．また， k – 次元確率変数 $\boldsymbol{X} = (X_1, \ldots, X_k)^T$ とベクトル $\boldsymbol{t} = (t_1, \ldots, t_k)^T \in \boldsymbol{R}^k$ に対して，

$$\boldsymbol{t}^T\boldsymbol{X} = t_1X_1 + \cdots + t_kX_k$$

ときすことにする．確率変数 $\boldsymbol{X}$ の特性関数を

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \boldsymbol{E}[e^{\sqrt{-1}\boldsymbol{t}^T\boldsymbol{X}}]$$

で定義する．

<u>特性関数の性質</u>
(i) $\phi_X(0) = 1$
(ii) $|\phi_X(t)| \leq 1$ である[1)].
(iii) a, b を定数とする．$Y = aX + b$ ならば，

$$\phi_Y(t) = e^{\sqrt{-1}tb}\phi_X(at)$$

である．
(iv) $\phi_X(t) = \overline{\phi_X(-t)}$ （$\bar{z}$ は z の共役）
(v) $\phi_X(t)$ は一様連続[2)].
(vi) 確率変数 X が次数 k 次の絶対モーメント $\boldsymbol{E}|X|^k < \infty$ をもつとする．このとき，$\phi_X(t)$ は k 回連続微分可能で

$$\phi_X^{(k)}(0) = \{\sqrt{-1}\}^k\boldsymbol{E}[X^k]$$

である．
　ここで，今後よく使われるつぎの関係式をあげておく.

補題 1.8　：a, b を任意の実数とする.

---

1) 実際，$|\phi_X(t)| \leq \boldsymbol{E}|e^{\sqrt{-1}tX}| = 1$ よりわかる．

2) どんな h に対しても

$$|\phi_X(t+h)-\phi_X(t)| = |\boldsymbol{E}[e^{\sqrt{-1}tX}(e^{\sqrt{-1}th}-1)]| \leq \boldsymbol{E}[|e^{\sqrt{-1}tX}||e^{\sqrt{-1}th}-1|] = \boldsymbol{E}|e^{\sqrt{-1}th}-1|$$

となる．$h \downarrow 0$ のとき，有界収束定理から $\boldsymbol{E}|e^{\sqrt{-1}th}-1| \to 0$ がわかる．

(i)

$$\left|\int_a^b \frac{\sin x}{x}\,dx\right| \le C.$$

ただし，C は a, b に依存しない定数である．

(ii)

$$\lim_{T\to\infty}\int_0^T \frac{\sin x}{x}\,dx = \frac{\pi}{2}.$$

(iii)

$$\lim_{T\to\infty}\frac{2}{\pi}\int_0^T \frac{\sin ax}{x}dx = \begin{cases} -1 & (a<0) \\ 0 & (a=0) \\ 1 & (a>0) \end{cases}. \tag{1.24}$$

(iv)

$$\lim_{T\to\infty}\frac{2}{\pi}\int_0^T \frac{\sin ax}{x}\cos bx\,dx = \begin{cases} 0 & (|b|>a) \\ \frac{1}{2} & (|b|=a, a>0) \\ 1 & (|b|<a) \end{cases}.$$

(v)

$$\lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^T \frac{\sin^2 x}{x^2}e^{bx\sqrt{-1}}\,dx = \begin{cases} 0 & (|b|>2) \\ 1-\frac{|b|}{2} & (|b|\le 2) \end{cases}. \tag{1.25}$$

証明：(ii) は 例 1.3 からわかる．(iii) は (ii) よりわかる．
(i) および (iv) は「フーリエ解析と確率入門（河田龍夫著，日本評論社）の 83 ページ，(v) は同書のくページ 108 を参照．□

定理 1.32　：F を確率変数 X の分布関数とする．$a<b$（$a, b\in\boldsymbol{R}$）を F の連続点としたとき，

$$P\{a<X<b\} = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^T \frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}\phi_X(t)dt$$

が成り立つ．

証明：Fubini の定理と有界収束定理を用いた後に (1.24) を利用すれば,

$$
\begin{aligned}
\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\phi_X(t)dt &= \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\sqrt{-1}t(x-a)}-e^{-\sqrt{-1}t(x-b)}}{\sqrt{-1}t}dt\right]dF(x) \\
&= \int_{-\infty}^{\infty}\left[\frac{1}{\pi}\int_{0}^{T}\frac{\sin t(x-a)}{t}dt-\int_{0}^{T}\frac{\sin t(x-b)}{t}dt\right]dF(x) \\
&\to \int_{-\infty}^{\infty}\left[\frac{1}{\pi}\int_{0}^{\infty}\frac{\sin t(x-a)}{t}dt-\int_{0}^{\infty}\frac{\sin t(x-b)}{t}dt\right]dF(x) \\
&= \int_{(-\infty,a)}\left[\frac{1}{\pi}\frac{-\pi}{2}-\frac{1}{\pi}\frac{-\pi}{2}\right]dF(x)+\int_{\{a\}}\left[0-\frac{1}{\pi}\frac{-\pi}{2}\right]dF(x) \\
&\quad +\int_{(a,b)}\left[\frac{1}{\pi}\frac{\pi}{2}-\frac{1}{\pi}\frac{-\pi}{2}\right]dF(x)+\int_{\{b\}}\left[\frac{1}{\pi}\frac{\pi}{2}-0\right]dF(x) \\
&\quad +\int_{(b,\infty)}\left[\frac{1}{\pi}\frac{\pi}{2}-\frac{1}{\pi}\frac{\pi}{2}\right]dF(x) \\
&= \int_{(a,b)}dF(x)
\end{aligned}
$$

より定理は示された． □

系 1.2 ：2 つの確率測度が同じ特性関数をもてば，それらは等しい．

定理 1.33 ：$X_n \xrightarrow{L} X$ となるための必要十分条件は，すべての $\xi \in \boldsymbol{R}$ について

$$
\boldsymbol{E}e^{\sqrt{-1}\xi X_n} \to \boldsymbol{E}e^{\sqrt{-1}\xi X}
$$

が成立することである．

証明：<u>必要性</u>. $X_n \xrightarrow{L} X$ とする．それぞれの t に対して，$\cos tx$ と $\sin tx$ は有界連続関数なので，定理 1.9 を用いれば,

$$
\begin{aligned}
\phi_{X_n}(t) &= \boldsymbol{E}[\cos tX_n]+\sqrt{-1}\boldsymbol{E}[\sin tX_n] \\
&\to \boldsymbol{E}[\cos tX]+\sqrt{-1}\boldsymbol{E}[\sin tX]=\phi_X(t)
\end{aligned}
$$

からわかる．
<u>十分性</u>. F の任意の連続点 a,b（$a<b$）に対して,

$$
F_{X_n}(b)-F_{X_n}(a) \to F_X(b)-F_X(a)
$$

を示せばよい．ただし，F_{X_n} と F_X は X_n と X の分布関数である．定理 1.14 と有界収束定理を用いれば,

$$
\begin{aligned}
F_X(b) - F_X(a) &= \lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\phi_X(t)dt \\
&= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\left[\lim_{n\to\infty}\phi_{X_n}(t)\right]dt \\
&= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\left[\lim_{n\to\infty}\phi_{X_n}(t)\right]dt \\
&= \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\phi_{X_n}(t)dt \\
&= \lim_{n\to\infty}[F_{X_n}(b)-F_{X_n}(a)]
\end{aligned}
$$

となることより示せた． □

⋯⋯1.11⋯⋯

大数の法則

定理 1.34　：$X_1, X_2, \ldots, X_n$ は独立に同一の分布[1] 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 1.3　：$X_1, X_2, \ldots, X_n$ 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，

$$\begin{aligned} g(t) &= \log \boldsymbol{E}[\exp(tZ_1)] \\ &= -pt + \log(pe^t + (1-p)) \end{aligned}$$

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，

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

[1] 同一ではなくい独立な確率変数列にたいしては，2 次の積率が有限であれば，同様な結果が成立することが知られている．

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,

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

を意味している．$\epsilon > 0$ は任意だったので，これより

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

を得る．同様にして,

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

を得ることができる．したがって,

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

を得る． □

つぎに，独立な確率変数列に対する大数の法則を述べる．

定理 1.35　：$X_1, X_2, \ldots, X_n$ を独立な確率変数で有限の期待値をもつとする．
(A)（大数の強法則）

$$\sum_{i=1}^{\infty} \frac{\mathbb{VAR}(X_i)}{i^2} < \infty$$

ならば,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \xrightarrow{a.s.} 0$$

が成立する．
(B)（大数の弱法則）ある正の数 $\delta \in (0,1)$ が存在して，

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}|X_i|^{1+\delta} = 0$$

ならば，

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \xrightarrow{P} 0$$

が成立する．

証明. 略．

·····1.12·····

中心極限定理

1.12.1 独立同一な確率変数にたいする中心極限定理

定理 1.36　：$\{X_n\}_{n\geq 1}$ は独立同一の分布に従う確率変数の列とし，

$$\boldsymbol{E}(X_1) = \xi, \qquad \boldsymbol{var}(X_1) = \sigma^2$$

を満足するものとする．このとき，

$$\frac{\sqrt{n}(\bar{X}_n - \xi)}{\sigma} \xrightarrow{L} N(0,1)$$

が成立する．

証明：概略のみを示す．$Y \sim N(0,1)$ とすれば，

$$\boldsymbol{E}[e^{\sqrt{-1}tY}] = \exp(-\frac{t^2}{2})$$

である．また，

$$\begin{aligned}\boldsymbol{E}\left[\exp\left(\sqrt{-1}t\frac{\sqrt{n}(\bar{X}_n - \xi)}{\sigma}\right)\right] &= \Pi_{i=1}^n \boldsymbol{E}\left[\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_i - \xi}{\sigma}\right)\right\}\right] \\ &= \left[\boldsymbol{E}\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1 - \xi}{\sigma}\right)\right\}\right]^n\end{aligned}$$

となる．また，

$$\boldsymbol{E}\left[\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1 - \xi}{\sigma}\right)\right\}\right] = 1 - \frac{t^2}{2n} + o(\frac{t^2}{n})$$

より

$$\boldsymbol{E}\left[\exp\left(\sqrt{-1}t\frac{\sqrt{n}(\bar{X}_n - \xi)}{\sigma}\right)\right] = \left[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right]^n \to e^{-\frac{t^2}{2}}$$

となり，定理 1.15 からわかる． □

1.12.2 独立な確率変数にたいする中心極限定理

定理 1.37　：$\{X_n\}_{n\geq 1}$ は独立な確率変数の列で，それぞれは 3 次のモーメントをもつものとする．さらに，

$$\boldsymbol{E}(X_i) = \xi_i, \qquad \boldsymbol{var}(X_i) = \sigma_i^2, \qquad (i = 1, 2, \ldots)$$

を満足するものとする．このとき，

$$\left[\sum_{j=1}^{n} \boldsymbol{E}|X_j - \xi_j|^3\right]^2 = o\left[\left(\sum_{j=1}^{n} \sigma_j^2\right)^3\right] \tag{1.26}$$

ならば，

$$\frac{(\bar{X}_n - \boldsymbol{E}(\bar{X}_n))}{\sqrt{\boldsymbol{var}(\bar{X}_n)}} \xrightarrow{L} N(0,1)$$

が成立する．

証明：略．□

定理 1.38　：$\{X_n\}_{n=1}^{\infty}$ 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

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^{n} \int_{[|x-\mu_i|>\epsilon B_n]} (x-\mu_i)^2 dF_i(x) = 0 \tag{1.27}$$

ならば，

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{\sigma_i^2}{B_n^2} = 0 \tag{1.28}$$

かつ

$$\frac{S_n - \boldsymbol{E}(S_n)}{B_n} \xrightarrow{d} N(0,1) \tag{1.29}$$

である．

　逆に，(1.28) と (1.29) が成立するならば，(1.27) が成り立つ．

証明：略．□

注意 1.10　：(1.27) の十分条件で比較的確認が簡単のもととして Liapunov 条件がある．すなわち，ある正の数 $\delta > 0$ が存在して，

$$\sum_{i=1}^{n} \mathbb{E}|X_i - \mathbb{E}X_i|^{2+\delta} = o(B_n^{2+\delta}) \tag{1.30}$$

を満足すればよい．

例 1.6　：$X_1, X_2, \ldots, X_n$ を $P_\theta(X_1 \le x) = F(x-\theta)$ からのランダム標本とする．ただし，θ は未知の定数とし，$F(0) = 1/2$ とする．簡単のために，$n = 2m-1$ は奇数とする．順序統計量を $X_{(1)} \le \cdots \le X_{(n)}$ としたとき，

$$\tilde{X}_n = X_{(m)}$$

を**メディアン**という．

$f = dF/dx$ とし，$f(0) > 0$ の仮定もとで，$\tilde{X}_n$ の極限分布を求める：

$$P_\theta[\sqrt{n}(\tilde{X}_n - \theta) \le t] = P_0[\sqrt{n}\tilde{X}_n \le t] = P_0\left[X_{(m)} \le \frac{t}{\sqrt{n}}\right]$$

S_n を $t/\sqrt{n}$ を越える $X_i (i = 1, 2, \ldots, n)$ の個数とすれば，

$$X_{(m)} \le \frac{t}{\sqrt{n}} \Leftrightarrow S_n \le m - 1 = \frac{1}{2}(n-1)$$

となる．S_n は二項分布 $Bi(n, p_n)$ に従う．ただし，$p_n = 1 - F(t/\sqrt{n})$ である．したがって，

$$P_\theta\left[S_n \le \frac{1}{2}(n-1)\right] = P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{(1/2)(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right]$$

となる．$p_n \to 1 - F(0) = 1/2 (n \to \infty)$ より，

$$P_0\left[S_n \le \frac{1}{2}(n-1)\right] - \Phi\left(\frac{(1/2)(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right) \to 0$$

となる．ただし，$\Phi(\cdot)$ は標準正規分布の分布関数である．また，$p_n \to 1/2$ より，

$$\begin{aligned}
\frac{(1/2)(n-1) - np_n}{\sqrt{np_n(1-p_n)}} &= \frac{\sqrt{n}\left(\frac{1}{2} - p_n\right) - \frac{1}{2\sqrt{n}}}{\sqrt{p_n(1-p_n)}} \\
&\sim 2\sqrt{n}\left(\frac{1}{2} - p_n\right) \\
&= -2t\frac{F(t/\sqrt{n}) - F(0)}{t/\sqrt{n}} \to 2tf(0)
\end{aligned}$$

となることより，

$$P_\theta[\sqrt{n}(\tilde{X}_n - \theta)] \to \Phi(-2tf(0))$$

となる．したがって，

$$\sqrt{n}(\tilde{X}_n - \theta) \xrightarrow{L} N\left(0, \frac{1}{4f^2(0)}\right)$$

をえる．

問い 1.8　：$X_1, X_2, \ldots, X_n$ を分布関数 F からの実数値ランダム標本とし，f を F の確率密度関数とする．任意の $0 < p < 1$ に対して，ξ_p を

$$F(\xi_p) = p$$

で定義し，$f(\xi_p) > 0$ とする．このとき，

$$\frac{n_1}{n} = p + o\left(\frac{1}{\sqrt{n}}\right)$$

ならば，

$$\sqrt{n}(X_{(n_1)} - \xi_p) \xrightarrow{L} N\left(0, \frac{p(1-p)}{(f(\xi_p))^2}\right)$$

となることを示せ.

1.12.3 従属のある確率変数に対する中心極限定理

定義 1.29　：確率変数列 $\{Y_n\}_{n=1}^{\infty}$ が $m-$ dependent　であるとは，すべての自然数 s に対して，$\{Y_1, \ldots, Y_s\}$ と $\{Y_{m+s+1}, Y_{m+s+2}, \ldots\}$ は独立であることをいう．

定義 1.30　確率変数列 $\{Y_n\}_{n=1}^{\infty}$ が**定常**（stationary）であるとは，ある自然数 s と t にたいして，$(Y_t, Y_{t+1}, \ldots, Y_{t+s})$ の同時分布が t に依存しないことである．

確率変数列 $\{Y_n\}_{n=1}^{\infty}$ を定常な m – dependent 列としたとき，

$$S_n = \sum_{i=1}^{n} Y_i$$

の極限分布を求めることを考える．
記号：

$$\mu = \boldsymbol{E}Y_t, \qquad \sigma_{00} = \boldsymbol{var}(Y_t), \qquad \sigma_{0i} = \boldsymbol{cov}(Y_t, Y_{t+i})$$

と記す．
これより，

$$\boldsymbol{E}S_n = \mu$$

と $n \geq m$ のとき，

$$\begin{aligned}\boldsymbol{var}(S_n) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{cov}(Y_i, Y_j) \\ &= n\sigma_{00} + 2(n-1)\sigma_{01} + 2(n-2)\sigma_{02} + \cdots + 2(n-m)\sigma_{0m}\end{aligned}$$

となる．これより，

$$\frac{1}{n}\boldsymbol{var}(S_n) \to \sigma^2$$

となる．ただし，

$$\sigma^2 = \sigma_{00} + \sigma_{01} + \ldots + \sigma_{0m} \tag{1.31}$$

である.

定理 1.39　：確率変数列 $\{Y_n\}_{n=1}^{\infty}$ を定常な m – dependent 列で有限な分散を持つとする．$S_n = \sum_{i=1}^{n} Y_i$ とする．このとき，

$$\frac{S_n - \boldsymbol{E}(S_n)}{\sqrt{\boldsymbol{var}(S_n)}} \xrightarrow{L} N(0,1)$$

が成立[1)]する．ただし，$\mu = \boldsymbol{E}(Y_1)$，$\sigma^2$ は (1.31) で与えられる．

n が十分大きいとする．S_n を長さ $(k+m)$（$k > m$）の列に分割する．$n = s(k+m) + r$ とし，

$$S_n = S'_n + S''_n + R_n$$

とかく．ただし，

$$S'_n = \sum_{j=0}^{s-1} V_{kj}, \qquad S''_n = \sum_{j=1}^{s-1} W_{kj}, \qquad R_n = \sum_{i=1}^{r} Y_{s(m+k)+i}$$

$$V_{kj} = \sum_{i=1}^{k} Y_{j(k+m)+i}, \qquad W_{kj} = \sum_{i=k+1}^{k+m} Y_{j(k+m)+i}$$

である．$V_{kj}, j = 1, 2, \ldots, s-1$ は k に依存する独立同一の分布に従う確率変数列となる．したがって，S_n の極限分布を求めるためには，S''_n と R_n が確率的に無視できることを示さなければならない．

補題 1.9　：$T_n = Z_{nk} + X_{nk}$ とする．以下の条件が成立するとする．
(i) $k \to \infty$ のとき，n に関して一様に

$$X_{nk} \xrightarrow{P} 0$$

である．
(ii) それぞれの k に対して，$n \to \infty$ のとき，

$$Z_{nk} \xrightarrow{L} Z_k$$

である．
(iii) $k \to \infty$ のとき，

$$Z_k \xrightarrow{L} Z$$

である．このとき，$n \to \infty$ とすれば，

$$T_n \xrightarrow{L} Z$$

が成立する．

[1)] これは書き換えれば，

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{L} N(0, \sigma^2)$$

となる．

証明：確率変数 Z と Z_k の分布関数を F_Z と F_{Z_k} と記し，分布関数 F の連続点の集合を $C(F)$ と書くことにする．任意の正の数 $\epsilon > 0$ と $z \in C(F_Z)$ を取る．適当な正の数 $\delta > 0$ を取り，$z+\delta, z-\delta \in C(F_Z)$ かつ $z+\delta, z-\delta \in C(F_{Z_k})$（$k$ はすべて）で

$$P(|Z - z| < \delta) < \epsilon$$

とできる．条件 (i) より，ある K が存在して，どんな $k > K$ と n に対しても

$$P(|X_{nk}| \geq \delta) < \epsilon$$

とできる．また，条件 (iii) から，$K' \geq K$ が存在して，どんな $k > K'$ に対して

$$|P(Z_k \leq z+\delta) - P(Z \leq z+\delta)| < \epsilon, \qquad |P(Z_k \leq z-\delta) - P(Z \leq z-\delta)| < \epsilon$$

とできる．$K > K'$ を固定する．

$$P(T_n \leq z) = P(Z_{nk} + X_{nk} \leq z) \leq P(Z_{nk} \leq z+\delta) + \epsilon$$

である．いま，条件 (ii) を使えば，

$$\begin{aligned}\limsup_{n\to\infty} P(T_n \leq z) &\leq \lim_{n\to\infty} P(Z_{nk} \leq z+\delta) + \epsilon \\ &= P(Z_k \leq z+\delta) + \epsilon \leq P(Z \leq z+\delta) + 2\epsilon \\ &\leq P(Z \leq z) + 3\epsilon\end{aligned}$$

となる．同様な議論より，

$$P(T_n \leq z) = P(Z_{nk} + X_{nk} \leq z) \geq P(Z_{nk} \leq z-\delta) - P(|X_{nk}| \geq \delta) \geq P(Z_{nk} \leq z-\delta) - \epsilon$$

となるから，

$$\begin{aligned}\liminf_{n\to\infty} P(T_n \leq z) &\geq P(Z_k \leq z-\delta) - \epsilon \\ &\geq P(Z \leq z-\delta) - 2\epsilon \geq P(Z \leq z) - 3\epsilon\end{aligned}$$

を得る．これらより

$$P(Z \leq z) - 3\epsilon \leq \liminf_{n\to\infty} P(T_n \leq z) \leq \limsup_{n\to\infty} P(T_n \leq z) \leq P(Z \leq z) + 3\epsilon$$

が成り立つことがわかる．よって，補題は証明された．□

この補題を利用して，定理を証明する．
定理の証明：一般性を失わず，$\boldsymbol{E}Y_1 = 0$ とできる．

$$T_n = \frac{S_n}{\sqrt{n}}, \qquad Z_{nk} = \frac{S'_n + R_n}{\sqrt{n}}, \qquad X_{nk} = \frac{S''_n}{\sqrt{n}}$$

とおく．中心極限定理を使えば，固定した k に対して，

$$\frac{S'_n}{\sqrt{s}} \xrightarrow{L} N(0, \boldsymbol{var}(S_k))$$

が成り立つことがわかる．さらに，$s/n \to 1/k$ より

$$\frac{S'_n}{\sqrt{n}} = \left(\frac{\sqrt{s}}{\sqrt{n}}\right)\frac{S'_n}{\sqrt{s}} \xrightarrow{L} N(0, \frac{\boldsymbol{var}(S_k)}{k})$$

となり，補題の条件 (ii) が満たされることがわかる．$|\sigma_{0i}| \leq \sigma_{00}$ に注意すれば，固定した k に対して，$n \uparrow \infty$ のとき，

$$\boldsymbol{var}(\frac{R_n}{\sqrt{n}}) \leq \frac{r^2\sigma_{00}}{n} \leq \frac{(k+m)^2\sigma_{00}}{n} \to 0$$

となる．したがって，$\boldsymbol{E}(R_n/\sqrt{n}) = 0$ から

$$\frac{R_n}{n} \xrightarrow{P} 0$$

となる．よって，

$$Z_{nk} = \frac{1}{\sqrt{n}}(S'_n + R_n) \xrightarrow{L} N(0, \frac{\boldsymbol{var}(S_k)}{k}) = \boldsymbol{L}(Z_k)$$

となる[2)]．さらに，$k \uparrow \infty$ のとき，$\boldsymbol{var}(S_k)/k \to \sigma^2$ なので，$Z_k \xrightarrow{L} Z \sim N(0, \sigma^2)$ が成立する．したがって，補題の (iii) が満たされる．

最後に，(i) を確認する．

$$\boldsymbol{var}(X_{nk}) = \frac{s}{n}\boldsymbol{var}(S_m) < \frac{1}{k}\boldsymbol{var}(S_m)$$

より，n について一様に

$$P(|X_{nk}| > \delta) \leq \frac{1}{\delta^2}\boldsymbol{var}(X_{nk}) \leq \frac{1}{k\delta^2}\boldsymbol{var}(S_m) \longrightarrow 0, \qquad (k \to \infty)$$

が成立することからわかる．よって，補題を適用すれば，定理は示される． □

2) $\boldsymbol{L}(Z_k)$ は Z_k の分布を示す．

2

統計モデルと統計推測の枠組み

·····2.1·····

統計モデル

ランダムな試行を考える．試行の可能な結果の集まりを Ω と記し、これを標本空間と呼ぶ．標本空間上に確率ベクトル $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を定義する[1]．$\omega\,(\omega \in \Omega)$ を試行の結果としたとき、$\boldsymbol{X}(\omega)$ を観測値、実現値または標本と呼ぶ．観測できるのはランダム試行の結果である $\boldsymbol{X}(\omega)$ だけなので、$\boldsymbol{X}(\omega)$ の確率分布を考える必要がある．この確率分布は $\mathbb{R}^n$ 上の確率分布のある族 $\mathcal{P}$ の属していると仮定する．このとき、$\mathcal{P}$ を $\boldsymbol{X}(\omega)$ の**統計モデル**という．与えられた観測値は $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上のある（未知の）確率分布にしたがう確率変数 $\boldsymbol{X}$ の観測値とみなし，その観測値に基づきその確率分布についてのなんらかの判断を行うことを**統計的推測**という．

統計モデル $\mathcal{P}$ を記述するために、統計モデルに属する確率分布に添え字をつけることを考えよう．すなわち、添え字の空間 Θ から統計モデル $\mathcal{P}$ への写像 $\theta \to P_\theta\,(\theta \in \Theta)$ を考え，

$$\mathcal{P} = \{P_\theta \text{ は } \boldsymbol{X} \text{ の確率分布} : \theta \in \Theta\}$$

とする．このとき、θ を母数と呼び，Θ のことを母数空間と呼ぶ．

例 2.1　計測がスカラー値の n 個の観測 $x_1, x_2, \ldots, x_n$ を独立同一に未知の確率分布関数 F に従う確率変数 $X_1, X_2, \ldots, X_n$ の実現値としてモデル化することを考える．このことを

$$\mathbf{X} \equiv (X_1, X_2, \ldots, X_n) \sim P_\theta, \qquad \theta \in \Theta$$

と表記することにする．

(a)　φ を標準正規分布の確率密度関数とする．$\boldsymbol{X}$ の確率分布 $P_\theta\,(\theta = (\mu, \sigma))$ は確率密度関数 $\sigma^{-n}\Pi_{i=1}^n \varphi((x_i - \mu)/\sigma)$ を持ち，母数空間は $\Theta_1 = \mathbb{R} \times \mathbb{R}^+$ で $\mathcal{P}_1 = \{P_\theta : \theta \in \Theta_1\}$ とする．

(b)　母数空間を

$$\Theta_2 = \{(\mu, G) : \mu \in \mathbb{R}, G \text{ は確率密度関数 } g \text{ を持ち，} \textstyle\int xg(x)\,dx = 0 \text{ をみたす．}\}$$

とし，確率分布 P_θ は累積分布関数 $\Pi_{i=1}^n G(x_i - \mu)$ をもち，$\mathcal{P}_2 = \{P_\theta : \theta \in \Theta_2\}$ とする．

特に，$X_1, X_2, \ldots, X_n$ が互いに独立に同一の確率分布 P にしたがうとき，$X_1, X_2, \ldots, X_n$ は分布 P からの**大きさ n のランダム標本**という．

例 2.1 の (a) のように母数空間がユークリッド空間の部分空間で，母数 θ がわかれば，その確率分布が完全に特定できる場合，その統計モデルをパラメトリックモデルと呼ぶ．(b) のようなモデルをノンパラメトリックモデルという．

[1] 確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の確率変数を $X_1, X_2, \ldots, X_n$ とする．

(b) において，母数空間を

$$\Theta_3 = \{(\mu, G) : \mu \in \mathbb{R}, G \text{ は確率密度関数 } g \text{ を持つ.}\}$$

とする．このとき，$\mu_1 = 0, \mu_2 = 1$ とし，G_1, G_2 は確率密度関数 $\varphi(x), \varphi(x+1)$ を持つとすれば，$P_{(\mu_1, G_1)}$ と $P_{(\mu_2, G_2)}$ は標準正規分布となる．一般に，$\theta_1, \theta_2 \in \Theta$ に対し，$\theta_1 \neq \theta_2$ であるにも関わらず，$P_{\theta_1} = P_{\theta_2}$ であるとき，この母数化は**認定不可能**という．逆に，$\theta_1 \neq \theta_2$ ならば，$P_{\theta_1} \neq P_{\theta_2}$ であるとき，この母数化は**認定不可能**という．以後は認定可能な母数化のみを考える．

·····2.2·····

指数分布族

P_θ を $\mathcal{X} \subset \mathbb{R}^n$ 上の確率測度とし，$\{P_\theta : \theta \in \Theta \subset \mathbb{R}^l\}$ を確率分布族とする．また，$\lambda(x)$ を $\mathcal{X}$ 上の確率測度とする．

定義 2.1　確率測度族 $\{P_\theta : \theta \in \Theta \subset \mathbb{R}^l\}$ が k 母数指数分布族であるとは，Θ 上の関数 $\eta_1(\theta), \eta_2(\theta), \ldots, \eta_k(\theta)$ と $\tilde{A}(\theta)$，および $\mathcal{X}$ 上の関数 $T_1(\boldsymbol{x}), T_2(\boldsymbol{x}), \ldots, T_k(\boldsymbol{x})$ と $S(\boldsymbol{x})$) が存在して，P_θ の ($\mathbb{R}^n$ 上の測度 $\lambda(\boldsymbol{x})$ に関する) 確率密度関数 $p(\boldsymbol{x}|\boldsymbol{\theta})$ がつぎで与えられるとき， k 母数指数分布族になるという．

$$p(\boldsymbol{x}|\,\boldsymbol{\theta}) = S(\boldsymbol{x}) \exp\left[\sum_{i=1}^{k} \eta_i(\theta) T_i(\boldsymbol{x}) - \tilde{A}(\theta)\right] I_B(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n \tag{2.32}$$

と書けることをいう．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in B$ とし，B は未知の母数 $\boldsymbol{\theta}$ に依存しないとする．

例 2.2　：X は母数 n と未知の母数 θ の 2 項分布にしたがうとする．すなわち，X の確率関数は

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{1-x}$$

で与えられる．ただし，$x \in B = \{0, 1, \ldots, n\}$ である．これは

$$p(x|\theta) = \exp\left[\log\left(\frac{\theta}{1-\theta}\right) x + n \log(1-\theta) + \log\binom{n}{x}\right]$$

となり，1 母数指数分布族となる．

例 2.3　：$X_1, X_2, \ldots, X_n$ を未知の母数 (α, β) を持つガンマ分布からのランダム標本とする．ただし，$\alpha > 0, \beta > 0$ である．すなわち，その確率密度関数

$$p(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}, \qquad x \in (0, \infty)$$

からのランダム標本である．ただし，

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$

である．このとき，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$\begin{aligned} p(\boldsymbol{x}|\alpha, \beta) &= \prod_{i=1}^{n} \left[\frac{\beta^\alpha x_i^{\alpha-1} \exp(-\beta x_i)}{\Gamma(\alpha)} \right] \\ &= \exp\left[(\alpha-1) \sum_{i=1}^{n} \log x_i - \beta \sum_{i=1}^{n} x_i + n\alpha \log \beta - n \log \Gamma(\alpha) \right] \end{aligned}$$

となる．ただし，$B = (0, \infty)^n$ である．したがって，$\boldsymbol{X}$ の確率密度関数は 2 母数指数分布族となる．

例 2.4　：$P_\theta = N(\mu, \sigma^2), \theta \in \Theta$ である．ただし，

$$\Theta = \{(\mu, \sigma^2):\ -\infty < \mu < \infty, \sigma^2 > 0\}$$

である．P_θ の確率密度関数は

$$p(\boldsymbol{x}) = \exp\left[\frac{\mu}{\sigma^2} x - \frac{x^2}{2\sigma^2} - \frac{1}{2} \left(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2) \right) \right]$$

である．したがって，正規分布族は 2 母数指数分布族である．すなわち，(2.32) において

$$\begin{aligned} &n = 1, \quad \theta_1 = \mu, \quad \theta_2 = \sigma^2, \quad \eta_1(\theta) = \tfrac{\mu}{\sigma^2}, \quad T_1(x) = x, \\ &\eta_2(\theta) = -\tfrac{1}{2\sigma^2}, \quad T_2(x) = x^2, \quad B(\theta) = \tfrac{1}{2}\left(\tfrac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2) \right), \quad S(x) = 1 \end{aligned}$$

である．

指数分布族が $(\eta_1, \eta_2, \ldots, \eta_k)'$ で添え字つけられているとしよう．$\boldsymbol{T}$ と S で生成される指数分布族を

$$f(\boldsymbol{x}|\eta) = S(x) \exp[\boldsymbol{T}(\boldsymbol{x})\eta - A(\eta)], \qquad \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n \tag{2.33}$$

と書くことにする．ただし，

$$A(\eta) = \begin{cases} \log\{\int [S(\boldsymbol{x}) \exp\{\boldsymbol{T}(\boldsymbol{x})'\eta\}]\, d\mu(\boldsymbol{x})\} & (\text{連続型}) \\ \log\{\sum_x [S(\boldsymbol{x}) \exp\{\boldsymbol{T}(\boldsymbol{x})'\eta\}]\} & (\text{離散型}) \end{cases}$$

である．このモデルの自然母数空間を

$$\mathcal{E} = \{\eta \in \mathbb{R}^k:\ -\infty < A(\eta) < \infty\}$$

で定義する．これを正準 k 指数分布族という．

例 2.5　(正規分布族の続き)

$$\begin{aligned} &k = 2, \quad \boldsymbol{T} = (T_2(\boldsymbol{x}), T_2(\boldsymbol{x}))' = (x, x^2), \quad \eta_1 = \mu/\sigma^2, \\ &\eta_2 = -1/(2\sigma^2) \quad A(\eta) = \tfrac{1}{2}\left[-\tfrac{\eta_1^2}{2\eta_2} + \log\left(\tfrac{\pi}{-\eta_2} \right) \right] \end{aligned}$$

である．したがって，

$$\mathcal{E} = \{(\eta_1, \eta_2):\ -\infty < \eta_1 < \infty, -\infty < \eta_2 < 0\}$$

である．

例 2.6　(線形回帰モデル)　$Y_1, Y_2, \ldots, Y_n$ は互いに独立で各 $Y_i,\ i = 1, 2, \ldots, n$ は $N(\beta_1 + z_i\beta_2,\ \sigma^2)$ に従うとする．$z_1, z_2, \ldots, z_n$ は説明変数という．$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ の同時確率密度関数は

$$
\begin{aligned}
p(\boldsymbol{y}|\,\boldsymbol{\beta},\ \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\Pi_{i=1}^n \exp\left[-\frac{(y_i-\beta_1-z_i\beta_2)^2}{2\sigma^2}\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\beta_1}{\sigma^2}\sum_{i=1}^n y_i + \frac{\beta_2}{\sigma^2}\sum_{i=1}^n z_i y_i - \frac{1}{2\sigma^2}\sum_{i=1}^n(\beta_1+z_i\beta_2)^2 - \frac{n}{2}\log(2\pi\sigma^2)\right]
\end{aligned}
$$

である．ただし，$\boldsymbol{\beta} = (\beta_1,\ \beta_2)$ である．これより

$$
\begin{aligned}
&k = 3, \qquad \boldsymbol{T}(\boldsymbol{Y}) = \left(\textstyle\sum_{i=1}^n Y_i,\ \sum_{i=1}^n Y_i^2,\ \sum_{i=1}^n z_iY_i\right)', \qquad \eta_1 = \tfrac{\beta_1}{\sigma^2} \quad \eta_2 = \tfrac{\beta_2}{\sigma^2}, \\
&\eta_3 = -\tfrac{1}{(2\sigma^2)} \quad A(\eta) = -\tfrac{n}{4\eta_3}\left[\eta_1^2 + \hat{m}_2\eta_2^2 + \bar{z}\eta_1\eta_2 + 2\log\left(\tfrac{\pi}{-\eta_3}\right)\right]
\end{aligned}
$$

となる．ただし，$\bar{z} = (1/n)\sum_{i=1}^n z_i,\ \hat{m}_2 = n^{-1}\sum_{i=1}^n z_i^2$ である．したがって，

$$
\mathcal{E} = \{(\eta_1,\ \eta_2,\ \eta_3) :\ -\infty < \eta_1 < \infty,\ -\infty < \eta_2 < \infty,\ -\infty < \eta_3 < 0\}
$$

である．

2.2.1 指数分布族の性質

$\boldsymbol{T}$ の積率簿関数を

$$
M_{\boldsymbol{T}}(\boldsymbol{s}) = \mathbb{E}[e^{\boldsymbol{s}'\boldsymbol{T}}]
$$

とかく．ただし，$\boldsymbol{s}' = (s_1,\ s_2,\ \ldots,\ s_k)$ である．また，

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{T}] &= (\mathbb{E}(T_1),\ \mathbb{E}(T_2),\ \ldots, \mathbb{E}(T_k))' \\
\mathbb{VAR}[\boldsymbol{T}] &= \begin{bmatrix} \mathbb{COV}(T_1,\ T_1) & \mathbb{COV}(T_1,\ T_2) & \ldots & \mathbb{COV}(T_1,\ T_k) \\ \vdots & & \ddots & \vdots \\ \mathbb{COV}(T_k,\ T_1) & \mathbb{COV}(T_k,\ T_2) & \ldots & \mathbb{COV}(T_k,\ T_k) \end{bmatrix}
\end{aligned}
$$

と書く．

定理 2.1　$\mathcal{P}$ を (2.33) で与えらる正準 k 母数指数分布族とする．このとき，
(i) $\mathcal{E}$ は convex である．
(ii) $A : \mathcal{E} \to \mathbb{R}$ は convex である．
(iii) $\mathcal{E}$ は $\mathbb{R}^k$ の空でない内部集合を含み，$\boldsymbol{\eta}$ を $\mathcal{E}$ の内点とすれば，$\boldsymbol{\eta}$ のもとで $\boldsymbol{T}(\boldsymbol{X})$ の積率母数関数は，$\boldsymbol{\eta} + \boldsymbol{s} \in \mathcal{E}$ なるすべての $\boldsymbol{s}$ に対して

$$
M_{\boldsymbol{T}}(\boldsymbol{s}) = \exp\{A(\boldsymbol{\eta}+\boldsymbol{s}) - A(\boldsymbol{\eta})\}
$$

で与えらる．$\boldsymbol{\eta}$ は $\mathcal{E}$ の内部なので，$\{\boldsymbol{s} :\ \boldsymbol{\eta} + \boldsymbol{s} \in \mathcal{E}\}$ は原点を含むある球を含む．

証明：まず，(ii) を示す．$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{E}$ と $0 < \alpha < 1$ を取る．Hölder の不等式を用いる：$u(\boldsymbol{x}), v(\boldsymbol{x}), h(\boldsymbol{x}) \geq 0$, $r, s > 0$, $1/r + 1/s = 1$ に対して

$$\int u(\boldsymbol{x})v(\boldsymbol{x})h(\boldsymbol{x})\,d\boldsymbol{x} \leq \left\{\int u^r(\boldsymbol{x})h(\boldsymbol{x})\,d\boldsymbol{x}\right\}^{1/r} \left\{\int v^s(\boldsymbol{x})h(\boldsymbol{x})\,d\boldsymbol{x}\right\}^{1/s}$$

ここで

$$\frac{1}{r} = \alpha, \quad \frac{1}{s} = 1-\alpha, \quad u(\boldsymbol{x}) = \exp[\alpha\boldsymbol{\eta}_1'\boldsymbol{T}], \quad v(\boldsymbol{x}) = \exp[(1-\alpha)\boldsymbol{\eta}_2'\boldsymbol{T}], \quad h(\boldsymbol{x}) = S(\boldsymbol{x})$$

とおき，Hólder の不等式の両辺に対数をとれば，

$$A(\alpha\boldsymbol{\eta}_1 + (1-\alpha)\boldsymbol{\eta}_2)) \leq \alpha A(\boldsymbol{\eta}_1) + (1-\alpha)A(\boldsymbol{\eta}_2) \tag{2.34}$$

より (ii) は示せた．

また，$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{E}$ ならば，(2.34) より $\alpha\boldsymbol{\eta}_1 + (1-\alpha)\boldsymbol{\eta}_2 \in \mathcal{E}$ となる．また，すべての $\boldsymbol{\eta}$ に対して $\int \exp(\boldsymbol{\eta}'\boldsymbol{T}(x))S(x)\,dx > 0$ より $\log\mathbb{E}[S(\boldsymbol{X})\exp\{\boldsymbol{\eta}'\boldsymbol{T}(\boldsymbol{X})] > \infty$ となり，(i) も示せた．

連続型の場合について (iii) を示そう．

$$\begin{aligned}
M_{\boldsymbol{T}}(\boldsymbol{s}) &= \mathbb{E}[\exp\{\boldsymbol{s}'\boldsymbol{T}(\boldsymbol{X})\}] \\
&= \int\cdots\int \boldsymbol{S}(\boldsymbol{x})\exp\{(\boldsymbol{s}+\boldsymbol{\eta})\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\eta})\}\,dx_1\cdots d\boldsymbol{x}_q \\
&= \exp[A(\boldsymbol{s}+\boldsymbol{\eta}) - A(\boldsymbol{\eta})]\int\cdots\int \boldsymbol{s}(\boldsymbol{x})\exp\{(\boldsymbol{s}+\boldsymbol{\eta})\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{s}+\boldsymbol{\eta})\}\,dx_1\cdots dx_q \\
&= \exp[A(\boldsymbol{s}+\boldsymbol{\eta}) - A(\boldsymbol{\eta})]
\end{aligned}$$

□

系 2.1　：$\mathcal{E}$ は $\mathbb{R}^k$ の空でない内部集合を含み，$\boldsymbol{\eta}$ を $\mathcal{E}$ の内点とする．このとき

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\eta}}[\boldsymbol{T}(\boldsymbol{X})] &= \dot{A}(\boldsymbol{\eta}) \\
\mathbb{VAR}_{\boldsymbol{\eta}}[\boldsymbol{T}(\boldsymbol{X})] &= \ddot{A}(\boldsymbol{\eta})
\end{aligned}$$

となる．ただし，

$$\begin{aligned}
\dot{A}(\boldsymbol{\eta}) &= \left(\frac{\partial A}{\partial\eta_1}(\boldsymbol{\eta}), \cdots, \frac{\partial A}{\partial\eta_k}(\boldsymbol{\eta})\right)' \\
\ddot{A}(\boldsymbol{\eta}) &= \left(\frac{\partial^2 A}{\partial\eta_i\partial\eta_j}(\boldsymbol{\eta})\right)_{i=1,2,\ldots,k,\,j=1,2,\ldots,k} \\
\boldsymbol{\eta} &= (\eta_1, \ldots, \eta_k)'
\end{aligned}$$

である．

証明：定理 2.1 (iii) と積率母関数の性質

$$\mathbb{E}[T_j(\boldsymbol{X})] = M_{\boldsymbol{T}}(\boldsymbol{s})\frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})\Big|_{\boldsymbol{s}=\boldsymbol{0}} = \frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})\Big|_{\boldsymbol{s}=\boldsymbol{0}}, \qquad j=1,\,2\,\ldots,\,k$$

$$\begin{aligned}\mathbb{E}[T_j(\boldsymbol{X})T_i(\boldsymbol{X})] &= \Big[M_{\boldsymbol{T}}(\boldsymbol{s})\frac{\partial^2}{\partial s_j \partial s_i}A(\boldsymbol{s}+\boldsymbol{\eta}) + M_{\boldsymbol{T}}(\boldsymbol{s})\frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})M_{\boldsymbol{T}}(\boldsymbol{s})\frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})\Big]_{\boldsymbol{s}=\boldsymbol{0}} \\ &= \Big[\frac{\partial^2}{\partial s_j \partial s_i}A(\boldsymbol{s}+\boldsymbol{\eta}) + \frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})\frac{\partial}{\partial s_j}A(\boldsymbol{s}+\boldsymbol{\eta})\Big]_{\boldsymbol{s}=\boldsymbol{0}}, \qquad i,\,j=1,\,2\,\ldots,\,k\end{aligned}$$

□

例 2.7　可能な結果が k 個のカテゴリーの実験の n 回の独立試行を考える．結果に対応する確率ベクトルを $\boldsymbol{X} = (X_1,\,X_2,\,\ldots,\,X_n)$ とおく．ただし，$X_1,\,X_2,\,\ldots,\,X_n$ は独立同一の確率変数列で，各 $X_i,\,i=1,\,2,\,\ldots,\,n,$ は k 個のカテゴリー $\{1,\,2,\,\ldots,\,k\}$ のどれかをとるものとする．いま，

$$T_j(\boldsymbol{x}) = \sum_{i=1}^{n} 1\{X_i = j\},\ \lambda_j = \mathbb{P}(X_1 = j), \qquad j=1,\,2,\,\ldots,\,k$$

とおく．

このとき，確率関数は

$$p(\boldsymbol{x}|\,\boldsymbol{\lambda}) = \Pi_{j=1}^{k}\lambda_j^{T_j(\boldsymbol{x})}, \qquad \boldsymbol{\lambda}\in\Lambda$$

と書ける．ただし，

$$\Lambda = \{\boldsymbol{\lambda} = (\lambda_1,\,\lambda_2,\,\ldots,\,\lambda_k)\in\mathbb{R}^k :\, 0<\lambda_j<1,\, \sum_{j=1}^{k}\lambda_j = 1,\, j=1,\,2,\,\ldots,\,k\}$$

とする．制限なしの母数空間を考える．そのために

$$\lambda_j = \frac{e^{\alpha_j}}{\sum_{j=1}^{k} e^{\alpha_j}}, \qquad j=1,\,2,\,\ldots,\,k, \qquad \boldsymbol{\alpha} = (\alpha_1,\,\alpha_2,\,\ldots,\,\alpha_k)'\in\mathbb{R}^k$$

とおく．

このとき，確率関数は

$$p_1(\boldsymbol{x}|\,\boldsymbol{\alpha}) = \exp\big\{\sum_{j=1}^{k}\alpha_j T_j(\boldsymbol{x}) - n\log\sum_{j=1}^{k}e^{\alpha_j}\big\}$$

となる．しかし，この母数化は認定不可能性を持つので，さらに

$$\boldsymbol{T}_{(k-1)}(\boldsymbol{x}) = (T_1(\boldsymbol{x}),\,T_2(\boldsymbol{x}),\,\ldots,T_{k-1}(\boldsymbol{x}))', \qquad \eta_j = \log\frac{\lambda_j}{\lambda_k} = \alpha_j - \alpha_k, \qquad j=1,\,2,\,\ldots,k-1$$

とおく．$\sum_{i=1}^{k}T_i(\boldsymbol{x}) = n$ と $\sum_{i=1}^{k}\lambda_i = 1$ に注意すれば，確率関数は

$$p_2(\boldsymbol{x}|\,\boldsymbol{\eta}) = \exp\big\{\boldsymbol{T}'_{(k-1)}(\boldsymbol{x})\boldsymbol{\eta} - n\log(1+\sum_{j=1}^{k-1}e^{\eta_j})\big\}$$

となる．ただし，

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{k-1}), \qquad \lambda_j = \frac{e^{\eta_j}}{1+\sum_{j=1}^{k-1} e^{\eta_j}} = \frac{e^{\alpha_j}}{\sum_{j=1}^{k} e^{\alpha_j}}, \qquad j = 1, 2, \ldots, k-1$$

である．したがって，$p_2(\boldsymbol{x}|\boldsymbol{\eta})$ は $T_{(k-1)}(\boldsymbol{x})$ と $S(\boldsymbol{x})$ で生成される $k-1$ 母数指数分布族で，自然母数空間は $\mathcal{E} = \mathbb{R}^{k-1}$ となる．

また，

$$A(\boldsymbol{\eta}) = n \log(1 + \sum_{j=1}^{k-1} e^{\eta_j})$$

に注意して，系 2.1 を用いれば，

$$\mathbb{E}_{\boldsymbol{\eta}} T_j(\boldsymbol{X}) = \frac{\partial A}{\partial \eta_j} = n \frac{e^{\eta_j}}{1+\sum_{j=1}^{k-1} e^{\eta_j}} = n\lambda_j, \quad j = 1, 2, \ldots, k-1$$

となる．さらに，$1 \le j_1, j_2 \le k-1 (j_1 \ne j_2)$ に対し，

$$\mathbb{COV}_{\boldsymbol{\eta}}(T_{j_1}(\boldsymbol{X}), T_{j_2}(\boldsymbol{X})) = \frac{\partial^2 A}{\partial \eta_{j_1} \partial \eta_{j=2}} = -\frac{e^{\eta_{j_1}} e^{\eta_{j_2}}}{\{1+\sum_{j=1}^{k-1} e^{\eta_j}\}^2} = -n\lambda_{j_1}\lambda_{j_2},$$

$$\mathbb{VAR}_{\boldsymbol{\eta}}[T_{j_1}(\boldsymbol{X})] = \frac{\partial^2 A}{\partial \eta_{j_1}^2} = n\left[\frac{e^{\eta_{j_1}}}{1+\sum_{j=1}^{k-1} e^{\eta_j}} - \frac{e^{\eta_{j_1}^2}}{\{1+\sum_{j=1}^{k-1} e^{\eta_j}\}^2}\right] = n\lambda_{j_1}(1-\lambda_{j_1})$$

となる．最後に，$\mathbb{E}_{\boldsymbol{\eta}}[T_k(\boldsymbol{X})]$, $\mathbb{COV}_{\boldsymbol{\eta}}(T_j(\boldsymbol{X}), T_k(\boldsymbol{X}))$, $j = 1, 2, \ldots, k-1$, $\mathbb{VAR}_{\boldsymbol{\eta}}[T_k(\boldsymbol{X})]$ を求める．これらは，$\sum_{j=1}^{k} T_j(\boldsymbol{X}) = n$ と $\sum_{j=1}^{k} \lambda_j = 1$ を注意すれば，

$$\mathbb{E}_{\boldsymbol{\eta}}[T_k(\boldsymbol{X})] = \mathbb{E}_{\boldsymbol{\eta}}[n - \sum_{j=1}^{k-1} T_j(\boldsymbol{X})] = n(1 - \sum_{j=1}^{k-1} \lambda_j) = n\lambda_k$$

と

$$\begin{aligned}
\mathbb{COV}_{\boldsymbol{\eta}}(T_j(\boldsymbol{X}), T_k(\boldsymbol{X})) &= \mathbb{COV}_{\boldsymbol{\eta}}(T_j(\boldsymbol{X}), n - \sum_{i=1}^{k-1} T_i(\boldsymbol{X})) \\
&= -\sum_{i \ne j}^{k-1} \mathbb{COV}_{\boldsymbol{\eta}}(T_j(\boldsymbol{X}), T_i(\boldsymbol{X})) + \mathbb{VAR}_{\boldsymbol{\eta}}[T_j(\boldsymbol{X})] \\
&= n\big[\sum_{i \ne j}^{k-1} \lambda_j \lambda_i - \lambda_j(1-\lambda_j)\big] \\
&= -n\lambda_j\big[1 - \sum_{i=1}^{k-1} \lambda_i\big] = -n\lambda_j\lambda_k
\end{aligned}$$

となることがからわかる．また，$T_k(\boldsymbol{X})$ の分散は

$$\begin{aligned}\mathbb{VAR}_{\boldsymbol{\eta}}(T_k(\boldsymbol{X})) &= \mathbb{COV}_{\boldsymbol{\eta}}(T_k(\boldsymbol{X}), T_k(\boldsymbol{X}))\\ &= \mathbb{COV}_{\boldsymbol{\eta}}(T_k(\boldsymbol{X}), n - \sum_{i=1}^{k-1} T_i(\boldsymbol{X}))\\ &= -\sum_{i=1}^{k-1} \mathbb{COV}_{\boldsymbol{\eta}}(T_k(\boldsymbol{X}), T_i(\boldsymbol{X}))\\ &= n\lambda_k \sum_{i=1}^{k-1} \lambda_j = n\lambda_k(1-\lambda_k)\end{aligned}$$

よりわかる．

指数分布族の階数（rank）

指数分布族の階数が k であるとは，指数分布族 (2.33) を生成する $\boldsymbol{T}(\boldsymbol{x})$ は k - 次元で，$1, T_1(\boldsymbol{x}), T_2(\boldsymbol{x}), \ldots, T_k(\boldsymbol{x})$ は正の確率で線形独立であることをいう．すなわち，すべてのスカラー a_j, $j = 1, 2, \ldots, k+1$, がゼロでないとき，$P_{\boldsymbol{\eta}}(\sum_{j=1}^{k} a_j T_j(\boldsymbol{x}) = a_{k+1}) < 1$ となる．

定理 2.2　$\mathcal{P} = \{p(\boldsymbol{x}, \boldsymbol{\eta}) : \boldsymbol{\eta} \in \mathcal{E}\}$ は (2.33) で与えられる k 母数指数分布族とし，自然母数空間 $\mathcal{E}$ は集合とする．このとき，次は同値である．
(i) $\mathcal{P}$ の階数は k
(ii) $\boldsymbol{\eta}$ は認定可能な母数
(iii) $\mathbb{VAR}_{\boldsymbol{\eta}}(\boldsymbol{T})$ は正定値
(iv) $\boldsymbol{\eta} \to \dot{A}(\boldsymbol{\eta})$ は $\mathcal{E}$ 上で 1 対 1
(v) A は $\mathcal{E}$ 上で厳密に凸

証明[1]：$\underline{\text{(iii)} \Rightarrow \text{(i)}}$　(i) が成立しないと仮定する．すると，ある $\boldsymbol{a} \neq \boldsymbol{0}$ と c が存在して，ある $\boldsymbol{\eta}$ に対し，

$$\mathbb{P}_{\boldsymbol{\eta}}[\boldsymbol{a}'\boldsymbol{T} = c] = 1$$

が成立する．これより

$$\boldsymbol{a}'\mathbb{VAR}_{\boldsymbol{\eta}}[\boldsymbol{T}]\boldsymbol{a} = \mathbb{VAR}_{\boldsymbol{\eta}}[\boldsymbol{a}'\boldsymbol{T}] = 0$$

となる．したがって，$\mathbb{VAR}_{\boldsymbol{\eta}}[T]$ は正定値でなるなる．よって，(iii) $\Rightarrow$ (i) は示された．
$\underline{\text{(i)} \Rightarrow \text{(iii)}}$　(iii) が成立しないと仮定する．ある $\boldsymbol{\eta}$ において，ある $\boldsymbol{a}$ と c が存在して，

$$0 = \boldsymbol{a}'\mathbb{VAR}_{\boldsymbol{\eta}}[\boldsymbol{T}]\boldsymbol{a} = \mathbb{VAR}_{\boldsymbol{\eta}}[\boldsymbol{a}'\boldsymbol{T}]$$

となり，

$$\mathbb{P}_{\boldsymbol{\eta}}[\boldsymbol{a}'\boldsymbol{T} = c] = 1$$

[1] 証明は不完全：(i) $\Leftrightarrow$ (iii)，(i) $\Rightarrow$ (ii)，(iii) $\Rightarrow$ (iv)
と (iii) $\Rightarrow$ (v)　のみ示した．

となり，(i) $\Rightarrow$ (iii) は示せた．
(i) $\Rightarrow$ (ii)　まず，$k=1$ の場合に示す．(ii) が成立しないと仮定するとある $\eta_1 \neq \eta_2$ が存在し，$\mathbb{P}_{\eta_1} = \mathbb{P}_{\eta_2}$ をみたす．これは

$$\exp\{\eta_1 T(\boldsymbol{x}) - A(\eta_1)\}S(\boldsymbol{x}) = \exp\{\eta_2 T(\boldsymbol{x}) - A(\eta_2)\}S(\boldsymbol{x})$$

となり，両辺に対数をとれば，

$$(\eta_1 - \eta_2)T(\boldsymbol{x}) = A(\eta_2) - A(\eta_1)$$

となり，$1, T(\boldsymbol{x})$ は 1 次独立でなくなる．つぎに，$k>1$ の場合を考える．同様に，(ii) が成立しないと仮定するとある．このとき，$\boldsymbol{\eta}_1 \neq \boldsymbol{\eta}_2$ が存在し，$\mathbb{P}_{\boldsymbol{\eta}_1} = \mathbb{P}_{\boldsymbol{\eta}_2}$ をみたす．

$$\mathcal{Q} = \{P_{\boldsymbol{\eta}_1 + c(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)} : \boldsymbol{\eta}_1 + c(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \in \mathcal{E}\}$$

とおくと $\mathcal{Q}$ は $(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)'\boldsymbol{T}\boldsymbol{X}$ で生成される 1 母数指数分布族になる．$c_1 = 1$ と $c_2 = 0$ のとき，ふたつの分布は一致するので，$k=1$ の議論を用いることができる．よって，$(c_1 - c_2)(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)'\boldsymbol{T}(\boldsymbol{x})$ は定数となり，$1, T_1(\boldsymbol{x}), T_2(\boldsymbol{x}), \ldots, T_k(\boldsymbol{x})$ は 1 次独立でなくなる．
(iii) $\Rightarrow$ (iv) と (v)　系 2.1 より $\ddot{A}(\boldsymbol{\eta})$ は正定値となる．したがって，$\dot{A}(\boldsymbol{\eta})$ は狭義単調増加となり，(iv) が示せた．
(iv) $\Rightarrow$ (iii)　□

·····2.3·····

十分統計量

定義 2.2　：$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ を確率分布族とし，θ が未知の確率分布 P_θ からの標本を $\boldsymbol{X}$ とする．統計量 $T(\boldsymbol{X})$ が $\theta \in \Theta$ に対し**十分**であるとは，T が与えられたときにの $\boldsymbol{X}$ の条件付き分布が θ に依存せず既知となることである．

例 2.8　：$X_1, X_2, \ldots, X_n$ を二項分布

$$f_\theta(z) = \theta^z(1-\theta)^{1-z}1_{\{0,1\}}(z), \qquad \theta \in (0,1)$$

からのランダム標本とし，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とする．$T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ なる統計量は，直観から θ に対する情報をすべて含んでいると予想される．したがって，十分統計量となることが予想される．これを示すために，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とし，

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P(T = t)}$$

が $\theta \in (0,1)$ に依存しないことを示せばよい．

$$P(T = t) = \begin{pmatrix} n \\ t \end{pmatrix} \theta^t(1-\theta)^{n-t}1_{\{0,1,\ldots,n\}}(t)$$

と

$$\begin{aligned} P(\boldsymbol{X} = \boldsymbol{x}, T = t) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}1_{\{0,1\}}(x_i) \\ &= \theta^t(1-\theta)^{1-t}\prod_{i=1}^n 1_{\{0,1\}}(x_i) \end{aligned}$$

となることに注意する．$B_t = \{(x_1, x_2, \ldots, x_n) : x_i = 0, 1, \sum_{i=1}^n x_i = t\}$ とおけば，

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{1}{\begin{pmatrix} n \\ t \end{pmatrix}} 1_{B_t}(\boldsymbol{x})$$

となり，θ に依存しないので，$T(\boldsymbol{X})$ は $\theta \in (0,1)$ に対する十分統計量である．

　定義に従って十分統計量をみつけるには，あらかじめそれと思われるものが事前にわかっていなければならない．つぎの定理を用いると比較的容易に十分統計量を見つけることができる．

定理 2.3　：σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族を $\mathcal{P}=\{P_\theta : \theta \in \Theta\}$ とし，$\boldsymbol{X}$ を P_θ からのランダム標本とする．このとき，$T(\boldsymbol{X})$ が $\theta \in \Theta$ に対して十分であるための必要十分条件は $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の非負のボレロ可測関数 h と T の値域上の関数 g_θ（P に依存）が存在し，

$$\frac{dP_\theta}{d\nu}(\boldsymbol{x}) = g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{2.35}$$

と書けることである．

証明：まず，証明は離散型分布の場合は比較的簡単であるので，その場合にについて証明を与える．σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族に対する証明は後で示す．

はじめに T が十分であると仮定する．このとき，

$$\begin{aligned}
P_\theta(\boldsymbol{X}=\boldsymbol{x}) &= \sum_t P_\theta(\boldsymbol{X}=\boldsymbol{x}, T=t) \\
&= P_\theta(\boldsymbol{X}=\boldsymbol{x}, T=T(\boldsymbol{x})) \\
&= P_\theta(T=T(\boldsymbol{x}))P_\theta(\boldsymbol{X}=\boldsymbol{x}|T=T(\boldsymbol{x}))
\end{aligned}$$

となる[1]．T が十分統計量であることから $P_\theta(\boldsymbol{X}=\boldsymbol{x}|T=T(\boldsymbol{x}))$ は θ に依存しないので，

$$\begin{aligned}
g_\theta(T(\boldsymbol{x})) &= P_\theta(T=T(\boldsymbol{x})) \\
h(\boldsymbol{x}) &= P(\boldsymbol{X}=\boldsymbol{x}|T=T(\boldsymbol{x}))
\end{aligned}$$

とおけばよい．

つぎに，(2.35) が成立すると仮定する．$t=T(\boldsymbol{x})$ とおく．このとき，

$$\begin{aligned}
P_\theta(T=t) &= \sum_{\boldsymbol{y}:T(\boldsymbol{y})=t} P_\theta(\boldsymbol{X}=\boldsymbol{y}) \\
&= \sum_{\boldsymbol{y}:T(\boldsymbol{y})=t} g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y}) \\
&= g_\theta(t)\sum_{\boldsymbol{y}:T(\boldsymbol{y})=t} h(\boldsymbol{y})
\end{aligned}$$

となる．従って，

$$\begin{aligned}
P_\theta(\boldsymbol{X}=\boldsymbol{x}|T=t) &= \frac{P_\theta(\boldsymbol{X}=\boldsymbol{x},\, T=t)}{P_\theta(T=t)} \\
&= \frac{g_\theta(t)h(\boldsymbol{x})}{g_\theta(t)\sum_{\boldsymbol{y}:T(\boldsymbol{y})=t} h(\boldsymbol{y})} \\
&= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y}:T(\boldsymbol{y})=t} h(\boldsymbol{y})}
\end{aligned}$$

---

[1] 一番目の等式は

$$\{\boldsymbol{X}=\boldsymbol{x}\} = \{\boldsymbol{X}=\boldsymbol{x}\}\cap(\cup_t\{T=\}) = \cup_t(\{\boldsymbol{X}=\boldsymbol{x}\}\cap\{T=t\})$$

からわかる．また，二番目の $T(\boldsymbol{x})=t$ を満たさない $\boldsymbol{x}$ との積事象は空事象なので，$T(\boldsymbol{x})=t$ 以外の和の t に関する項の事象は空事象となるので，和はなくなることがわかる．

となり，θ に依存しないことがわかる．□

例 2.9　：$X_1, X_2, \ldots, X_n$ を確率密度関数

$$f_X(x|\theta) = \frac{1}{\theta}\mathbf{1}_{(0,\theta)}(x)$$

からのランダム標本とする．ただし，$\theta > 0$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$\begin{aligned} p(\boldsymbol{x}|\theta) &= \frac{1}{\theta^n}\mathbf{1}_{(0,\theta)^n}(\boldsymbol{x}) \\ &= \frac{1}{\theta^n}\mathbf{1}\{\max_{1\le i\le n} x_i < \theta\}\mathbf{1}\{\min_{1\le i\le n} x_i > 0\} \\ &= g_\theta(\max(x_1, x_2, \ldots, x_n))h(\boldsymbol{x}) \end{aligned}$$

となる．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．従って，$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ は $\theta \in (0, \infty)$ の十分統計量である．

例 2.10　：$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の分布が k 母数指数分布族に属するとする．すなわち，その同時確率関数または確率密度関数が

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{k}\eta_i(\theta)T_i(\boldsymbol{x}) - \tilde{A}(\theta) + S(\boldsymbol{x})\right]\mathbf{1}\{\boldsymbol{x} \in B\}$$

で与えられる．このとき，$h(\boldsymbol{x}) = \exp[S(\boldsymbol{x})]\mathbf{1}\{\boldsymbol{x} \in B\}$ とみれば，因数分解定理から

$$\boldsymbol{T} = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X}))$$

は $\boldsymbol{\theta}$ の十分統計量となる．

与えられた分布族に対して十分統計量は必ず存在[2)]する．また，複数存在する．たとえば，例 2.8 においては，$m(1 \le m \le n)$ を固定された自然数とすれば，$(\sum_{i=1}^{m} X_i, \sum_{i=m+1}^{n} X_i)$ は θ に対する十分統計量になる．もし，ある可測関数とある統計量を用いて十分統計量 T が $T = h(S)$ と書ければ，S も十分統計量になることが因数分解定理より直ちにわかる．このとき，$\sigma(T) \subset \sigma(S)$ なので，T の方が S より有用である．データの情報を最大限に縮約する統計量とはどのようなものであろうか？そのために以下の記号と概念を導入する．

すべての $P \in \mathcal{P}$ に対し，$P(A) = 0$ を満足する事象 A を除いてあ，る命題が成立するとき，その命題は $\mathcal{P}$ **に関してほとんどいたるところ成立する**といい，$a.e.\mathcal{P}$ とかく．

定義 2.3　：T は $P \in \mathcal{P}$ の十分統計量とする．T が**最小十分統計量**であるとは，$P \in \mathcal{P}$ の任意の十分統計量 U に対し，ある可測関数 h が存在し，

$$T = h(U), \quad a.e. \quad \mathcal{P}$$

と書けることをいう．

[2)] すなわち，与えられた統計量自体である．

もし，T と U がともに最小十分統計量ならば，一対一写像 h が存在し，

$$T = h(U), \quad a.e. \quad \mathcal{P}$$

となる．

定理 2.4　：(i) $\mathcal{P}$ をある分布族とし，その部分分布族を $\mathcal{P}_0$ とする．$\mathcal{P}_0$ に関してほとんどいたるところ成立するならば，$\mathcal{P}$ に関してもほとんどいたるところ成立すると仮定する．このとき，統計量 T が $P \in \mathcal{P}$ に関して十分で，$\mathcal{P}_0$ に関して最小十分ならば，T は $\mathcal{P}$ に関しても最小十分である．
(ii) $\mathcal{P}$ を σ-有限な測度に関する $(k+1)$ 個の密度関数 $f_0,\, f_1,\, \ldots,\, f_k$ からなる分布族とする．さらに，$f_i,\, i = 1,\, 2,\, \ldots,\, k$ の台は

$$\{\boldsymbol{x} : f_i(\boldsymbol{x}) > 0\} \subset \{\boldsymbol{x} : f_0(\boldsymbol{x}) > 0\}$$

を満たし，$f_0(\boldsymbol{x}) > 0$ 上で

$$T_i(\boldsymbol{X}) = \frac{f_i(\boldsymbol{X})}{f_0(\boldsymbol{X})}, \quad i = 1,\, 2,\, \ldots,\, k$$

なる統計量 $(T_1,\, T_2,\, \ldots, T_k)$ は $P \in \mathcal{P}$ に関して最小十分である．

証明：(i) S を $P \in \mathcal{P}$ に関して十分統計量とすれば，明らかに $P \in \mathcal{P}_0$ に関しても十分である．T が $P \in \mathcal{P}_0$ に関して最小十分であることから，ある可測関数 h が存在し，$T = h(S), a.s., \mathcal{P}_0$ と表現できる．さらに，仮定から，これは $T = h(S), a.s., \mathcal{P}$ を意味するので，(i) は示された．
(ii) $f_0 > 0,\, a.s.\mathcal{P}$ であることに注意する．いま，

$$g_0(T) = 1, \quad , g_i(T) = T_i, \quad i = 1,\, 2,\, \ldots, k$$

とおく．このとき，T_i の定義から

$$f_i(\boldsymbol{x}) = g_i(T(\boldsymbol{x}))f_0(\boldsymbol{x}). \quad a.s.\mathcal{P}$$

となる．因数分解定理から，T は $P \in \mathcal{P}$ に関して十分であることがわかる．T の最小性を示すために，S を他の十分統計量とする．因数分解定理を再度用いれば，ある可測関数 h と $\tilde{g}_i$ が存在し，

$$f_i(\boldsymbol{x}) = \tilde{g}_i(S(\boldsymbol{x}))h(\boldsymbol{x}), \quad i = 0,\, 1, \ldots, k$$

と表現できる．また，T_i の定義から

$$T_i(\boldsymbol{x}) = \frac{\tilde{g}_i(S(\boldsymbol{x}))}{\tilde{g}_0(S(\boldsymbol{x}))}$$

と $\{\boldsymbol{x} : f_0(\boldsymbol{x}) > 0\}$ 上で表現できる．したがって，最小性の定義から T は $P \in \mathcal{P}$ に関して最小十分であることがわかる．　□

例 2.11　（例 2.10 の続き）：$\Theta_0 = \{\theta_0, \theta_1, \ldots, \theta_k\} \subset \Theta$ とする．$\boldsymbol{c} = (c_1(\theta), \ldots, c_k(\theta))' \subset \mathbb{R}^k$ に対し，

$$\tilde{\boldsymbol{c}}_i(\theta) = \boldsymbol{c}(\theta_i) - \boldsymbol{c}(\theta_0), \quad i = 1, 2, \ldots, k$$

としたとき，$\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立になるように Θ_0 をさだめることができると仮定[3)]する．これは，k 母数指数分布族の階数が k であるならば，この仮定をみたす．例 2.10 から T は $\theta \in \Theta$ に関して十分であるので，最小性をつぎに示す．$\mathcal{P}_0 = \{f_\theta : \theta \in \Theta_0\}$ とおけば，定理 2.4 (ii) より

$$S(\boldsymbol{x}) = (\exp\{T(\boldsymbol{x})\tilde{\boldsymbol{c}}_1'(\theta) - \tilde{d}_1\}, \ldots, \exp\{T(\boldsymbol{x})\tilde{\boldsymbol{c}}_k'(\theta) - \tilde{d}_k\})$$

は $\theta \in \Theta_0$ に関して最小十分である．ただし，$\tilde{d}_i = d(\theta_i) - d(\theta_0), i = 1, 2, \ldots, k$ である．$\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立なので，１対１の可測関数 $h : \mathbb{R}^k \to \mathbb{R}^k$ が存在して，$T(\boldsymbol{x}) = h(S(\boldsymbol{x})), a.s., \mathcal{P}_0$ と書ける．したがって，$\theta \in \Theta_0$ に関して最小十分である．さらに，これは $\mathcal{P}$ においてもほとんどいたるところ成立することが $\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立からわかる．したがって，定理 2.4 (i) から T は $\theta \in \Theta$ に関して最小十分であることが示せた．

例 2.12　$X_1, X_2, \ldots, X_n\,(n \geq 2)$ を正規分布 $N(\mu, \sigma^2)$ からのランダム標本とする．$\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ は未知とする．このとき，

$$\begin{aligned} p(\boldsymbol{x}|\,(\mu, \sigma^2)) &= \frac{1}{(\sqrt{2\pi}\sigma)^2} \exp\Big[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\Big] \\ &= \exp\Big[\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 - \frac{n}{2}\Big(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\Big)\Big] \end{aligned}$$

となる．したがって，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ として，$T_1(\boldsymbol{X}) = \sum_{i=1}^n X_i, T_2(\boldsymbol{X}) = \sum_{i=1}^n X_i^2$ とした 2 母数指数分布族となる．よって，$(T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))$ は (μ, σ^2) の十分統計量になることがわかる．さらに，

$$\mathbb{COV}(T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) = \mathbb{COV}(\sum_{k=1}^n X_k, \sum_{l=1}^n X_l^2) = \sum_{k=1}^n \mathbb{COV}(X_k, X_k^2)$$

となる．したがって，

$$\mathbb{VAR}(T_1, T_2) = n\begin{pmatrix} \mathbb{VAR}(X_1) & \mathbb{COV}(X_1, X_1^2) \\ \mathbb{COV}(X_1, X_1^2) & \mathbb{VAR}(X_1^2) \end{pmatrix} = 4\sigma^2\begin{pmatrix} 1 & 2\mu \\ 2\mu & 2\sigma^2 + 4\mu^2 \end{pmatrix} \tag{2.36}$$

となり，正定値であることがわかる．したがって，$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ は (μ, σ^2) の最小十分統計量となる．さらに，$\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ としたとき，$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ と $(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i - \bar{X}_n)^2)$ は 1 対 1 対応するので，$(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i - \bar{X}_n)^2)$ も (μ, σ^2) の最小十分統計量となる．

[3)] これは $\{\boldsymbol{c}(\theta) : \theta \in \Theta\}$ が $\mathbb{R}^k$ の開集合ならば可能で　　ある．

最後に，(2.36) を確認する．X_1 の積率母関数が $m_X(t) = \exp(\mu t + \sigma^2 t^2/2)$ で与えられることに注意する．これより

$$\begin{aligned}
\mathbb{E}[X_1] &= \left[(\mu + \sigma^2 t) m_X(t)\right]_{t=0} = \mu, \\
\mathbb{E}[X_1^2] &= \left[\sigma^2 m_X(t) + (\mu + \sigma^2 t)^2 m_X(t)\right]_{t=0} = \mu^2 + \sigma^2, \\
\mathbb{E}[X_1^3] &= \left[\sigma^2(\mu + \sigma^2 t) m_X(t) + 2\sigma^2(\mu + \sigma^2 t) m_X(t) + (\mu + \sigma^2 t)^3 m_X(t)\right]_{t=0} = 3\mu\sigma^2 + \mu^3, \\
\mathbb{E}[X_1^4] &= \left[3\sigma^4 m_X(t) + 6\sigma^2(\mu + \sigma^2 t)^2 m_X(t) + (\mu + \sigma^2 t)^4 m_X(t)\right]_{t=0} = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4
\end{aligned}$$

となる．したがって，

$$\begin{aligned}
\mathbb{COV}(X_1, X_1^2) &= \mathbb{E}[(X_1 - \mu)(X_1^2 - \mu^2 - \sigma^2)] = 2\mu\sigma^2, \\
\mathbb{VAR}[X_1^2] &= \mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2 = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 - (\mu^2 + \sigma^2)^2 = 2\sigma^4 + 4\mu^2\sigma^2
\end{aligned}$$

からわかる．

2.3.1 完備統計量

Θ を母数空間とする統計モデル $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ を考えよう．

定義 2.4　：確率変数 X は分布 P_θ にしたがうとする．統計量 $V(X)$ が**補助統計量**（ancillary statistics）であるとは，$V(X)$ の分布が θ に依存しないときをいう．さらに，$V(X)$ が**一次のオーダーで補助統計量**（first-order ancillary statistics）であるとは，$\mathbb{E}[V(X)]$ が θ に依存しないことをいう．

自明な補助統計量は $V(X)$ が定数であるような統計量である．しかし，もし $V(X)$ が自明でない補助統計量であれば，$\sigma(X)$ の部分 σ - 集合族 $\sigma(V(X))$ は P_θ の情報を含まない自明でない σ - 集合族となる．したがって，十分統計量 $T(X)$ のすべての定数ではない関数が補助統計量でないのならば，$T(X)$ は最もうまく情報縮約できたことになる．

定義 2.5　：統計量 $T(X)$ が $\theta \in \Theta$ に対し**完備**（*complete*）であるとはつぎの条件を満たすことである：
任意のボレル関数 f に対し，すべての $P \in \mathcal{P}$ について

$$\mathbb{E}[f(T)] = 0 \quad ならば \quad f = 0, \quad a.s., \mathcal{P}$$

が成立する．
$T(X)$ が有界な関数に対して完備であるとは，有界ボレロ関数に対して，上の条件が成立することである．

つぎの定理は指数分布族においては比較的容易に完備十分統計量をみつけることができることを示している．

定理 2.5　確率分布 P の密度関数が (2.32) で与えられる指数分布族で，自然母数空間がフルランクであるとする．このとき，$T(X)$ は $c \in \Xi$ に対して完備十分統計量となる．

証明：例 2.10 から十分性はわかる．完備性については Lehmann の TSH の 4.2 節を参照せよ．
□

つぎに完備統計量と補助統計量の独立性についての定理をあげる．

定理 2.6　（Basu の定理）：確率変数 X は $P \in \mathcal{P}$ にしたがい，$V(X)$ を補助統計量とし，$T(X)$ を有界関数にたいして完備な十分統計量とする．このとき，V と T は独立となる

証明：B を V の値域の事象とする．V は補助統計量なので，$P \in \mathcal{P}$ を動かしても $P(V^{-1}(B))$ は定数となる．また，T は十分統計量なので，$\mathbb{E}[1_B|T]$ は T のみの関数である．条件付き確率の性質から

$$0 = \mathbb{E}[1_B(V)] - P(V^{-1}(B)) = \mathbb{E}\{\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B))\}$$

となり，上の式の右辺の期待値の中 $\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B))$ は T の有界ボレロ関数となるので，完備性から

$$\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B)) = 0$$

となる．さらに，A を T の値域の事象とすれば，

$$\begin{aligned}
P\{T^{-1}(A) \cap V^{-1}(B)\} &= \mathbb{E}\{\mathbb{E}(1_A(T)1_B(V)\ T)\} = \mathbb{E}\{1_A(T)\mathbb{E}(1_B(V)|T)\} \\
&= \mathbb{E}\{1_A(T)P(V^{-1}(B)\} = P(T^{-1}(A))\mathbb{E}\{1_A(T)\} \\
&= P(T^{-1}(A))P(V^{-1}(B))
\end{aligned}$$

となり，独立性が示せた． □

・・・・・2.4・・・・・

最尤法

確率ベクトル $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の統計モデルを $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ とする．$\mathcal{P}$ に含まれる P_θ に対応する確率密度関数もしくは確率関数を $p(\boldsymbol{x}|\,\theta)$ と記すことにする．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．$\boldsymbol{X} = \boldsymbol{x}$ が観測されたときの尤度関数を

$$L_n(\theta|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta), \qquad \theta \in \Theta$$

で定めることにする．$L_n(\cdot\,|\,\boldsymbol{x})$ は標本空間から $\{\theta \mapsto p(\boldsymbol{x}|\,\theta) : \boldsymbol{x} \in S\}$ なる関数族への対応となる．$\boldsymbol{x}$ があたられたとき，$L_n(\theta|\,\boldsymbol{x})$ は θ の関数とみなす．これを簡単に $L_n(\theta)$ と書くことにする．$L_n(\theta)$ は $\boldsymbol{x}$ が与えられたとき，いろいろな θ の「確からしさ」もしくは「尤もらしさ」を表現するものである．

特に，$X_1, X_2, \ldots, X_n$ が独立同一に確率密度関数 $f(x|\,\theta)$ に従うならば，尤度関数は

$$L_n(\theta|\,\boldsymbol{x}) = \Pi_{i=1}^n f(x_i|\,\theta)$$

で与えられる．

最尤法とは，与えられたデータを実現させるために「尤もらしい」母数の値を母数の推定値として用いる手法である．すなわち，$\boldsymbol{X} = \boldsymbol{x}$ が与えられたとき，尤度関数を最大にする値 $\widehat{\theta}(\boldsymbol{x})$ を見つけることである：

$$L_n(\widehat{\theta}(\boldsymbol{x})|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\widehat{\theta}(\boldsymbol{x})) = \max\{p(\boldsymbol{x}|\,\theta) : \theta \in \Theta\} = \max\{L_n(\theta|\,\boldsymbol{x}) : \theta \in \Theta\}$$

$\widehat{\theta}(\boldsymbol{x})$ を θ の最尤推定値といい，$\widehat{\theta}(\boldsymbol{X})$ を θ の最尤推定量という．

例 2.13　確率変数 X が正規分布 $N(\theta, \sigma^2)$ に従うとする．ただし，σ^2 は既知とする．このとき，尤度関数は

$$L_1(\theta|\,x) = \frac{1}{\sigma}\psi\left(\frac{x-\theta}{\sigma}\right)$$

となる．ただし，$\psi(x) = (1/\sqrt{2\pi})e^{x^2/2}$ である．このとき，最大は

$$\widehat{\theta}(x) = x$$

のとき唯一達成される．したがって，$\widehat{\theta}(X) = X$ は最尤推定量となる．

つぎに，$X_1, X_2, \ldots, X_n$ は独立同一に正規分布 $N(\theta, \sigma^2)$ に従うとする．ここでも σ^2 は既知とする．このとき，尤度関数は

$$\begin{aligned} L_n(\theta|\,\boldsymbol{x}) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^n \frac{x_i-\theta}{\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2}\frac{(\bar{x}_n-\theta)^2}{\sigma^2/n}\right] \exp\left[-\frac{1}{2}\sum_{i=1}^n \frac{(x_i-\bar{x}_n)^2}{\sigma^2}\right] \end{aligned}$$

となる．よって，最大は

$$\widehat{\theta}(\boldsymbol{x}) = \bar{x}_n, \qquad \bar{x}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

で達成される．したがって，最尤推定量は $\widehat{\theta}(\boldsymbol{X}) = \bar{X}_n$ となる．ただし，$\bar{X}_n = (1/n)(X_1 + X_2 + \cdots + X_n)$ である．

尤度関数に対数をとったものを対数尤度とよび，

$$l_n(\theta) = \log L_n(\theta|\,\boldsymbol{x})$$

と記す[1]．特に，$X_i,\ i = 1,\ 2,\ \ldots,\ n,$ が独立同一に確率密度関数 $f(x|\,\theta)$ に従う場合には

$$l_n(\theta) = \log \Pi_{i=1}^n f(x_i|\,\theta) = \sum_{i=1}^n \log f(x_i|\,\theta)$$

となる．

もし，Θ が開集合で $l_n(\theta)$ が θ に関して微分可能で $\widehat{\theta}(\boldsymbol{x})$ が存在するならば，$\widehat{\theta}(\boldsymbol{x})$ は方程式

$$\frac{\partial}{\partial \theta_i} l_n(\theta) = 0, \qquad i = 1,\ 2,\ \ldots,\ k$$

をみたす．この方程式を尤度方程式という．

例 2.14　標識 1, 2, 3 のどれかをもつ個体から構成される母集団を考える．それぞれの標識の出現確率は Hardy-Weinberg 比率で与えられるとする：

$$p(1|\theta) = \theta^2, \qquad p(2|\theta) = 2\theta(1-\theta), \qquad p(3|\theta) = (1-\theta)^2, \qquad 0 < \theta < 1$$

たとえば，3 つの個体を観測し，$x_1 = 1,\ x_2 = 2,\ x_3 = 1$ を得たとする．このとき

$$L_3(\theta|\,\boldsymbol{x}) = p(1|\,\theta)p(2|\,\theta)p(1|\,\theta) = 2\theta^5(1-\theta)$$

となる．尤度方程式は

$$\frac{\partial}{\partial \theta} l_3(\theta) = \frac{5}{\theta} - \frac{1}{1-\theta} = 0$$

となり，唯一の解 $\widehat{\theta} = 5/6$ を得る．これは

$$\frac{\partial^2}{\partial \theta^2} l_3(\theta) = -\frac{5}{\theta^2} - \frac{1}{(1-\theta)^2} < 0, \qquad 0 < \theta < 1$$

よりわかる．

一般に，n 個の観測 $x_1,\ x_2,\ \ldots,\ x_n$ を得たとする．いま

$$n_j = \#\{x_i = j : i = 1,\ 2,\ \ldots,\ n\}, \qquad j = 1,\ 2,\ 3$$

とする．尤度関数は

$$L_n(\theta|\,\boldsymbol{x}) = \theta^{2n_1}\{2\theta(1-\theta)\}^{n_2}(1-\theta)^{2n_3} = 2^{n_2}\theta^{2n_1+n_2}(1-\theta)^{n_2+2n_3}$$

[1]形式的に，$0/0 = 0,\ 0 \times \infty = 0$ とする．

より

$$\frac{\partial}{\partial\theta}l_n(\theta) = \frac{2n_1+n_2}{\theta} - \frac{n_2+2n_3}{1-\theta} = \frac{1}{\theta(1-\theta)}\{(2n_1+n_2) - 2(n_1+n_2+n_3)\theta\}$$

より，$2n_1+n_2>0$, $n_2+2n_3>0$ のとき，最尤推定値は唯一存在して，

$$\widehat{\theta}(\boldsymbol{x}) = \frac{2n_1+n_2}{2(n_1+n_2+n_3)}$$

となる．もし，$2n_1+n_2=0$ のとき，尤度関数は

$$2^{n_2}(1-\theta)^{(2n_1+n_2)+(n_2+2n_3)} = 2^{n_2}(1-\theta)^{2(n_1+n_2+n_3)}$$

となり，$\theta=0$ のとき，尤度関数は最大になり，$\Theta=(0,1)$ なので，最尤推定値は存在しない．また，$n_2+2n_3=0$ のときは，尤度関数は $2^{n_2}\theta^{2(n_1+n_2+n_3)}$ なり，最尤推定値は存在しない．

例 2.15　$X_1, X_2, \ldots, X_n$ は独立同一分布に従い，各 X_i, $i=1,2,\ldots,n$, は $1,2,\ldots,k$ の値をとり，その確率は $\theta_j = \mathbb{P}\{X_i=j\}$, $j=1,2,\ldots,k$, で与えられるとする．ここで，$n \geq k-1$ を仮定する．いま，$N_j=\sum_{i=1}^n I\{X_i=j\}$ おく．$\boldsymbol{X}=\boldsymbol{x}$ が与えられたとする．このとき，$n_j=\sum_{i=1}^n I\{x_i=j\}$ とおけば，対数尤度は

$$l_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log p(\boldsymbol{x}|\,\boldsymbol{\theta}) = \sum_{j=1}^k n_j \log\theta_j$$

となる．ただし，$\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)$ で

$$\theta_k = 1 - \sum_{j=1}^{k-1}\theta_j \tag{2.37}$$

である．

まず，$n_j>0$, $j=1,2,\ldots,k$, を仮定する．このとき，θ_j のどれかがゼロならば，$p(\boldsymbol{x}|\,\boldsymbol{\theta})=0$ となる．したがって，最尤推定値は $\theta_j>0$ となるので，上の仮定のもとでは，最尤推定値は $[0,1]^k$ の内点である．したがって，最尤推定値は尤度方程式

$$\frac{\partial}{\partial\theta_j}l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}\sum_{\ell=1}^k n_\ell\log\theta_\ell = \sum_{\ell=1}^k \frac{n_\ell}{\theta_\ell}\frac{\partial\theta_\ell}{\partial\theta_j} = 0, \qquad j=1,2,\ldots,k-1$$

となる．(2.37) から $\partial\theta_k/\partial\theta_j=-1$, $j=1,2,\ldots,k-1$, となる．よって，尤度方程式は

$$\frac{\widehat{\theta}_k}{\widehat{\theta}_j} = \frac{n_k}{n_j}, \qquad j=1,2,\ldots,k-1$$

となる．これを再度 (2.37) に代入すれば，$\widehat{\theta}_k=n_k/n$ となり，

$$\widehat{\theta}_j = \frac{n_j}{n}, \qquad j=1,2,\ldots,k$$

となる．ただし，$n=\sum_{\ell=1}^{k} n_\ell$ とした．つぎに，$\theta_j = n_j/n,\ j=1,\,2,\,\ldots,\,k$ が実際に $l_n(\boldsymbol{\theta})$ を最大にしていることを確認するために，$l_n(\boldsymbol{\theta})$ は $(\theta_1,\,\theta_2,\,\ldots,\theta_{k-1})$ に関して concave であることを示す．$1\le r\le k-1,\ 1\le j\le k-1$ に対して，

$$\frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_j} l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_r}\left(\frac{n_j}{\theta_j}-\frac{n_k}{\theta_k}\right) = \begin{cases} -\left(\frac{n_r}{\theta_r^2}+\frac{n_k}{\theta_k^2}\right)<0, & r=j \\ -\frac{n_k}{\theta_k^2}<0, & r\neq j \end{cases}$$

となる．

ある j に対して，$n_j=0$ のとき，$\widehat{\theta}_j = n_j/n$ が最尤推定値であることも確認することができる．

定理 2.7　　$T(\boldsymbol{X})$ を未知母数 θ の十分統計量とする．このとき，θ の最尤推定量が一意に存在するならば，θ の最尤推定量は T の関数である．

証明　$p(\boldsymbol{x}|\,\theta)$ を確率関数または確率密度関数とする．因子分解定理から θ と T を通してのみ $\boldsymbol{x}$ に依存する関数 g と $\boldsymbol{x}$ のみに依存する関数 h が存在して，

$$p(\boldsymbol{x}|\,\theta) = h(\boldsymbol{x})g(T(\boldsymbol{x})|\,\theta)$$

と書ける．これより θ に関して $p(\boldsymbol{x}|\,\theta)$ を最大化することは θ に関して $g(T(\boldsymbol{x})|\,\theta)$ を最大化することと同値になる．また，$T(\boldsymbol{x})=t$ と与えられたとき，$g(t|\,\theta)$ が 2 つ以上 θ で最大になるとすれば，それに対応する $\boldsymbol{x}$ において $g(T(\boldsymbol{x})|\,\theta)$ も 2 つ以上の θ で最大化されるので，仮定と矛盾する．したがって，$g(t|\,\theta)$ を θ について最大にする点はひとつである．　□

定理 2.8　　関数 $g:\mathbb{R}\to\mathbb{R}$ が 1 対 1 のとき，$\widehat{\theta}$ が θ の最尤推定量であれば，$g(\widehat{\theta})$ は $g(\theta)$ の最尤推定量である．

証明　g は 1 対 1 だから，逆関数 g^{-1} が存在し，$\tau=g(\theta)$ のとき，$\theta=g^{-1}(\tau)$ となる．これより

$$L_n(\theta|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta) = p(\boldsymbol{x}|\,g^{-1}(\tau)) = \tilde{L}_n(\tau|,\boldsymbol{x})$$

と書けるので，

$$\sup_{\tau}\tilde{L}_n(\tau|,\boldsymbol{x}) = \sup_{\tau} L_n(g^{-1}(\tau)|,\boldsymbol{x}) = \sup_{\theta} L_n(\theta|\,\boldsymbol{x})$$

よって，$\widehat{\theta}=g^{-1}(\widehat{\tau})$ のとき最大化される．したがって，$\widehat{\tau}=g(\widehat{\theta})$ のとき最大化される．　□

例 2.16　（指数分布族の最尤推定量）　$\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_n)'$ は k 母数指数分布族

$$f(x|\,\boldsymbol{\eta}) = S(x)\exp[\sum_{j=1}^{k}\eta_j T_j(x) - A(\boldsymbol{\eta})]$$

からの大きさ n のランダム標本とする．ただし，$\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ とし，自然母数空間 $\eta \in \mathcal{E}$ は $\mathbb{R}^k$ の開集合とする．である．このとき，$\boldsymbol{X}$ の同時確率（密度）関数は

$$\begin{aligned} p(\boldsymbol{x}|\,\boldsymbol{\eta}) &= \Pi_{i=1}^n f(x_i|\,\boldsymbol{\eta}) \\ &= \Pi_{i=1}^n S(x_i) \exp[\sum_{j=1}^k n\eta_j \bar{T}_j(\boldsymbol{x}) - nA(\boldsymbol{\eta})] \end{aligned}$$

となる．ただし，$\bar{T}_j(\boldsymbol{x}) = \sum_{i=1}^n T_j(x_i)$, $j = 1, 2, \ldots, k$ である．したがって，対数尤度は

$$l_n(\boldsymbol{\eta}) = n\sum_{j=1}^k \eta_j \bar{T}_j(\boldsymbol{x}) - nA(\boldsymbol{\eta}) + (\text{定数項})$$

となる．系 2.1 を用いれば，

$$\frac{\partial}{\partial \eta_j} l_n(\boldsymbol{\eta}) = n\bar{T}_j(\boldsymbol{x}) - n\frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta}) = n\bar{T}_j(\boldsymbol{x}) - n\mathbb{E}_\eta[T_j(X)]$$

と

$$\frac{\partial^2}{\partial \eta_j \eta_m} l_n(\boldsymbol{\eta}) = -n\frac{\partial^2}{\partial \eta_j \eta_m} A(\boldsymbol{\eta}) = -n\mathbb{COV}_\eta(T_j(X), T_m(X))$$

となるので，行列 $((\partial^2/\partial\boldsymbol{\eta}'\boldsymbol{\eta})A(\boldsymbol{\eta}))$ は負の定符号となる．したがって，尤度方程式

$$\frac{1}{n}\sum_{i=1}^n T_j(x_i) = \frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta})$$

または

$$\frac{1}{n}\sum_{i=1}^n T_j(x_i) = \mathbb{E}_\eta[T_j(X)], \qquad j = 1, 2, \ldots, k$$

は唯一の解を持ち，これは $\boldsymbol{\eta}$ の最尤推定値となる．

·····2.5·····

最尤法とその計算アルゴリズム

ここでは，最尤推定値を数値計算で求める方法を 3 つ紹介する．

2.5.1 ニュートン・ラプソン法

いま，$g:\mathbb{R}\to\mathbb{R}$ を 2 階微分可能な関数とし，方程式 $g(x)=0$ をみたす解 $x=c$ をみつけたい．そのために，c に近い x に対して，テーラー展開をする：

$$0=g(c)\approx g(x)+\dot{g}(x)(x-c)$$

ただし，$\dot{g}(x)=dg/dx$ である．$\dot{g}(x)\neq 0$ のとき，これを c について解けば

$$c\approx x-\frac{g(x)}{\dot{g}(x)}$$

を得る．初期値 x_0 を取り，点列 $\{x_n\}_{n=1}^{\infty}$ を逐次的に

$$x_{n+1}=x_n-\frac{g(x_n)}{\dot{g}(x_n)},\qquad n=0,\,1,\,\ldots \tag{2.38}$$

で定義する．そして，$|g(x_n)/\dot{g}(x_n)|$ が十分小さくなるまで操作を繰り返すとする．区間 $I=[a,\,b]$ 上で $\ddot{g}(x)>0$ で，$g(a)g(b)<0$ のとき，$g(x_0)>0$ となる $x_0\in(a,\,b)$ をひとつ見つければ，(2.38) で得られる点列 $\{x_n\}_{n=1}^{\infty}$ は I 上における零点 c に収束することが知られている[1)]．ニュートン・ラプソン法を用いて，尤度方程式の解として定義される最尤推定値を求めることができる．

いま，$\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_n)$ は同時確率密度関数または確率関数 $p(\boldsymbol{x}|\,\theta)$ を持つとする．ただし，$\theta\in\Theta\subset\mathbb{R}$ とする．$\boldsymbol{X}=\boldsymbol{x}$ が与えられたときの尤度関数は $L_n(\theta|\,\boldsymbol{x})=p(\boldsymbol{x}|\,\theta)$ である．最尤推定値 $\widehat{\theta}$ は尤度方程式

$$S(\theta)=\frac{d}{d\theta}\log L_n(\theta|\,\boldsymbol{x})=0 \tag{2.39}$$

の解とする．さらに，k 回操作を行ったのちの θ の推定値を $\widehat{\theta}^{(k)}$ とする：

$$\widehat{\theta}^{(k+1)}=\widehat{\theta}^{(k)}+\frac{S(\widehat{\theta}^{(k)})}{H(\widehat{\theta}^{(k)})}$$

ただし，

$$H(\theta)=-\frac{d^2}{d\theta^2}\log L_n(\theta)$$

である．この操作を $\widehat{\theta}^{(k+1)}$ と $\widehat{\theta}^{(k)}$ との差が十分小さくなるまで繰り返す．

---

1) 杉浦「解析入門 I (東京大学出版会)」p.105 を参照．

ニュートン・ラプソン法を用いるために初期推定値 $\widehat{\theta}^{(0)}$ が必要である．初期推定値に何を用いるかによって，アルゴリズムは収束したりしなかったりする．また，尤度方程式 $S(\theta)=0$ が複数の解を持つ場合には，尤度方程式 (2.39) の解は尤度関数の極小点，極大点，鞍馬点 (saddle-point) に対応するので，$\widehat{\theta}^{(k)}$ は最尤推定値とは異なる点に収束する可能性がある．収束先が最尤推定値と異なるかどうかが不明な場合には，複数の初期値で試すとよい．また，初期値として，別の推定値 (別の推定量の実現値) を用いることもできる．たとえば，$\widehat{\theta}_n^{(0)}(\boldsymbol{X})$ が θ の十分よい推定量ならば，一段階推定量

$$\widehat{\theta}_n^{(1)} = \widehat{\theta}_n^{(0)} + \frac{S(\widehat{\theta}_n^{(0)})}{H(\widehat{\theta}_n^{(0)})}$$

は，最尤推定量と同じ性質を漸近的には同じ性質をもつことが知られている．正確に言えば，$\sqrt{n}(\widehat{\theta}_n^{(0)}-\theta)$ は正規分布に分布収束するならば，$\sqrt{n}(\widehat{\theta}_n^{(1)}-\widehat{\theta}_n) \xrightarrow{P} 0$ となる．ただし，$\widehat{\theta}_n$ は θ の最尤推定量である．

例 2.17　確率密度関数

$$f_X(x|\,\theta) = \frac{1}{\pi\{1+(x-\theta)^2\}} I_{(-\infty,\infty)}(x), \qquad (-\infty<\theta<\infty)$$

を持つコーシー分布からの大きさ n のランダム標本を $X_1, X_2, \ldots, X_n$ とする．このとき，実現値 $x_1, x_2, \ldots, x_n$ に対する対数尤度関数は

$$\log L_n(\theta) = -\sum_{i=1}^{n} \log\{1+(x_i-\theta)^2\} - n\log\pi$$

となる．最尤推定値 $\widehat{\theta}_n$ は尤度方程式

$$S(\widehat{\theta}_n) = \sum_{i=1}^{n} \frac{2(x_i-\widehat{\theta}_n)}{1+(x_i-\widehat{\theta}_n)^2} = 0$$

の解である．$S(\theta)$ は θ の単調関数でないので，与えられた $(x_1, x_2, \ldots, x_n)$ に対して，尤度方程式は複数の解を持つ可能性がある．したがって，適切な初期値 $\widehat{\theta}^{(0)}$ を選ぶことが重要である．コーシー分布は $\mathbb{E}[X_1]$ が定義されないので，初期値として，標本平均を用いるのは適当ではない．X_1 の分布は θ に関して対称であることに注目して，標本中央値を初期値 $\widehat{\theta}^{(0)}$ として用いる．これを用いて，逐次的に $\widehat{\theta}^{(k)}, k=1, 2, \ldots$ を

$$\widehat{\theta}^{(k)} = \widehat{\theta}^{(k-1)} + \frac{S(\widehat{\theta}^{(k-1)})}{H(\widehat{\theta}^{(k-1)})}$$

で定める．ただし，

$$H(\theta) = 2\sum_{i=1}^{n} \frac{1-(x_i-\theta)^2}{\{1+(x_i-\theta)^2\}^2}$$

である．$\theta=10$ のコーシー分布から標本の大きさ $n=100$ のランダム標本に基づいて最尤推定値を求めた例が次である．

k	$\widehat{\theta}^{(k)}$	$logL_n(\widehat{\theta}^{(k)}) + 100\log(\pi)$
0	9.932387	11.95144
1	9.98055	11.9517
2	9.980323	11.9517
3	9.980323	11.9517

つぎに，母数の次元が p の場合を考える．母数 $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ が p - 次元のとき，$\boldsymbol{\theta}$ の最尤推定値 $\widehat{\boldsymbol{\theta}}$ は尤度方程式 $\boldsymbol{S}(\boldsymbol{\theta}) = \boldsymbol{0}$ の解として定義される．ただし，

$$\boldsymbol{S}(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial\theta_1}\log L_n(\boldsymbol{\theta}),\ \frac{\partial}{\partial\theta_2}\log L_n(\boldsymbol{\theta}),\ \ldots,\ \frac{\partial}{\partial\theta_p}\log L_n(\boldsymbol{\theta})\right)'$$

である．このとき，k 回目の逐次回 $\widehat{\boldsymbol{\theta}}^{(k)}$ は

$$\widehat{\boldsymbol{\theta}}^{(k)} = \widehat{\boldsymbol{\theta}}^{(k-1)} + [H(\widehat{\boldsymbol{\theta}}^{(k-1)})]^{-1} S(\widehat{\boldsymbol{\theta}}^{(k-1)}) \tag{2.40}$$

で定義される．ただし，$H(\boldsymbol{\theta})$ の (i, j) - 成分は

$$H_{ij}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L_n(\boldsymbol{\theta})$$

で定義される．

2.5.2 Fisher のスコア法

Newton-Raphson アルゴリズムの簡単な修正として，Fisher のスコアアルゴリズムがある．Fisher のスコアアルゴリズムは (2.40) の中の H の代わりに

$$H^*(\boldsymbol{\theta}) = \mathbb{E}_\theta[H(\boldsymbol{\theta})] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log L_n(\boldsymbol{\theta})\right]$$

である．ただし，

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log L_n(\boldsymbol{\theta})$$

の (i, j) - 成分は

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log L_n(\boldsymbol{\theta})$$

である．したがって，k 回目の逐次解 $\widehat{\boldsymbol{\theta}}^{(k-1)}$ は

$$\widehat{\boldsymbol{\theta}}^{(k)} = \widehat{\boldsymbol{\theta}}^{(k-1)} + \left[H^*(\widehat{\boldsymbol{\theta}}^{(k-1)})\right]^{-1}\boldsymbol{S}(\widehat{\boldsymbol{\theta}}^{(k-1)})$$

で定義される．

例 2.18　(例 2.17 の続き)

$$H(\theta) = 2\sum_{i=1}^{n}\frac{1-(x_i-\theta)^2}{\{1+(x_i-\theta)^2\}^2}$$

から

$$H^*(\theta) = \frac{2n}{\pi}\int_{-\infty}^{\infty} \frac{1-(x-\theta)^2}{\{1+(x-\theta)^2\}^3}\,dx = \frac{n}{2} \tag{2.41}$$

から Fisher のスコアアルゴリズムは

$$\widehat{\theta}^{(k)} = \widehat{\theta}^{(k-1)} + \frac{4}{n}\sum_{i=1}^{n} \frac{x_i - \widehat{\theta}^{(k-1)}}{1+(x_i-\widehat{\theta}^{(k-1)})^2}$$

となる．最後に，(2.41) の計算をする．$z = x-\theta$ とおき，さらに $y = \tan\gamma$ とおけば，

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1-y^2}{(1+y^2)^3}\,dy &= \int_{-\infty}^{\infty} \frac{2}{(1+y^2)^3}\,dy - \int_{-\infty}^{\infty} \frac{1}{(1+y^2)^2}\,dy \\
&= 2\int_{-\pi/2}^{\pi/2} \frac{1}{(1+\tan\gamma^2)^3}\frac{1}{\cos\gamma^2}\,d\gamma - \int_{-\pi/2}^{\pi/2} \frac{1}{(1+\tan\gamma^2)^2}\frac{1}{\cos\gamma^2}\,d\gamma \\
&= 2\int_{-\pi/2}^{\pi/2} \cos^4\gamma\,d\gamma - \int_{-\pi/2}^{\pi/2} \cos^2\gamma\,d\gamma, \\
&= 2\int_{-\pi/2}^{\pi/2} \left(\frac{\cos(4\gamma)+1}{8} + \cos(2\gamma) + \frac{1}{4}\right) d\gamma - \int_{-\pi/2}^{\pi/2} \left(\frac{\cos(2\gamma)+1}{2}\right) d\gamma \\
&= 2\left[\frac{\sin(4\gamma)}{32} + \frac{\gamma}{8} + \frac{\sin(2\gamma)}{2} + \frac{\gamma}{4}\right]_{-\pi/2}^{\pi/2} - \left[\frac{\sin(2\gamma)}{4} + \frac{\gamma}{4}\right]_{-\pi/2}^{\pi/2} \\
&= \frac{\pi}{4}
\end{aligned}$$

となることがわかる．

2.5.3 EM アルゴリズム

$(\mathcal{X}, \mathcal{A}, \mu)$ を測度空間とする．ここで，μ は σ - 有限な測度[2)]とする．$\mathcal{X}$ - 値確率変数 $\boldsymbol{X}$ は母数 $\theta\,(\theta \in \Theta)$ の確率測度 P_θ を持つ：

$$\mathbb{P}(\{\boldsymbol{X} \in A\}) = P_\theta(A), \qquad A \in \mathcal{A}$$

さらに，P_θ は μ に関する確率密度関数 $p(\cdot\,|\,\theta)$ をもつとする：

$$P_\theta(A) = \int_A p(\boldsymbol{x}|\,\theta)\,\mu(d\boldsymbol{x}), \qquad A \in \mathcal{A}$$

いま，$(\mathcal{X}, \mathcal{A})$ は隠れた空間で $\boldsymbol{X}$ のすべてを観測できないとする．実際には，ある可測空間 $(\mathcal{Y}, \mathcal{B})$ と可測関数 $T : \mathcal{X} \mapsto \mathcal{Y}$ が存在して，$\boldsymbol{Y} = T(\boldsymbol{X})$ のみが観測できるとする．ν を $(\mathcal{Y}, \mathcal{B})$ 上の σ - 有限な測度とする．すると T は $(\mathcal{Y}, \mathcal{B})$ 上の測度を誘導する：

$$Q_\theta(B) = P_\theta T^{-1}(B) = P_\theta(T^{-1}(B)), \qquad B \in \mathcal{B}$$

---

2) ある部分集合の列 $\{G_n\}_{n=1}^{\infty}$, $G_n \in \mathcal{A}$ で $\cup_{n=1}^{\infty} G_n = \mathcal{X}$ かつ各 n に対して $\mu(G_n) < \infty$ なるものが存在することである．

さらに，確率測度 Q_θ は測度 ν に関して確率密度関数 $q(\cdot\,|\,\theta)$ を持つとする：

$$Q_\theta(B) = \int_B q(\boldsymbol{y}|\,\theta)\,\nu(d\boldsymbol{y}), \qquad B \in \mathcal{B}$$

EM アルゴリズムは，観測 $\boldsymbol{Y} = \boldsymbol{y}$ が与えられたとき，θ の関数として $q(\boldsymbol{y}|\,\theta)$ を最大化することで θ の最尤推定値を求める方法である．

EM アルゴリズムは以下のように行う：$\boldsymbol{y}$ が与えられたとき，θ の初期推定値の $\widehat{\theta}^{(0)}$ から始める．$\widehat{\theta}^{(0)}$ より $P_{\widehat{\theta}^{(0)}}$ と $Q_{\widehat{\theta}^{(0)}} = P_{\widehat{\theta}^{(0)}}T^{-1}$ が初期の推定された確率測度となる．

<u>E - 段階 (E - Step)</u>：各 $\theta \in \Theta$ に対し，条件付き期待値

$$\phi_1(\theta) = \mathbb{E}_{P_{\widehat{\theta}^{(0)}}}\{\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\} \tag{2.42}$$

を求める．$P_{\widehat{\theta}^{(0)}}\{T(\boldsymbol{X}) = \boldsymbol{y}\} > 0$ の場合には，$T(\boldsymbol{X}) = \boldsymbol{y}$ が与えられたときの $\boldsymbol{X}$ の条件付分布を求め，それに関して関数 $\boldsymbol{x} \mapsto \log p(\boldsymbol{x}|\,\theta)$ の期待値を求めればよい．$P_{\widehat{\theta}^{(0)}}\{T(\boldsymbol{X}) = \boldsymbol{y}\} = 0$ のときは注意が必要であるが，(2.42) の正当化は可能である．以後は簡単のために，$P_{\widehat{\theta}^{(0)}}\{T(\boldsymbol{X}) = \boldsymbol{y}\} > 0$ の場合を考える．

<u>M - 段階 (M - Step)</u>：θ に関して

$$\phi_1(\theta)$$

を最大化する．最大を与える点 (存在すれば) を $\widehat{\theta}^{(1)}$ とおく．つぎに，$P_{\widehat{\theta}^{(0)}}$ の代わりに $P_{\widehat{\theta}^{(1)}}$ を用いる．

<u>E - 段階 (E - Step)</u>：各 $\theta \in \Theta$ に対して

$$\phi_2(\theta) = \mathbb{E}_{P_{\widehat{\theta}^{(1)}}}\{\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\}$$

を求める．

<u>M - 段階 (M - Step)</u>：θ に関して

$$\phi_2(\theta)$$

を最大化する．最大を与える点 (存在すれば) を $\widehat{\theta}^{(2)}$ とおく．

一般には，m - 段階 (m $= 1, 2, \ldots$) で

<u>E - 段階 (E - Step)</u>：各 $\theta \in \Theta$ に対し，条件付き期待値

$$\phi_m(\theta) = \mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\} \tag{2.43}$$

を求める．

<u>M - 段階 (M - Step)</u>：θ に関して

$$\phi_m(\theta)$$

を最大化する．最大を与える点 (存在すれば) を $\widehat{\theta}^{(m)}$ とおく．

この操作を $\widehat{\theta}^{(m)}$ が $\widehat{\theta}^{(m-1)}$ とほとんど変化がなくなるまで繰り返す．

例 2.19　$X_1, X_2, \ldots, X_n$ は独立同一に指数分布

$$f(x|\,\theta) = \frac{1}{\theta}e^{-x/\theta}I_{(0,\infty)}(x)$$

に従うとする．ただし，$\theta > 0$ である．しかし，各 X_i, $i = 1, 2, \ldots, n$, は直接観測されず，各 X_i の整数部分のみが観測されるとする．すなわち，$Y_i = \lfloor X_i \rfloor$ である．$Y_1, Y_2, \ldots, Y_n$ の観測に基づいて θ の最尤推定値を求めよう．

この場合，$\mathcal{X} = (\mathbb{R}^+)^n$，$\mathcal{A}$ は $(\mathbb{R}^+)^n$ 上のボレル可測集合体であり，P_θ はルベーグ測度に関する確率密度関数

$$p(\boldsymbol{x}|\,\theta) = \Pi_{i=1}^n f(x_i|\,\theta) = \theta^{-n}\exp\bigl(-\sum_{i=1}^n x_i/\theta\bigr), \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

を持つ．また，$\mathcal{Y} = \{0, 1, \ldots\}^n$ で $\mathcal{B}$ は $\mathcal{Y}$ のすべての部分集合の集まりからなる σ - 集合体である．さらに，関数 $T : \mathcal{X} \mapsto \mathcal{Y}$ は

$$T(\boldsymbol{x}) = (\lfloor x_1 \rfloor, \lfloor x_2 \rfloor, \ldots, \lfloor x_n \rfloor)$$

で定義される．

いま，

$$P(Y_i = y) = \int_y^{y+1} f(x|\,\theta)\,dx = e^{-y/\theta}(1 - e^{-1/\theta})$$

となるので

$$q(\boldsymbol{y}|\,\theta) = \Pi_{i=1}^n e^{-y_i/\theta}(1 - e^{-1/\theta}), \qquad \boldsymbol{y} = (y_1, y_2, \ldots, y_n)$$

となる．直接 $q(\boldsymbol{y}|\,\theta)$ を θ に関して最大化して，θ の最尤推定値を求めることはできるが，EM アルゴリズムを用いるとどうなるかを観てみよう．

まず，$\lfloor X_i \rfloor = y$ が与えられたとき，θ のもとでの X_i の条件付確率密度関数を求める：

$$k_\theta(x|\,y) = \frac{P(X_i = x,\, \lfloor X_i \rfloor = y)}{P(\lfloor X_i \rfloor = y)} = \frac{\theta^{-1}e^{-x/\theta}I_{[y,\,y+1)}(x)}{e^{-y/\theta}(1 - e^{-1/\theta})}$$

これより

$$\phi_1(\theta) = \mathbb{E}_{P_{\widehat{\theta}(0)}}[\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X}) = \boldsymbol{y}] = -\frac{1}{\theta}\mathbb{E}_{P_{\widehat{\theta}(0)}}\bigl[\sum_{i=1}^n X_i|T(\boldsymbol{X}) = \boldsymbol{y}\bigr] - n\log\theta$$

と

$$
\begin{aligned}
\mathbb{E}_{P_{\widehat{\theta}^{(0)}}}[X_i|T(\boldsymbol{X})=\boldsymbol{y}] &= \frac{1}{\widehat{\theta}^{(0)}e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\int_y^{y+1} xe^{-x/\widehat{\theta}^{(0)}}\,dx \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{\left[-xe^{-x/\widehat{\theta}^{(0)}}\right]_y^{y+1} + \int_y^{y+1} e^{-x/\widehat{\theta}^{(0)}}\,dx\right\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{-(y+1)e^{-(y+1)/\widehat{\theta}^{(0)}} + ye^{-y/\widehat{\theta}^{(0)}}\right. \\
&\quad \left. -\widehat{\theta}^{(0)}e^{-(y+1)/\widehat{\theta}^{(0)}} + \widehat{\theta}^{(0)}e^{-y/\widehat{\theta}^{(0)}}\right\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{(y+\widehat{\theta}^{(0)})e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}}) - e^{-(y+1)/\widehat{\theta}^{(0)}}\right\} \\
&= y+\widehat{\theta}^{(0)} - \frac{e^{-1/\widehat{\theta}^{(0)}}}{1-e^{-1/\widehat{\theta}^{(0)}}} \\
&= y+\widehat{\theta}^{(0)} - \frac{1}{e^{1/\widehat{\theta}^{(0)}}-1}
\end{aligned}
$$

となる．よって，E 段階は

$$
\phi_1(\theta) = n\left(-\log\theta + \frac{1}{\theta(e^{1/\widehat{\theta}^{(0)}}-1)} - \frac{\bar{y}_n+\widehat{\theta}^{(0)}}{\theta}\right)
$$

となる．ただし，$\bar{y}_n=(1/n)\sum_{i=1}^n y_i$ である．次に，M 段階は上の式を θ に関して最大化する：

$$
\widehat{\theta}^{(1)} = \mathrm{argmax}_\theta\phi_1(\theta) = \widehat{\theta}^{(0)} + \bar{y}_n - \frac{1}{e^{1/\widehat{\theta}^{(0)}}-1}
$$

となる．したがって，EM アルゴリズムは

$$
\widehat{\theta}^{(m)} = \mathrm{argmax}_\theta\phi_m(\theta) = \widehat{\theta}^{(m-1)} + \bar{y}_n - \frac{1}{e^{1/\widehat{\theta}^{(m-1)}}-1}
$$

で与えられる．

つぎに，EM アルゴリズムがどうしてうまく働くかを観る．$\boldsymbol{Y}=\boldsymbol{y}$ が与えられたとき，$\widehat{\theta}$ を θ の最尤推定値とし，Θ の内部上で関数

$$
\theta \mapsto q(\boldsymbol{y}|\,\theta), \qquad \theta\in\Theta
$$

は微分可能とする．このとき，

$$
\left.\frac{\partial}{\partial\theta}q(\boldsymbol{y}|\,\theta)\right|_{\theta=\widehat{\theta}} = 0
$$

である．

いま，$\widehat{\theta}^{(\infty)}$ を Θ の内点とし，EM アルゴリズムの収束先とする．このとき，$\widehat{\theta}^{(\infty)}$ は関数

$$
\theta \mapsto \mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\} \tag{2.44}
$$

を最大化する．(2.44) は Θ の内部で微分可能とし，期待値と微分記号の交換が可能とすれば，$\theta = \widehat{\theta}^{(\infty)}$ において

$$\frac{\partial}{\partial\theta}\mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X}) = \boldsymbol{y}\}\Big|_{\theta=\widehat{\theta}^{(\infty)}} = \mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\left\{\frac{\partial}{\partial\theta}\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\right\}\Big|_{\theta=\widehat{\theta}^{(\infty)}} = 0 \tag{2.45}$$

となる．ここで

$$\dot{l}_n(\theta|\,\boldsymbol{x}) = \frac{\partial}{\partial\theta}\log p(\boldsymbol{x}|\,\theta)$$

とおけば，適当な仮定のもとで

$$\begin{aligned}
\mathbb{E}_{P_\theta}\{\dot{l}_n(\theta|\,\boldsymbol{X})|T(\boldsymbol{X}) = \boldsymbol{y}\} &= \left\{\int_{T(\boldsymbol{x})=\boldsymbol{y}} \dot{l}_n(\theta|\,\boldsymbol{x})p(\boldsymbol{x}|\,\theta)\mu(d\boldsymbol{x})\right\} \\
&= \left\{\int_{T(\boldsymbol{x})=\boldsymbol{y}} \left(\frac{\partial}{\partial\theta}p(\boldsymbol{x}|\,\theta)\right)\mu(d\boldsymbol{x})\right\} \\
&= \left\{\frac{\partial}{\partial\theta}\int_{T(\boldsymbol{x})=\boldsymbol{y}} p(\boldsymbol{x}|\,\theta)\mu(d\boldsymbol{x})\right\} \\
&= \left(\frac{\partial}{\partial\theta}q(\boldsymbol{y}|\,\theta)\right) \\
&= \frac{\partial}{\partial\theta}q(\boldsymbol{y}|\,\theta)
\end{aligned}$$

となる．よって，$q(\boldsymbol{y}|\,\theta) > 0$ となる $\boldsymbol{y}$ に対して，(2.45) は

$$\frac{\partial}{\partial\theta}\log q(\boldsymbol{y}|\,\theta)\Big|_{\theta=\widehat{\theta}^{(\infty)}} = 0 \tag{2.46}$$

を意味する．したがって，最尤推定値が一意に存在するならば，$\widehat{\theta} = \widehat{\theta}^{(\infty)}$ となる．

$\theta = \widehat{\theta}^{(\infty)}$ なる解をもつ方程式

$$\mathbb{E}_{P_\theta}\{\dot{l}_n(\theta|\,\boldsymbol{X})|T(\boldsymbol{X}) = \boldsymbol{y}\} = 0$$

を自己一致方程式 (self - consistency equation) という．

以上の議論からつぎの場合には，EM アルゴリズムはうまく機能されるかは保障されていない．

1. 最大値を与える点が Θ の内点に含まれない．
2. 尤度関数がその最大を取る点で微分可能ではない．
3. スコア方程式 (2.46) が複数の解を持ち，そのいくつかは尤度関数を最大にしない．

最後に，EM アルゴリズムの各段階で，対数尤度関数 $\log q(\boldsymbol{y}|\,\theta)$ は非減少であることを示す：

$$\log q(\boldsymbol{y}|\,\widehat{\theta}^{(m)}) > \log q(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)}), \qquad m = 0, 1, \ldots \tag{2.47}$$

$T(\boldsymbol{X})=\boldsymbol{y}$ を与えたとき，$\boldsymbol{X}$ の条件付確率密度関数 $k_\theta(\boldsymbol{x}|\,\boldsymbol{y})$ は次で与えられる：

$$k_\theta(\boldsymbol{x}|\,\boldsymbol{y})=\frac{p(\boldsymbol{x}|\,\theta)}{q(\boldsymbol{y}|\,\theta)}I_{T^{-1}(\boldsymbol{y})}(\boldsymbol{x})$$

$T(\boldsymbol{x})=\boldsymbol{y}$, $p(\boldsymbol{x}|\,\theta)>0$, $q(\boldsymbol{y}|\,\theta)>0$ の場合，

$$\log q(\boldsymbol{y}|\,\theta)=\log p(\boldsymbol{x}|\,\theta)-\log k_\theta(\boldsymbol{x}|\,\boldsymbol{y})$$

となる．

以下では，最尤推定値の候補は $q(\boldsymbol{y}|\,\theta)>0$ をみたしていなければいけないので，$q(\boldsymbol{y}|\,\theta)>0$ を仮定して議論を進める．これは，そうでなければ， $\log q(\boldsymbol{y}|\,\theta)>0$ を最大にしないことからわかる．

各 m に対し

$$\begin{aligned}\log q(\boldsymbol{y}|\,\theta)&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log q(\boldsymbol{y}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\}\\&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)-\log k_\theta(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\\&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_\theta(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\end{aligned}\tag{2.48}$$

となる．上式の最右辺の各項を別々に評価していく．

まず，$\widehat{\theta}^{(m)}$ の定義から

$$\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\widehat{\theta}^{(m)})|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\widehat{\theta}^{(m-1)})|T(\boldsymbol{X})=\boldsymbol{y}\}\geq 0\tag{2.49}$$

がわかる．

次に，(2.48) の最右辺の 2 項目を評価する：

$$\begin{aligned}&\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\\&\quad=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\log\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|T(\boldsymbol{X})=\boldsymbol{y}\right\}\\&\quad\leq\log\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|T(\boldsymbol{X})=\boldsymbol{y}\right\}\end{aligned}\tag{2.50}$$

となる．最後の不等号は Jensen の不等式よりわかる．いま，

$$g(\boldsymbol{y})=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|T(\boldsymbol{X})=\boldsymbol{y}\right\}$$

とおく．条件付き期待値の定義から任意の $B\in\mathcal{B}$ に対し

$$\int_B g(\boldsymbol{y})\,dQ_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y})=\int_{T^{-1}(B)}\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\,dP_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x})\tag{2.51}$$

となる．しかし

$$\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{x}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x}|\,\boldsymbol{y})}=\frac{p(\boldsymbol{x}|\,\widehat{\theta}^{(m)})}{q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)})}\cdot\frac{q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)})}{p(\boldsymbol{x}|\,\widehat{\theta}^{(m-1)})}$$

に注意すれば，(2.51) の右辺は

$$\begin{aligned}
\int_{T^{-1}(B)} \frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\,dP_{\widehat{\theta}^{(m-1)}} &= \int_{T^{-1}(B)} \frac{p(\boldsymbol{x}|\,\widehat{\theta}^{(m)})}{q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)})} q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)})\,\mu(d\boldsymbol{x}) \\
&= \int_{T^{-1}(B)} \frac{q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m-1)})}{q(T(\boldsymbol{x})|\,\widehat{\theta}^{(m)})}\,dP_{\widehat{\theta}^{(m)}}(\boldsymbol{x}) \\
&= \int_B \frac{q(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)})}{q(\boldsymbol{y}|\,\widehat{\theta}^{(m)})}\,dQ_{\widehat{\theta}^{(m)}}(\boldsymbol{y}) \\
&= \int_B 1\cdot\, dQ_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y})
\end{aligned}$$

となる．これより

$$g(\boldsymbol{y}) = \mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}|T(\boldsymbol{X})=\boldsymbol{y}\right\} = 1, \qquad a.e. \quad Q_{\widehat{\theta}^{(m-1)}}$$

となる．この式と (2.50) から

$$\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\} - \mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\} \leq 0 \qquad (2.52)$$

がわかる．よって，(2.49) と (2.52) から (2.48) は示せた．

·····2.6·····

Cramér – Rao の不等式

　ここでは、情報不等式を証明する．まず、1 次元母数空間（これを Θ と記す）の場合を考える．確率変数 X は測度 ν に関する確率密度関数 $f(x|\theta)$ （$\theta \in \Theta$）を持つとする．$(\partial/\partial\theta)f(x|\theta)$ が存在すれば、Fisher 情報量は

$$I(\theta) = \boldsymbol{var}_\theta\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)$$

で定義される．

定理 2.9　：$\hat{\theta}(X)$ を θ の推定量とし、$\boldsymbol{E}|\hat{\theta}(X)| < \infty$ とし、$g(\theta) = \boldsymbol{E}_\theta[\hat{\theta}(X)]$ とおく．以下の正則条件を仮定する．
(D1) $f(x|\theta)$ の台 $\{x : f(x|\theta) > 0\}$ は θ に依存しない
(D2) $(\partial/\partial\theta)f(x|\theta)$ が存在する
(D3) 微分記号と $\int f(x|\theta)d\nu(x)$ と $\int \hat{\theta}(x)f(x|\theta)d\nu(x)$ の中の積分記号とが交換できる
(D4) すべての $\theta \in \Theta$ に対して、$0 < I(\theta) < \infty$ となる．
このとき、すべての θ について、

$$\boldsymbol{var}_\theta(\hat{\theta}(X)) \geq \frac{\{\dot{g}(\theta)\}^2}{I(\theta)} \tag{2.53}$$

が成立する．さらに、(2.53) の等号は $\Psi(x,\theta) = (\partial/\partial\theta)\log f(x|\theta)$ と $\hat{\theta}(x)$ が線形関係にあるときである．

証明：正則条件 (D3) より、任意の $\theta \in \Theta$ に対して

$$\boldsymbol{E}_\theta\Psi(X,\theta) = \int \frac{\partial}{\partial\theta}f(x|\theta)d\nu(x) = 0$$

と

$$\dot{g}(\theta) = \int \hat{\theta}(x)\frac{\partial}{\partial\theta}f(x|\theta)d\nu(x) = \boldsymbol{E}_\theta[\hat{\theta}(X)\Psi(X,\theta)] = \boldsymbol{cov}_\theta[\hat{\theta}(X),\Psi(X,\theta)]$$

が成り立つ．2 乗可積分な確率変数 U, V に対して

$$\{\boldsymbol{cov}(U,V)\}^2 \leq \boldsymbol{var}U \cdot \boldsymbol{var}V$$

が成り立ち、等号成立は U と V が線形関係にある場合のみであることに注意すれば、

$$\{\dot{g}(\theta)\}^2 \leq \boldsymbol{var}_\theta\hat{\theta}(X) \cdot \boldsymbol{var}_\theta\Psi(X,\theta)$$

となり、$I(\theta) = \boldsymbol{var}_\theta\Psi(X,\theta)$ から定理は証明された．　□

注意 2.1　：$\hat{\theta}$ のバイアスを $b(\theta) = g(\theta) - \theta$ とかけば、(2.53) は

$$\boldsymbol{var}_\theta(\hat{\theta}(X)) \geq \frac{\{1+\dot{b}(\theta)\}^2}{I(\theta)}$$

となり、$\hat{\theta}$ が θ の不偏推定量であれば、

$$\boldsymbol{var}_\theta(\hat{\theta}(X)) \geq \frac{1}{I(\theta)}$$

を得る．

また、$X_1, \ldots, X_n$ を $f(x|\theta)$ からのランダム標本とすれば、$\Pi_{i=1}^n f(x_i|\theta)$ の Fisher 情報量は

$$I_n(\theta) = \boldsymbol{var}_\theta\left((\partial/\partial\theta)\log \Pi_{i=1}^n f(X_i|\theta)\right) = \sum_{i=1}^n \boldsymbol{var}_\theta\big((\partial/\partial\theta)\log f(X_i|\theta)\big) = nI(\theta)$$

となり、$X_1, \ldots, X_n$ に基づく推定量 $\hat{}(X_1, \ldots, X_n)$ の情報不等式は

$$\boldsymbol{var}_\theta(\hat{\theta}(X)) \geq \frac{\{1+b(\theta)\}^2}{I(\theta)}$$

となる．

情報不等式 k – 次元母数空間 Θ の場合に拡張する．ただし、Θ は $\boldsymbol{R}^k$ の開集合とする．共分散行列を比較するために、$r \times r$ の共分散行列 Σ_1 と Σ_2 に対して、$\Sigma_1 - \Sigma_2$ が非負定値行列[1)]ならば、$\Sigma_1 \geq \Sigma_2$ と記すことにする．

定理 2.10　：$\boldsymbol{X}$ を k – 次元確率変数とし、測度 ν に関する確率密度関数 $f(\boldsymbol{x}|\theta)$ を持つとする．$\hat{\theta}(\boldsymbol{X})$ を k 次元ベクトル θ の推定量とし、$\boldsymbol{E}|\hat{\theta}(\boldsymbol{X})| < \infty$ とし、$\boldsymbol{g}(\theta) = \boldsymbol{E}\hat{\theta}(\boldsymbol{X})$ とおく．以下の正則条件を仮定する．
(D1') $f(\boldsymbol{x}|\theta)$ の台 $\{\boldsymbol{x} : f(\boldsymbol{x}|\theta) > 0\}$ は θ に依存しない
(D2') $(\partial/\partial\theta) f(\boldsymbol{x}|\theta)$ が存在する
(D3') 微分記号と $\int f(\boldsymbol{x}|\theta) d\nu(\boldsymbol{x})$ と $\int \hat{\theta}(\boldsymbol{x}) f(\boldsymbol{x}|\theta) d\nu(\boldsymbol{x})$ の中の積分記号とが交換できる
(D4') すべての $\theta \in \Theta$ に対して、$0 < \boldsymbol{I}(\theta) < \infty$ となる．ただし、

$$\boldsymbol{I}(\theta) : k \times k = \left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) \frac{\partial}{\partial\theta^T}\log f(\boldsymbol{x}|\theta)\right)$$

である．
このとき、すべての θ について、

$$\boldsymbol{var}_\theta(\hat{\nu}(X)) \geq \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T\theta)$$

が成立する．ただし、

$$\dot{\boldsymbol{g}}(\theta) : k \times k = \frac{\partial}{\partial\theta}\boldsymbol{g}^T(\theta)$$

とする．

[1)]すなわち、すべての r – 次元ベクトルに対して、
$a'(\Sigma_1 - \Sigma_2)a \geq$ が成立することである．

証明：前の定理の証明中の議論と同様にすれば、

$$\boldsymbol{E}_\theta \Psi(\boldsymbol{X},\theta) : k\times 1 = 0 \qquad \text{と} \qquad \dot{\boldsymbol{g}}(\theta) : k\times k = \boldsymbol{cov}_\theta[\hat{\theta}(\boldsymbol{X}), \Psi(\boldsymbol{X},\theta)]$$

を得る．したがって、

$$\begin{aligned}
&\boldsymbol{var}_\theta(\hat{\theta}(\boldsymbol{X}) - \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\Psi(\boldsymbol{X},\theta)) \\
&\quad = \boldsymbol{var}_\theta\hat{\theta}(\boldsymbol{X}) - 2\boldsymbol{cov}_\theta[\hat{\theta}(\boldsymbol{X}), \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\Psi(\boldsymbol{X},\theta)] + \boldsymbol{var}_\theta(\dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\Psi(\boldsymbol{X},\theta)) \\
&\quad = \boldsymbol{var}_\theta\hat{\theta}(\boldsymbol{X}) - 2\boldsymbol{cov}_\theta[\hat{\theta}(\boldsymbol{X}), \Psi(\boldsymbol{X},\theta)]\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T(\theta) \\
&\qquad + \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)(\boldsymbol{var}_\theta\Psi(\boldsymbol{X},\theta))\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T(\theta) \\
&\quad = \boldsymbol{var}_\theta\hat{\theta}(\boldsymbol{X}) - 2\dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T(\theta) + \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\boldsymbol{I}(\theta)\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T(\theta) \\
&\quad = \boldsymbol{var}_\theta\hat{\theta}(\boldsymbol{X}) - \dot{\boldsymbol{g}}(\theta)\boldsymbol{I}^{-1}(\theta)\dot{\boldsymbol{g}}^T(\theta) \geq 0
\end{aligned}$$

からわかる．　□

·····2.7·····

Neyman-Pearon の基本定理について

定理 2.11　(Neyman-Peason の基本定理)：P_0 と P_1 を確率分布とし，それぞれはある測度 μ[1]に関する確率密度関数 p_0 と p_1 を持つこととする．
(i) 存在：検定問題 $H_0 : p_0$，$H_1 : p_1$ に対して，ある検定関数 ψ と定数 k が存在し，

$$\mathbb{E}_0\psi(X) = \alpha \tag{2.54}$$

と

$$\psi(x) = \begin{cases} 1, & (p_1(x) > kp_0(x) \text{ のとき}) \\ 0, & (p_1(x) < kp_0(x) \text{ のとき}) \end{cases} \tag{2.55}$$

を満足する．
(ii) 最強力検定のための十分条件：ある検定関数がある定数 k に対して，(2.54) と (2.55) を満足するならば，その検定関数は最強力検定になる．
(iii) 最強力検定のための必要条件：もし，ψ が検定問題 $H_0 : p_0$，$H_1 : p_1$ の有意水準 α の最強力検定ならば，ψ はある k に対して (2.55) を満足する．また，検定のサイズが α より小さく検出力が *1* なる検定が存在しなければ，ψ は (2.54) も満足する．

証明：(i) $\alpha(c) = P_0\{p_1(X) > cp_0(X)\}$ とおく．確率は P_0 のもとで計算されるので，上の式の不等式は $\{x : p_0(x) > 0\}$ 上で定義されればよい．すると，$1 - \alpha(c)$ は確率変数 $p_1(X)/p_0(X)$ の分布関数となる[2]．さらに，$\alpha(c)$ は非増加，右連続で

$$\alpha(c-0) - \alpha(c) = P_0\{\frac{p_1(X)}{p_0(X)} = c\}, \qquad \alpha(-\infty) = 1, \qquad \alpha(+\infty) = 0$$

となる．与えられた α_0 に対して c_0 を

$$\alpha(c_0) \leq \alpha_0 \leq \alpha(c_0)$$

を満足するようにとり，検定関数 ψ を

$$\psi(x) = \begin{cases} 1, & (p_1(x) > c_0p_0(x) \text{ のとき}) \\ \frac{\alpha_0 - \alpha(c_0)}{\alpha(c_0-0) - \alpha(c_0)}, & (p_1(x) = c_0p_0(x) \text{ のとき}) \\ 0, & (p_1(x) < c_0p_0(x) \text{ のとき}) \end{cases}$$

---

1) 一般には，μ を $P_0 + P_1$ とすればよい．

2) $P_0\{p_0(X) = 0\} = 0$ より．

で定義する．上の式の 2 番目の項は $\alpha(c_0-0)-\alpha(c_0)=0$ の場合以外は定義される．また，$\alpha(c_0-0)-\alpha(c_0)=0$ の場合は $P_0\{p_1(X)=c_0p_0(X)\}=0$ となるので，ψ はほとんどいたるところで定義される．さらに，ψ のサイズは

$$\mathbb{E}_0\psi(X)=P_0\{\frac{p_1(X)}{p_0(X)}>c_0\}+\frac{\alpha_0-\alpha(c_0)}{\alpha(c_0-0)-\alpha(c_0)}P_0\{\frac{p_1(X)}{p_0(X)}=c_0\}=\alpha_0$$

となるように c_0 をとればよい[3)]．
(ii) $\mathbb{E}_0\psi^*(X)$ なるどんな検定関数 ψ^* にたいしても

$$\mathbb{E}_0\psi(X)\leq\mathbb{E}\psi^*(X)$$

が成立することを示せばよい．いま，

$$S^+=\{x:\psi(x)-\psi^*(x)>0\},\qquad S^-=\{x:\psi(x)-\psi^*(x)<0\}$$

とおく．$x\in S^+$ ならば[4)]，$p_1(x)>kp_0(x)$ となる．同様に，$x\in S^-$ ならば[5)]，$p_1(x)<kp_0(x)$ となる．これらから $S^*\cup S^-$ 上で

$$(\psi(x)-\psi^*(x))(p_1(x)-kp_0(x))\geq 0$$

となる．したがって

$$\int(\psi(x)-\psi^*(x))(p_1(x)-kp_0(x))\,d\mu=\int_{S^*\cup S^-}(\psi(x)-\psi^*(x))(p_1(x)-kp_0(x))\,d\mu\geq 0$$

となる．これから

$$\int(\psi(x)-\psi^*(x))p_1(x)\,d\mu\geq k\int(\psi(x)-\psi^*(x))p_0(x)\,d\mu=k\{\mathbb{E}_0[\psi(X)-\psi^*(X)]\geq 0$$

よって，(ii) は証明された．
(iii) ψ^* を最強力検定とする．また，ψ を (2.54) と (2.55) を満足する検定とする．

$$S=(S^*\cup S^-)\cap\{x:p_1(x)\not kp_0(x)\}$$

とする．すなわち，$S^*\cup S^-$ 上で ψ と ψ^* の値がことなる点である．さらに，$\mu(S)>0$ とする．すると

$$\int_{S^+\cup S^-}(\psi-\psi^*)(p_1-kp_0)\,d\mu=\int_S(\psi-\psi^*)(p_1-kp_0)\,d\mu>0$$

となり，ψ が最強力検定であることに矛盾する．したがって，$\mu(S)=0$ となる．　□

---

3) これは α の非増加性と $\alpha(-\infty)=1$, $\alpha(+\infty)=0$ から保障される．もし，α_0 を挟んで α がジャンプしていれば，上の式の右辺の 2 項目がジャンプする確率の内分点になっていることに注意.

4) $\psi(x)>\psi^*(x)$ となり，$\psi^*(x)$ は 0 または 1 なので，$psi(x)=1$ となり，

$$\psi(x)=1\Longleftrightarrow p_1(x)>kp_0(x)$$

となる.

5) $\psi(x)<\psi^*(x)$ となり，$\psi^*(x)$ は 0 または 1 なので，$psi(x)=0$ となり，

$$\psi(x)=0\Longleftrightarrow p_1(x)<kp_0(x)$$

となる.

3

母数モデルにおける有効推定と検定

·····3.1·····

一様強一致性

$X_1, X_2, \ldots, X_n$ を独立な確率変数列とし, 同一の分布 $F(x)$ をもつとする．Θ を母数空間とし, $U(x, \theta)$ を x の実数値関数とする．推定や検定において,

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta)$$

の評価が重要である．θ にたいし,

$$\mu(\theta) = \boldsymbol{E}U(X_1, \theta) = \int U(x, \theta) dF(x) < \infty$$

とく．大数の法則から, 各 θ に対し,

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) \xrightarrow{as} \mu(\theta) \qquad (n \to \infty) \tag{3.56}$$

が成立する．しかし,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta) \right| \xrightarrow{as} 0 \qquad (n \to \infty) \tag{3.57}$$

が成立すると便利である．

(3.57) が成立すると仮定し, 統計量の列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ が $\hat{\theta}_n \xrightarrow{as} \theta_0$ を満足するとする． さらに, $\mu(\theta)$ は θ の連続関数であるとする．すると, (3.57) は

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) \xrightarrow{as} \mu(\theta_0) \qquad (n \to \infty)$$

を保障することがわかる．実際,

$$\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\theta_0) \right| &\leq \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\hat{\theta}_n) \right| + \left| \mu(\hat{\theta}_n) - \mu(\theta_0) \right| \\
&\leq \sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta) \right| + \left| \mu(\hat{\theta}_n) - \mu(\theta_0) \right| \\
&\xrightarrow{as} 0, \qquad (n \to \infty)
\end{aligned}$$

からわかる．

以下, (3.57) が成立するための十分条件を求めることにする．

Θ 上の実数値関数 $g(\theta)$ が**上半連続**（upper semicontineous）であるとは, $\theta_n \to \theta$ なるどんな列 $\{\theta_n\}_{n=1}^{\infty}$ に対しても,

$$\limsup_{n\to\infty} g(\theta_n) \le g(\theta)$$

が成立すること[1]である．

上半連続関数の例

補題 3.1　：つぎを仮定する．

(A1)　Θ はコンパクト

(A2)　すべての x に対し, $U(x,\theta)$ は θ について上半連続

(A3)　ある関数 $K(x)$ で $\boldsymbol{E}_F K(X) < \infty$ なるものが存在し, すべての θ に対し,

$$|U(x,\theta)| \le K(x)$$

を満足

(A4)　すべての θ と十分小さなすべての $\rho > 0$ に対し,

$$\sup_{|\theta'-\theta|<\rho} U(x,\theta')$$

は x の可測関数である．

このとき,

$$P\left\{\limsup_{n\to\infty}\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}U(X_i,\theta) \le \sup_{\theta\in\Theta}\mu(\theta)\right\} = 1$$

が成立する．

証明：

$$\varphi(x,\theta,\rho) = \sup_{|\theta'-\theta|<\rho} U(x,\theta')$$

とおく．仮定から, 十分小さな $\rho > 0$ に対し, φ は x の可測関数である．また, $\varphi(x,\theta,\rho) \le K(x)$ で $\rho \downarrow 0$ のとき, $\rho(x,\theta,\rho) \downarrow U(x,\theta)$ である．したがって, 単調収束定理を用いれば, $\rho \downarrow 0$ のとき,

$$\int \varphi(x,\theta,\rho)dF(x) \downarrow \int U(x,\theta)dF(x) = \mu(\theta)$$

となる．任意の正の数 $\epsilon > 0$ を取る．それぞれの $\theta \in \Theta$ に対し, $\rho_\theta > 0$ が存在して,

$$\int \varphi(x,\theta,\rho_\theta)dF(x) < \mu(\theta) + \epsilon$$

[1] これは $\sup_{|\theta'-\theta|<\rho} f(\theta') \to f(\theta)(\rho \to 0)$ と同値であ　．
る．また, $-f$ が上半連続のとき, f は下連続という

とできる．各 $\theta \in \Theta$ に対して, 開球

$$S(\theta, \rho_\theta) = \{\theta' : |\theta' - \theta| < \rho_\theta\}$$

をとれば, Θ はコンパクトより有限開被覆の存在が保障される：すなわち, ある正の整数 m が存在して,

$$\Theta \subset \cup_{i=1}^m S(\theta_j, \rho_{\theta_j})$$

とできる．どんな $\theta \in \Theta$ に対してもある $i \in \{1, 2, \ldots, m\}$ が存在し, $\theta \in S(\theta_i, \rho_{\theta_i})$ となる．φ の定義からどんな x に対しても

$$U(x, \theta) \leq \varphi(x, \theta_i, \rho_{\theta_i})$$

である．したがって,

$$\frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \leq \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

となることから

$$\sup_{\theta \in \Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \leq \sup_{1 \leq i \leq m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

が成り立つ．上の式の右辺に対して大数の法則を適用すれば,

$$P\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i}) \leq \mu(\theta_i) + \epsilon\right\} = 1$$

が成り立つ．よって,

$$P\left\{\lim_{n\to\infty} \sup_{1 \leq i \leq m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i}) \leq \sup_{1 \leq i \leq m} \mu(\theta_i) + \epsilon\right\} = 1$$

となる．さらに,

$$\sup_{1 \leq i \leq m} \mu(\theta_i) \leq \sup_{\theta \in \Theta} \mu(\theta)$$

と

$$\limsup_{n\to\infty} \sup_{\theta \in \Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \leq \lim_{n\to\infty} \sup_{1 \leq i \leq m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

に注意すれば,

$$P\left\{\limsup_{n\to\infty} \sup_{\theta \in \Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \leq \sup_{\theta \in \Theta} \mu(\theta) + \epsilon\right\} = 1$$

が成り立つことがわかる．よって, 補題は証明された． □

定理 3.1　：補題の仮定 $(A1)$, $(A3)$ の他に, つぎを仮定する．
(A5) すべての x に対し, $U(x,\theta)$ は θ の連続関数である．
このとき,

$$P\left\{\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}U(X_j,\theta)-\mu(\theta)\right|=0\right\}=1$$

が成立する．

証明：(A5) から補題の (A2) は保障される．また, (A3) と U の連続性に注意して有界収束定理を用いれば,

$$\lim_{\theta'\to\infty}\mu(\theta')=\lim_{\theta'\to\theta}\int U(x,\theta')dF(x)=\int U(x,\theta)dF(x)=\mu(\theta)$$

となり, μ は連続となることがわかる．さらに,

$$\tilde{U}(x,\theta)\equiv U(x,\theta)-\mu(\theta)$$

とすれば, $\boldsymbol{E}\tilde{U}(X,\theta)=0$ となる．また, すべての固定した x に対して, $\tilde{U}(x,\theta)$ は θ について連続なので, $\tilde{U}(x,\theta)$ と $-\tilde{U}(x,\theta)$ に対して前の補題を適用すれば,

$$P\left\{\limsup_{n\to\infty}\sup_{\theta\in\Theta}\frac{1}{n}\sum_{j=1}^{n}\tilde{U}(X_j,\theta)\le 0\right\}=1$$

と

$$P\left\{\limsup_{n\to\infty}\sup_{\theta\in\Theta}-\frac{1}{n}\sum_{j=1}^{n}\tilde{U}(X_j,\theta)\le 0\right\}=1$$

を得る．任意の関数 g に対して,

$$0\le\sup_{\theta}|g(\theta)|=\max\{\sup_{\theta}g(\theta),\sup_{\theta}(-g(\theta))\}$$

であることに注意すれば,

$$P\left\{\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}\tilde{U}(X_j,\theta)\right|=0\right\}=1$$

がわかる．よって, 定理は示された．　□

·····3.2·····

最尤推定量の強一致性と漸近正規性

$f(x|\theta)$ を $\sigma-$ 有限[1] な測度（通常は Lebesgue 測度か counting measure）に関する確率密度関数とする．$X_1, X_2, \ldots, X_n$ を $f(x|\theta)$ からのランダム標本としたとき, **尤度関数**を

$$L_n(\theta) = \Pi_{i=1}^n f(x_i|\theta)$$

で定義する．また, **対数尤度関数**を

$$\Lambda_n(\theta) = \log L_n(\theta)$$

と記すことにする．θ の**最尤推定値** $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$ を

$$L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta)$$

で定義する．$\hat{\theta}$ は一般には存在するかどうかはわからないが, Θ がコンパクトで $f(x|\theta)$ が θ について上半連続ならば, 存在することがわかる．

3.2.1 強一致性

定義 3.1　：母数 $\theta \in \Theta$ の推定量の列 $\{\hat{\theta}_n\}_{n=1}^\infty$ が θ の**弱一致推定量**（**強一致推定量**）であるとは, すべての $\theta \in \Theta$ に対して, $\hat{\theta}_n \xrightarrow{P} \theta$ （$\hat{\theta}_n \xrightarrow{as} \theta$ ）を満足することである．

一致性の議論を行うために以下の記号を準備する．f_0 と f_1 を $\sigma-$ 有限な測度 λ に関する確率密度関数とする．**カルバック – ライブラー情報量**を

$$K(f_0, f_1) = \boldsymbol{E}_{f_0} \log \frac{f_0(X)}{f_1(X)} = \int \log\left(\frac{f_0(x)}{f_1(x)}\right) f_0(x) d\lambda(x)$$

で定義する．ただし,

$$\log \frac{f_0(x)}{f_1(x)} = \begin{cases} \log \frac{f_0(x)}{f_1(x)} & (f_1 > 0) \\ 0 & (f_0 = f_1 = 0) \\ \infty & (f_0 > f_1 = 0) \end{cases}$$

と定義する． したがって, $K(f_0, f_1) = \infty$ となることもある． また, $f_0 = 0, f_1 > 0$ の場合, $\int \log \frac{f_0(x)}{f_1(x)} f_0(x) d\lambda(x) = 0$ とする．

[1] 測度 ν が $\sigma-$ 有限であるとは, $\Omega = \cup_{i=1}^\infty A_i, A_i \in \mathcal{A}, \nu(A_i) < \infty$ を満足することである．ルベーグ測度 $\lambda((a, b]) = b - a$ は $A_i = (-i, i]$ ととれば, $\sigma-$ 有限であることがわかる．

補題 3.2　：f_0 と f_1 を λ に関する確率密度関数とする．このとき,

$$K(f_0, f_1) = \int \log\left(\frac{f_0(x)}{f_1(x)}\right) f_0(x) d\lambda(x) \geq 0$$

が成立する．ただし, 等号成立は $f_1(x) = f_0(x)(a.e.\lambda)$ のときである．

証明：$\log x$ は凸関数だから, Jensen の不等式を用いれば,

$$-K(f_0, f_1) = \boldsymbol{E}_{f_0} \log \frac{f_1(X)}{f_0(X)} \leq \log \boldsymbol{E}_{f_0} \frac{f_1(X)}{f_0(X)}$$

である．ただし, 等号成立は, $X \sim f_0$ としたとき, $f_1(X)/f_0(X)$ が定数（a.e.）である場合のみである．しかし,

$$\boldsymbol{E}_{f_0} \frac{f_1(X)}{f_0(X)} = \int \frac{f_1(x)}{f_0(x)} f_0(x) d\lambda(x) = \int_{\{x: f_0(x) > 0\}} f_1(x) d\lambda(x) \leq 1$$

となる．ただし, 等号成立は $\boldsymbol{E}_1 1\{f_0(X) > 0\} = 1$ である．このふたつの式を合わせれば,

$$-K(f_0, f_1) \leq 0$$

が成り立つことがわかる． □

定理 3.2　：$\theta_0 \in \Theta$ とし, $X_1, X_2, \ldots, X_n$ を確率密度関数 $f(x|\theta_0)$ からのランダム標本とする．θ_0 を真の母数とし, 以下を仮定する．

(B1) Θ はコンパクト

(B2) すべての x について, $f(x|\theta)$ は θ について上半連続とする

(B3) 関数 $K(x)$ が存在し, $\boldsymbol{E}_{\theta_0}|K(X)| < \infty$ で, すべての x と θ に対し,

$$U(x, \theta) = \log f(x|\theta) - \log f(x|\theta_0) \leq K(x)$$

を満足する

(B4) すべての $\theta \in \Theta$ と十分小さな $\rho > 0$ に対して, $\sup_{|\theta' - \theta| < \rho} f(x|\theta')$ は x の可測関数である

(B5) $f(x|\theta) = f(x|\theta_0)(a.e.\lambda)$ ならば, $\theta = \theta_0$ が成立する

このとき, θ の最尤推定量の列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ は

$$\hat{\theta}_n \xrightarrow{as} \theta_0$$

を満足する．

証明：$\rho > 0$ を取り,

$$S = \{\theta : |\theta - \theta_0| \geq \rho\}$$

とする．条件 (B1) から S もコンパクトとなり, 補題 2.1 より

$$P\left\{\limsup_{n\to\infty} \sup_{\theta \in S} \frac{1}{n} \sum_{j=1}^{n} U(X_j, \theta) \leq \sup_{\theta \in S} \mu(\theta)\right\} = 1$$

である．ただし, $\theta \in S$ に対して,

$$\mu(\theta) = -K(\theta_0, \theta) = \int U(x,\theta) f(x|\theta_0) d\lambda(x) < 0$$

である．有界収束定理を使えば, $\mu(\theta)$ は上半連続であることがわかる．したがって, S 上で最大値を取る．$\delta = \sup_{\theta\in S} \mu(\theta) < 0$ とする．このとき,

$$P\left\{\limsup_{n\to\infty} \sup_{\theta\in S} \frac{1}{n}\sum_{j=1}^{n} U(X_j,\theta) \le \delta\right\} = 1$$

と書ける．したがって, ある正の整数 N が存在して, どんな $n > N$ に対しても

$$\sup_{\theta\in S} \frac{1}{n}\sum_{j=1}^{n} U(X_j,\theta) \le \frac{\delta}{2} < 0 \tag{3.58}$$

とほとんど確実にできる．しかし, $\hat{\theta}_n$ の定義から

$$\frac{1}{n}\sum_{j=1}^{n} U(X_j,\hat{\theta}_n) = \sup_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^{n} U(X_j,\theta) \ge 0 \tag{3.59}$$

であり[2], (3.58) と矛盾する．これより, $n > N$ に対して, $\hat{\theta}_n \notin S$ となり, $|\hat{\theta}_n - \theta| < \rho$ となる．ρ は任意なので, 定理は証明された．□

注意 3.1　：この定理は一様分布 $U(\theta,\theta+1)$ の場合も含む．すなわち, $f(x|\theta) = 1_{[\theta,\theta+1]}$ よりわかる．

3.2.2 漸近正規性

$\theta \in \Theta \subset \boldsymbol{R}^k$ とし, モデルを $\boldsymbol{P} = \{f(x|\theta) : \theta \in \Theta\}$ とする．$(\partial/\partial\theta)\log f(x|\theta)$ が存在すれば, 最尤推定量 $\hat{\theta}_n$ は方程式

$$\dot{l}_n(\theta) = \frac{\partial}{\partial\theta}\log L_n(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(x_i|\theta) = 0$$

の解である．ただし, 解は複数あるかもしれない．

いま,

$$\Psi(x,\theta) : k\times 1 = \left(\frac{\partial}{\partial\theta_i}\log f(x|\theta)\right)$$

[2] (3.59) は

$$\sup_{\theta\in S}\left\{\frac{1}{n}\sum_{j=1}^{n}\log f(X_j,\theta) - \frac{1}{n}\sum_{j=1}^{n}\log f(X_j,\theta_0)\right\} \ge \frac{1}{n}\sum_{j=1}^{n}\log f(X_j,\theta_0) - \frac{1}{n}\sum_{j=1}^{n}\log f(X_j,\theta_0) = 0$$

よりわかる.

と

$$\dot{\Psi}(x,\theta) : k \times k = \left(\frac{\partial^2}{\partial\theta_i \partial\theta_j^T} \log f(x|\theta)\right)$$

とおく．ただし, $\theta = (\theta_1, \ldots, \theta_k)^T$ とした．このとき, **Fisher 情報量**を

$$I(\theta) : k \times k = \boldsymbol{E}_\theta[\Psi(x,\theta)\Psi^T(x,\theta)]$$

で定義する．積分と微分記号の交換[3)]が保障されるとき,

$$\boldsymbol{E}_\theta \Psi(x,\theta) = \int \frac{(\partial/\partial\theta) f(x|\theta)}{f(x|\theta)} f(x|\theta) d\lambda(x) = \int \frac{\partial}{\partial\theta} f(x|\theta) d\lambda(x) = 0$$

となる．したがって,

$$I(\theta) = \boldsymbol{var}[\Psi(X,\theta)]$$

である．再度, 積分と微分記号の交換を許せば, $\int (\partial^2/\partial\theta\partial\theta^T) f(x|\theta)\lambda(x) = 0$ となることに注意すれば,

$$\begin{aligned}
\boldsymbol{E}_\theta[\dot{\Psi}(x,\theta)] &= \int \left[\frac{\partial}{\partial\theta} \frac{(\partial/\partial\theta^T) f(x|\theta)}{f(x|\theta)}\right] f(x|\theta) d\lambda(x) \\
&= \int \frac{f(x|\theta)(\partial^2/\partial\theta\partial\theta^T) f(x|\theta) - \{(\partial/\partial\theta) f(x|\theta)\}\{(\partial/\partial\theta) f(x|\theta)\}^T}{f(x|\theta)^2} f(x|\theta) d\lambda(x) \\
&= 0 - \int \Psi(x,\theta)\Psi^T(x,\theta) f(x|\theta) d\lambda(x)
\end{aligned}$$

となる．すなわち,

$$I(\theta) = -\boldsymbol{E}_\theta \dot{\Psi}(x,\theta)$$

である．

定理 3.3　：$X_1, X_2, \ldots, X_n$ を（測度 λ に関する）確率密度関数 $f(x|\theta_0)$ からのランダム標本とする．また, 真の母数を θ_0 と記すことにする．前定理の仮定 (B5) と以下を仮定する．
(C1) Θ は $\boldsymbol{R}^k$ の開集合とする
(C2) θ に関する $f(x|\theta)$ の 2 階偏導関数は存在し, すべての x に関して, θ の連続関数となり, 微分記号 $(\partial/\partial\theta)$ と 積分記号 $\int f(x|\theta) d\lambda(x)$ の交換が可能である
(C3) ある関数 $K(x)$ が存在し, $\boldsymbol{E}_{\theta_0}|K(X)| < \infty$ であり, θ_0 のある近傍で一様に,

$$|\dot{\Psi}_{ij}(x,\,\theta)| \le K(x)$$

が成立する．ただし, $\dot{\Psi}(x,\,\theta) = (\dot{\Psi}_{ij}(x,\,\theta))$ である．
(C4) $I(\theta_0) = -\boldsymbol{E}_{\theta_0}[\dot{\Psi}(x,\,\theta_0)]$ は正定値である
このとき, 尤度方程式の解 $\hat{\theta}_n$ は θ_0 の強一致推定量で,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0))$$

が成立する．

---

3) すなわち, $(\partial/\partial\theta)\int f(x|\theta) d\lambda(x) = \int (\partial/\partial\theta) f(x|\theta) d\lambda(x)$ が成立することである．

証明：一致性．ある正の数 $\rho >$ に対し, $S_\rho = \{\theta : |\theta - \theta_0| \le \rho\}$ とし, S_ρ 上で $\dot{\Psi}$ の成分が $K(x)$ で有界になるように ρ をとる．定理 2.2 の (B1), (B2), (B4) は自動的に保障される．(B4) は $f(x|\theta)$ の連続性から確認できる．(B3) を確かめるために, $U(x,\theta)$ を θ_0 のまわりで展開する：

$$U(x,\theta) = U(x,\theta_0) + \Psi^T(x,\theta_0)(\theta-\theta_0) + (\theta-\theta_0)^T \int_0^1 \int_0^1 t\dot{\Psi}(x,\theta_0 + u(\theta-\theta_0))dtdu(\theta-\theta_0)$$

となる．$\boldsymbol{E}_{\theta_0}|\Psi(x,\theta_0)| < \infty$ であり, S_ρ 上で $\dot{\Psi}$ の各成文も $K(x)$ で有界である．したがって,

$$|U(x,\,\theta)| \le \rho^2 |\Psi(x,\,\theta_0)| + \rho^2 K(x)$$

となり, 右辺は θ_0 のもとで可積分である．よって, (B3) は確認されたので, 定理 2.2 より, 強一致性がわかる．

漸近正規性．$\dot{l}_n(\theta) = \sum_{j=1}^n \Psi(X_j,\theta)$ を θ_0 のまわりで展開する：

$$\dot{l}_n(\theta) = \dot{l}_n(\theta_0) + \int_0^1 \sum_{j=1}^n \dot{\Psi}(X_j, \theta_0 + t(\theta-\theta_0))dt(\theta-\theta_0)$$

$\theta = \hat{\theta}_n$ を上の式に代入し, 両辺を $\sqrt{n}$ で割れば,

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) = -B_n\sqrt{n}(\hat{\theta}_n - \theta_0)$$

を得る．ただし,

$$B_n = \frac{1}{n}\int_0^1 \sum_{j=1}^n \dot{\Psi}(X_j, \theta_0 + t(\theta-\theta_0))dt$$

である．$\boldsymbol{E}_{\theta_0}\Psi(X,\theta_0) = 0$ と $\boldsymbol{var}_{\theta_0}(\Psi(X,\theta_0)) = I(\theta_0)$ に注意して, 中心極限定理を適用すれば,

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^n \Psi(X_j,\theta_0)\right) \xrightarrow{L} Z \sim N(0, I(\theta_0))$$

となる．もし, $B_n \xrightarrow{as} -I(\theta_0) < 0$ ならば, ほとんど確実に B_n^{-1} は有界なので, Slutsky の定理を用いれば,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -B_n^{-1}\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) \xrightarrow{L} I^{-1}(\theta_0)Z \sim N(0, I^{-1}(\theta_0))$$

がわかる．

最後に, $B_n \xrightarrow{as} -I(\theta_0)$ を示す. まず, $\boldsymbol{E}_{\theta_0}\dot{\Psi}(X,\theta)$ が θ_0 に関して連続であることを示す．実際, (C3) と有界収束定理から

$$\lim_{\theta'\to\theta_0} \boldsymbol{E}_{\theta_0}\dot{\Psi}(X,\theta') = \boldsymbol{E}_{\theta_0}\lim_{\theta'\to\theta_0}\dot{\Psi}(X,\theta') = \boldsymbol{E}_{\theta_0}\dot{\Psi}(X,\theta_0)$$

よりわかる．これより, どんな正の数 $\epsilon > 0$ に対してもある正の数 $\rho > 0$ が存在して, $|\theta - \theta_0| < \rho$ ならば,

$$|\boldsymbol{E}_{\theta_0}\dot{\Psi}(X,\theta) + I(\theta_0)| < \epsilon$$

とできる[4)]．定理 2.1 と大数の法則から，ある正の整数 N が存在して, $n > N$ にならば

$$\sup_{\theta \in S_\rho} \left| \frac{1}{n} \sum_{j=1}^{n} \dot{\Psi}(X_j, \theta) - \boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta) \right| < \epsilon$$

となることがわかる．必要ならば, N をさらに大きくすることで, $n > N$ ならば, $|\hat{\theta}_n - \theta_0| < \rho$ とできることが $\hat{\theta}_n$ の一致性から保障される．したがって, $n > N$ のとき,

$$\begin{aligned}
|B_n + I(\theta_0)| &\leq \int_0^1 \left| \frac{1}{n} \sum_{j=1}^{n} \dot{\Psi}(X_j, \theta_0 + t(\hat{\theta}_n - \theta_0)) + I(\theta_0) \right| dt \\
&\leq \int_0^1 \left[\sup_{\theta \in S_\rho} \left| \frac{1}{n} \sum_{j=1}^{n} \dot{\Psi}(X_j, \theta) - \boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta) \right| + |\boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta) + I(\theta_0)| \right] dt \\
&\leq 2\epsilon
\end{aligned}$$

となることがわかる．よって, 定理は証明された．　□

4) $I(\theta_0) = -\boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta_0)$ であることに注意せよ．

·····3.3·····

One-Step 推定量の漸近分布

適当な条件のもとで、最尤推定量は有効であることがわかるが、尤度方程式を陽に解くことができない場合もある．その場合には、初期推定量として、モーメント推定量を使い、Newton 法により、近似的に尤度方程式を解くことを考える．

まず、$\hat{\theta}_n^{(0)}$ を初期推定量とする．繰り返し

$$\hat{\theta}_n^{(k+1)} = \hat{\theta}_n^{(k)} - \ddot{l}_n(\hat{\theta}_n^{(k)})\dot{l}(\hat{\theta}_n^{(k)}), \qquad k = 0, 1, 2, \ldots$$

によって、有効推定量を得ることが期待できる．通常、$n \uparrow \infty$ とすれば、$(1/n)\ddot{l}_n(\hat{\theta}_n^{(k)}) \to -I(\theta)$ に収束することが期待できるので、

$$\hat{\theta}_n^{(k+1)} = \hat{\theta}_n^{(k)} + I^{-1}(\hat{\theta}_n^{(k)})(1/n)\dot{l}_n(\hat{\theta}_n^{(k)}), \qquad k = 0, 1, 2, \ldots$$

なる繰り返し法により推定量をもとめることを考える．これを**スコア法**とよぶ．

定理 3.4　：$\tilde{\theta}_n$ を θ の強一致推定量とし、

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{L} N(0, \sigma^2(\theta))$$

が成立するものとする．ただし、すべての $\theta \in \Theta$ に対し、$\Sigma(\theta) < \infty$ である．このとき、前定理の仮定のもとで、

$$\tilde{\theta}_n^{(1)} = \tilde{\theta}_n - \{\ddot{l}_n(\tilde{\theta})\}^{-1}\dot{l}(\tilde{\theta}_n)$$
$$\theta_n^* = \tilde{\theta}_n + I^{-1}(\tilde{\theta}_n)\frac{1}{n}\dot{l}(\tilde{\theta}_n)$$

は最尤推定量と漸近同値である．したがって、これらのふたつの有効推定量となる．

証明：$\hat{\theta}_n$ を θ の強一致推定量とし、

$$\dot{l}_n(\hat{\theta}_n) = 0$$

を満足すると仮定する．$\dot{l}_n(\tilde{\theta}_n)$ を $\hat{\theta}_n$ のまわりで展開する：

$$\dot{l}_n(\tilde{\theta}) = \dot{l}_n(\hat{\theta}_n) + \int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\tilde{\theta}_n - \hat{\theta}_n))dt(\tilde{\theta}_n - \hat{\theta})$$

を得る．ここで、$\tilde{\theta}_n^{(1)} - \hat{\theta}_n = \tilde{\theta}_n - \hat{\theta}_n - \{\ddot{l}_n(\tilde{\theta}_n)\}^{-1}\dot{l}_n(\tilde{\theta}_n)$ を使えば、

$$\sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) = \left[\mathrm{id} - \{\ddot{l}_n(\tilde{\theta}_n)\}^{-1}\int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\tilde{\theta}_n - \hat{\theta}_n))dt\right]\sqrt{n}(\tilde{\theta}_n - \hat{\theta})$$

となる．定理 2.3 の証明の議論から

$$\ddot{l}_n(\tilde{\theta}_n) \xrightarrow{as} -I(\theta)$$

と

$$\int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\tilde{\theta}_n - \hat{\theta}_n))dt \xrightarrow{as} -I(\theta)$$

が成り立つことがわかる．これらより

$$\mathrm{id} - \{\ddot{l}_n(\tilde{\theta}_n)\}^{-1} \int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\tilde{\theta}_n - \hat{\theta}_n))dt \xrightarrow{as} 0$$

を得る．また、

$$\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) = \sqrt{n}(\hat{\theta}_n - \theta) - \sqrt{n}(\tilde{\theta}_n - \theta) = O_P(1)$$

である．これらから

$$\sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n) \xrightarrow{P} 0$$

を得る．したがって、$\tilde{\theta}_n^{(1)}$ と $\hat{\theta}_n$ が漸近同値であることが示せた．θ_n^* と $\hat{\theta}_n$ も漸近同値であることは同様に示すことができる．□

·····3.4·····

尤度比統計量の漸近分布

$X_1, X_2, \ldots, X_n$ を確率密度関数 $f(x|\theta)$ からのランダム標本とする．ただし, $\theta \in \Theta \subset \boldsymbol{R}^k$ とする．Θ_0 を Θ の部分集合とし, 仮説検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

を考える．尤度比統計量

$$\Lambda_n = \frac{\sup_{\theta\in\Theta_0} \Pi_{i=1}^n f(x_j|\theta)}{\sup_{\theta\in\Theta} \Pi_{i=1}^n f(x_j|\theta)} = \frac{L_n(\theta_n^*)}{L_n(\hat{\theta}_n)}$$

が小さい時に H_0 を棄却する．ただし, θ_n^* は Θ_0 に母数空間を制限したときの最尤推定量で, $\hat{\theta}_n$ は制約のない最尤推定量である．

簡単のために, $\theta = (\theta_1, \ldots, \theta_k)^T$ とし, 帰無仮説が

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_r = 0 \quad (1 \le r < k)$$

と表現される場合[1)]を考える．

定理 3.5　：定理 2.3 の仮定が成立するとする．真の母数を θ_0 は H_0 を満足するとする．このとき,

$$-2\log\Lambda_n \xrightarrow{L} \chi_r^2$$

が成立する．

証明：前節の記号を踏襲する. $l_n(\theta_n^*) = \log L_n(\theta_n^*)$ を $\hat{\theta}_n$ のまわりで展開する：

$$l_n(\theta_n^*) = l_n(\hat{\theta}_n) + \dot{l}_n^T(\hat{\theta}_n)(\theta_n^* - \hat{\theta}_n) + \frac{n}{2}(\theta_n^* - \hat{\theta}_n)^T B_n(\theta_n^*)(\theta_n^* - \hat{\theta}_n)$$

となる[2)]. ただし, 定理 2.3 から

$$B_n(\theta_n^*) = \frac{1}{n}\int_0^1 2(1-t)\ddot{l}_n(\hat{\theta}_n + t(\theta_n^* - \hat{\theta}_n))dt \xrightarrow{as} -I(\theta_0)$$

---

1) このようにすると θ_n^* のはじめの r 成分はゼロになることに注意せよ.

2) まず, $\hat{\theta}_n$ のまわりで展開：

$$\dot{l}_n(\theta_n^*) = \dot{l}_n(\hat{\theta}_n) + \int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\theta_n^* - \hat{\theta}_n))dt(\theta_n^* - \hat{\theta}_n).$$

さらに, 部分積分から

$$\frac{n}{2}(\theta_n^* - \hat{\theta}_n)^T B_n(\theta_n^*)(\theta_n^* - \hat{\theta}_n) = (1-t)\dot{l}_n^T(\hat{\theta}_n + t(\theta_n^* - \hat{\theta}_n))(\theta_n^* - \hat{\theta}$$
$$+ \int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\theta_n^* - \hat{\theta}_n))dt$$

となることに注意すればよい.

となる[3)]．簡単のために, おおきな n のとき, $\dot{l}_n(\hat{\theta}_n)=0$ が成り立つとする．定理 2.3 の証明中の議論から

$$-2\log\Lambda_n=-n(\theta_n^*-\hat{\theta}_n)^TB_n(\theta_n^*)(\theta_n^*-\hat{\theta}_n)\sim n(\theta_n^*-\hat{\theta}_n)^TI(\theta_0)(\theta_n^*-\hat{\theta}_n)$$

となる．ここでは, $\sim$ は漸近同値を示す．この証明中は,以後同様な意味で用いる．H_0 のもとで

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{L}N(0,I^{-1}(\theta_0))$$

となる．

つぎに, $\sqrt{n}(\theta_n^*-\hat{\theta}_n)$ の漸近分布をもとめる．$\dot{l}_n(\theta^*)$ を $\hat{\theta}_n$ のまわりで展開する：

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*)=\frac{1}{\sqrt{n}}\dot{l}_n(\hat{\theta}_n)+\frac{1}{n}\int_0^1\ddot{l}_n(\hat{\theta}_n+t(\theta_n^*-\hat{\theta}_n))dt\sqrt{n}(\theta_n^*-\hat{\theta}_n)\sim-I(\theta_0)\sqrt{n}(\theta_n^*-\hat{\theta}_n)$$

したがって,

$$\sqrt{n}(\theta_n^*-\hat{\theta}_n)\sim-I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\dot{l}_n(\theta^*)$$

と

$$-2\log\Lambda_n\sim\frac{1}{\sqrt{n}}\dot{l}_n^T(\theta_n^*)I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*)$$

を得る．$\dot{l}_n(\theta_n^*)$ の漸近分布を求めるために $\dot{l}_n(\theta_n^*)$ を θ_0 のまわりで展開する：

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*)=\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)+\frac{1}{n}\int_0^1\ddot{l}_n(\theta_0+t(\theta_n^*-\theta_0))dt\sqrt{n}(\theta_n^*-\theta_0)$$

これと

$$\frac{1}{n}\int_0^1\ddot{l}_n(\theta_0+t(\theta_n^*-\theta_0))dt\xrightarrow{a.s.}-I(\theta_0)$$

に注意すれば,

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*)\sim\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)-I(\theta_0)\sqrt{n}(\theta_n^*-\theta_0)\tag{3.60}$$

Fisher 情報量 $I(\theta_0)$ を分割する：

$$I(\theta_0)=\begin{bmatrix}G_1 & G_2\\ G_2^T & G_3\end{bmatrix},\qquad G_1:r\times r,\quad G_2:r\times(k-r),\quad G_3:(k-r)\times(k-r).$$

また,

$$H=\begin{bmatrix}0 & 0\\ 0 & G_3^{-1}\end{bmatrix}$$

[3)] なぜならば, $\int_0^1 2(1-t)dt=1$ に注意し, さらに有界　　　からわかる.
収束定理を使えば,

$$\int_0^1 2(1-t)\{\frac{1}{n}\ddot{l}_n(\hat{\theta}_n+t(\theta_n^*-\hat{\theta}_n))+I(\theta_0)\}dt\le 2\int_0^1\{\frac{1}{n}\ddot{l}_n(\hat{\theta}_n+t(\theta_n^*-\hat{\theta}_n))+I(\theta_0)\}dt\to 0$$

とおく. $\dot{l}_n(\theta_n^*)$ の $(k-r)$ 成分はゼロなので, $H\dot{l}_n(\theta_n^*)=0$ であることに注意して, (3.60) 両辺に左から H を作用させれば,

$$H\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) \sim \sqrt{n}(\theta_n^*-\theta_0) \tag{3.61}$$

を得る[4)]．ただし, θ_n^* と θ_0 のはじめの r 成分はゼロである．(3.61) を (3.60) に代入すれば,

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*) \sim [\mathrm{id}-I(\theta_0)H]\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0)$$

となる．中心極限定理から

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) = \sqrt{n}\left(\frac{1}{n}\dot{l}_n(\theta_0)\right) \xrightarrow{L} N(0, I(\theta_0))$$

となる．したがって,

$$\frac{1}{\sqrt{n}}\dot{l}_n(\theta_n^*) \xrightarrow{L} [\mathrm{id}-I(\theta_0)H]Y$$

を得る. ただし, Y は $N(0,I(\theta_0))$ に従う確率変数である．$HI(\theta_0)H=H$ に注意してこれらの式をあわせれば,

$$\begin{aligned}-2\log\Lambda_n &\xrightarrow{L} Y^T[\mathrm{id}-I(\theta_0)H]^T I^{-1}(\theta_0)[\mathrm{id}-I(\theta_0)H]Y = Y^T[I^{-1}(\theta_0)-H]Y\\ &= Z^T I^{1/2}(\theta_0)[I^{-1}(\theta_0)-H]I^{1/2}(\theta_0)Z\end{aligned}$$

となる．ただし, $Z=I^{1/2}(\theta_0)Y$ は $N(0,\mathrm{id})$ に従う確率変数である．$I^{1/2}(\theta_0)[I^{-1}(\theta_0)-H]I^{1/2}(\theta_0)$ は idempotent でランクが r なので, Cochran の定理[5)]を使えば,

$$-2\log\Lambda_n \xrightarrow{L} \chi_r^2$$

を得る． □

---

4)
$$H\frac{1}{\sqrt{n}}\dot{l}_n(\theta_0) \sim HI(\theta_0)\sqrt{n}(\theta_n^*-\theta_0) = \begin{bmatrix} 0 & 0 \\ G_3^{-1}G_2^T & \mathrm{id} \end{bmatrix}\sqrt{n}(\theta_n^*-\theta_0) = \sqrt{n}(\theta_n^*-\theta_0)$$

からわかる. 上の式の最後の等式は $\sqrt{n}(\theta^*-\theta_0)$ ははじめの r 成分がゼロであることをもちいた.

5) $Y=(Y_1,Y_2,\ldots,Y_n)^T$ で, 各 Y_i は独立な平均 0, 分散 1 の正規分布にしたがっているとする. このとき, A が階数 $k(k\le n)$ の対称なベキ等行列とすると, Y^TAY の分布は χ_k 分布に従うことがわかる. 証明は数理統計学の基礎（野田著, 共立）の 152 ページを参照.

·····3.5·····

適合度検定：多項分布に基づく手法

ここでは, 得られた観測に特定の分布がうまく適合しているかを調べる手法である適合度検定のうちでも多項分布に基づく手法についての漸近分布について述べる. 帰無仮説はパラメトリックなモデルで対立仮説はノンパラメトリックなモデルとなるので, 厳密にはパラメトリックモデルにおける推測手法ではないが, 多項分布を利用すれば, パラメトリックモデルにおける検定問題に帰着できる. 経験分布に基づく適合度検定が他によく知られているが, ここではこれにはふれない.

$X_1, X_2, \ldots, X_n$ を未知の分布からのランダム標本とする. ここで, 検定問題

$$H_0 : X_i \sim f(x;\theta) \quad \text{v.s.} \quad H_1 : X_i \sim \text{任意の確率分布}$$

を考える. ただし, $f(x;\theta)$ は確率密度関数とし, $\theta \in \Theta \subset \mathbb{R}$ とする.

S を標本空間とし, $P(X_1 \in S) = 1$ とし, S の排反な部分集合 $A_1, A_2, \ldots, A_k$ で

$$S = \cup_{j=1}^{k} A_j$$

なるものを取る. さらに,

$$Y_j = \sum_{i=1}^{n} 1\{X_i \in A_j\}$$

とおく. ただし, n は k よりもかなり大きいものとする. すると, $Y = (Y_1, Y_2, \ldots, Y_k)$ は多項分布：

$$P(Y = y) = \frac{n}{y_1!y_2! \times \cdots \times y_k!}\Pi_{j=1}^{k}\phi_j^{y_j}, \qquad y = (y_1, y_2, \ldots, y_k)$$

に従うことが容易にわかる.

いま, X_1 は確率密度関数 $f(x;\theta)$ を持つとする. ただし, $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ は未知とする. よって,

$$\phi_j = P(X_1 \in A_j) = p_j(\theta) \tag{3.62}$$

と表現できる. ここで, 帰無仮説は

$$H_0 : \phi_j = p_j(\theta) \qquad \text{for some } \theta \in \Theta$$

と書き直せる.

例 3.1　：$X_1, X_2, \ldots, X_n$ は確率密度関数

$$f(x;\theta) = \theta x^{\theta-1} 1_{[0,1]}(x)$$

からのランダム標本とする．ただし，$\theta > 0$ は未知とする．このとき，$S = [0,1]$ であり，$A_1 = [0,1/4), A_2 = [1/4.1/2), A_3 = [1/2.3/4), A_4 = [3/4,1]$ とする．ここで，

$$p_j(\theta) = \int_{(j-1)/4}^{j/4} \theta x^{\theta-1}\,dx = \left(\frac{j}{4}\right)^{\theta} - \left(\frac{j-1}{4}\right)^{\theta}$$

となることに注意すれば，

$$P(Y_1 = y_1, \ldots, Y_4 = y_4) = \frac{n!}{y_1!y_2!y_3!y_4!}\Pi_{j=1}^{4}\left[\left(\frac{j}{4}\right)^{\theta} - \left(\frac{j-1}{4}\right)^{\theta}\right]^{y_j}$$

となる．

観測 $(X_1, X_2, \ldots, X_n)$ を $Y = (Y_1, Y_2, \ldots, Y_k)$ に置き換え，$\phi = (\phi_1, \phi_2, \ldots, \phi_k)$ を未知母数とみなせば，H_0 を制約 (3.62) を満たすモデルとし，H_1 を制約のないモデルで $\phi_i \geq 0, \sum_{j=1}^{k}\phi_j = 1$ のみだけを満たすものとし，検定問題を考える．$Y = y(y = (y_1, y_2, \ldots, y_k))$ が与えられたとき，対数尤度を

$$\log L_1(\phi) = \sum_{j=1}^{k} y_j \log\phi_j + \log\left(\frac{n!}{y_1!y_2! \times \cdots \times y_k!}\right)$$

と記し，H_0 のもとでの対数尤度を

$$\log L_0(\theta) = \sum_{j=1}^{k} y_j \log(p_j(\theta)) + \log\left(\frac{n!}{y_1!y_2! \times \cdots \times y_k!}\right)$$

と記すことにする．H_1 のもとでは，ϕ の最尤推定量は Y/n となる．一方，H_0 のもとでは，θ の最尤推定量 $\hat{\theta}_n$ は方程式

$$\sum_{j=1}^{k} \frac{Y_j}{p_j(\theta)}\dot{p}_j(\theta) = 0$$

を満たすものである．ただし，$\dot{p}_j(\theta) : p \times 1 = (\partial p_j(\theta)/\partial\theta_l)$ である．

よって，尤度比統計量は

$$\log\Lambda_n = \log L_0(\hat{\theta}_n) - \log L_1(Y/n) = -\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\hat{\theta}_n)}\right)$$

で与えられる．

定理 3.6　：Θ を $\mathbb{R}^p$ の開部分集合とし，$Y = (Y_1, Y_2, \ldots, Y_k)'$ を多項分布

$$Mult(n; p_1(\theta), p_2(\theta), \ldots, p_k(\theta))$$

に従う確率変数ベクトルとする．$p_j(\theta), j = 1, 2, \ldots, k$ は Θ 上で 2 階連続偏微分可能とする．また，$p \leq k-2$ とする．このとき，

$$-2\log\Lambda_n \xrightarrow{L} \chi^2_{k-1-p}$$

が成立する．

証明：定理 2.4 からわかる. □

注意 3.2　：$C(\theta) : k \times k = (c_{ij})$ とし,

$$c_{ij} = \begin{cases} p_i(\theta)(1 - p_i(\theta)) & (i = j) \\ -p_i(\theta)p_j(\theta) & (i \neq j) \end{cases}$$

とする. $0 < p_j(\theta) < 1$ ならば, $C(\theta)$ のランクは $k-1$ になることに注意せよ. ここで,

$$P(\theta) : k \times p = \left(\frac{\partial \log p_i(\theta)}{\partial \theta_j}\right)_{i=1,\ldots,k,j=1,\ldots,p}$$

とすれば, 正則条件（$p \leq k-1, rankP(\theta) = p$ 等）のもとで

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, I^{-1}(\theta))$$

が成立する. ただし,

$$I(\theta) = P'(\theta)C(\theta)P(\theta)$$

である.

尤度比統計量の他に Peason の chi^2 統計量が知られている：

$$K_n^2 = \sum_{j=1}^{k} \frac{(Y_j - np_j(\hat{\theta}_n))^2}{np_j(\hat{\theta}_n)}$$

定理 3.7　：前の定理と同じ仮定もとで,

$$K_n^2 + 2\log\Lambda_n \xrightarrow{P} 0$$

が成立する. したがって,

$$K_n^2 \xrightarrow{L} \xi_{k-1-p}^2$$

が成立する.

証明：CLT から

$$\sqrt{n}\left(\frac{Y_j}{n} - p_j(\theta)\right) \xrightarrow{L} N(0, c_{jj}(\theta))$$

$$\sqrt{n}(p_j(\hat{\theta}_n) - p_j(\theta)) \xrightarrow{L} N(0, \dot{p}_j(\theta)^T I^{-1} \dot{p}_j(\theta))$$

を得る. ただし,

$$\dot{p}_j : p \times 1 = (\partial p_j(\theta)/\partial \theta_l)$$

である. これらより,

$$\frac{Y_j}{n} - p_j(\theta) = o_P(1) \tag{3.63}$$

と

$$n\left|\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right|^r = o_P(1) \tag{3.64}$$

がわかる[1]. ただし, $r > 2$ である. さらに, Taylor 展開から

$$\begin{aligned}\log\frac{Y_j}{n} - \log p_j(\hat{\theta}_n) &= \frac{1}{p_j(\hat{\theta}_n))}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right) - 2\frac{1}{2p_j^2(\hat{\theta}_n))}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)^2 \\ &\quad + \frac{1}{p_j^3(\theta_n^*))}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)^3\end{aligned} \tag{3.65}$$

となる. ただし, θ_n^* は θ と $\hat{\theta}_n$ の間にあるものである. さらに,

$$\begin{aligned}-2\log\Lambda_n = 2n\sum_{j=1}^{k}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)\log\left(\frac{Y_j}{np_j(\hat{\theta}_n)}\right) \\ +2n\sum_{j=1}^{k}p_j(\hat{\theta}_n)\log\left(\frac{Y_j}{np_j(\hat{\theta}_n)}\right)\end{aligned}$$

となる. これに (3.65) を代入し, (3.63) と (3.64) を利用すれば,

$$\begin{aligned}-2\log\Lambda_n &= 2K_n^2 - \sum_{j=1}^{k}\frac{n}{2p_j^2(\hat{\theta}_n)}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)^3 + \sum_{j=1}^{k}\frac{n}{3p_j^2(\theta_n^*)}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)^4 \\ &\quad +2n\sum_{j=1}^{k}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right) - K_n^2 + \sum_{j=1}^{k}\frac{p_j(\hat{\theta}_n)n}{3p_j^2(\theta_n^*)}\left(\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right)^4 \\ &= K_n^2 + o_P(1)\end{aligned}$$

となること[2]がわり, 定理は証明された. □

[1] まず,

$$n\left|\frac{Y_j}{n} - p_j(\hat{\theta}_n)\right|^r \leq \text{ constant } \left(n\left|\frac{Y_j}{n} - p_j(\theta)\right|^r + n|p_j(\hat{\theta}_n) - p_j(\theta)|^r\right)$$

となることに注意する. さらに,

$$n|p_j(\hat{\theta}_n) - p_j(\theta)|^r = n^{(2-r)/2}\{\sqrt{n}|p_j(\hat{\theta}_n) - p_j(\theta)|\}^r = n^{(2-r)/2}O_P($$
$$n\left|\frac{Y_j}{n} - p_j(\theta)\right|^r = n^{(2-r)/2}\left\{\sqrt{n}\left|\frac{Y_j}{n} - p_j(\theta)\right|\right\}^r = n^{(2-r)/2}O_P(1$$

からわかる.

[2] $\sum_{j=1}^{n}\{Y_j/n - p_j(\hat{\theta}_n)\} = 0$ になることに注意せよ.

·····3.6·····

推定量の漸近有効性

定義 3.2　：$X_1, X_2, \ldots, X_n$ を母数 $\theta \in \Theta$ に依存する分布からのランダム標本とする．このランダム標本に基づく推定量の列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ はどんな θ に対しても、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, \Sigma(\theta))$$

を満足するものとする．このとき、$\hat{\theta}_n$ が θ の**漸近有効**であるとは、どんな $\theta \in \Theta$ に対しても、$\Sigma(\theta) = I^{-1}(\theta)$ が成立することである．

注意 3.3　(super efficient estimator)：$\theta \in \Theta \subset \boldsymbol{R}$ とする．$\hat{\theta}_n$ を θ の最尤推定量とし、

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, \frac{1}{I(\theta)})$$

が成立すると仮定する．$\theta_0 \in \Theta$ を任意の点とし、$\hat{\theta}_n$ の修正推定量を

$$\tilde{\theta}_n = \begin{cases} \theta_0 & (n^{1/4}|\hat{\theta}_n - \theta_0| \leq 1) \\ \hat{\theta}_n & \text{その他} \end{cases}$$

で定める．

$\theta \neq \theta_0$ のとき、

$$\begin{aligned} P_\theta(\hat{\theta}_n \neq \tilde{\theta}_n) &= P_\theta(n^{1/4}|\hat{\theta}_n - \theta_0| \leq 1) \\ &\leq P_\theta(|\theta - \theta_0| - |\hat{\theta}_n - \theta_0| \leq n^{-1/4}) \\ &= P_\theta(|\hat{\theta}_n - \theta| \geq |\theta - \theta_0| - n^{-1/4}) \\ &= P_\theta(\sqrt{n}(\hat{\theta}_n - \theta) \geq \sqrt{n}|\theta - \theta_0| - n^{1/4}) \to 0 \end{aligned}$$

が $\sqrt{n}(\hat{\theta}_n - \theta) = O_P(1)$ からわかる．したがって、

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, \frac{1}{I(\theta)})$$

となる．

一方、$\theta = \theta_0$ の場合、

$$\begin{aligned} P_{\theta_0}(\sqrt{n}(\hat{\theta}_n - \theta_0) = 0) &= P_{\theta_0}(n^{1/4}|\hat{\theta}_n - \theta_0| \leq 1) \\ &= P_{\theta_0}(\sqrt{n}|\hat{\theta}_n - \theta_0| \leq n^{1/4}) \to 1 \end{aligned}$$

となる．したがって、$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0,0)$ となる．これより、$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, \sigma^2(\theta))$ が成立することがわかる．ただし、

$$\sigma^2(\theta) = \begin{cases} 1/I(\theta) & (\theta \neq \theta_0) \\ 0 & (\theta = \theta_0) \end{cases}$$

である．

議論を簡単にするために、以後は $\theta \in \Theta \subset \boldsymbol{R}$ とする．

定理 3.8　：T_n を θ の推定量とし、

$$\sqrt{n}(T_n - \theta) \xrightarrow{L} N(0, \sigma^2(\theta))$$

が成り立つとき、つぎの不等式が成り立つ：

$$\liminf_{n\to\infty} \boldsymbol{E}\{\sqrt{n}(T_n - \theta)\}^2 \geq \sigma^2(\theta)$$

証明：部分積分と Fatou の補題から

$$\begin{aligned}\liminf_{n\to\infty} \boldsymbol{E}\{\sqrt{n}(T_n - \theta)\}^2 &= \liminf_{n\to\infty} \int_0^\infty P\bigl[\{\sqrt{n}(T_n - \theta)\}^2 > t\bigr]dt \\ &\geq \int_0^\infty \liminf_{n\to\infty} P\bigl[\{\sqrt{n}(T_n - \theta)\}^2 > t\bigr]dt\end{aligned}$$

となる．さらに、仮定より、

$$\liminf_{n\to\infty} P\bigl[\{\sqrt{n}(T_n - \theta)\}^2 > t\bigr] = 2\bigl\{1 - \Phi(\frac{\sqrt{t}}{\sigma(\theta)})\bigr\}$$

となる[1]. これを積分すれば、

$$\begin{aligned}2\int_0^\infty \bigl[1 - \Phi(\frac{\sqrt{t}}{\sigma(\theta)})\bigr]dt &= 4\sigma^2(\theta)\int_0^\infty u\{1 - \Phi(u)\}du \\ &= 2\sigma^2(\theta)\int_0^\infty u^2\phi(u)du = \sigma^2(\theta)\end{aligned}$$

より[2]、定理は示せた． □

例 3.2　：$X_1, \ldots, X_n$ を正規分布 $N(\theta, \tau^2)$ からのランダム標本とする．中心極限定理から、$n \uparrow \infty$ のとき、

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{\tau} \xrightarrow{L} N(0,1)$$

が成り立つ．ただし，$\bar{X}_n$ は標本平均である．したがって、標本平均は母平均の漸近正規推定量である．また、$\boldsymbol{E}\{\sqrt{n}(\bar{X}_n - \theta)^2\} = 1$ であり、定理において等号が成立する場合となる．

つぎに、$X_1, \ldots, X_n$ とは独立な 2 値の確率変数 Y_n を

$$Y_n = \begin{cases} 1 & (\text{ 確率 } \ \frac{1}{n^2}) \\ 0 & (\text{ 確率 } \ 1 - \frac{1}{n^2}) \end{cases}$$

---

[1] $P\bigl[\sqrt{n}(T_n - \theta) > \sqrt{t}\bigr] = P\bigl[\{\sqrt{n}(T_n - \theta)\}^2 > t\bigr] + P\bigl[\sqrt{n}(T_n - \theta) < -\sqrt{t}\bigr]$ と $\liminf_{n\to\infty} P\bigl[\{\sqrt{n}(T_n - \theta)\}^2 > t\bigr] > t\bigr] \to 1 - \Phi(\sqrt{t}/\sigma(\theta))$ および $\liminf_{n\to\infty} P\bigl[\sqrt{n}(T_n - \theta) < -\sqrt{t}\bigr] > t\bigr] \to \Phi(-\sqrt{t}/\sigma(\theta))$ からわかる.

[2] ただし、Φ と ϕ は標準正規分布の分布関数と確率密度関数である．

をとる．推定量 T_n を

$$T_n = (1-Y_n)\bar{X}_n + Y_n\sqrt{n}M$$

と定める．ただし、M は正の定数とする．$n \uparrow \infty$ のとき、

$$\begin{aligned}
\boldsymbol{E}(T_n) &= \boldsymbol{E}(1-Y_n)\boldsymbol{E}\bar{X}_n + \sqrt{n}M\boldsymbol{E}T_n \\
&= (1-\frac{1}{n^2})\theta + \sqrt{n}M\frac{1}{n^2} \to \theta \\
\boldsymbol{var}(\sqrt{n}(T_n-\theta)) &= n\boldsymbol{var}\{(1-Y_n)(\bar{X}_n-\theta) + Y_n(\sqrt{n}M-\theta)\} \\
&= n\big[\boldsymbol{E}\{(1-Y_n)(\bar{X}_n-\theta) + Y_n(\sqrt{n}M-\theta)\}^2 - \frac{1}{n^4}(\sqrt{n}M-\theta)^2\big] \\
&= n\big[\boldsymbol{E}(1-Y_n)(\bar{X}_n-\theta)^2 + (\sqrt{n}M-\theta)^2\boldsymbol{E}Y_n - \frac{1}{n^4}(\sqrt{n}M-\theta)^2\big] \\
&= (1-\frac{1}{n^2})\sigma^2 + (M-\frac{\theta}{\sqrt{n}})(1-\frac{1}{n^2}) \to \sigma^2 + M^2
\end{aligned}$$

となる．一方、どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned}
P\{|\sqrt{n}(T_n-\bar{X}_n)| > \epsilon\} &= P\{|\sqrt{n}Y_n(\sqrt{n}M-\bar{X}_n)| > \epsilon\} \\
&= P\{Y_n = 1 \text{かつ} |\sqrt{n}(\sqrt{n}M-\bar{X}_n)| > \epsilon\} \\
&= P(Y_n=1)P\{|\sqrt{n}(\sqrt{n}M-\bar{X}_n)| > \epsilon\} \to 0
\end{aligned}$$

となる[3)]．したがって、$\sqrt{n}(T_n-\theta)$ と $\sqrt{n}(\bar{X}_n-\theta)$ は漸近同値となる．よって、$\sqrt{n}(T_n-\theta)$ の漸近分散は σ^2 となり、定理の不等式が成り立つことがわかる．

定義 3.3　：任意の $\theta + h/\sqrt{n} \in \Theta$ に対して、

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}\big\{\sqrt{n}(T_n-\theta-\frac{h}{\sqrt{n}}) \le x\big\} = \Phi(\frac{x}{\sigma(\theta)})$$

が成り立つ[4)]とき、漸近正規推定量 T_n は**正則**であるという．

補題 3.3　：検定仮説 $H_0 : P_\theta \quad v.s. H_1 : P_{\theta+h/\sqrt{n}}$ に対する対数尤度比

$$\log\Lambda_n(h) = \log\Pi_{j=1}^n \frac{f(X_i|\theta+h/\sqrt{n})}{f(X_i|\theta)}$$

は

$$\log\Lambda_n(h) \xrightarrow{L} \begin{cases} N(-\frac{1}{2}h^2I(\theta), h^2I(\theta)) & (P_\theta \text{ のもと}) \\ N(\frac{1}{2}h^2I(\theta), h^2I(\theta)) & (P_{\theta+h\sqrt{n}} \text{ のもと}) \end{cases}$$

が成り立つ．

---

3) なぜならば、$n \uparrow \infty$ のとき、

$$P(Y_n=1) = \frac{1}{n^2} \to 0, \qquad P\{|\sqrt{n}(\sqrt{n}M-\bar{X}_n)| > \epsilon\} \to 1$$

からわかる．

4) ただし、Φ は標準正規分布の分布関数である．

証明：

$$l_n(\theta) = \log \Pi_{j=1}^n f(X_j|\theta)$$

とおく．$l_n(\theta + h/\sqrt{n})$ を θ のまわりで展開する：

$$l_n(\theta + \frac{h}{\sqrt{n}}) = l_n(\theta) + \dot{l}_n(\theta)\frac{h}{\sqrt{n}} + \frac{1}{2}h^2 B_n(\theta + \frac{h}{\sqrt{n}})$$

となる．ただし、

$$B_n(\theta + \frac{h}{\sqrt{n}}) = \frac{1}{n}\int_0^1 2(1-t)\ddot{l}_n(\theta + \frac{h}{\sqrt{n}} + t\frac{h}{\sqrt{n}})dt$$

である．また、定理 2.3 の証明中の議論から、P_θ のもとで、

$$B_n \xrightarrow{as} -I(\theta)$$

である．したがって、P_θ のもとで、

$$\begin{aligned} \log \Lambda_n(h) &= \frac{h}{\sqrt{n}}\dot{l}_n(\theta) - \frac{1}{2}h^2 I(\theta) + o_P(1) \\ &\xrightarrow{L} N(-\frac{1}{2}h^2 I(\theta), h^2 I(\theta)) \end{aligned}$$

となる．これを特性関数でかけば、

$$\phi_n(t) \equiv \boldsymbol{E}_\theta[\exp(\sqrt{-1}tS_n)] \to \exp\{-\frac{1}{2}\sqrt{-1}th^2 I(\theta) + \frac{1}{2}t^2 h^2 I(\theta)\}$$

となる．ただし，簡単のために，$S_n = \log \Lambda_n(h)$ とおいた．したがって、$P_{\theta+h/\sqrt{n}}$ のもとで、$\Lambda_n(h)$ の特性関数は

$$\begin{aligned} \boldsymbol{E}_{\theta+h/\sqrt{n}}[\exp(\sqrt{-1}tS_n)] &= \boldsymbol{E}_\theta[\exp(\sqrt{-1}tS_n)\exp(S_n)] \\ &= \boldsymbol{E}_\theta[\exp(\sqrt{-1}(t-\sqrt{-1})S_n)] \\ &= \phi_n(t-\sqrt{-1}) \\ &\longrightarrow \exp\{\frac{1}{2}\sqrt{-1}th^2 I(\theta) + \frac{1}{2}t^2 h^2 I(\theta)\} \end{aligned}$$

となるので、$P_{\theta+h/\sqrt{n}}$ のもとで、

$$\log \Lambda_n \xrightarrow{L} N(\frac{1}{2}h^2 I(\theta), h^2 I(\theta))$$

が成り立つことがわかる．　□

定義 3.4　：確率分布の列 $\{Q_n\}_{n=1}^\infty$ が確率分布の列 $\{P_n\}_{n=1}^\infty$ に**接触**するとは、任意の（可測な）事象の列 $\{B_n\}_{n=1}^\infty$ に対して、$n \uparrow \infty$ のとき、$P_n(B_n) \to 0$ ならば、$Q_n(B_n) \to 0$ が成り立つことである．

定理 3.9　：$P_{\theta+h/\sqrt{n}}$ は P_θ に接触する．

証明：どんな正の数 $\epsilon > 0$ に対しても、十分大きな数 K をとれば、十分大きなすべての n に対して、

$$P_{\theta+h/\sqrt{n}}\{\log \Lambda_n(h) > K\} < \epsilon$$

とできる.

いま、$\{B_n\}_{n=1}^{\infty}$ を事象の列とし、$n \uparrow \infty$ のとき、$P_\theta(B_n) \to 0$ となるようなものとする．$n \uparrow \infty$ のとき、

$$\begin{aligned} P_{\theta+h/\sqrt{n}}(B_n) &\leq P_{\theta+h/\sqrt{n}}\{\log \Lambda_n(h) > K\} + \boldsymbol{E}_\theta[\exp(\log \Lambda_n(h))]1\{\log \Lambda_n(h) \leq K\}1\{B_n\}] \\ &\leq \epsilon + \exp(K) P_\theta(B_n) \to 0 \end{aligned} \tag{3.66}$$

となり、補題は示された．　□

定理 3.10　：最尤推定量は正則である．

証明：$\hat{\theta}_n$ を θ の最尤推定量とする．P_θ のもとで、

$$\sqrt{n}(\hat{\theta}_n - \theta) = I^{-1}(\theta)\frac{1}{\sqrt{n}}\dot{l}_n(\theta) + o_P(1)$$

$$\log \Lambda_n(h) = \frac{h}{\sqrt{n}}\dot{l}_n(\theta) - \frac{1}{2}h^2 I(\theta) + o_P(1)$$

が成り立つ．これらふたつの式を合わせれば、

$$\sqrt{n}(\hat{\theta}_n - \theta - \frac{h}{\sqrt{n}}) = h^{-1}I^{-1}(\theta)\log \Lambda_n(h) - \frac{1}{2}h + o_P(1) \tag{3.67}$$

を得る．$P_{\theta+h/\sqrt{n}}$ は P_θ に接触しているので、(3.67) は $P_{\theta+h/\sqrt{n}}$ のもとでも成り立つ．しかし、補題 2.3 から、$P_{\theta+h/\sqrt{n}}$ のもとで、

$$\log \Lambda_n \xrightarrow{L} N(\frac{1}{2}h^2 I(\theta), h^2 I(\theta))$$

であるから、$P_{\theta+h/\sqrt{n}}$ のもとで、

$$\sqrt{n}(\hat{\theta}_n - \theta - \frac{h}{\sqrt{n}}) = h^{-1}I^{-1}(\theta)\log \Lambda_n(h) - \frac{1}{2}h + o_P(1) \xrightarrow{L} N(0, I^{-1}(\theta))$$

となる．すなわち、$\hat{\theta}_n$ は θ の正則推定量である．　□

定理 3.11　：漸近正規推定量 T_n が正則で漸近分散 $\sigma^2(\theta)$ をもつとき、

$$\sigma^2(\theta) \geq I^{-1}(\theta)$$

が成り立つ．これは、漸近分散に関する Cramér – Rao の不等式と考えることができる.

証明：T_n の正則性より

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}\big\{\sqrt{n}(T_n-\theta-\frac{h}{\sqrt{n}})\le x\big\}=\Phi(\frac{x}{\sigma(\theta)}) \tag{3.68}$$

となる．これより，有意水準 1/2 の検定の棄却域を

$$W_n=\{\sqrt{n}(T_n-\theta)>k_n\}$$

で定める．すると、$n\uparrow\infty$ のとき、$P_\theta(W_n)=1/2$ だから、$k_n\downarrow 0$ となる．一方、有意水準 1/2 の尤度比検定の棄却域を

$$W_n^*=\{\log\Lambda_n>k_n^*\}$$

とすれば、$n\uparrow\infty$ のとき、$P_\theta(W_n^*)\to 1/2$ となり、補題 2.3 から、$k_n^*\to-(1/2)h^2I(\theta)$ となる．Neymann – Pearson の補題から

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}(W_n^*)\ge\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}(W_n) \tag{3.69}$$

となる．しかし、補題 2.3 から、(3.69) の左辺は

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}\big\{\log\Lambda_n(h)-\frac{1}{2}h^2I(\theta)>k_n^*-\frac{1}{2}h^2I(\theta)\big\}=1-\Phi(-h\sqrt{I(\theta)})$$

となる. また、正則性から、(3.69) の右辺は

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}\big\{\sqrt{n}(T_n-\theta)>0\big\}=1-\Phi(-\frac{h}{\sigma(\theta)})$$

となる[5]. したがって、

$$\Phi(h\sqrt{I(\theta)})\ge\Phi(\frac{h}{\sigma(\theta)})$$

となり[6]、

$$I(\theta)\ge\frac{1}{\sigma^2(\theta)}$$

を得る. よって、定理は示された．□

[5] これは (3.68) において $x=-h$ とおくことにより得られる.

[6] $1-\Phi(-x)=\Phi(x)$ に注意せよ.

4

線形回帰モデルと一般化線形モデル

·····4.1·····

回帰モデルにおける大標本理論

回帰モデルは説明変数（独立変数）と被説明変数（従属変数）の関係を記述するものである．
単回帰線形モデル

$$Y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{4.70}$$

ただし，α と β は未知の回帰係数，x_i は Y_i に対応する（既知の）説明変数で，誤差項 ϵ_i は未観測の確率変数で独立同一の分布に従い，平均と分散は 0 と $\sigma^2\,(0 < \sigma^2 < \infty)$ である．
重回帰線形モデル

$$\boldsymbol{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n \tag{4.71}$$

ただし，$\boldsymbol{Y}_n = (Y_1, Y_2, \ldots, Y_n)'$ は観測ベクトルで，$\boldsymbol{X}_n = (x_{ij})_{i=1,2,\ldots,n,\, j=1,2,\ldots,q}$ は既知の行列で，$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)'$ は未知の回帰係数ベクトルである．ここで，誤差ベクトル $\boldsymbol{\epsilon}_n = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ は未観測の確率変数であり，各 $\epsilon_i\,(i = 1, 2, \ldots, n)$ は独立同一の分布に従い，平均と分散は 0 と $\sigma^2\,(0 < \sigma^2 < \infty)$ である．したがって，$\mathbb{E}[\boldsymbol{\epsilon}_n] = \boldsymbol{0}$, $\mathbb{VAR}[\boldsymbol{\epsilon}_n] = \sigma^2 I_n$ である．行列 $\boldsymbol{X}_n' = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_q)$ とおけば，モデル (4.71) は

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \\ &= x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iq}\beta_q + \epsilon_i, \qquad i = 1, 2, \ldots, n \end{aligned}$$

と書き直すことができる．

4.1.1 最小 2 乗推定量

統計解析では，$\boldsymbol{\beta}$ の推定や仮説検定を行う．$\boldsymbol{\beta}$ の推定手法でもっとも知られているものは最小 2 乗法である．すなわち，$\boldsymbol{\beta}$ の推定量を

$$\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n = (\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta})'(\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) \tag{4.72}$$

を最小にするもので定義する．単回帰線形モデルでは,

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2$$

の最小化になる．(4.72) は $\boldsymbol{\beta}$ の 2 次関数なので，(4.70) を最小にする $\boldsymbol{\beta}$ は方程式

$$\boldsymbol{X}_n'\boldsymbol{X}_n\boldsymbol{\beta} = \boldsymbol{X}_n'\boldsymbol{Y}_n \tag{4.73}$$

または

$$\sum_{i=1}^n (\boldsymbol{Y}_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \boldsymbol{x}_i = \boldsymbol{0} \tag{4.74}$$

の解となる．この方程式を正規方程式と呼び，$\hat{\boldsymbol{\beta}}_n$ を $\boldsymbol{\beta}$ の最小 2 乗推定量という．さらに，$\mathrm{rank}(\boldsymbol{X}_n) = q$ ならば，正規方程式は一意的に解をもち，最小 2 乗推定量は

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}'_n\boldsymbol{X}_n)^{-1}\boldsymbol{X}'_n\boldsymbol{Y}_n \tag{4.75}$$

で与えられる．

また，単回帰線形モデル (4.70) における (α, β) の正規方程式は

$$\sum_{i=1}^n Y_i = n\alpha + \beta\sum_{i=1}^n x_i$$
$$\sum_{i=1}^n x_iY_i = \alpha\sum_{i=1}^n x_i + \beta\sum_{i=1}^n x_i^2$$

となるので，すべての x_i が一致する場合を除いて，

$$\hat{\beta}_n = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}, \qquad \hat{\alpha}_n = \bar{Y}_n - \hat{\beta}_n\bar{x}_n,$$

である．ただし，$\bar{x}_n = (1/n)\sum_{i=1}^n x_i$, $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i$ である．

重回帰線形モデル (4.71) の仮定のもとで

$$\mathbb{E}[\hat{\boldsymbol{\beta}}_n] = \boldsymbol{\beta}, \qquad \mathbb{VAR}[\hat{\boldsymbol{\beta}}_n] = \sigma^2(\boldsymbol{X}'_n\boldsymbol{X}_n)^{-1}$$

となる．また，単回帰線形モデル (4.70) では

$$\mathbb{VAR}\begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}\begin{pmatrix} \sum_{i=1}^n x_i^2/n & -\bar{x}_n \\ -\bar{x}_n & 1 \end{pmatrix}$$

となる[1]．

$\hat{\boldsymbol{\beta}}_n$ は次のような最適性を持っている．

定義 4.1　標本の線形関数になっている不偏推定量を線形不偏推定量と呼ぶ．線形不偏推定量の中で，すべての母数に対してその分散を最小にするものを最良線形不偏推定量 (Best Linear Unbiased Estimator) という．略して，BLUE という．

定理 4.1 (Gauss - Markov)　$\boldsymbol{c}\,(\neq \boldsymbol{0})$ を q 次元ベクトルとする．このとき，$\boldsymbol{c}'\hat{\boldsymbol{\beta}}_n$ は $\boldsymbol{c}'\boldsymbol{\beta}$ の BLUE である．

証明：$\boldsymbol{a}$ を 零ベクトルでない任意の n 次元ベクトルとする．$\boldsymbol{a}'\boldsymbol{Y}_n$ が $\boldsymbol{c}'\boldsymbol{\beta}$ の不偏推定量となるためには

$$\boldsymbol{c}'\boldsymbol{\beta} = \mathbb{E}[\boldsymbol{a}'\boldsymbol{Y}_n] = \boldsymbol{a}'\mathbb{E}[\boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n] = \boldsymbol{a}'\boldsymbol{X}_n\boldsymbol{\beta}$$

---

[1] これは (4.71) において，

$$\boldsymbol{X}_n = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

とすれば，(4.70) を得ることからわかる．

とならなければならない．$\boldsymbol{\beta}$ は任意なので，$\boldsymbol{a}'\boldsymbol{X}_n = \boldsymbol{c}'$ である．また，

$$\mathbb{VAR}[\boldsymbol{a}'\boldsymbol{Y}_n] = \boldsymbol{a}'\mathbb{VAR}(\boldsymbol{Y}_n)\boldsymbol{a} = \sigma^2\boldsymbol{a}'\boldsymbol{a}$$

となる．$\boldsymbol{a}'\boldsymbol{X}_n = \boldsymbol{c}'$ の制約下で $\mathbb{VAR}[\boldsymbol{a}'\boldsymbol{Y}_n]$ を $\boldsymbol{a}$ に関して最小化する：

$$\begin{aligned}\boldsymbol{a}'\boldsymbol{a} - \boldsymbol{c}'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c} &= \boldsymbol{a}'\boldsymbol{a} - 2(\boldsymbol{X}_n'\boldsymbol{a} - \boldsymbol{c})'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c} - \boldsymbol{c}'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c}\\ &= \{\boldsymbol{a} - \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c}\}'\{\boldsymbol{a} - \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c}\}\end{aligned}$$

となるので，$\mathbb{VAR}[\boldsymbol{a}'\boldsymbol{Y}_n]$ は $\boldsymbol{a} = \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{c}$ で最小化されることがわかる．　□

いま，誤差項 $\boldsymbol{\epsilon}_n$ が正規分布 $N(\boldsymbol{0},\, \sigma^2 I_n)$ に従うとする．$Y_1 = y_1,\, Y_2 = y_2,\, \ldots,\, Y_n = y_n$ が与えられたときの対数尤度関数は

$$L_n(\boldsymbol{\beta},\, \sigma^2 |\, \boldsymbol{y}) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}'\boldsymbol{\beta})^2 - \frac{n}{2}\log(2\pi)$$

となる．ただし，$\boldsymbol{y} = (y_1,\, y_2,\, \ldots,\, y_n)'$ とした．これを母数に関して偏微分すれば，方程式

$$\frac{\partial}{\partial\boldsymbol{\beta}}\log L_n(\boldsymbol{\beta},\, \sigma^2 |\, \boldsymbol{y}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i = \boldsymbol{0}$$

$$\frac{\partial}{\partial\sigma}\log L_n(\boldsymbol{\beta},\, \sigma^2 |\, \boldsymbol{y}) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 = 0$$

を得る．したがって，$\boldsymbol{\beta}$ の最尤推定量 $\widehat{\boldsymbol{\beta}}_n$ は正規方程式

$$\sum_{i=1}^{n}(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n)\boldsymbol{x}_i = \boldsymbol{0}$$

をみたす．一方，σ^2 の最尤推定量 $\widehat{\sigma}_n^2$ は

$$\widehat{\sigma}_n^2 = \frac{1}{n}(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n)^2$$

となる．誤差分布が多変量正規分布 $N(\boldsymbol{0},\, \sigma^2 I_n)$ に従うとき，$\boldsymbol{\beta}$ の最尤推定量 $\widehat{\boldsymbol{\beta}}_n$ は $\boldsymbol{\beta}$ の最小 2 乗推定量と一致する．

rank$(\boldsymbol{X}_n) = q$ のとき，(4.75) より σ^2 の最尤推定量はつぎのようになる：

$$\begin{aligned}\widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n)^2\\ &= \frac{1}{n}\|\boldsymbol{Y} - \boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_n\|^2\\ &= \frac{1}{n}\|\boldsymbol{Y} - \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{Y}\|^2\\ &= \frac{1}{n}\|(I_n - \boldsymbol{H})\boldsymbol{Y}\|^2\end{aligned}$$

となる．ただし，$\boldsymbol{H} = \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'$ とした．また，$\boldsymbol{x}' = (x_1,\, x_2,\, \ldots,\, x_n)$ に対して $\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ とした．行列 $\boldsymbol{H}$ をハット行列と呼ぶことにしよう．$\boldsymbol{H}$ は n 次元ベクト

ルを $\boldsymbol{X}_n$ の列で張られた空間への射影行列となっている．すなわち，$\boldsymbol{X}_n = (\boldsymbol{z}_1,\, \boldsymbol{z}_2,\, \ldots,\, \boldsymbol{z}_q)$ としたとき，

$$\text{span}\ (\boldsymbol{X}_n) = \{\boldsymbol{z} \in \mathbb{R}^n |\, \boldsymbol{z} = a_1\boldsymbol{z}_1 + a_2\boldsymbol{z}_2 + \cdots + a_q\boldsymbol{z}_q,\, a_i \in \mathbb{R},\, i = 1,\, 2,\, \ldots,\, q\}$$

としたとき，任意のベクトル $\boldsymbol{w} \in \mathbb{R}^n$ は $\boldsymbol{H}\boldsymbol{w} \in$ span $(\boldsymbol{X}_n)$ をみたし，任意のベクトル $\boldsymbol{z} \in$ span $(\boldsymbol{X}_n)$ は $\boldsymbol{H}\boldsymbol{z} = \boldsymbol{z}$ をみたす．すると，$\boldsymbol{H} = \boldsymbol{H}' = \boldsymbol{H}^2$ となり，$\boldsymbol{H}$ のランクは tr $(\boldsymbol{H})$ となる．

$\widehat{\boldsymbol{\beta}}_n$ と $\widehat{\sigma}_n^2$ の同時分布を求めるために必要な補題がまず述べる．

補題 4.1　$\boldsymbol{Y} \sim N(\boldsymbol{0},\, I_n)$ とし，$\boldsymbol{Q} = \boldsymbol{Q}' = \boldsymbol{Q}^2$ とし，$tr\,\boldsymbol{Q} = q$ とする．このとき，$\boldsymbol{Y}'\boldsymbol{Q}\boldsymbol{Y} \sim \chi^2(q)$ となる．

証明：$n \times n$ の直交行列 $\boldsymbol{R}$ が存在して

$$\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}' = \begin{bmatrix} I_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

とできる．一方，

$$\boldsymbol{Z} = \boldsymbol{R}\boldsymbol{Y} \sim N(\boldsymbol{0},\, I_n), \qquad \boldsymbol{Z} = (Z_1,\, Z_2,\, \ldots,\, Z_n)'$$

となる．これより

$$\boldsymbol{Y}'\boldsymbol{Q}\boldsymbol{Y} = (\boldsymbol{R}\boldsymbol{Y})'\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{Y}) = \boldsymbol{Z}' \begin{bmatrix} I_q & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{Z} = Z_1^2 + Z_2 + \cdots + Z_q^2 \sim \chi^2(q)$$

より補題は証明された．□

定理 4.2　$\boldsymbol{X}_n'\boldsymbol{X}_n$ の逆行列が存在するとする．このとき，

(i) $\widehat{\boldsymbol{\beta}}_n \sim N_q(\boldsymbol{\beta},\, \sigma^2(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1})$
(ii) $n\widehat{\sigma}_n^2/\sigma^2 \sim \chi^2(n-q)$
(iii) $\widehat{\boldsymbol{\beta}}_n$ と $\widehat{\sigma}_n^2$ は独立である．

証明：(i) $\boldsymbol{Y} \sim N(\boldsymbol{X}_n\boldsymbol{\beta},\, \sigma^2 I_n)$ に注意すれば，$\widehat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{Y} = \boldsymbol{A}_n\boldsymbol{Y}$ と書ける．したがって，$\widehat{\boldsymbol{\beta}}_n \sim N(\boldsymbol{A}_n\boldsymbol{X}_n\boldsymbol{\beta},\, \sigma^2\boldsymbol{A}_n\boldsymbol{A}_n')$ となる．ただし，

$$\begin{aligned} \boldsymbol{A}_n\boldsymbol{X}_n\boldsymbol{\beta} &= (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{\beta} = \boldsymbol{\beta} \\ \boldsymbol{A}_n\boldsymbol{A}_n' &= (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1} = (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1} \end{aligned}$$

である．

(ii) $\boldsymbol{H}_n\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{H}\boldsymbol{\epsilon}$ に注意すれば，

$$n\frac{\widehat{\sigma}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\|(I_N - \boldsymbol{H})\boldsymbol{Y}\|^2 = \frac{1}{\sigma^2}\boldsymbol{\epsilon}'(I_n - \boldsymbol{H})\boldsymbol{\epsilon}$$

となる．$\boldsymbol{H}$ のランクは q なので，$I_n - \boldsymbol{H}$ のランクは $(n-q)$ となるので $(1/\sigma^2)\boldsymbol{\epsilon}'(I_n - \boldsymbol{H})\boldsymbol{\epsilon} \sim \chi^2(n-q)$ を得る．

(iii) $\widehat{\boldsymbol{\beta}}_n$ と $\widehat{\sigma}_n^2$ は独立であることを示すためには，$\widehat{\boldsymbol{\beta}}_n$ と $\boldsymbol{Y} - \boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_n$ が独立であることを示せばよい．これら二つは正規分布に従うので，独立性を示すためには $\mathbb{COV}(\widehat{\boldsymbol{\beta}}_n, \boldsymbol{Y} - \boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_n) = \boldsymbol{0}$ を示せばよい．$\boldsymbol{X}_n'\boldsymbol{H} = \boldsymbol{X}_n'$ に注意すれば，

$$
\begin{aligned}
\mathbb{COV}(\widehat{\boldsymbol{\beta}}_n, \boldsymbol{Y} - \boldsymbol{X}_n\boldsymbol{\beta}) &= \mathbb{COV}(\boldsymbol{A}_n\boldsymbol{Y}, (I_n - \boldsymbol{HY}) = \boldsymbol{A}_n\mathbb{COV}(\boldsymbol{Y}, \boldsymbol{Y})(I_n - \boldsymbol{H})' = \sigma^2\boldsymbol{A}_n(I_n - \boldsymbol{H}) \\
&= \sigma^2\{(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n' - (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\} = \boldsymbol{0}
\end{aligned}
$$

となる．□

4.1.2 最小 2 乗推定量の漸近分布

単回帰線形モデル (4.71) において，(α, β) の最小 2 乗推定量 $(\hat{\alpha}_n, \hat{\beta}_n)$ の漸近分布を求める．そのために以下の仮定を述べる．

いま，

$$
t_n = \sum_{i=1}^n (x_i - \bar{x}_n)^2, \qquad c_{ni} = \frac{x_i - \bar{x}_n}{\sqrt{t_n}}
$$

とし，以下を仮定する：

(N1)　$\max_{1\le i\le n} c_{ni}^2 \to 0 (n \to \infty)$　(Noether 条件)

(N2)　$\lim_{n\to\infty} \bar{x}_n \equiv \bar{x} < \infty$

(N3)　$0 < \lim_{n\to\infty} \frac{t_n}{n} \equiv t < \infty$

定理 4.3　単回帰線形モデル (4.71) において (α, β) の最小 2 乗推定量を $(\hat{\alpha}_n, \hat{\beta}_n)$ とする．このとき，仮定 (N1) - (N3) のもとで

$$
\sqrt{n}\begin{pmatrix} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{pmatrix} \xrightarrow{L} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2\begin{pmatrix} 1 + \bar{x}^2/t & -\bar{x}/t \\ -\bar{x}/t & 1/t \end{pmatrix}\right)
$$

が成立する．

証明：まず，

$$
\begin{aligned}
\hat{\beta}_n &= \frac{1}{t_n}\sum_{i=1}^n (x_i - \bar{x}_n)\{\alpha + \beta(x_i - \bar{x}_n) + \epsilon_i\} \\
&= \beta + \frac{1}{t_n}\sum_{i=1}^n (x_i - \bar{x}_n)\epsilon_i
\end{aligned}
$$

となること[2)]に注意する．これより

$$
\sqrt{t_n}(\hat{\beta}_n - \beta) = \sum_{i=1}^n \frac{x_i - \bar{x}_n}{\sqrt{t_n}}\epsilon_i = \sum_{i=1}^n c_{ni}\epsilon_i
$$

[2)] $\sum_{i=1}^n Y_i(x_i - \bar{x}) = \sum_{i=1}^n (\alpha + \beta x_i + \epsilon_i)(x_i - \bar{x}_n)$ と　$\sum_{i=1}^n (x_i - \bar{x}_n)\beta\bar{x}_n = 0$ からわかる．

となる．ただし，$c_{ni} = (x_i - \bar{x}_n)/\sqrt{t_n}$ である．定義から

$$\sum_{i=1}^{n} c_{ni} = 0, \qquad \sum_{i=1}^{n} c_{ni}^2 = 1$$

となる．Lindeberg 条件を確認する：まず，$\mathbb{VAR}[\sum_{i=1}^{n} c_{ni}\epsilon_i] = \sum_{i=1}^{n} c_{ni}^2 \mathbb{VAR}[\epsilon_i] = \sigma^2 \sum_{i=1}^{n} c_{ni}^2 = \sigma^2$ に注意する．任意の正の数 δ に対して

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbb{E}\left[|c_{ni}\epsilon_i|^2 I\{|c_{ni}\epsilon_i| > \delta\sigma\}\right] &\le \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbb{E}\left[|c_{ni}\epsilon_i|^2 I\{\epsilon_i^2 > \frac{\delta^2\sigma^2}{\max_{1\le i\le n} c_{ni}^2}\}\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbb{E}\left[c_{ni}^2\epsilon_i^2 I\{\epsilon_i^2 > \frac{\delta^2\sigma^2}{\max_{1\le i\le n} c_{ni}^2}\}\right] \\
&= \frac{1}{\sigma^2}\mathbb{E}\left[\epsilon_1^2 I\{\epsilon_1^2 > \frac{\delta^2\sigma^2}{\max_{1\le i\le n} c_{ni}^2}\}\right] \to 0, \qquad n \nearrow \infty
\end{aligned}$$

となる[3]．よって

$$\sqrt{t_n}(\hat{\beta}_n - \beta) \xrightarrow{L} N(0, \sigma^2)$$

を得る．さらに，$(N3)$ に注意して，Slutsky の補題を用いれば，

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{L} N(0, \sigma^2/t)$$

がわかる．

つぎに，$\hat{\alpha}_n$ の極限分布を求める：

$$\hat{\alpha}_n = \bar{Y}_n - \hat{\beta}_n\bar{x}_n = \alpha + \beta\bar{x}_n + \bar{\epsilon}_n - \hat{\beta}_n\bar{x}_n = \alpha - (\hat{\beta}_n - \beta)\bar{x}_n + \bar{\epsilon}_n$$

である．ただし，$\bar{\epsilon}_n = (1/n)\sum_{i=1}^{n}\epsilon_i$ である．これより

$$\begin{aligned}
\sqrt{n}(\hat{\alpha}_n - \alpha) &= \sqrt{n}\bar{\epsilon}_n - \sqrt{n}(\hat{\beta}_n - \beta)\bar{x}_n \\
&= \sqrt{n}\bar{\epsilon}_n - \sqrt{\frac{n}{t_n}}\bar{x}_n\sqrt{t_n}(\hat{\beta}_n - \beta) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i - \sqrt{\frac{n}{t_n}}\bar{x}_n\sum_{i=1}^{n} c_{ni}\epsilon_i \\
&= \sum_{i=1}^{n} d_{ni}\epsilon_i
\end{aligned}$$

となる．ただし，

$$d_{ni} = \frac{1}{\sqrt{n}} - \sqrt{\frac{n}{t_n}}\bar{x}_n c_{ni}$$

である．また，

$$\sum_{i=1}^{n} d_{ni} = \sqrt{n}, \qquad \sum_{i=1}^{n} d_{ni}^2 = 1 + \frac{n\bar{x}_n^2}{t_n} \equiv r_n^2$$

---

[3] これは，$\mathbb{E}[\epsilon_i^2] < \infty$ と $\delta^2\sigma^2/\max_{1\le i\le n} c_{ni}^2 \to \infty\,(n \to \infty)$ からわかる．

である．$\mathbb{VAR}[\sum_{i=1}^{n} d_{ni}\epsilon_i] = r_n^2\sigma^2$ に注意する．Lindeberg 条件を確認する．任意の正の数 δ に対して

$$\begin{aligned}
\frac{1}{r_n^2\sigma^2}\sum_{i=1}^{n}\mathbb{E}\left[|d_{ni}\epsilon_i|^2 I\{|d_{ni}\epsilon_i| > \delta r_n\sigma\}\right] &\le \frac{1}{r_n^2\sigma^2}\sum_{i=1}^{n}\mathbb{E}\left[|d_{ni}\epsilon_i|^2 I\{\epsilon_i^2 > \frac{(\delta r_n\sigma)^2}{\max_{1\le i\le n} d_{ni}^2}\}\right] \\
&\le \frac{1}{r_n^2\sigma^2}\sum_{i=1}^{n}\mathbb{E}\left[|d_{ni}\epsilon_i|^2 I\left\{\epsilon_i^2 > \frac{(\delta r_n\sigma)^2}{2\left(\frac{1}{n} + \frac{n}{t_n}\bar{x}_n^2\max_{1\le i\le n} c_{ni}^2\right)}\right\}\right] \\
&= \frac{1}{r_n^2\sigma^2}\sum_{i=1}^{n} d_{ni}^2\mathbb{E}\left[\epsilon_i^2 I\left\{\epsilon_i^2 > \frac{(\delta r_n\sigma)^2}{2\left(\frac{1}{n} + \frac{n}{t_n}\bar{x}_n^2\max_{1\le i\le n} c_{ni}^2\right)}\right\}\right] \\
&= \frac{1}{\sigma^2}\mathbb{E}\left[\epsilon_1^2 I\left\{\epsilon_1^2 > \frac{(\delta r_n\sigma)^2}{2\left(\frac{1}{n} + \frac{n}{t_n}\bar{x}_n^2\max_{1\le i\le n} c_{ni}^2\right)}\right\}\right] \to 0, \qquad n \to \infty
\end{aligned}$$

となる．$(N1) - (N3)$ より

$$r_n^2 \to 1 + \frac{\bar{x}^2}{t}, \qquad \bar{x}_n^2 \to \bar{x}^2, \qquad 2\left(\frac{1}{n} + \frac{n}{t_n}\bar{x}_n^2\max_{1\le i\le n} c_{ni}^2\right) \to \infty$$

と $\mathbb{E}[\epsilon_i^2] < \infty$ となることから保障される．よって，

$$\sqrt{n}(\hat{\alpha}_n - \alpha) = \sum_{i=1}^{n} d_{ni}\epsilon_i \xrightarrow{L} N\big(0,\ \big(1 + \frac{\bar{x}^2}{t}\big)\sigma^2\big)$$

を得る．

最後に，$\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in \mathbb{R}^2$, $\boldsymbol{\lambda} \neq \mathbf{0}$ をとり，

$$\lambda_1\sqrt{n}(\hat{\alpha}_n - \alpha) + \lambda_2\sqrt{n}(\hat{\beta} - \beta) = \sum_{i=1}^{n} f_{ni}\epsilon_i$$

の漸近分布を調べる．ただし，$f_{ni} = \lambda_1 d_{ni} + \lambda_2\frac{n}{t_n}c_{ni}$ である．これは

$$\sum_{i=1}^{n} f_{ni} = \lambda_1\sqrt{n}, \qquad \sum_{i=1}^{n} f_{ni}^2 = \lambda_1^2 r_n^2 + \lambda_2\frac{n}{t_n} - 2\lambda_1\lambda_2\bar{x}_n\frac{n}{t_n}$$

を満足する．ここで

$$V_n = \begin{bmatrix} r_n^2 & -n\bar{x}_n/t_n \\ -n\bar{x}_n/t_n & n/t_n \end{bmatrix}$$

とおく．再度，Lindeberg 条件を確認する：任意の正の数 δ に対して，

$$
\begin{aligned}
&\frac{1}{\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}} \sum_{i=1}^{n} \mathbb{E}\left[|f_{ni}\epsilon_i|^2 I\{|f_{ni}\epsilon_i| > \delta\sigma\sqrt{\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}\}\right] \\
&\quad \leq \frac{1}{\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}} \sum_{i=1}^{n} \mathbb{E}\left[|f_{ni}\epsilon_i|^2 I\left\{|\epsilon_i|^2 > \frac{\delta^2\sigma^2\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}{\max_{1\leq i\leq} |f_{ni}|^2}\right\}\right] \\
&\quad = \frac{1}{\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}} \sum_{i=1}^{n} f_{ni}^2 \mathbb{E}\left[\epsilon_1^2 I\left\{|\epsilon_i|^2 > \frac{\delta^2\sigma^2\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}{\max_{1\leq i\leq} |f_{ni}|^2}\right\}\right] \\
&\quad = \mathbb{E}\left[\epsilon_1^2 I\left\{|\epsilon_i|^2 > \frac{\delta^2\sigma^2\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}{\max_{1\leq i\leq} |f_{ni}|^2}\right\}\right] \to 0, \qquad n \nearrow \infty
\end{aligned}
$$

となる．これは (N1)-(N3) より

$$
\frac{\delta^2\sigma^2\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}{\max_{1\leq i\leq} |f_{ni}|^2} \geq \frac{\delta^2\sigma^2\boldsymbol{\lambda}' V_n \boldsymbol{\lambda}}{2(\lambda_1^2+\lambda_2^2)\max_{1\leq i\leq n}(d_{ni}^2 + nc_{ni}/t_n)} \to \infty
$$

となることよりわかる．任意の $\boldsymbol{\lambda}$ なので，Cramér - Wold の補題よりこの定理の主張は示される．□

補遺

定理 4.4　　各 $n = 1, 2, \ldots$ に対して，$\{X_{nj}\}_{j=1}^{n}$ を独立な確率変数列で，$\mathbb{E}[X_{nj}] = 0$，$\mathbb{VAR}[X_{nj}] = \sigma_{nj}^2$ とする．$Z_n = \sum_{j=1}^{n} X_{nj}$ と $B_n^2 = \mathbb{VAR}[X_{nj}] = \sum_{j=1}^{n}\sigma_{nj}^2$ とおく．このとき，任意の正数 $\epsilon > 0$ に対して，

$$
\frac{1}{B_n^2}\sum_{j=1}^{n}\mathbb{E}\left[X_{nj}^2 I\{|X_{nj}| \geq \epsilon B_n\}\right] \to 0, \qquad (n\to\infty) \tag{4.76}
$$

ならば，

$$
\frac{Z_n}{B_n} \xrightarrow{d} N(0,\,1)
$$

が成立する．

　逆に，

$$
\frac{1}{B_n^2}\max_{1\leq j\leq n}\sigma_{nj}^2 \to 0,\ (n\to\infty) \qquad \text{かつ} \qquad \frac{Z_n}{B_N} \xrightarrow{d} N(0,\,1)
$$

ならば，(4.76) が成立する．

定理 4.5　Cramér - Wold の工夫)　　$\{\boldsymbol{Y}_n\}_{n=1}^{\infty}$ を $\mathbb{R}^p$ - 値確率変数列とし，$\boldsymbol{Y}$ を確率ベクトルとする．このとき，

$$
\boldsymbol{Y}_n \xrightarrow{L} \boldsymbol{Y}
$$

となるための必要十分条件は，任意の固定された実数ベクトル $\boldsymbol{\alpha}$ に対して

$$\boldsymbol{\alpha}'\boldsymbol{Y}_n \xrightarrow{L} \boldsymbol{\alpha}'\boldsymbol{Y}$$

が成り立つことである．

・・・・・4.2・・・・・

一般化線形モデル

$Y_1, Y_2, \ldots, Y_n$ を独立な確率変数とし，$Y_i\,(i = 1, 2, \ldots, n)$ は $\mathbb{R}$ 上の測度 $\nu(\cdot)$ に関する確率 (密度) 関数

$$f(y_i|\,\theta_i,\,\phi_i) = c(y_i,\,\phi_i)\exp\left[\frac{y_i\theta_i - b(\theta_i)}{\phi_i}\right] \tag{4.77}$$

を持つとする．ここで，$\theta_1, \theta_2, \ldots, \theta_n$ は興味のある母数，$\phi_i(> 0)$ は局外母数，$b(\cdot), c(\cdot,\cdot)$ は既知の関数である．さらに，

$$\theta_i \in \Theta = \{\theta \in \mathbb{R}:\, 0 < \int c(y,\,\phi)e^{\theta y}\nu(dy),\, \phi > 0\}$$

とし，Θ^0 を Θ の内部とする．$b(\cdot)$ は Θ^0 上の 2 階連続微分可能関数とする．また，すべての $\theta \in \Theta^0$ に対し，$\ddot{b}(\theta) > 0$ とする．確率 (密度) 関数 (4.77) を持つ分布を指数分布族という．(4.77) より，各 i に対し

$$\int c(y,\,\phi_i)\exp\left[\frac{y\theta_i}{\phi_i}\right]\,d\nu(y) = \exp\left[\frac{b(\theta_i)}{\phi_i}\right]$$

となる．$\mathbb{E}[\,|Y_i|] < \infty$ のとき，上の式の両辺を θ_i に関して微分すれば

$$\int c(y,\,\phi_i)\left\{\frac{y}{\phi_i}\right\}\exp\left[\frac{y_i\theta_i}{\phi_i}\right]\,d\nu(y) = \frac{\dot{b}(\theta_i)}{\phi_i}\exp\left[\frac{b(\theta_i)}{\phi_i}\right] \tag{4.78}$$

となることから

$$\mathbb{E}[Y_i] = \int c(y,\,\phi_i)y\exp\left[\frac{y\theta_i}{\phi_i}\right]\nu(dy) = \dot{b}(\theta_i)$$

を得る．ただし，$\dot{b}(\theta) = (\partial/\partial\theta)b(\theta)$ である．さらに，$\mathbb{VAR}[Y_i] < \infty$ のとき，(4.78) の両辺を θ_i に関して微分すれば，

$$\int c(y,\,\phi_i)\left\{\frac{y}{\phi_i}\right\}^2\exp\left[\frac{y\theta_i}{\phi_i}\right]\,d\nu(y) = \left[\frac{\ddot{b}(\theta_i)}{\phi_i} + \left\{\frac{\dot{b}(\theta_i)}{\phi_i}\right\}^2\right]\exp\left[\frac{b(\theta_i)}{\phi_i}\right]$$

となることから

$$\mathbb{E}[Y_i^2] = \int c(y,\,\phi_i)y^2\exp\left[\frac{y\theta_i}{\phi_i}\right]\nu(dy) = \phi_i\ddot{b}(\theta_i) + \{\dot{b}(\theta_i)\}^2 = \phi_i\ddot{b}(\theta_i) + \{\mathbb{E}[Y_i]\}^2$$

となる．したがって，

$$\mathbb{E}[Y_i] = \mu_i \equiv \mu(\theta_i) = \dot{b}(\theta_i), \quad \mathbb{VAR}[Y_i] = \phi_i\ddot{b}(\theta_i) \tag{4.79}$$

となる．さらに，$\theta_i = \dot{b}^{-1}(\mu_i)$ より

$$\mathbb{VAR}[Y_i] = \phi_i \ddot{b}(\dot{b}^{-1}(\mu_i)) \equiv \phi_i V(\mu_i) \tag{4.80}$$

と書くこと[1]にする．(4.80) の V を分散関数とよぶことにする．

いま，Y_i に対応する q 次元の共変量 $\boldsymbol{x}_i$ が

$$g(\mu(\theta_i)) = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{4.81}$$

と関連付けられていると仮定する．ただし，$\boldsymbol{\beta}$ は q 次元未知母数とし，$g : \mathbb{R} \to \mathbb{R}$ は既知の関数で $\{\mu(\theta) : \theta \in \Theta^0\}$ 上で一対一対応で 3 階連続微分可能とする．g は平均 $\mathbb{E}[Y_i]$ と線形予測母数 $\boldsymbol{x}_i'\boldsymbol{\beta}$ を結ぶ関数でリンク関数と呼ぶ．また，

$$\theta_i = (g \circ \mu)^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}) \equiv h(\boldsymbol{x}_i'\boldsymbol{\beta})$$

と表現でき，$g = \mu^{-1}$ のとき，正準リンク関数と呼ぶ．

一般化線形モデルにおいては，$\boldsymbol{\beta}$ が興味ある母数である．共変量 $\boldsymbol{x}_i$ が動く範囲を $\mathcal{X}$ としたとき，

$$\mathcal{B} = \{\boldsymbol{\gamma} \in \mathbb{R} : \text{すべての } \boldsymbol{x} \in \mathcal{X} \text{ に対し}, (g \circ \mu)^{-1}(\boldsymbol{\gamma}'\boldsymbol{x}_i) \in \Theta^0\} \tag{4.82}$$

とする．このとき，$\boldsymbol{\beta} \in \mathcal{B}$ とする．多くのモデルでは

$$\phi_i = \frac{\phi}{t_i}, \qquad i = 1, 2, \ldots, n \tag{4.83}$$

と仮定できる．ただし，t_i は既知の正数，$\phi > 0$ は未知の正数である．

例 4.1　　$Z_1, Z_2, \ldots, Z_n$ は独立な確率変数で，各 $Z_i\,(i = 1, 2, \ldots, n)$ は二項分布 $\mathrm{Bi}(n, p_i)$, $0 < p_i < 1$ に従うものとする．$Y_i = Z_i/n$ とおいたとき，Y_i の確率関数は

$$f(y|\, n, p_i) = \begin{pmatrix} n \\ ny \end{pmatrix} p_i^{ny}(1-p_i)^{n(1-y)}, \qquad y = 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$$

となる．これは

$$\begin{aligned} f(y|\, n, p_i) &= \begin{pmatrix} n \\ ny \end{pmatrix} \exp\bigl\{ny \log p_i + n(1-y)\log(1-p_i)\bigr\} \\ &= \begin{pmatrix} n \\ ny \end{pmatrix} \exp\Bigl[\Bigl\{y \log\frac{p_i}{1-p_i} + \log(1-p_i)\Bigr\}n\Bigr] \\ &= c(y, \phi_i)\exp\left[\frac{y\theta_i - b(\theta_i)}{\phi_i}\right] \end{aligned}$$

と表現できる．ただし，

$$\theta_i = \log\{p_i/(1-p_i)\},\ b(\theta_i) = \log(1 + e^{\theta_i}),\ \phi_i = n^{-1},\ c(y, \phi_i) = \begin{pmatrix} n \\ ny \end{pmatrix},\ \phi_i = n^{-1}$$

---

[1] すなわち，$V(s) = (\ddot{b} \circ \dot{b}^{-1})(s)$ である．

である．

ここで 2 つの一般化線形モデルを考えることができる．

(i) 正準リンク関数

$$\theta_i = \log \frac{p_i}{1-p_i} = \alpha + \beta x_i, (i = 1,\, 2), \qquad x_1 = 1,\, x_2 = -1$$

で α と β は未知関数である．

(ii) リンク関数が $g(z) = z$

$$p_i = \frac{e^{\theta_i}}{1+e^{\theta_i}} = \alpha + \beta x_i, (i = 1,\, 2), \qquad x_1 = 1,\, x_2 = -1$$

で α と β は未知関数である．

どちらのモデルもモデルの同一性 $(\beta = 0)$ を調べるのが目的である．

例 4.2　$Y_1, Y_2, \ldots, Y_n$ は独立な確率変数列で，各 $Y_i\,(i = 1, 2, \ldots, n)$ はポアソン (Poisson) 分布に $Po(\lambda_i)$ に従うものとする．すなわち，各 Y_i の確率関数は

$$f(y|\,\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^y}{y!}, \qquad y = 0,\, 1,\, \ldots$$

で与えられる．これを書き換えると

$$f(y|\,\lambda_i) = \frac{1}{y!}\exp\{y\log\lambda_i - \lambda_i\} = c(y,\,\phi_i)\exp\{y\theta_i - b(\theta_i)\}$$

となる．ただし，

$$\theta_i = \log\lambda_i, \qquad b(\theta_i) = e^{\theta_i}, \qquad \phi_i = 1, \qquad c(y,\,\phi_i) = \frac{1}{y!}, \qquad \phi_i = 1$$

である．

各 Y_i は q 次元ベクトル $\boldsymbol{x}_i$ と q 次元説明変数 $\boldsymbol{\beta}$ で表現されると仮定する．

(i) 正準リンク関数

$$\theta_i = \log\lambda_i = \boldsymbol{x}_i'\boldsymbol{\beta}, \qquad i = 1,\, 2,\, \ldots,\, n$$

ただし，$\boldsymbol{\beta}$ は未知母数である．

(ii) リンク関数が $g(z) = z$

$$\lambda_i = e^{\theta_i} = \boldsymbol{x}_i'\boldsymbol{\beta}, \qquad i = 1,\, 2,\, \ldots,\, n$$

例 4.3　$Y_1, Y_2, \ldots, Y_n$ は独立な確率変数列で，各 $Y_i\,(i = 1, 2, \ldots, n)$ はベルヌーイ分布 $Bin(1,\, p_i)$ に従うものとする．すなわち，各 Y_i の確率関数は

$$f(y|\,p_i) = c(y,\,\phi_i)\exp\{y\theta_i - b(\theta_i)\}$$

と表現できる．ただし，

$$\theta_i = \log\frac{p_i}{1-p_i}, \qquad b(\theta_i) = \log(1+e^{\theta_i}), \qquad \phi_i = 1, \qquad c(y,\,\phi_i) = 1, \qquad a(\phi_i) = 1$$

である．

ここで例 4.1 同様に $\boldsymbol{x}_i$ の回帰モデルを考える．

$$\theta_i = \log \frac{p_i}{1-p_i} = \boldsymbol{x}_i'\boldsymbol{\beta}, \qquad i = 1,\, 2,\, \ldots,\, n$$

と仮定する．ただし，$\boldsymbol{\beta}$ は未知母数である．これは

$$p_i = \frac{1}{1+\exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\}}$$

となる．

仮定 (4.81) と (4.83) のもとの一般化線形モデルモデルの $\boldsymbol{\beta}$ の最尤推定量を求める．対数尤度は

$$\log L_n(\boldsymbol{\beta},\, \phi) = \sum_{i=1}^{n} \left[\log c(y_i,\, \phi/t_i) + \frac{h(\boldsymbol{x}_i'\boldsymbol{\beta})y_i - b\{h(\boldsymbol{x}_i'\boldsymbol{\beta})\}}{\phi/t_i} \right]$$

となる．(4.79) に注意すれば，尤度方程式は

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log L_n(\boldsymbol{\beta},\, \phi) = \frac{1}{\phi} \sum_{i=1}^{n} [y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))] \dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta}) t_i \boldsymbol{x}_i = 0$$

と

$$\frac{\partial}{\partial \phi} \log L_n(\boldsymbol{\beta},\, \phi) = \sum_{i=1}^{n} \left[\frac{\partial \log c(y_i,\, \phi/t_i)}{\partial \phi} - \frac{t_i[h(\boldsymbol{x}_i'\boldsymbol{\beta})y_i - b\{h(\boldsymbol{x}_i'\boldsymbol{\beta})\}]}{\phi^2} \right] = 0$$

となる．$\boldsymbol{\beta}$ の最尤推定量が存在すれならば，一番目の尤度方程式より，ϕ を推定することなく，$\boldsymbol{\beta}$ の最尤推定量を求めることができるので，

$$\dot{l}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))] \dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta}) t_i \boldsymbol{x}_i$$

とおくことにする．$h(s) = (g \circ \mu)^{-1}(s) = (g \circ \dot{b})^{-1}(s)$ と (4.80) に注意する．$y = (g \circ \dot{b})^{-1}(s) = \dot{b}^{-1}(g^{-1}(s)$ とおけば，

$$\begin{aligned} \dot{h}(s) &= \frac{1}{\frac{d}{dy}(g \circ \dot{b})(y)} = \frac{1}{\dot{g}(\dot{b}(y))\ddot{b}(y)} = \frac{1}{\dot{g}(\dot{b}(\dot{b}^{-1}(g^{-1}(s)))\ddot{b}(\dot{b}^{-1}(g^{-1}(s))} \\ &= \frac{1}{\dot{g}(g^{-1}(s))V(g^{-1}(s))} \end{aligned}$$

となる．これより $\boldsymbol{\beta}$ の正規方程式は

$$\dot{l}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\{y_i - g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})\}t_i}{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))} \boldsymbol{x}_i = \boldsymbol{0} \tag{4.84}$$

となる．

いま

$$M_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 \ddot{b}(h(\boldsymbol{x}_i'\boldsymbol{\beta})) t_i \boldsymbol{x}_i \boldsymbol{x}_i' \tag{4.85}$$

と

$$R_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} [Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))]\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})t_i\boldsymbol{x}_i\boldsymbol{x}_i' \tag{4.86}$$

とおく．このとき

$$\begin{aligned}\mathbb{VAR}\left[\frac{\partial}{\partial\boldsymbol{\beta}}\log L_n(\boldsymbol{\beta},\,\phi)\right] &= \frac{1}{\phi^2}\sum_{i=1}^{n}\mathbb{VAR}[\{Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))\}\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})t_i\boldsymbol{x}_i] \\ &= \frac{1}{\phi^2}\sum_{i=1}^{n}\mathbb{VAR}[Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))][\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2t_i^2\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{\phi^2}\sum_{i=1}^{n}\frac{\phi}{t_i}\ddot{b}(h(\boldsymbol{x}_i'\boldsymbol{\beta}))[\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2t_i^2\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{\phi}M_n(\boldsymbol{\beta}) \end{aligned}\tag{4.87}$$

となる．また

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\log L_n(\boldsymbol{\beta},\,\phi) = \frac{1}{\phi}\ddot{l}_n(\boldsymbol{\beta}) = \frac{R_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\beta})}{\phi} \tag{4.88}$$

となる．

定理 4.6　(4.77), (4.81), (4.82), (4.83) をみたす一般化線形モデルを考える．さらに，以下を仮定する．
(GLM 1)　(4.82) で定義された $\boldsymbol{\beta}$ の定義域は $\mathbb{R}^q$ の開集合
(GLM 2)　真の母数 $\boldsymbol{\beta}$ において

$$0 < \inf_{i=1,\,2,\,\ldots,\,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta}) \le \sup_{i=1,\,2,\,\ldots,\,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta}) < \infty$$

とする．ただし，$\varphi(t) = [\dot{h}(t)]^2\ddot{b}(h(t))$ と $h(t) = (g\circ\mu)^{-1}(t)$ とする．
(GLM 3)　$n\to\infty$ のとき

$$\max_{i=1,\,2,\,\ldots,\,n}\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \to 0, \qquad \mathrm{ch}_{\min}(\boldsymbol{X}_n'\boldsymbol{X}_n)\to 0$$

とする．ただし，$\boldsymbol{X}_n' = (\boldsymbol{x}_1,\,\boldsymbol{x}_2,\,\ldots,\,\boldsymbol{x}_n)$ とし，$\mathrm{ch}_{\min}(\boldsymbol{X}_n'\boldsymbol{X}_n)$ は $\boldsymbol{X}_n'\boldsymbol{X}_n$ の最小の固有根とする．
(GLM 4)　(4.79) において $t_i\in(t_0,\,t_\infty)$ とする．ただし，$0<t_0<t_\infty<\infty$ とする．
<u>一致性</u>　このとき

$$\mathbb{P}(\dot{l}_n(\hat{\boldsymbol{\beta}}_n) = 0)\to 1, \qquad \hat{\boldsymbol{\beta}}_n\to\boldsymbol{\beta} \tag{4.89}$$

をみたす推定量列 $\{\hat{\boldsymbol{\beta}}_n\}_{n=1}^{\infty}$ が唯一存在する
<u>漸近正規性</u>　$I_n(\boldsymbol{\beta}) = \mathbb{VAR}[\dot{l}_n(\boldsymbol{\beta})]$ とおく．このとき

$$[I_n(\boldsymbol{\beta})]^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{L} N(0,\,I_p)$$

が成立する．

証明

一致性　正数 a に対し

$$B_n(a) = \{\boldsymbol{\gamma} \in \mathcal{B} :\ \|[I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})\| \le a\}$$

とおく．$\mathcal{B}$ は開集合なので，n を十分大きく取れば，$B_n(a) \subset \mathcal{B}$ とできる．また，$\cap_{n=1}^{\infty} B_n(a) = \{\boldsymbol{\beta}\}$ となる．これらから (4.89) をみたす推定量列 $\{\hat{\boldsymbol{\beta}}\}_{n=1}^{\infty}$ の存在を示すためには，任意の正数 ϵ に対し，a と正の整数 n_0 をうまく取れば

$$\begin{aligned}\mathbb{P}\{\ &\text{すべての } \boldsymbol{\gamma} \in \partial B_n(a)\,(n \ge n_0) \text{ とすべての } \phi > 0 \text{ に対し} \\ &\log L_n(\boldsymbol{\gamma}, \phi) - \log L_n(\boldsymbol{\beta}, \phi) < 0\} \ge 1 - \epsilon\end{aligned} \tag{4.90}$$

とできることを示せばよい．ただし，$\partial B_n(a)$ は $B_n(a)$ の境界である．

テーラー展開を用いれば，各 $\boldsymbol{\gamma} \in \partial B_n(a)$ とすべての $\phi > 0$ に対し

$$\log L_n(\boldsymbol{\gamma}, \phi) - \log L_n(\boldsymbol{\beta}, \phi) = a\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi} + \frac{a^2}{2}\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\frac{\ddot{l}_n(\boldsymbol{\gamma}^*)}{\phi}[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{\lambda}$$

となる．ただし，$\boldsymbol{\lambda} = [I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})/a$, $\ddot{l}_n(\boldsymbol{\gamma}) = \partial\dot{l}_n(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^*$ は $\boldsymbol{\gamma}$ と $\boldsymbol{\beta}$ を結んだ直線上の点である．$\boldsymbol{\gamma} \in \partial B_n(a)$ より $\|\boldsymbol{\lambda}\| = 1$ である．

まず,

$$\max_{\boldsymbol{\gamma} \in B_n(a)}\left\|[I_n(\boldsymbol{\beta})]^{-1/2}\frac{\ddot{l}_n(\boldsymbol{\gamma}^*)}{\phi}[I_n(\boldsymbol{\beta})]^{-1/2}\right\| \xrightarrow{P} 0 \tag{4.91}$$

を示すことにする．(4.85) と (4.86) に注意すれば，(4.91) を示すためには

$$\max_{\boldsymbol{\gamma} \in \partial B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\gamma}) - M_n(\boldsymbol{\beta})][M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.92}$$

と

$$\max_{\boldsymbol{\gamma} \in \partial B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}R_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.93}$$

を示せばよいことがわかる．まず，(4.92) を評価する．そのために，

$$\Phi(\boldsymbol{\gamma}) = \mathrm{diag}(t_1\varphi(\boldsymbol{x}_1'\boldsymbol{\beta}),\ t_2\varphi(\boldsymbol{x}_2'\boldsymbol{\beta}),\ \ldots,\ t_n\varphi(\boldsymbol{x}_n'\boldsymbol{\beta}))$$

とおく．さらに，H を $n \times q$ の半直交行列で $H'H = I_q$ をみたすものとし，G を $q \times q$ の正則行列で $H'\boldsymbol{X}_n = G$ をみたすものとする．このとき

$$\begin{aligned}&\mathrm{tr}\{[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\gamma}) - M_n(\boldsymbol{\beta})][M_n(\boldsymbol{\beta})]^{-1/2}\} \\ &\quad = \mathrm{tr}\{(\boldsymbol{X}_n'\Phi(\boldsymbol{\beta})\boldsymbol{X}_n)^{-1}[\boldsymbol{X}_n'\Phi(\boldsymbol{\gamma})\boldsymbol{X}_n - \boldsymbol{X}_n'\Phi(\boldsymbol{\beta})\boldsymbol{X}_n]\}^2 \\ &\quad = \mathrm{tr}\{(G'H'\Phi(\boldsymbol{\beta})HG)^{-1}[G'H'(\Phi(\boldsymbol{\gamma}) - \Phi(\boldsymbol{\beta}))HG]\}^2 \\ &\quad \le q\frac{t_\infty^2 \max_{\boldsymbol{\gamma} \in B_n(a)} \sup_{i=1,2,\ldots,n} |\varphi(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{t_0^2 \min_{i=1,2,\ldots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|^2}\end{aligned}$$

となるので，(4.92) の上限は

$$\sqrt{q}\frac{t_\infty}{t_0}\frac{\max_{\boldsymbol{\gamma} \in B_n(a)} \sup_{i=1,2,\ldots,n} |\varphi(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})|}{\inf_{i=1,2,\ldots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|} \to 0, \qquad (n \to \infty)$$

となる．なぜならば，$\boldsymbol{\gamma} \in B_n(a)$ に対し，$n \to \infty$ のとき

$$
\begin{aligned}
|\boldsymbol{x}_i'\boldsymbol{\gamma} - \boldsymbol{x}_i'\boldsymbol{\beta}|^2 &\le |\boldsymbol{x}_i'[I_n(\boldsymbol{\beta})]^{-1/2}[I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})|^2 \\
&\le \|[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{x}_i\|^2 \|[I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})\|^2 \\
&\le a^2 \max_{i=1,2,\dots,n} \boldsymbol{x}_i'[I_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i \qquad (4.94) \\
&\le a^2 \max_{i=1,2,\dots,n} \boldsymbol{x}_i'[M_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i \\
&\le a^2 \frac{1}{t_0 \min_{i=1,2,\dots,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})} \max_{i=1,2,\dots,n} \boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \to 0, \qquad (4.95)
\end{aligned}
$$

となることと φ の連続性よりわかる．このことより，(4.92) は示せた．

つぎに，(4.93) を示すために以下のようにおく．

$$
\begin{aligned}
e_i &= Y_i - \mu(h(\boldsymbol{x}_i\boldsymbol{\beta})) \\
U_n(\boldsymbol{\gamma}) &= \sum_{i=1}^n [\mu(h(\boldsymbol{x}_i'\boldsymbol{\beta})) - \mu(h(\boldsymbol{x}_i'\boldsymbol{\gamma}))]\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})t_i\boldsymbol{x}_i\boldsymbol{x}_i' \\
V_n(\boldsymbol{\gamma}) &= \sum_{i=1}^n e_i[\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]t_i\boldsymbol{x}_i\boldsymbol{x}_i' \\
W_n &= \sum_{i=1}^n e_i\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})t_i\boldsymbol{x}_i\boldsymbol{x}_i'
\end{aligned}
$$

このとき

$$
R_n(\boldsymbol{\gamma}) = U_n(\boldsymbol{\gamma}) + V_n(\boldsymbol{\gamma}) + W_n
$$

となる．まず

$$
\max_{\boldsymbol{\gamma} \in B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}U_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \qquad (4.96)
$$

がわかる．これは

$$
\mathrm{tr}\{[M_n(\boldsymbol{\beta})]^{-1/2}U_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\}^2 \le q\frac{t_\infty^2 \max_{\boldsymbol{\gamma} \in B_n(a)} \sup_{i=1,2,\dots,n} |\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{t_0^2 \inf_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|^2}
$$

になることに注意すれば，(4.95) と $\ddot{h}$ の連続性よりわかる．つぎに

$$
\max_{\boldsymbol{\gamma} \in B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}V_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \qquad (4.97)
$$

を示す．まず

$$
\begin{aligned}
\mathbb{E}\left[\left\{\sum_{i=1}^n e_i[\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]t_i\boldsymbol{x}_i\boldsymbol{x}_i'\right\}^2\right] &= \sum_{i=1}^n [\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 t_i^2 (\boldsymbol{x}_i\boldsymbol{x}_i')^2 \mathbb{E}[e_i^2] \\
&= \sum_{i=1}^n \phi_i t_i^2 [\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 t_i^2 (\boldsymbol{x}_i\boldsymbol{x}_i')^2 \ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})
\end{aligned}
$$

となることに注意する．これから，$n \to \infty$ の

$$
\begin{aligned}
\mathbb{E}[\mathrm{tr}\{M_n^{-1}(\boldsymbol{\beta})V_n(\boldsymbol{\gamma})M_n^{-1}(\boldsymbol{\beta})V_n(\boldsymbol{\gamma})\}] &\le \mathbb{E}\left[\frac{q^2\mathrm{tr}\{V_n^2(\boldsymbol{\gamma})\}}{\{\mathrm{tr}M_n(\boldsymbol{\beta})\}^2}\right] \\
&\le q^2\phi t_\infty \frac{\sum_{i=1}^n \ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'\boldsymbol{x}_i)^2 \max_{\boldsymbol{\gamma}\in B_n(a)} |\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{\{\sum_{i=1}^n [\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})t_i\boldsymbol{x}_i'\boldsymbol{x}_i\}^2} \\
&\le \frac{q^2\phi t_\infty}{t_0^2} \sup_{i=1,2,\dots,n} \max_{\boldsymbol{\gamma}\in B_n(a)} \left|\frac{\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})}{\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})} - 1\right|^2 \frac{\sup_{i=1,2,\dots,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}{\inf_{i=1,2,\dots,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})} \frac{\sum_{i=1}^n(\boldsymbol{x}_i'\boldsymbol{x}_i)^2}{(\sum_{i=1}^n \boldsymbol{x}_i'\boldsymbol{x}_i)^2} \\
&\to 0
\end{aligned} \tag{4.98}
$$

となることが $\ddot{h}$ の連続性と (4.95) よりわかる．(4.98) と Markov の不等式より (4.97) は示せた．つぎに

$$
\|[M_n(\boldsymbol{\beta})]^{-1/2}W_n[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.99}
$$

を示す．

$$
\begin{aligned}
&\mathbb{E}[\|[M_n(\boldsymbol{\beta})]^{-1/2}W_n[M_n(\boldsymbol{\beta})]^{-1/2}\|^2] \\
&\quad\le \inf_{i=1,2,\dots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbb{E}[\{\sum_{i=1}^n h(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i\}^2] \\
&\quad= \inf_{i=1,2,\dots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n h^2(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2\mathbb{E}[e_i] \\
&\quad= \inf_{i=1,2,\dots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n h^2(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2\frac{\phi}{t_i}\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta}) \\
&\quad\le \frac{\phi}{t_0}\inf_{i=1,2,\dots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sup_{i=1,2,\dots,n}\varphi^2(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2 \\
&\quad\le \frac{q\phi}{t_0}\inf_{i=1,2,\dots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sup_{i=1,2,\dots,n}\varphi^2(\boldsymbol{x}_i'\boldsymbol{\beta})\max_{i=1,2,\dots,n}\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \to 0
\end{aligned}
$$

となる．したがって，Markov の不等式を用いれば，(4.99) は示せた．よって，(4.96), (4.97), (4.99) より (4.93) は示せた．したがって，(4.92) と (4.93) が示せたので，(4.91) が示せた．よって

$$
\log L_n(\boldsymbol{\gamma}, \phi) - \log L_n(\boldsymbol{\beta}, \phi) = a\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi} - \frac{a^2}{2} + o_P(1)
$$

となる．Cauchy - Schwarz の不等式から

$$
\max_{\boldsymbol{\lambda}\in\partial B_n(a)}|\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{\lambda}|^2 \le \max_{\boldsymbol{\lambda}\in\partial B_n(a)}\|\boldsymbol{\lambda}\|^2\dot{l}_n'(\boldsymbol{\beta})[I_n(\boldsymbol{\beta})]^{-1}\dot{l}_n(\boldsymbol{\beta})
$$

と

$$
\mathbb{E}[\dot{l}_n'(\boldsymbol{\beta})[I_n(\boldsymbol{\beta})]^{-1}\dot{l}_n(\boldsymbol{\beta})] = \mathrm{tr}\{[I_n(\boldsymbol{\beta})]^{-1}\mathbb{E}[\dot{l}_n(\boldsymbol{\beta})\dot{l}_n'(\boldsymbol{\beta})]\} = q
$$

となる．これより a を十分大きくとれば

$$
\mathbb{P}\{\|\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta}]^{-1/2}\boldsymbol{\lambda}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi}\| \le \frac{a}{4}\} \ge 1 - \left(\frac{4}{a\phi}\right)^2\mathbb{E}[\dot{l}_n'(\boldsymbol{\beta})[I_n(\boldsymbol{\beta})]^{-1}\dot{l}_n(\boldsymbol{\beta})] = 1 - \frac{16q}{(a\phi)^2} \ge 1 - \epsilon
$$

となる．したがって，(4.91) は示せた．
一意性

$$\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\dot{l}_n(\boldsymbol{\gamma}) = \boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\dot{l}_n(\boldsymbol{\beta}) + a\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\ddot{l}_n(\boldsymbol{\gamma}^*)[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{\lambda}$$

に注意する．(4.91) より，$n \to \infty$ のとき

$$\mathbb{P}\left(\sup_{\boldsymbol{\gamma}\in B_n(a)} \ddot{l}_n(\boldsymbol{\gamma}) < 0\right) \to 1$$

となるので，一意性は示された．
漸近正規性　正数 ϵ に対し

$$A_\epsilon = \{\boldsymbol{\gamma} \in \mathcal{B} :\ \|\boldsymbol{\gamma} - \boldsymbol{\beta}\| \le \epsilon\}$$

とおく．平均値の定理を用いれば

$$\dot{l}_n(\hat{\boldsymbol{\beta}}_n) = \dot{l}_n(\boldsymbol{\beta}) + \int_0^1 \ddot{l}_n(\boldsymbol{\beta} + t(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\,dt(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$$

となる．$\mathbb{P}(\dot{l}_n(\hat{\boldsymbol{\beta}}_n) = 0) \to 0\,(n \to \infty)$ より

$$\begin{aligned}&\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi}\\ &\quad = -[M_n(\boldsymbol{\beta})]^{-1/2}\int_0^1 \ddot{l}_n(\boldsymbol{\beta} + t(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\,dt[M_n(\boldsymbol{\beta})]^{-1/2}\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + o_P(1)\end{aligned}$$

となる．また，$\hat{\boldsymbol{\beta}}_n \in A_\epsilon$ のとき

$$\begin{aligned}&\left\|[M_n(\boldsymbol{\beta})]^{-1/2}\int_0^1 \ddot{l}_n(\boldsymbol{\beta} + t(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\,dt[M_n(\boldsymbol{\beta})]^{-1/2} + I_q\right\|\\ &\quad \le \max_{\boldsymbol{\gamma}\in A_\epsilon}\left\|[M_n(\boldsymbol{\beta})]^{-1/2}[R_n(\boldsymbol{\gamma}) + M_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\gamma})][M_n(\boldsymbol{\beta})]^{-1/2}\right\|\\ &\quad \le \max_{\boldsymbol{\gamma}\in A_\epsilon}\left\|[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\gamma})][M_n(\boldsymbol{\beta})]^{-1/2}\right\|\\ &\quad + \max_{\boldsymbol{\gamma}\in A_\epsilon}\left\|[M_n(\boldsymbol{\beta})]^{-1/2}R_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\right\|\\ &\quad \xrightarrow{P} 0\end{aligned}$$

となることが (4.92) と (4.93) からわかる．さらに，$\mathbb{P}(\hat{\boldsymbol{\beta}}_n \in A_\epsilon) \to 1\,(n \to \infty)$ に注意すれば，

$$\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi} = \left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + o_P(1) \tag{4.100}$$

がわかる．つぎに，任意のベクトル $\boldsymbol{\alpha} = (a_1, a_2, \ldots, a_q)(\neq 0) \in \mathbb{R}^q$ に対して

$$\boldsymbol{\alpha}'\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi} = \sum_{i=1}^n \boldsymbol{\alpha}'\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{t_i\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})\sqrt{\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}}{\phi}\boldsymbol{x}_i\frac{Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))}{\sqrt{\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}}$$

に Hájek-Šidak の中心極限定理 (Sen and Singer (1993), p.119) を用いる：

$$\sum_{i=1}^{n}\boldsymbol{\alpha}'\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{t_i\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})\sqrt{\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}}{\phi}\boldsymbol{x}_i\boldsymbol{x}_i'\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\boldsymbol{\alpha}=\phi\boldsymbol{\alpha}'\boldsymbol{\alpha}$$

と

$$\begin{aligned}
&\max_{i=1,\,2,\,\dots,\,n}\left\{\frac{t_i^2\dot{h}^2(\boldsymbol{x}_i'\boldsymbol{\beta})\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}{\phi}[\boldsymbol{\alpha}'[M_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i]^2\right\}\Bigg/\sum_{i=1}^{n}\left\{\frac{t_i^2\dot{h}^2(\boldsymbol{x}_i'\boldsymbol{\beta})\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}{\phi}[\boldsymbol{\alpha}'[M_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{x}_i]^2\right\}\\
&\qquad\le\max_{i=1,\,2,\,\dots,\,q}a_i^2\left\{\frac{t_\infty^2\sup_{i=1,\,2,\,\dots,\,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}{t_0\inf_{i=1,\,2,\,\dots,\,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}\right\}^2\max_{i=1,\,2,\,\dots,\,n}\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i\Big/\boldsymbol{\alpha}'\boldsymbol{\alpha}\phi I_q\\
&\qquad\to 0
\end{aligned}$$

ととなるので

$$\boldsymbol{\alpha}'\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{\ddot{l}_n(\boldsymbol{\beta})}{\phi}\xrightarrow{L}N(0,\,\phi\boldsymbol{\alpha}'\boldsymbol{\alpha})$$

となる．したがって，Cramér - Wold の工夫から

$$\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2}\frac{\ddot{l}_n(\boldsymbol{\beta})}{\phi}\xrightarrow{L}N(0,\,\phi I_q)$$

最後に，(4.100) に注意して Slutsky の定理を用いれば，漸近正規性は証明される．□

最尤推定値の計算アルゴリズム

データが与えられたとき，$\boldsymbol{\beta}$ の最尤推定値 $\widehat{\boldsymbol{\beta}}_n$ 値を具体的に求めるための手続きを考える．$\boldsymbol{\beta}$ の最尤推定値は正規方程式

$$S(\boldsymbol{\beta})\equiv\frac{\partial}{\partial\boldsymbol{\beta}}\log L_n(\boldsymbol{\beta},\,\phi)=\frac{1}{\phi}\sum_{i=1}^{n}\frac{\{y_i-g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})\}t_i}{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))}\boldsymbol{x}_i=\boldsymbol{0}$$

の解であった．この解を Fisher の scoring アルゴリズムを用いて求めよう．(4.80) と (4.81) から

$$\begin{aligned}
H(\boldsymbol{\beta})\equiv\mathbb{VAR}[S(\boldsymbol{\beta})]&=\frac{1}{\phi^2}\sum_{i=1}^{n}\left\{\frac{t_i^2}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2}\mathbb{VAR}[Y_i-g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})]\right\}\boldsymbol{x}_i\boldsymbol{x}_i'\\
&=\frac{1}{\phi^2}\sum_{i=1}^{n}\left\{\frac{t_i^2}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2}\frac{\phi}{t_i}V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\right\}\boldsymbol{x}_i\boldsymbol{x}_i'\\
&=\frac{1}{\phi}\sum_{i=1}^{n}\frac{t_i}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))}\boldsymbol{x}_i\boldsymbol{x}_i'
\end{aligned}$$

となる．k 回目の繰り返し計算で求められた $\boldsymbol{\beta}$ の最尤推定値の値を $\widehat{\boldsymbol{\beta}}_n^{(k)}$ とおけば，$(k+1)$ 回目の繰り返し計算で得られる $\boldsymbol{\beta}$ の最尤推定値の値は

$$\widehat{\boldsymbol{\beta}}_n^{(k+1)}=\widehat{\boldsymbol{\beta}}_n^{(k)}+H^{-1}(\widehat{\boldsymbol{\beta}}_n^{(k)})S(\widehat{\boldsymbol{\beta}}_n^{(k)})\tag{4.101}$$

となる．(4.101) を書き換えると

$$\widehat{\boldsymbol{\beta}}_n^{(k+1)} = (\boldsymbol{X}'\boldsymbol{W}^{(k)}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(k)}\boldsymbol{z}^{(k)}$$

となる．ただし，$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)'$ とし，$\boldsymbol{W}^{(k)}$ は対角行列でその i-番目の対角成分を $w_i^{(k)}, i = 1, 2, \ldots, n,$ とし，$\boldsymbol{z}^{(k)} = (z_1^{(k)}, z_2^{(k)}, \ldots, z_n^{(k)})$ とし，

$$w_i^{(k)} = \frac{t_i}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))\}^2 V(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))}$$

$$z_i^{(k)} = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)} + \dot{g}(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))(y_i - g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))$$

である．したがって，$\widehat{\boldsymbol{\beta}}_n^{(k+1)}$ は $\boldsymbol{\beta}$ に関する目的関数

$$\sum_{i=1}^{n} w_i^{(k)}(z_i^{(k)} - \boldsymbol{x}_i'\boldsymbol{\beta})^2$$

を最小にする重み付き最小乗推定量となる．

補遺

定理 4.7　(Hájek-Šidak の中心極限定理)　$\{Y_n\}_{n=1}^{\infty}$ を平均が μ，分散が σ^2 の分布に独立同一に従う確率変数列とする．さらに，$\{\boldsymbol{c}_n\}_{n=1}^{\infty}$ を実ベクトル $\boldsymbol{c}_n = (c_{n1}, c_{n2}, \ldots, c_{nn})$ の列とする．このとき

$$\frac{\max_{1\le i\le n} c_{ni}^2}{\sum_{i=1}^n c_{ni}^2} \longrightarrow 0, \qquad (n \to \infty)$$

ならば

$$Z_n = \frac{\sum_{i=1}^n c_{ni}(Y_i - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n c_{ni}^2}} \xrightarrow{L} N(0, 1)$$

が成立する．

定理 4.8　Cramér - Wold の工夫)　$\{\boldsymbol{Y}_n\}_{n=1}^{\infty}$ を $\mathbb{R}^p$ - 値確率変数列とし，$\boldsymbol{Y}$ を確率ベクトルとする．このとき，

$$\boldsymbol{Y}_n \xrightarrow{L} \boldsymbol{Y}$$

となるための必要十分条件は，任意の固定された実数ベクトル $\boldsymbol{\alpha}$ に対して

$$\boldsymbol{\alpha}'\boldsymbol{Y}_n \xrightarrow{L} \boldsymbol{\alpha}'\boldsymbol{Y}$$

が成り立つことである．

5

ノンパラメトリック推定

補題 5.1 (Cramér)　$g : \mathbb{R}^d \to \mathbb{R}^k$ を写像とし，$\dot{g}(\boldsymbol{x})$ は $\boldsymbol{\mu} \in \mathbb{R}^d$ の近傍で連続微分可能とする[2)]．$\{\boldsymbol{X}_n\}_{n=1}^{\infty}$ は d 次元確率ベクトルの列で $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{L} \boldsymbol{X}$ を満足する．このとき，

$$\sqrt{n}(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{L} \dot{g}(\boldsymbol{\mu})\ \boldsymbol{X}$$

が成立する．特に，$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{L} N(\boldsymbol{0}, \Sigma)$ ならば，

$$\sqrt{n}(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{L} N(\boldsymbol{0}, \dot{g}(\boldsymbol{\mu})\,\Sigma\,\dot{g}(\boldsymbol{\mu})')$$

が成立する．

証明：$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \boldsymbol{X}$ より，$\boldsymbol{X}_n - \boldsymbol{\mu} = o_P(1)$ がわかる．また，g の $\boldsymbol{\mu}$ の近傍での連続微分可能性から十分小さな $\delta > 0$ に対して，$|\boldsymbol{x} - \boldsymbol{\mu}| < \delta$ ならば

$$g(\boldsymbol{x}) = g(\boldsymbol{\mu}) + \int_0^1 \dot{g}(\boldsymbol{\mu} + t(\boldsymbol{x} - \boldsymbol{\mu}))\,dt\,(\boldsymbol{x} - \boldsymbol{\mu})$$

となる．したがって，$|\boldsymbol{X}_n - \boldsymbol{\mu}| < \delta$ に対して

$$\sqrt{n}(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})) = \int_0^1 \dot{g}(\boldsymbol{\mu} + t(\boldsymbol{X}_n - \boldsymbol{\mu}))\,dt\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})$$

[2)] $\dot{g}(\boldsymbol{x}) = (\partial g_i/\partial x_j)_{i=1,\ldots,k,\,j=1,\ldots,d}$ である．ただし，　$g = (g_1, \ldots, g_k)'$ と $\boldsymbol{x} = (x_1, \ldots, x_d)'$ とした．

となる．$\dot{g}$ の連続性と $\mathbb{P}\{|\boldsymbol{X}_n - \boldsymbol{\mu}| < \delta\} \to 1\,(n \to \infty)$ から

$$\int_0^1 \dot{g}(\boldsymbol{\mu} + t(\boldsymbol{X}_n - \boldsymbol{\mu}))\,dt \xrightarrow{P} \dot{g}(\boldsymbol{\mu})$$

を得る[3)]．Slutsky の定理を用いれば，補題は証明された．□

---

3)
$$h(\boldsymbol{y}) = \int_0^1 \dot{g}(\boldsymbol{\mu} + t\boldsymbol{y})\,dt$$
とおく．$h(\boldsymbol{y})$ が $\boldsymbol{y} = 0$ で連続であることがわかれば，$\boldsymbol{X}_n - \boldsymbol{\mu} \xrightarrow{P} \boldsymbol{0}$ から
$$h(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{P} h(\boldsymbol{0}) = \int_0^1 \dot{g}(\boldsymbol{\mu} + t\boldsymbol{0})\,dt = \dot{g}(\boldsymbol{\mu})$$
がわかる．最後に，$h(\boldsymbol{y})$ の連続性を示そう．$\boldsymbol{\mu}$ の近傍での $\dot{g}$ の連続性から，ある正の数 M が存在して，すべての $|\boldsymbol{y}| \le \delta$ に対して $|\dot{g}(\boldsymbol{\mu} + t\boldsymbol{y})| \le M$ とできる．有界収束定理から $h(\boldsymbol{y})$ の連続性は示せた．

·····5.1·····

標本分位点の漸近分布

$X_1, X_2, \ldots, X_n$ を $\mathbb{R}$ 上の分布関数 F からのランダム標本とする．ただし，F は連続とする．$X_{(n:1)} < X_{(n:2)} < \cdots < X_{(n:n)}$ を $X_1, X_2, \ldots, X_n$ の順序統計量とする．簡単に，$X_{(k)} = X_{(n:k)}, k = 1, 2, \ldots, n$ と記そう．正の数 $p\,(0 < p < 1)$ に対して

$$x_p = F^{-1}(p) \equiv \inf\{t : F(t) \geq p\}$$

を p の分位点とし，$k = \lceil np \rceil$（np 以上の最小の整数）とし，$X_{(k)}$ を標本 p 分位点とする．もし，F は確率密度関数 $f(\cdot)$ を持ち，$f(\cdot)$ は連続で分位点の近傍で正ならば，対応する（複数の）の標本分位点の同時分布は漸近的に正規分布に従う．いま，$U_{(k)} = F(X_{(k)}), k = 1, 2, \ldots, n$ とおけば，$U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ は一様分布 $U(0, 1)$ からの順序統計量と同じ分布に従う．一様分布からのランダム標本に基づく標本分位点の漸近分布を求め，変換 $g(u) = F^{-1}(u)$ を用いて，標本分位点の漸近分布を求める．

補題 5.2　　$Y_1, Y_2, \ldots, Y_{n+1}$ を指数分布 $f_Y(y) = e^{-y} I_{(0,\infty)}(y)$ からの大きさ $n+1$ のランダム標本とし，$S_k = \sum_{i=1}^{k} Y_i, k = 1, 2, \ldots, n+1$ とおく．このとき，S_{n+1} と

$$\left[\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots \frac{S_n}{S_{n+1}}\right] \tag{5.102}$$

は独立で，S_{n+1} を与えたときの (5.102) の条件付き分布は，一様分布からの n 個のランダム標本に基づく順序統計量と同一分布である．

証明：$Y_1, Y_2, \ldots, Y_{n+1}$ の同時確率密度関数は

$$f_{Y_1, Y_2, \ldots, Y_{n+1}}(y_1, y_2, \ldots, y_{n+1}) = \exp\left\{-\sum_{k=1}^{n+1} y_k\right\} I\{\Pi_{k=1}^{n+1} y_k > 0\}$$

となる．また，$(S_1, S_2, \ldots, S_{n+1})$ の同時確率密度関数は

$$f_{S_1, S_2, \ldots, S_{n+1}}(s_1, s_2, \ldots, s_{n+1}) = \exp\{-s_{n+1}\}\, I\{0 < s_1 < s_2 < \cdots < s_{n+1} < \infty\}$$

となる．$Z_k = S_k / S_{n+1}$ と $W = S_{n+1}$ とおけば，Jacobian $= W^n$ となるので，$(Z_1, Z_2, \ldots, Z_n, W)$ の同時確率密度関数は

$$f_{(Z_1, Z_2, \ldots, Z_n, W)}(z_1, z_2, \ldots, z_n, w) = w^n e^{-w} I\{0 < z_1 < z_2 < \ldots < z_n < 1, w > 0\}$$

となる．したがって，$Z_1, Z_2, \ldots, Z_n$, と W は独立で

$$f_{(Z_1, Z_2, \ldots, Z_n)}(z_1, z_2, \ldots, z_n) = n! I\{0 < z_1 < z_2 < \ldots < z_n < 1\}$$

となることがわかる．　□

この補題から $(U_{(k_1)}, U_{(k_2)})$, $1 \le k_1 < k_2 \le n$ の同時分布は

$$\left[\frac{S_{k_1}}{S_{n+1}}, \frac{S_{k_2}}{S_{n+1}}\right]$$

の分布と同じになる．さらに，

$$g(x_1, x_2, x_3) = \frac{1}{x_1 + x_2 + x_3}\left[\begin{array}{c} x_1 \\ x_1 + x_2 \end{array}\right] \tag{5.103}$$

とおく．このとき，

$$g\left(\frac{S_{k_1}}{n+1}, \frac{S_{k_2} - S_{k_1}}{n+1}, \frac{S_{n+1} - S_{k_2}}{n+1}\right) = \frac{1}{S_{n+1}}\left[\begin{array}{c} S_{k_1} \\ S_{k_2} \end{array}\right]$$

であるので，

$$\left(\frac{S_{k_1}}{n+1}, \frac{S_{k_2} - S_{k_1}}{n+1}, \frac{S_{n+1} - S_{k_2}}{n+1}\right)$$

の漸近分布に注目する．

補題 5.3　$Y_1, Y_2, \ldots, Y_{n+1}$ を指数分布 $f_Y(y) = e^{-y} I_{(0,\infty)}(y)$ からの大きさ $n+1$ のランダム標本とする．$k_1 < k_2$ は $n \to \infty$ のとき，$\sqrt{n}\,((k_1/n) - p_1) \to 0$,　$\sqrt{n}\,((k_2/n) - p_2) \to 0$ とする．このとき

$$\sqrt{n+1}\left[\frac{S_{k_1}}{n+1} - p_1, \frac{S_{k_2} - S_{k_1}}{n+1} - (p_2 - p_1), \frac{S_{n+1} - S_{k_2}}{n+1} - (1 - p_2)\right] \xrightarrow{L} N(0, \Sigma)$$

となる．ただし，

$$\Sigma = \left[\begin{array}{ccc} p_1 & 0 & 0 \\ 0 & p_2 - p_1 & 0 \\ 0 & 0 & 1 - p_1 \end{array}\right]$$

である．

証明：中心極限定理から

$$\sqrt{k_1}\left(\frac{S_{k_1}}{k_1} - 1\right) \xrightarrow{d} N(0, 1)$$

が成立する．$k_1/n \to p_1 (n \to \infty)$ から

$$\sqrt{n+1}\left[\frac{S_{k_1}}{n+1} - \frac{k_1}{n+1}\right] = \sqrt{\frac{k_1}{n+1}}\sqrt{k_1}\left[\frac{S_{k_1}}{k_1} - 1\right] \xrightarrow{L} \sqrt{p_1} N(0, 1) = N(0, p_1)$$

となる．$k_1/n \to p_1$ と $k_2/n \to p_2 (n \to \infty)$ に注意すれば，

$$\begin{aligned}\sqrt{n+1}\left[\frac{S_{k_2} - S_{k_1}}{n+1} - \frac{k_2 - k_1}{n+1}\right] &= \sqrt{\frac{k_2 - k_1}{n+1}}\sqrt{k_2 - k_1}\left[\frac{1}{k_2 - k_1}\sum_{i=k_1+1}^{k_2} Y_i - 1\right] \\ &\xrightarrow{L} N(0, p_2 - p_1)\end{aligned}$$

と

$$\sqrt{n+1}\left[\frac{S_{n+1}-S_{k_2}}{n+1}-\frac{n+1-k_2}{n+1}\right]\xrightarrow{L}N(0,\,1-p_2)$$

を得る．さらに，

$$\sqrt{n+1}\left[\frac{S_{k_1}}{n+1}-\frac{k_1}{n+1}\right]-\sqrt{n+1}\left[\frac{S_{k_1}}{n+1}-p_1\right]=o_P(1)$$

と $(S_{k_1},\,S_{k_2}-S_{k_1},\,S_{n+1}-S_{k_2})$ は独立であることに注意すれば，補題は証明される． □

定理 5.1　　$U_{(1)}<U_{(2)}<\cdots<U_{(n)}$ を一様分布 $f_U(u)=I_{(0,\,1)}(u)$ からの大きさ n のランダム標本に基づく順序統計量とし，$n\to\infty,\,k_1\to\infty,\,k_2\to\infty$ で

$$\sqrt{n}\left(\frac{k_1}{n}-p_1\right)\to 0,\qquad \sqrt{n}\left(\frac{k_2}{n}-p_2\right)\to 0$$

を満足するものとする．ただし，$k_1<k_2,\,0<p_1<p_2<1$ である．このとき，

$$\sqrt{n}\begin{bmatrix} U_{(k_1)}-p_1 \\ U_{(k_2)}-p_2 \end{bmatrix}\xrightarrow{L}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2)\end{bmatrix}\right)$$

が成立する．

証明：補題 5.3 を用いる．まず，(5.103) から

$$\dot{g}(x_1,\,x_2,\,x_3)=\frac{1}{(x_1+x_2+x_3)^2}\begin{bmatrix} x_2+x_3 & -x_1 & -x_1 \\ x_3 & x_3 & -(x_1+x_2)\end{bmatrix}$$

となることに注意すれば，

$$\dot{g}(p_1,\,p_2-p_1,\,p_2)=\begin{bmatrix} 1-p_1 & -p_1 & -p_1 \\ 1-p_2 & 1-p_2 & -p_2 \end{bmatrix}$$

となることがわかる．さらに，

$$\dot{g}(p_1,\,p_2-p_1,\,p_2)\begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2-p_1 & 0 \\ 0 & 0 & 1-p_2\end{bmatrix}\dot{g}'(p_1,\,p_2-p_1,\,p_2)=\begin{bmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2)\end{bmatrix}$$

に注意すればよい． □

系 5.1　　$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ を累積分布関数 $F(x)$ からの大きさ n のランダム標本とする．さらに，F は連続な確率密度関数 $f(x)$ を持ち，$f(x)$ は x_{p_1} と x_{p_2} の近傍で正とする．ただし，x_{p_1} と x_{p_2} は p_1 分位点と p_2 分位点 $(p_1<p_2)$ とする[1)]．このとき，

$$\sqrt{n}\begin{bmatrix} X_{([np_1])}-x_{p_1} \\ X_{([np_2])}-x_{p_2}\end{bmatrix}\xrightarrow{L}N(0,\,\Sigma)$$

---

1) $x_{p_i}=F^{-1}(x_i),\,i=1,\,2$ である．

が成立する．ただし，

$$\Sigma = \begin{bmatrix} \frac{p_1(1-p_1)}{f^2(x_{p_1})} & \frac{p_1(1-p_2)}{f(x_{p_1})f(x_{p_2})} \\ \frac{p_1(1-p_2)}{f(x_{p_1})f(x_{p_2})} & \frac{p_2(1-p_2)}{f^2(x_{p_2})} \end{bmatrix}$$

である．

証明：$g(y_1, y_2) = [F^{-1}(y_1), F^{-1}(y_2)]'$ とおく．

$$\dot{g}(y_1, y_2) = \begin{bmatrix} \frac{1}{f(F^{-1}(y_1))} & 0 \\ 0 & \frac{1}{f(F^{-1}(y_2))} \end{bmatrix}$$

と

$$\sqrt{n}\begin{bmatrix} U_{([np_1])} - p_1 \\ U_{([np_2])} - p_2 \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{bmatrix}\right)$$

に注意して，補題 5.1 を用いればよい．□

例 5.1　$X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの大きさ n のランダム標本とし，m_n を標本メデアンとする．$x_{1/2} = \mu$ と $f(x_{1/2}) = 1/(\sqrt{2\pi}\sigma)$ に注意すれば，

$$\sqrt{n}(m_n - \mu) \xrightarrow{L} N\left(0, \frac{1}{4f^2(\mu)}\right) = N\left(0, \frac{\pi\sigma^2}{2}\right)$$

を得る．

注意 5.1　例 5.1 において，別の推定量 $\bar{X}_n$ を考える．$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2)$ になり，$\bar{X}_n$ は漸近的有効推定量になっているので，$\sigma^2 < \pi\sigma^2/2$ になっていることに注意せよ．

例 5.2　$X_1, X_2, \ldots, X_n$ をコーシー分布

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1 + [(x-\mu)/\sigma]^2} I_{(-\infty, \infty)}$$

からの大きさ n のランダム標本とする．ただし，$-\infty < \mu < \infty$, $\sigma > 0$ である．メデアンは μ である．また，$x_{1/4} = \mu - \sigma$, $x_{3/4} = \mu + \sigma$ となるので，$\sigma = (x_{3/4} - x_{1/4})/2$ である．標本メデアンの漸近分布は

$$\sqrt{n}(X_{[n/2]} - \mu) \xrightarrow{L} N\left(0, \frac{\pi^2\sigma^2}{4}\right)$$

となる．また，σ の推定量

$$\hat{\sigma} = \frac{X_{[3n/4]} - X_{[n/4]}}{2}$$

の漸近分布を求めよう．系 5.1 から

$$\sqrt{n}\begin{bmatrix} X_{[n/4]} - (\mu - \sigma) \\ X_{[3n/4]} - (\mu + \sigma) \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{\pi^2\sigma^2}{4}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\right)$$

となることに注意する．さらに，補題 (5.1)　($p_1 = 1/4,\, p_2 = 3/4$) を用いれば[2)],

$$\sqrt{n}\left[\frac{X_{[3n/4]} - X_{[n/4]}}{2} - \sigma\right] \xrightarrow{L} N\left(0,\, \frac{\pi^2\sigma^2}{4}\right)$$

を得る．

[2)] $g(x_1,\, x_2) = (x_2 - x_1)/2$ とおく．すると，　よりわかる．
$\dot{g}(x_1,\, x_2) = [-1/2,\, 1/2]$ となり，

$$\dot{g}(x_1,\, x_2)\frac{\pi^2\sigma^2}{4}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\dot{g}'(x_1,\, x_2) = \frac{\pi^2\sigma^2}{4}$$

·····5.2·····

極値の漸近分布

$X_1, X_2, \ldots, X_n$ を連続な累積分布関数 $F(x)$ からの大きさ n のランダム標本とし，$M_n = \max_{1\le j\le n} X_j$ とおく．このとき，

$$\mathbb{P}\{M_n \le x\} = \{F(x)\}^n$$

となる．

ここで，M_n の漸近分布が存在するかを調べる．すなわち，ある数列 $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ とある累積分布関数 $G(\cdot)$ が存在して，G の連続点 x に対して

$$\mathbb{P}\left\{\frac{M_n - a_n}{b_n} \le x\right\} = \mathbb{P}\{M_n \le b_n x + a_n\} = \{F(b_n x + a_n)\}^n \to G(x)$$

となるかを調べよう．

定義 5.1　関数 $c : [0, \infty) \to \mathbb{R}$ が緩変動であるとは，すべての $x > 0$ に対して

$$\frac{c(tx)}{c(t)} \to 1. \qquad (t \to \infty)$$

を満足することである．

例 5.3　$c(x) = \log x$ は緩変動だが，$c(x) = \cos x^r (r \neq 0)$ はそうでない．これは，$(tx)^r/t^r \to x^r$ $(t \to \infty)$ からわかる．

定理 5.2　$F(x)$ を確率変数 X の累積分布関数とし，$x_0 = \sup\{x : F(x) < 1\}$ とする[1)]．

(a)　$x_0 = \infty$ とし，ある定数 $r > 0$ と緩変動関数 $c(x)$ が存在し，$1 - F(x) = x^{-r}c(x)$ と書けるとする．このとき，

$$\{F(b_n x)\}^n \to G_{1,r}(x) = \begin{cases} \exp\{-x^{-r}\}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

となる．ただし，b_n は $1 - F(b_n) = 1/n$ を満足する．

(b)　$x_0 < \infty$ とし，ある定数 $r > 0$ と緩変動関数 $c(x)$ が存在して

$$1 - F(x) = (x_0 - x)^r c(1/(x_0 - x))$$

[1)] x_0 は ∞ かもしれない．

をみたすとする．このとき，

$$\{F(x_0+b_nx)\}^n \to G_{2,r}(x) = \begin{cases} \exp\{-(-x)^r\}, & x<0 \\ 1, & x \ge 0 \end{cases}$$

となる．ただし，$1-F(x_0-b_n)=1/n$ で F は x_0 で jump しない．

(c)　ある関数 $R(t)$ が存在して，すべての x に対して

$$\frac{1-F(t+xR(t))}{1-F(t)} = e^{-x}, \qquad t \to x_0$$

とする．ただし，$x_0=\infty$ でも $x_0<\infty$ でもよい．このとき，

$$\{F(a_n+b_nx)\}^n \to G_3(x) = e^{-x}$$

となる．ただし，$1-F(a_n)=1/n$ と $b_n=R(a_n)$ である．

証明:(a) まず, b_n の定義より，$b_n \to \infty (n \to \infty)$ に注意すると $\{F(b_nx)\}^n = \{1-(b_nx)^{-r}c(b_nx)\}^n$ となることがわかる．ここで，$c(\,c\,)$ が緩変動関数であることを使えば，

$$\begin{aligned} \lim_{n\to\infty} n(b_nx)^{-r}c(b_nx) &= x^{-r}\lim_{n\to\infty} nb_n^{-r}c(b_n)\frac{c(b_nx)}{c(b_n)} = x^{-r}\lim_{n\to\infty} nb_n^{-r}c(b_n) \\ &= x^{-r}\lim_{n\to\infty} n[1-F(b_n)] = x^{-r} \end{aligned}$$

となることに注意する．

$$\lim_{n\to\infty}\{1-(b_nx)^{-r}c(b_nx)\}^n = \exp\{-\lim_{n\to\infty} n(b_nx)^{-r}c(b_nx)\}$$

となることがわかる[2)].

(b) $n\to\infty$ のとき，$b_n \to 0$ に注意する．$x<0$ に対し

$$\{F(x_0+b_nx)\}^n = \{1-(-b_nx)^r c\left(\frac{1}{-b_nx}\right)\}^n$$

となる．さらに，

$$\lim_{n\to\infty} n(-b_nx)^r c\left(\frac{1}{-b_nx}\right) = (-x)^r \lim_{n\to\infty} nb_n^r c(1/b_n)\frac{c(1/(-b_nx))}{c(1/b_n)} = (-x)^r\lim_{n\to\infty} nb_n^r c(1/b_n) = (-x)^r$$

となる[3)]．このことより

$$\{F(x_0+b_nx)\}^n = \exp\{-(-x)^r\}$$

を得る．

(c) いま，

$$\lim_{n\to\infty} n(1-F(a_n+b_nx)) = \lim_{n\to\infty} n(1-F(a_n+R(a_n)x)) = e^{-x}\lim_{n\to\infty} n(1-F(a_n)) = e^{-x}$$

[2)] $\lim_{n\to\infty}(1-d_n)^n = \lim_{n\to\infty}\left(1-\frac{nd_n}{n}\right)^n = \exp\{-\lim_{n\to\infty} nd_n\}$ に注意すればよい．

[3)] 最後の等号は $b_n^r c(1/b_n) = 1-F(x_0-b_n) = 1/n$ からわかる．

より

$$\begin{aligned}\{F(a_n+b_nx)\}^n &= [1-\{1-F(a_n+b_nx)\}]^n \\ &\to \exp[-\lim_{n\to\infty} n(1-F(a_n+b_nx))] = \exp(-\exp(-x))\end{aligned}$$

を得る．□

例 5.4　T_ν は自由度 ν の t 分布

$$f(x) = \frac{k}{(\nu+x^2)^{(\nu+1)/2}}$$

に従うものとする．ただし

$$k = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi}\Gamma(\nu/2)}$$

である．このとき，

$$\lim_{x\to\infty} \frac{f(x)}{kx^{-(\nu+1)}} = 1$$

となる．

$$\begin{aligned}\lim_{x\to\infty}\left|x^\nu(1-F(x)) - x^\nu\int_x^\infty \frac{k}{t^{\nu+1}}\,dt\right| &= \lim_{x\to\infty} kx^\nu\left|\int_x^\infty \left\{\frac{1}{(1+\nu/t^2)^{(\nu+1)/2}}-1\right\}\frac{1}{t^{\nu+1}}\,dt\right| \\ &\le \lim_{x\to\infty} kx^\nu\left|\frac{1}{(1+\nu/x^2)^{(\nu+1)/2}}-1\right|\int_x^\infty \frac{dt}{t^{\nu+1}} \\ &= \lim_{x\to\infty} kx^\nu\left|\frac{1}{(1+\nu/x^2)^{(\nu+1)/2}}-1\right|\frac{1}{\nu x^\nu} = 0\end{aligned}$$

から，ある $c(x)$ が存在して

$$1-F(x) = x^{-\nu}c(x) \qquad と \qquad c(x) \to \frac{k}{\nu}, \qquad x\to\infty$$

を満足する．$r=\nu$ として，

$$1-F(b_n) \sim \frac{c(b_n)}{b_n^\nu} \sim \frac{k}{\nu b_n^\nu} = \frac{1}{n}$$

より

$$b_n = \left(\frac{kn}{\nu}\right)^{1/\nu}$$

とすれば，(a) の場合になる．

特に，$\nu=1$ のときは $d=1/\pi$ となり，$b_n=n/\pi$ となるので

$$\frac{\pi}{n}M_n \xrightarrow{L} G_{1,1}$$

を得る．

例 5.5　　X はベーター分布

$$f(x) = dx^{\alpha-1}(1-x)^{\beta-1}I_{(0,\,1)}(x)$$

に従うとする．ただし，$\alpha > 0,\, \beta > 0,\, d = \Gamma(\alpha+\beta)/(\Gamma(\alpha)\Gamma(\beta))$ である．したがって，$x_0 = 1$ となり，$n \nearrow 1$ のとき，

$$f(x) \sim d(1-u)^{\beta-1}$$

と

$$1 - F(x) \sim d\int_x^1 (1-u)^{\beta-1}\,du = \frac{d}{\beta}(1-x)^\beta \tag{5.104}$$

である[4)]．よって，(b) の場合で $r = \beta$ と $x_0 = 1$ である[5)]．方程式 $1 - F(1-b_n) = 1/n$ と (5.104) より $1 - F(1-b_n) \sim \frac{d}{\beta}b_n^\beta$ となることに注意すれば，

$$b_n^\beta \sim \frac{\beta}{nd}$$

となる[6)]．したがって，

$$b_n = \left\{\frac{\Gamma(\alpha)\Gamma(\beta+1)}{n\Gamma(\alpha+\beta)}\right\}^{1/\beta}$$

となる．

いま，$\alpha = 1,\, \beta = 1$ とすれば X は一様分布 $U(0,\,1)$ となるので，このときの M_n の漸近分布は

$$n(M_n - 1) \xrightarrow{L} G_{2,\,1} = \mathrm{Gamma}(1,\,1)$$

となる．

例 5.6　(指数分布)　 X が指数分布に従うとき，

$$\mathbb{P}(X > t + x | X > t) = e^{-x}$$

となる．

$$\mathbb{P}(X > t + xR(t) | X > t) = \frac{1 - F(t + xR(t))}{1 - F(t)} \to e^{-x}$$

となるので，$R(t) = 1$ と $b_n = 1$ となる．また，$1 - F(x) = e^{-x}$ より，a_n を求めれば，

$$\exp\{-a_n\} = \frac{1}{n} \Longrightarrow a_n = \log n$$

---

4)

$$\lim_{x\nearrow 1}\left|(1-F(x)) - d\int_x^1 (1-u)^{\beta-1}\,du\right| = d\lim_{x\nearrow 1}\left|\int_x^1 \{u^{\alpha-1}-1\}(1-u)^{\beta-1}\,du\right|$$
$$\le d\lim_{x\nearrow 1}\left[|x^{\alpha-1}-1|\int_x^1 (1-u)^{\beta-1}\,du\right] \to 0$$

5) $c(1/(1-x)) = (1-F(x))/(1-x)^\beta$ とおけば，

$$c(x) = \frac{1-F(x)}{(1-x)^\beta}$$

となるので，(5.104) を用いてれば

$$\lim_{x\to\infty}\frac{c(tx)}{c(x)} = \frac{1-F(1/(1-tx))}{(-tx)^\beta}\Bigg/\frac{1-F(1/(1-x))}{(-x)^\beta} \sim \frac{1}{(-t)^\beta}\frac{\{1-}{\{1}$$

となり，緩変動関数となる．

6) $\beta\Gamma(\beta) = \Gamma(\beta+1)$ を用いた．

となる．よって

$$M_n - \log n \xrightarrow{L} G_3$$

を得る．

·····5.3·····

分布関数の推定

$X_1, X_2, \ldots, X_n$ を $\boldsymbol{R}^d$ 上の分布関数 F からのランダム標本[1]とする．確率変数 $X_1, \ldots, X_n$ に基づく F の**経験分布関数**を

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq t\}$$

で定義する．ただし，$X_i \leq t$ は成分ごとの比較である．t を固定するごとに，$F(t)$ の推定量 $F_n(t)$ は以下のような性質をもつ：

(i) どんな $t \in \boldsymbol{R}^d$ に対しても，$nF_n(t)$ は二項分布 $Bi(F(t), n)$ に従う．したがって，$F_n(t)$ は $F(t)$ の不偏推定量で，その分散は $F(t)(1-F(t))/n$ で与えられる．

(ii) $F_n(t)$ は $F(t)$ の $\sqrt{n}$ – 一致推定量[2]である．

(iii) どんな固定した正整 m と固定した異なる m 個の点 $t_1, \ldots, t_n \in \boldsymbol{R}^d$ に対しても，$n \to \infty$ のとき，

$$\sqrt{n}\{(F_n(t_1), \ldots, F_n(t_m)) - (F(t_1), \ldots, F_n(t_n))\} \xrightarrow{L} N_m(0, \Sigma)$$

となる．ただし，Σ の (i,j) – 成分は

$$F(t_i \wedge t_j) - F(t_i)F(t_j) \tag{5.105}$$

で与えられる．$t_i \wedge t_j$ は成分ごとに比較して小さい方の成分をその成分としてもつベクトルである．

5.3.1 分布関数上の距離

$\mathcal{F}$ を $\boldsymbol{R}^d$ 上の分布関数の集まりとする．F_n の大域的な性質をしらべるために，$\mathcal{F}$ 上の距離を導入する．

定義 5.2　：$\mathcal{F}_0$ を $\mathcal{F}$ の部分集合とする．関数

$$\rho : \mathcal{F}_0 \times \mathcal{F}_0 \longrightarrow [0, \infty)$$

がつぎを満たすとき，ρ を $\mathcal{F}_0$ の**距離**という．任意の $G_1, G_2, G_3 \in \mathcal{F}_0$ に対して，

(i) $\rho(G_1, G_2) = 0 \Leftrightarrow G_1 = G_2$

(ii) $\rho(G_1, G_2) = \rho(G_2, G_1)$

(iii) $\rho(G_1, G_2) \leq \rho(G_1, G_3) + \rho(G_3, G_2)$

---

[1] すなわち，$X_1, \ldots, X_n \sim i.i.d. F$ である．

[2] すなわち，どんな正の数 $\epsilon > 0$ に対しても

$$P(|F_n(t) - F(T)| > \epsilon) \to 0, \qquad (n \to \infty)$$

かつ $\sqrt{n}(F_n(t) - F(t)) = O_P(1)$ を満足する．

最も頻繁に用いられる距離は $\mathcal{F}$ 上の sup – norm である：

$$\rho_\infty(G_1, G_2) = \sup_{t \in R^d} |G_1(t) - G_2(t)|$$

$p \geq 1$ に対して，

$$\mathcal{F}_p = \{G \in \mathcal{F} : \int \|t\|^p dG < \infty\}$$

とする．$G_1, G_2 \in \mathcal{F}_p$ に対して，Mallow の距離を

$$\rho_{M_p}(G_1, G_2) = \inf\{\boldsymbol{E}\{\|Y_1 - Y_2\|^p\}^{1/p}\}$$

で定義する．ただし，infimum は周辺分布 G_1 と G_2 をもつ確率変数 Y_1 と Y_2 の同時分布に関してである．

定理 5.3　：$\{G_j; j = 0, 1, 2, \ldots\} \subset \mathcal{F}_p$ とする．このとき，つぎは同値である．
(i) $\rho_{M_p}(G_j, G_0) \to 0 \qquad (n \to \infty)$
(ii) $j \uparrow \infty$ のとき，$\int \|t\|^p dG_j \to \int \|t\|^p dG_0$ とすべての $t \in \boldsymbol{R}^d$ に対して，$G_j(t) \to G_0(t)$ が成立する．

証明：(i) $\Rightarrow$ (ii) をまず示す．確率変数 $Y_j (j = 1, 2, \ldots)$ と Y_0 は分布関数 G_j と G_0 をもつとする．周辺分布が G_j と G_0 である同時分布の列に関する期待値の列で

$$\lim_{n \to \infty} \boldsymbol{E}_n \|Y_j - Y_0\|^2 = d_{M_p}(G_j, G_0)$$

となるようなのものが存在する．すべての同時分布に対する周辺分布は G_j と G_0 になることに注意すれば，

$$\begin{aligned}\left[\int \|t\|^p dG_j\right]^{1/p} - \left[\int \|t\|^p dG_0\right]^{1/p} &= \{\boldsymbol{E}\|Y_j\|\}^{1/p} - \{\boldsymbol{E}\|Y_0\|\}^{1/p} \\ &\leq \lim_{n \to \infty} \{\boldsymbol{E}_n \|Y_j - Y_0\|\}^{1/p} = d_{M_p}(G_j, G_0) \\ &\to 0, \qquad (j \to \infty)\end{aligned}$$

となる．f を Lipschitz 連続関数[3)]とする．このとき，

$$\begin{aligned}\left| \int f(y) dG_j - \int f(y) dG_0 \right| &= |\boldsymbol{E} f(Y_j) - \boldsymbol{E} f(Y_0)| \\ &\leq \lim_{n \to \infty} \boldsymbol{E}_n |f(Y_j) - f(Y_0)| \leq K \lim_{n \to \infty} \boldsymbol{E}_n \|Y_j - Y_0\| \\ &\leq K \lim_{n \to \infty} \{\boldsymbol{E}_n \|Y_j - Y_0\|^p\}^{1/p} = d_{M_p}(G_j, G_0) \to 0, \qquad (j \to \infty)\end{aligned}$$

となる．よって，ただし，最後からふたつ目の不等式は Jensen の不等式を用いた．Portmanteau の定理から $G_j(t) \to G_0(t)$ となることがわかる．

つぎに，(ii) $\Rightarrow$ (i) を示す．まず，G_j と G_0 があるコンパクト集合 $C (\subset \boldsymbol{R}^d)$ に集中している場合を考える．半径 $\epsilon/2$ の C の開被覆から有限個の排反な開集合 $C_1, \ldots, C_l$ を選び出し C を

3) すなわち，ある定数 K が存在して，どんな x と y に対しても $|f(x) - f(y)| \leq K\|x - y\|$ とできる．

覆うことができる．また，$G(\partial C_i) = 0(i = 1, 2, \ldots, l)$ となるようにできる．ただし，∂C_i は C_i の境界である．$x_i \in C_i$ を選び固定し，$\tilde{G}_j$ を

$$\tilde{G}_j\{x_i\} = \int 1_{C_i}(y) dG_j(y), \qquad (j = 1, 2, \ldots)$$

で定義する．明らかに，

$$d_{M_p}(\tilde{G}_j, G_j) \leq \{\boldsymbol{E}\|\tilde{Y}_j - Y_j\|^p\}^{1/p} = \epsilon$$

である．ただし，$\tilde{Y}_j$ は分布 $\tilde{G}_j$ をもつ確率分布であり，上の式の期待値は周辺分布は $\tilde{G}_j$ と $\tilde{G}_0$ となる任意の同時分布に関してである．このとき，

$$\begin{aligned}
\sup_j \boldsymbol{E}\|\tilde{Y}_j\|^p &= \sup_j \sum_{i=1}^{l} \|x_i\|^p \int 1_{C_i}(y) dG_j \\
&\leq \sum_{i=1}^{l} \|x_i\|^p \sup_j \int 1_{\overline{C_i}}(y) dG_j \\
&\leq \sum_{i=1}^{l} \|x_i\|^p \sup_j \int 1_{\overline{C_i}}(y) dG_0 \qquad \text{（Portmantau の定理より）} \\
&= \boldsymbol{E}\|\tilde{Y}_0\|^p < \infty
\end{aligned}$$

から，$\{\tilde{Y}_j\}_{j=1}^{\infty}$ は一様可積分となる．よって，$\tilde{G}_j(t) \to \tilde{G}_0(t)$ から

$$\boldsymbol{E}\|\tilde{Y}_j - \tilde{Y}_0\|^p \to 0$$

がわかる．したがって，$j \uparrow \infty$ のとき，

$$d_{M_p}(\tilde{G}_j, \tilde{G}_0) \leq \{\boldsymbol{E}\|\tilde{Y}_j - \tilde{Y}_0\|^p\}^{1/p} \to 0$$

となる．よって，$j \uparrow \infty$ のとき，

$$d_{M_p}(G_j, G_0) \leq d_{M_p}(G_j, \tilde{G}_j) + d_{M_p}(\tilde{G}_j, \tilde{G}_0) + d_{M_p}(\tilde{G}_0, G_0) \leq 2\epsilon$$

となる．あと，一般の場合が測度がコンパクト集合 C に集中している場合に帰着できることを示せばよい． □

5.3.2 経験分布関数の大域的な性質

F_n と F の sup − norm に関する評価式として，以下の補題が有効である．

補題 5.4　（DKW 不等式）：
(i) $d = 1$ のとき,（F には依存しない）ある正の定数 C が存在して，$z > 0$ と $n = 1, 2, \ldots$ に対して，

$$P(\rho_\infty(F_n, F) > z) \leq Ce^{-2nz^2}$$

が成り立つ．
(ii) $d \geq 2$ のとき，どんな正の数 $\epsilon > 0$ に対しても，（F には依存しない）ある正の定数 $C_{\epsilon,d}$ が存在して，$z > 0$ と $n = 1, 2, \ldots$ に対して，

$$P(\rho_\infty(F_n, F) > z) \leq C_{\epsilon,d} e^{-(2-\epsilon)nz^2}$$

が成り立つ．

証明：略．　□

定理 5.4　：$X_1, \ldots, X_n$ を $\boldsymbol{R}^k$ 上の分布関数 F からのランダム標本とし，F_n を $X_1, \ldots, X_n$ に基づく経験分布とする．このとき，
(i) $\rho_\infty(F_n, F) \xrightarrow{as} 0 \qquad (n \to \infty)$
(ii) どんな $s > 0$ に対しても，

$$\boldsymbol{E}[\sqrt{n}\rho_\infty(F_n, F)]^s = O(1).$$

証明：(i) の証明．DKW 不等式から

$$\sum_{n=1}^{\infty} P(\rho_\infty(F_n, F) > z) \leq C_{\epsilon,d} \sum_{n=1}^{\infty} \exp(-(2-\epsilon)nz^2) = C_{\epsilon,d} \lim_{n\to\infty} \frac{1 - \exp(-(2-\epsilon)nz^2)}{1 - \exp(-(2-\epsilon)z^2)} < \infty$$

を得る．したがって，Borel – Cantelli の補題から (i) は示される．
(ii) の証明．$z = y^{1/s}/\sqrt{n}$ として，DKW 不等式を用いれば，

$$\begin{aligned} \boldsymbol{E}[\sqrt{n}\rho_\infty(F_n, F)]^s &= \int_0^\infty P(\sqrt{n}\rho_\infty(F_n, F) > y^{1/s}) dy \\ &\leq C_{\epsilon,d} \int_0^\infty e^{-(2-\epsilon)y^{1/s}} dy = O(1) \end{aligned}$$

が $2 - \epsilon > 0$ ならば，成り立つ．　□

注意 5.2　：定理 3.1(i) より，$t \in \boldsymbol{R}^d$ に関して，一様に $F_n(t) \xrightarrow{as} F(t)$ が成立することがわかり，F_n は F の強一致性より強い性質をもつことがわかる．また，$\sqrt{n}\rho_\infty(F_n, F) = O_P(1)$ は F_n の $\sqrt{n}$ – 一致性より強い結果である．

·····5.4·····

統計的汎関数

分布関数 F そのものよりも F を記述する特徴に興味あることが多い．ここでは，ある汎関数 $T:\mathcal{F}\longrightarrow \boldsymbol{R}^s$ を用いて表現される $T(F)$ がある場合を考える．ランダム標本に基づく経験分布関数 F_n で F を推定したとき，$T(F)$ の自然な推定量として，$T(F_n)$ を考える．これを**統計的汎関数**（statistical functional）とよぶ．

多くの統計量はある汎関数 T を用いて，$T(F_n)$ と表現できる．

例 5.7　：ある関数 ψ を用いて，$T(F)=\int \psi(x)dF(x)$ と表現できたとする．このとき，$T(F_n)=(1/n)\sum_{i=1}^n \psi(X_i)$ となる．

例 5.8　：F を $\boldsymbol{R}$ 上の分布関数とし，$p\in(0,1)$ に対して，

$$T(F)=F^{-1}(p)=\inf\{x:F(x)\geq p\}$$

とおけば，$T(F_n)=F_n^{-1}(p)$ は pth sample quantiles になる．

5.4.1 可微分性の定義

T の可微分性についてのいくつかの定義をあげておく．

定義 5.3　：$\mathcal{F}_0$ を $\mathcal{F}=\{F:F$ は $\boldsymbol{R}^d$ 上の分布関数$\}$ の部分集合とし，$\mathcal{D}=\{c(G_1-G_2):c\in\boldsymbol{R},G_1,G_2\in\mathcal{F}\}$ とする．　また，ρ を $\mathcal{F}_0$ 上の距離とする．$\mathcal{D}$ 上の距離 $\|\cdot\|_D$ を $c\in\boldsymbol{R}$ と $G_1,G_2\in\mathcal{F}_0$ に対して，$\|c(G_1-G_1)\|_D=|c|\rho(G_1,G_2)$ で定義[1]する．

(i) $\mathcal{F}_0$ 上の汎関数 T が $G\in\mathcal{F}_0$ において **Gâteaux 微分可能**とは，$\mathcal{D}$ 上の線形汎関数[2] $\dot{T}_G(\cdot)$ が存在し，$\Delta\in\mathcal{D}$ と $G+t\Delta\in\mathcal{F}_0$ に対して，

$$\lim_{t\to 0}\left[\frac{T(G+t\Delta)-T(G)}{t}-\dot{T}_G(\Delta)\right]=0$$

を満足することである．

---

[1] したがって，

- (i)$\|\Delta\|_D\geq 0$ *and* $\|\Delta\|_D=0$ *iff* $\Delta=0$
- $\|c\Delta\|_D=|c|\|\Delta\|_D$
- $\|\Delta_1+\Delta_2\|_D\leq\|\Delta_1\|_D+\|\Delta_2\|_2$

が成立する．

[2] すなわち，$\dot{T}_G(c_1\Delta_1+c_2\Delta_2)=c_1\dot{T}_G(\Delta_1)+c_2\dot{T}_G(\Delta_2)$ が成立する．

(ii) $\mathcal{F}_0$ 上の汎関数 T が $G \in \mathcal{F}_0$ において $\rho-$ **Hadamard 微分可能**とは，$\mathcal{D}$ 上の線形汎関数 $\dot{T}_G(\cdot)$ が存在し，どんな実数列 $\{t_j\}_{j=1}^{\infty}$ と $\mathcal{D}$ の列 $\{\Delta, \Delta_j : j = 1, 2, \ldots\}$ で $j \uparrow \infty$ のとき，$\|\Delta_j - \Delta\|_D \to 0$ となり，$G + t_j\Delta_j \in \mathcal{F}_0$ なるものに対して，

$$\lim_{j\to\infty}\left[\frac{T(G+t_j\Delta_j)-T(G)}{t_j} - \dot{T}_G(\Delta)\right] = 0$$

を満足することである．
(iii) $\mathcal{F}_0$ 上の汎関数 T が $G \in \mathcal{F}_0$ において $\rho-$ **Fréchet 微分可能**とは，$\mathcal{D}$ 上の線形汎関数 $\dot{T}_G(\cdot)$ が存在し，どんな列 $\{G_j : G_j \in \mathcal{F}_0\}_{j=1}^{\infty}$ で $j \uparrow \infty$ のとき $\rho(G_j, G) \to 0$ なるものに対して，

$$\lim_{j\to\infty}\left[\frac{T(G_j)-T(G)-\dot{T}_G(G_j-G)}{\rho(G_j,G)}\right] = 0$$

を満足することである．

$\dot{T}_G$ を G における T の**微分**という．$h(t) = T(G + t\Delta)$ とおけば，Gâteaux の意味での可微分性は $t = 0$ における h の可微分性と同値となる．特に，δ_x を x に退化した分布関数を δ_x と書く．G における T の**影響関数**を

$$\phi_G(x) = \dot{T}_G(\delta_x - G) \tag{5.106}$$

で定義する．

5.4.2 可微分性と漸近正規性

$X_1, \ldots, X_n$ を分布関数 F からのランダム標本とし，F_n をそれらに基づく経験分布関数とする．

T が $F \in \mathcal{F}_0$ において，Gâteaux 微分可能のとき，$t = n^{-1/2}$ と $\Delta = \sqrt{n}(F_n - F)$ とおけば，つぎの展開を得る：

$$\sqrt{n}(T(F_n) - T(F)) = \dot{T}_F(\sqrt{n}(F_n - F)) + R_n$$

ここで，(5.106) にたいして，

$$\boldsymbol{E}[\phi_F(X_1)] = 0 \qquad \boldsymbol{var}[\phi_F(X_1)] = \sigma_F^2 < \infty$$

と仮定すれば，$\dot{T}_F$ の線形性と中心極限定理より

$$\dot{T}_F(\sqrt{n}(F_n - F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_F(X_i) \xrightarrow{L} N(0, \sigma_F^2) \tag{5.107}$$

を得る．さらに，

$$R_n = o_P(1) \tag{5.108}$$

が成立すれば，Slutsky の定理より

$$\sqrt{n}(T(F_n) - T(F)) \xrightarrow{L} N(0, \sigma_F^2) \tag{5.109}$$

を得る．

しかし，(5.108) を保障するためには，Gâteaux 可微分性では弱すぎる．そこで，他の可微分性の概念が必要となる．

定理 5.5　：
(i) T は F において ρ_∞ – Hadamard 微分可能ならば，(5.108) が成立する．
(ii) T は F において ρ – Fréchet 微分可能で，ρ は $\sqrt{n}\rho(F_n, F) = O_P(1)$ を満足するならば，(5.108) が成立する．

証明：(i) については，Prakasa Rao (1987) のページ 394 を参照のこと．

(ii) の証明．Fréchet 微分の定義から，どんな正の数 $\epsilon > 0$ に対してもある正の数 $\delta > 0$ が存在して，$\rho(F_n, F) < \delta$ ならば，$|R_n| < \epsilon\sqrt{n}\rho(F_n, F)$ とできる[3)]．このとき，どんな正の数 $\eta > 0$ に対しても

$$
\begin{aligned}
P(|R_n| > \eta) &= P(|R_n| > \eta, \rho(F_n, F) < \delta) + P(|R_n| > \eta, \rho(F_n, F) \geq \delta) \\
&\leq P(|R_n| > \eta, |R_n| < \epsilon\sqrt{n}\rho(F_n, F)) + P(\rho(F_n, F) \geq \delta) \\
&\leq P(\sqrt{n}\rho(F_n, F) > \frac{\eta}{\epsilon}) + P(\rho(F_n, F) \geq \delta)
\end{aligned}
$$

となる．これと条件より，

$$
\limsup_{n\to\infty} P(|R_n| > \eta) \leq \limsup_{n\to\infty} P(\sqrt{n}\rho(F_n, F) > \frac{\eta}{\epsilon}) \longrightarrow 0, \qquad (\epsilon \to 0)
$$

となる．よって，(ii) は示された． □

3) $R_n = \sqrt{n}\{T(F_n) - T(F) - \dot{T}_F(F_n - F)\}$ とした．

·····5.5·····

射影法

$X_1, \ldots, X_n$ を確率変数とし，T_n をこれに基づく統計量とする．k_n 個の確率要素（random element） $Y_1, \ldots, Y_{k_n}$ 上への T_n の射影を

$$\check{T}_n = \boldsymbol{E}(T_n) + \sum_{i=1}^{k_n}[\boldsymbol{E}(T_n|Y_i) - \boldsymbol{E}(T_n)]$$

で定義する．

いま，

$$\psi_n(X_i) = \boldsymbol{E}(T_n|Y_i)$$

とおく．T_n が対称[1] ならば，$\psi_n(Y_1), \ldots, \psi_n(Y_n)$ は独立同一の分布に従う確率変数で，その平均が $\boldsymbol{E}(\psi_n(Y_1)) = \boldsymbol{E}[\boldsymbol{E}(T_n|Y_1)] = \boldsymbol{E}(T_n)$ で与えられることがわかる．したがって，$\boldsymbol{E}(T_n^2) < \infty$ かつ $\boldsymbol{var}(\psi_n(Y_1)) > 0$ ならば，中心極限定理から

$$\frac{1}{\sqrt{k_n \boldsymbol{var}(\psi_n(Y_1))}}\sum_{i=1}^{n}[\psi_n(X_i) - \boldsymbol{E}(T_n)] \xrightarrow{L} N(0,1)$$

がわかる．さらに，

$$T_n - \check{T}_n = T_n - \boldsymbol{E}(T_n) - \sum_{i=1}^{n}[\psi_n(Y_i) - \boldsymbol{E}(T_n)] = o_P(1) \tag{5.110}$$

が成立すれば，Slutsky の定理から T_n の漸近分布がわかる．

つぎに，補題は (5.110) を示すときに有効である．以下では，$k_n = n$ とする．

補題 5.5　T_n を対称とし，すべての n に対して，$\boldsymbol{var}(T_n) < \infty$ とし，$\check{T}_n$ を $Y_1, \ldots, Y_n$ 上への T_n の射影とする．このとき，$\boldsymbol{E}(T_n) = \boldsymbol{E}(\check{T}_n)$ と

$$\boldsymbol{E}(T_n - \check{T}_n)^2 = \boldsymbol{var}(T_n) - \boldsymbol{var}(\check{T}_n)$$

が成り立つ．

証明：$\boldsymbol{E}(T_n) = \boldsymbol{E}(\check{T}_n)$ より

$$\boldsymbol{E}(T_n - \check{T}_n)^2 = \boldsymbol{var}(T_n) + \boldsymbol{var}(\check{T}_n) - 2\boldsymbol{cov}(T_n, \check{T}_n) \tag{5.111}$$

[1] すなわち，$Y_1, \ldots, Y_n$ の入れ替えに関して不変である．

となる．$\check{T}_n$ の定義から

$$\boldsymbol{var}(\check{T}_n) = n\boldsymbol{var}(\boldsymbol{E}(T_n|Y_1)) \tag{5.112}$$

となる．また，$\check{T}_n = \sum_{i=1}^{n} \boldsymbol{E}(T_n|Y_i) - (n-1)\boldsymbol{E}(T_n)$ より

$$\begin{aligned}
\boldsymbol{cov}(T_n, \check{T}_n) &= \boldsymbol{E}(T_n\check{T}_n) - \{\boldsymbol{E}(T_n)\}^2 \\
&= n\boldsymbol{E}[T_n\boldsymbol{E}(T_n|Y_1)] - n\{\boldsymbol{E}(T_n)\}^2 \\
&= n\boldsymbol{E}[\{\boldsymbol{E}(T_n|Y_1)\}^2] - n\{\boldsymbol{E}[\{\boldsymbol{E}(T_n|Y_i)\}^2] \\
&= n\boldsymbol{var}(\boldsymbol{E}(T_n|Y_1))
\end{aligned}$$

と (5.112) を (5.111) に代入すれば，補題は示される．□

·····5.6·····

U – 統計量とその漸近分布

$X_1, \ldots, X_n$ をモデル $\boldsymbol{P}$ に含まれる確率測度 P からのランダム標本とする．$\boldsymbol{P}$ 上で定義される母数 $\theta(\cdot)$ の推定を考える．多くの推定問題において，$P \in \boldsymbol{P}$ に対して，

$$\theta(P) = \boldsymbol{E}_P[h(X_1, \ldots, X_m)]$$

の形で表現できる．ただし，m は正の整数で，h は対称な関数[1)]で，$\boldsymbol{E}|h(X_1, \ldots, X_m)| < \infty$ とする．このとき，推定量

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i_1}, \ldots, X_{i_m})$$

は θ の不偏推定量となる．ただし，$\sum_c$ は $\{1, 2, \ldots, n\}$ からの m 個の異なる要素 $\{i_1, \ldots, i_m\}$ を取り出すすべての組合せに関する和である．統計量 U_n をオーダー m の核 h を持った **U – 統計量** という．

例 5.9　：$\theta = \sigma^2 = \boldsymbol{var}(X_1)$ とする．

$$\sigma^2 = \frac{1}{2}\{\boldsymbol{var}(X_1) + \boldsymbol{var}(X_2)\} = \boldsymbol{E}\left[\frac{(X_1 - X_2)^2}{2}\right]$$

から核を $h(x_1, x_2) = (x_1 - x_2)^2/2$ とおけば，

$$U_n = \frac{2}{n(n-1)} \sum_{i \le i < j \le n} \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

となる[2)]．

例 5.10　：$\theta = \boldsymbol{E}|X_1 - X_2|$ の推定を考える．核を $h(x_1, x_2) = |x_1 - x_2|$ ととれば，θ の U – 統計推定量として，

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} |X_i - X_j|$$

を得る． Gini 平均差である．

[1)] すなわち，$x_1, \ldots, x_m$ の入れ替えにたいして，$h(x_1, \ldots, x_m)$ は不変であること．

[2)]

$$\sum_{i<j}(X_i - X_j)^2 = \sum_{i<j}(X_i^2 + X_j^2) - 2\sum_{i<j} X_i X_j = (n-1)\sum_i X_i^2 - 2$$

$$= n\sum_i X_i^2 - \{\sum_i X_i^2 + 2\sum_{i<j} X_i X_j\} = n\sum_i X_i^2 -$$

よりわかる．

例 5.11　：$\theta = P(X_1 + X_2 \le 0)$ とする．核を

$$h(x_1, x_2) = 1_{(-\infty,0]}(x_1 + x_2)$$

とおけば，θ の U 統計量

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} 1_{(-\infty,0]}(X_i + X_j)$$

を得る．ここで，

$$1\{X_i + X_j > 0\} = 1\{X_i > 0\}1\{|X_j| < X_i\} + 1\{X_j > 0\}1\{|X_i| < X_j\}$$

に注意すれば，

$$\sum_{1 \le i < j \le n} 1\{X_i + X_j > 0\} + \sum_{i=1}^{n} 1\{X_i > 0\} = \sum_{i=1}^{n}\sum_{j=1}^{n} 1\{X_i > 0\}1\{|X_j| < X_i\}$$

となる．上の式の最右辺は Wilcoxon 統計量である．

5.6.1 U − 統計量の分散

$\boldsymbol{E}|h(X_1, \ldots, X_m)|^2 < \infty$ の場合には，U_n の分散が陽に与えられる．
分散を導出するために，以下の記号が必要である．$k = 1, 2, \ldots, m$ に対して，

$$\begin{aligned} h_k(x_1, \ldots, x_k) &= \boldsymbol{E}[h(X_1, \ldots, X_m)|X_1 = x_1, \ldots, X_k = x_k] \\ &= \boldsymbol{E}[h(x_1, \ldots, x_k, X_{k+1}, \ldots, X_m)] \end{aligned}$$

とおく．すると

$$h_k(x_1, \ldots, x_k) = \mathbb{E}[h_{k+1}(x_1, \ldots, x_k, X_{k+1})]$$

となる．さらに，

$$\tilde{h}_k = h_k - \boldsymbol{E}[h_k(X_1, \ldots, X_m)]$$

とする．このとき，

$$U_n - \boldsymbol{E}(U_n) = \begin{pmatrix} n \\ m \end{pmatrix}^{-1} \sum_c \tilde{h}_m(X_{i_1}, \ldots, X_{i_m})$$

と書ける．

定理 5.6　：$\boldsymbol{E}|h(X_1, \ldots, X_m)|^2 < \infty$ のとき，U_n の分散は以下で与えられる．

$$\boldsymbol{var}(U_n) = \begin{pmatrix} n \\ m \end{pmatrix}^{-1} \sum_{k=1}^{m} \begin{pmatrix} m \\ k \end{pmatrix} \begin{pmatrix} n-m \\ m-k \end{pmatrix} \zeta_k^2.$$

ただし，

$$\zeta_k^2 = \boldsymbol{var}[h_k(X_1, \ldots, X_k)]$$

である．

証明：$\{1,2,\ldots,n\}$ の部分集合 $\{i_1,\ldots,i_m\}$ と $\{j_1,\ldots,j_m\}$ は k 個の共通な整数をもつ異なる m 個の整数でそれぞれ構成されるとする．このような2つの集合の組の数は

$$\begin{pmatrix} n \\ m \end{pmatrix}\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} n-m \\ m-k \end{pmatrix}$$

となる．$\tilde{h}_m$ の対称性と $\boldsymbol{E}\tilde{h}_m=0$，$X_1,\ldots,X_n$ の独立性より，$k=1,2,\ldots,m$ に対して，

$$\boldsymbol{E}[\tilde{h}_m(X_{i_1},\ldots,X_{i_m})\tilde{h}_m(X_{j_1},\ldots,X_{j_m})]=\zeta_k^2$$

となる[3)]．したがって，

$$\begin{aligned}\boldsymbol{var}U_n &= \begin{pmatrix} n \\ m \end{pmatrix}^{-2}\sum_c\sum_c \boldsymbol{E}[\tilde{h}_1(X_{i_1},\ldots,X_{i_m})\tilde{h}_1(X_{j_1},\ldots,X_{j_m})] \\ &= \begin{pmatrix} n \\ m \end{pmatrix}^{-2}\sum_{k=1}^{n}\begin{pmatrix} n \\ m \end{pmatrix}\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} n-m \\ m-k \end{pmatrix}\zeta_k^2 \qquad (5.113)\end{aligned}$$

から定理は証明された．　□

系 5.2　：$\boldsymbol{E}|h(X_1,\ldots,X_m)|^2<\infty$ ならば，

(i)

$$\frac{m^2}{n}\zeta_1^2 \le \boldsymbol{var}(U_n) \le \frac{m}{n}\zeta_m^2$$

(ii)

$$(n+1)\boldsymbol{var}(U_{n+1}) \le n\boldsymbol{var}(U_n) \qquad (n>m)$$

(iii) 固定した m と $k(1\le k\le m)$ に対して，$j<k$ の時，$\zeta_j=0$ で，$\zeta_k^2>0$ ならば，

$$\boldsymbol{var}(U_n)=\frac{k!\begin{pmatrix} m \\ k \end{pmatrix}^2\zeta_k^2}{n^k}+O\left(\frac{1}{n^{k+1}}\right)$$

である．

証明：　□

5.6.2 U – 統計量の漸近分布

射影法を用いて漸近分布を求める．

---

3) $\{i_1,\ldots,i_m\}$ と $\{j_1,\ldots,j_m\}$ は k 個の共通な整数をもつのでこうなる．たとえば，$k=1$ と $m=2$ の場合には簡単にわかる：

$$\begin{aligned}\boldsymbol{E}[\tilde{h}_1(X_1,X_2)\tilde{h}_1(X_1,X_3)] &= \boldsymbol{E}[\boldsymbol{E}[\tilde{h}_1(X_1,X_2)\tilde{h}_1(X_1,X_3)|X_1]] \\ &= \boldsymbol{E}[\boldsymbol{E}[\tilde{h}_1(X_1,X_2)|X_1]\boldsymbol{E}[\tilde{h}_1(X_1,X_3)|X_1\end{aligned}$$

定理 5.7　：$\boldsymbol{E}|h(X_1,\ldots,X_m)|^2<\infty$ かつ $\zeta_1^2>0$ のとき，

$$\sqrt{n}(U_n-\boldsymbol{E}(U_n))\xrightarrow{L}N(0,m^2\zeta_1^2)$$

となる．

証明：U_n に対して，

$$\check{U}_n=\boldsymbol{E}(U_n)+\frac{m}{n}\sum_{i=1}^{n}\tilde{h}_1(X_i)$$

と定義する．系 3.1 より

$$\boldsymbol{var}(\check{U}_n)=m^2\frac{\zeta_1^2}{n}$$

となることがわかる．また，定理 3.4 から

$$\boldsymbol{var}(U_n)==m^2\frac{\zeta_1^2}{n}+O(\frac{1}{n^2})$$

となるので，補題 3.2 から

$$\boldsymbol{E}(U_n-\check{U}_n)=O(\frac{1}{n^2})$$

となる．これより，$\sqrt{n}(U_n-\tilde{U}_n)=o_P(1)$ が成立する[4]．$\tilde{U}_n$ は独立同一の確率変数の和あることから

$$\sqrt{n}(\tilde{U}_n-\mathbb{E}(\tilde{U}_n))\xrightarrow{L}N(0,m^2\zeta_1^2)$$

となり，Slutsky の補題を用いれば，定理は証明される．□

定理 5.8　：$\boldsymbol{E}|h(X_1,\ldots,X_m)|^2<\infty$，$\zeta_1^2=0$ かつ $\zeta_2^2>0$ のとき，

$$n(U_n-\boldsymbol{E}(U_n))\xrightarrow{L}\frac{m(m-1)}{2}\sum_{j=1}^{\infty}\lambda_j(\xi_{1j}^2-1)$$

となる．ただし，ξ_{1j}^2 は自由度 1 のカイ自乗分布で，λ_j は積分方程式

$$\boldsymbol{E}[h_2(x_1,X_2)f(X_2)]=\lambda_j f(x_1)$$

を満足する定数[5]でる．

証明：Lee（1990) の pages 80 – 83 または Serfling (1980) の paes 193 – 199 を参照されたい．□

[4] なぜならば，任意の $\epsilon>0$ に対し

$$P(\sqrt{n}|U_n-\tilde{U}_n|\geq\epsilon)\leq n\mathbb{E}(U_n-\tilde{U}_n)^2=O(\frac{1}{n}$$

からわかる．

[5] f は固有関数で，$\sum_{j=1}^{\infty}\lambda_j=\zeta_2^2$ を満足する．

・・・・・5.7・・・・・

R－統計量とその漸近分布

$R_{n1},\dots,R_{nn}$ を $\{1,2,\dots,n\}$ の置換とする．ここでは，

$$S_n=\sum_{j=1}^{n} z_{nj}a_n(R_{nj}) \tag{5.114}$$

の漸近分布を調べる．ただし，$z_{n1},\dots,z_{nn}$ と $a_n(1),\dots,a_n(n)$ はある与えられた定数である．

まず，S_n の平均と分散を求める．

補題 5.6　：
(i) $\boldsymbol{E}(S_n)=n\bar{z}_n\bar{a}_n$
(ii) $\boldsymbol{var}(S_n)=\frac{n^2}{n-1}\sigma_z^2\sigma_a^2$.
ただし，

$$\bar{z}_n=(1/n)\sum_{j=1}^{n} z_{nj},\qquad \sigma_z^2=(1/n)\sum_{j=1}^{n}(z_{nj}-\bar{z}_n)^2$$
$$\bar{a}_n=(1/n)\sum_{j=1}^{n} a_{nj},\qquad \sigma_a^2=(1/n)\sum_{j=1}^{n}(a_{nj}-\bar{a}_n)^2$$

である．

証明：以下，n を略してこの証明中では，$a_n(\cdot)$ や R_{nj} を $a(\cdot)$ や R_j のようにかく．$\boldsymbol{E}a(R_j)=(1/n)\sum_{j=1}^{n}a(j)=\bar{a}_n$ と $\boldsymbol{var}a(R_j)=(1/n)\sum_{j=1}^{n}(a(j)-\bar{a}_n)^2=\sigma_a^2$ になることに注意すれば，

$$\boldsymbol{E}S_n=\sum_{j=1}^{n} z_j\boldsymbol{E}a(R_j)=\sum_{j=1}^{n} z_j\bar{a}_n=n\bar{z}_na_n$$

がわかる．$j\neq k$ の場合，$\boldsymbol{cov}(a(R_j),a(R_k))$ は j と k に依存せず，すべての $j\neq k$ に対して，

$$P(R_1=j,R_2=k)=\frac{1}{n(n-1)}$$

であるので，

$$\begin{aligned}\boldsymbol{cov}(a(R_1),a(R_2))&=\frac{1}{n(n-1)}\sum_{j\neq k}(a(j)-\bar{a}_n)(a(k)-\bar{a}_n)\\&=\frac{1}{n(n-1)}\Big[\{\sum_{j=1}^{n}(a(j)-\bar{a}_n)\}^2-\sum_{j=1}^{n}(a(j)-\bar{a}_n)^2\Big]\\&=-\frac{1}{n(n-1)}\sum_{j=1}^{n}(a(j)-\bar{a}_n)^2=-\frac{1}{n-1}\sigma_a^2\end{aligned}$$

となる．これより，

$$\begin{aligned}\boldsymbol{var}S_n &= \sum_{j=1}^{n} z_j^2 \boldsymbol{var}(a(R_j)) + \sum_{j\neq k} z_j z_k \boldsymbol{cov}(a(R_j), a(R_k)) \\ &= \sigma_a^2 [\sum_{j=1}^{n} z_j^2 - \frac{1}{n-1}\sum_{j\neq k} z_j z_k] \\ &= \sigma_a^2 [\sum_{j=1}^{n} z_j^2 - \frac{1}{n-1}\{\sum_{j=1}^{n} z_j\}^2 + \frac{1}{n-1}\sum_{j=1}^{n} z_j^2] \\ &= \sigma_a^2 [n\sigma_z^2 + \frac{n}{n-1}\sigma_z^2] = \frac{n^2}{n-1}\sigma_z^2\sigma_a^2\end{aligned}$$

より補題は示された．□

S_n の漸近正規性を示すために，独立な確率変数からなる和 S'_n を構成し，これが S_n と漸近同値であることを示す．

そのために，一様分布 $U(0,1)$ からのランダム標本 $U_1,\ldots,U_n$ を考える．R_{nj} を U_j の順位とすれば，R_{nj}/n と U_j は非常に近い値をとることが期待できる．したがって，(5.114) において R_{nj} を nU_j に置き換えたものは独立な確率変数の和であり，また，$S_n - \boldsymbol{E}(S_n)$ に漸近同値となることが期待できる．ここで，

$$S_n - \boldsymbol{E}(S_n) = \sum_{j=1}^{n}(z_{nj} - \bar{z}_n)(a_n(R_{nj}) - \bar{a}_n)$$

に注意して，

$$S'_n = \sum_{j=1}^{n}(z_{nj} - \bar{z}_n)(a_n([nU_j]) - \bar{a}_n)$$

と定める．ただし，$[b]$ は b を越えない整数である．このとき，$\boldsymbol{E}(S'_n) = 0$ と

$$\boldsymbol{var}(S'_n) = n\sigma_z^2 \boldsymbol{var}(a_n(R_{n1})) \tag{5.115}$$

となる．なぜならば，$P([nU_j] = i) = P(R_{n1} = i) = (1/n)$ よりわかる．

補題 5.7　：

$$\boldsymbol{cov} = \sqrt{\frac{n}{n-1}}\boldsymbol{corr}(a_n(R_{n1}), a_n([nU_1]))$$

証明：

$$\boldsymbol{cov}(S_n, S'_n) = \sum_{j=1}^{n}\sum_{k=1}^{n}(z_j - \bar{z}_n)(z_k - \bar{z}_n)\boldsymbol{cov}(a(R_j), a(R_k))$$

に注意する．値 $c_1 = \boldsymbol{cov}(a(R_j), a([nU_j]))$ は j によらない．また，$j \neq k$ のとき，値 $c_2 = \boldsymbol{cov}(a(R_j), a([nU_k]))$ も j と k には依存しない．したがって，

$$\boldsymbol{cov}(S_n, S'_n) = c_1\sum_{j=1}^{n}(z_j - \bar{z}_n)^2 + c_2\sum_{j\neq k}(z_j - \bar{z}_n)(z_k - \bar{z}_n) = (c_1 - c_2)\sum_{j=1}^{n}(z_j - \bar{z}_n)^2 \tag{5.116}$$

となる[1]．$\sum_{j=1}^{n} a(R_j)$ は定数なので，

$$0 = \boldsymbol{cov}\bigl(\sum_{j=1}^{n} a(R_j), a([nU_k])\bigr) = \sum_{j=1}^{n} \boldsymbol{cov}(a(R_j), a([nU_k])) = c_1 + (n-1)c_2$$

となることから，$c_2 = -c_1/(n-1)$ がわかる．これを (5.116) に代入すれば，

$$\boldsymbol{cov}(S_n, S'_n) = \frac{nc_1}{n-1} \sum_{j=1}^{n} (z_j - \bar{z}_n)^2$$

となる．補題 3.3(i) および (ii) を使えば，

$$\begin{aligned}\boldsymbol{corr}(S_n, S'_n) &= \frac{\boldsymbol{cov}(S_n, S'_n)}{\sqrt{\boldsymbol{var}(S_n)}\sqrt{\boldsymbol{var}(S'_n)}} = \sqrt{\frac{n}{n-1}} \frac{\boldsymbol{cov}(a(R_1), a([nU_1]))}{\sqrt{\boldsymbol{var}(a(R_1))}\sqrt{\boldsymbol{var}(a([nU_1]))}} \\ &= \sqrt{\frac{n}{n-1}} \boldsymbol{corr}(a(R_1), a([nU_1]))\end{aligned}$$

となることから補題は証明された．　□

$$\frac{\boldsymbol{E}\{a_n(R_{n1}) - a_n([nU_1])\}^2}{\boldsymbol{var}(a_n(R_{n1}))} = 2(1 - \boldsymbol{corr}(a_n(R_{n1}), a_n([nU_1])))$$

から

$$\boldsymbol{corr}(S_n, S'_n) \to 1 \qquad (n \to \infty)$$

を示すためには，

$$\frac{\boldsymbol{E}\{a_n(R_{n1}) - a_n([nU_1])\}^2}{\boldsymbol{var}(a_n(R_{n1}))} \to 0 \qquad (n \to \infty)$$

を示せばよいことがわかる． つぎの補題は上の式の上限を与える．

補題 5.8　：$a_n(\cdot)$ は単調と仮定する．このとき，

$$\boldsymbol{E}\{a_n(R_{n1}) - a_n([nU_1])\}^2 \le \frac{2\sqrt{2}}{n} \max_j |a_n(j) - \bar{a}_n| \sqrt{\sum_{j=1}^{n} (a(j) - \bar{a}_n)^2}$$

が成立する．

証明：略．　□

定理 5.9　：

$$\delta_n = \frac{n \max_j (z_{nj} - \bar{z}_n)^2 \max_j (a_n(j) - \bar{a}_n)^2}{\sum_{j=1}^{n} (z_{nj} - \bar{z}_n)^2 \sum_{j=1}^{n} (a_n(j) - \bar{a}_n)^2} \tag{5.117}$$

[1] なぜならば，$\sum_{j \ne k} (z_j - \bar{z}_n)(z_k - \bar{z}_n) = \{\sum_{j=1}^{n} (z_j - \bar{z}_n)\}^2 - \sum_{j=1}^{n} (z_j - \bar{z}_n)^2$ よりわかる．

ならば,

$$\frac{S_n - \boldsymbol{E}(S_n)}{\boldsymbol{var}(S_n)} \xrightarrow{L} N(0,1)$$

が成立する.

証明：

$$\frac{n \max_j (z_j - \bar{z}_n)^2}{\sum_{j=1}^n (z_j - \bar{z}_n)^2} > 1$$

と条件から，$n \uparrow \infty$ のとき,

$$\frac{\max_j (a_i - \bar{a}_n)^2}{\sum_{j=1}^n (a_j - \bar{a}_n)^2} \longrightarrow 0$$

となる．したがって,

$$\frac{2\sqrt{2}}{n} \max_j |a_i - \bar{a}_n| \sqrt{\sum_{j=1}^n (a_j - \bar{a}_n)^2} = 2\sqrt{2} \sqrt{\frac{\max_j (a_i - \bar{a}_n)^2}{\sum_{j=1}^n (a_j - \bar{a}_n)^2}} \frac{1}{n} \sum_{j=1}^n (a_j - \bar{a}_n)^2 \longrightarrow 0, \qquad (n \uparrow \infty)$$

となる．補題 3.5 から

$$\frac{\boldsymbol{E}\{a(R_1) - a([nU_1])\}^2}{\boldsymbol{var}(a(R_1))} \longrightarrow 0, \qquad (n \uparrow \infty)$$

となり，補題 3.4 から

$$\boldsymbol{corr}(S_n, S_n') = \sqrt{\frac{n}{n-1}} \boldsymbol{corr}(a(R_1), a([nU_1])) \longrightarrow 1, \qquad (n \uparrow \infty)$$

となる．したがって,

$$\boldsymbol{E}\left\{ \frac{S_n'}{\sqrt{\boldsymbol{var} S_n'}} - \frac{S_n}{\sqrt{\boldsymbol{var} S_n}} \right\}^2 = 1 - \boldsymbol{corr}(S_n, S_n') \longrightarrow 0, \qquad (n \uparrow \infty)$$

より,

$$\frac{S_n'}{\sqrt{\boldsymbol{var} S_n'}} - \frac{S_n}{\sqrt{\boldsymbol{var} S_n}} = o_P(1)$$

を得る．あとは，Lindeberg – Feller の中心極限定理が (5.117) のもとで成立することを示せば,

$$\frac{S_n'}{\sqrt{\boldsymbol{var} S_n'}} \xrightarrow{L} N(0,1) \tag{5.118}$$

がわかる．そのために,

$$B_n^2 = \boldsymbol{var} S_n', \qquad X_j = (z_j - \bar{z}_n)(a([nU_j]) - \bar{a}_n)$$

とおく[2]．Lindeberg 条件 (1.27) を確認するために，正の数 $\epsilon > 0$ をとる．

$$
\begin{aligned}
&\frac{1}{B_n^2}\sum_{j=1}^{n}\boldsymbol{E}\big[X_j^2 1\{|X_j| \geq \epsilon B_n\}\big] \\
&\quad = \frac{1}{B_n^2}\sum_{j=1}^{n}\boldsymbol{E}\big[(z_j - \bar{z}_n)^2(a_n([nU_1]) - \bar{a}_n)^2 1\{(z_j - \bar{z}_n)^2(a_n([nU_1]) - \bar{a}_n)^2 \geq \epsilon^2 B_n^2\}\big] \\
&\quad \leq \frac{1}{B_n^2}\sum_{j=1}^{n}\boldsymbol{E}\big[(z_j - \bar{z}_n)^2(a_n([nU_1]) - \bar{a}_n)^2 1\{\max_j(z_j - \bar{z}_n)^2 \max_j(a_n([nU_1]) - \bar{a}_n)^2 \geq \epsilon^2 B_n^2\}\big] \\
&\quad \leq \frac{1}{B_n^2}\sum_{j=1}^{n}(z_j - \bar{z}_n)^2\boldsymbol{E}\big[(a_n([nU_1]) - \bar{a}_n)^2 1\{\delta_n \geq \epsilon^2\}\big] \\
&\quad \leq 1\{\delta_n \geq \epsilon^2\}\frac{1}{B_n^2}\sum_{j=1}^{n}(z_j - \bar{z}_n)^2\boldsymbol{var}(a([nU_1])) \\
&\quad \leq 1\{\delta_n \geq \epsilon^2\} \longrightarrow 0, \qquad (n \to \infty)
\end{aligned}
$$

となるので，(5.118) が示せた．よって，定理が証明された．　□

[2] $\boldsymbol{E}X_j = 0$ になっていることに注意せよ．

・・・・・5.8・・・・・

M – 推定量とその漸近分布

F を $\boldsymbol{R}^d$ 上の分布関数とし，$\rho(x,t)$ を $\boldsymbol{R}^d \times \boldsymbol{R}$ 上のボレロ可測関数とする．ρ に対応する M – 汎関数 T をつぎの方程式の解で定義する：

$$\int \rho(x, T(G))dG(x) = \inf_{t\in\Theta} \int \rho(x,t)dG(x), \qquad (G \in \mathcal{F})$$

ただし，Θ は $\boldsymbol{R}$ の開区間とする．$T(F_n)$ を $T(F)$ の **M – 推定量**とよぶ．$\psi(x,t) = \partial\rho(x,t)/\partial t$ がほとんど確実に存在するとき，T はつぎの方程式の解ともなる：

$$\int \psi(x, T(G))dG(x) = 0$$

例 5.12　(The least absolute value estimate)：$d = 1$ とし，$\rho(x,t) = |x - t|$ とおく．対応する ψ は

$$\psi(x,t) = \begin{cases} -1 & (x < t) \\ 0 & (x = t) \\ 1 & (x > t) \end{cases}$$

となる．

例 5.13　(The trimmed sample mean)：$d = 1$ とする．

$$\rho(x,t) = \begin{cases} \frac{1}{2}(x-t)^2 & (|x-t| \le k) \\ k|x-t| - \frac{k^2}{2} & (\text{その他}) \end{cases}$$

とおけば,

$$\psi(x,t) = \begin{cases} x - t & (|x-t| \le k) \\ k \ \mathrm{sign}(x-t) & (\text{その他}) \end{cases}$$

となる．

$\psi(x,t)$ に対して,

$$\lambda_F(t) = \int \psi(x,t)dF(x) \qquad \lambda_n(t) = \int \psi(x,t)dF_n(x)$$

とおく．

定理 5.10　：Θ を $\boldsymbol{R}$ の開区間とする．どんな $t \in \Theta$ に対しても $\lambda_n(t) \xrightarrow{P} \lambda_F(t)$ が成立するとする．さらに，$t \mapsto \lambda_n(t)$ は連続でただひとつの解 $\hat{\theta}_n$ をもつか非減少で $\lambda_n(\hat{\theta}_n) = o_P(1)$ であるとする．θ_0 をどんな $\epsilon > 0$ に対しても $\lambda_F(\theta_0 - \epsilon) < 0 < \lambda_F(\theta_0 + \epsilon)$ を満足するような点とする．このとき，$\hat{\theta} \xrightarrow{P} \theta_0$ である．

証明：$t \mapsto \lambda_n(t)$ は連続で，解 $\hat{\theta}_n$ を唯一もつので，

$$P(\lambda_n(\theta_0 - \epsilon) < 0, \lambda_n(\theta_0 + \epsilon) > 0) \leq P(\theta_0 - \epsilon < \hat{\theta}_n < \theta_0 + \epsilon)$$

である．$\lambda_n(\theta_0 \pm \epsilon) \xrightarrow{P} \lambda_F(\theta_0 \pm \epsilon)$ より，上の左辺は 1 に収束する．これより，$\hat{\theta}_n$ の一致性がわかる．

つぎに，$t \mapsto \lambda_n(t)$ の非減少性を仮定する．どんな $\eta > 0$ に対しても，$\lambda_n(\theta_0 - \epsilon) < -\eta$ かつ $\hat{\theta}_n \leq \theta_0 - \epsilon$ ならば，$\lambda_n(\hat{\theta}_n) < -\eta$ となるので，$\lambda_n(\hat{\theta}_n) = o_P(1)$ と矛盾する．したがって，$\lambda_n(\theta_0 - \epsilon) < -\eta$ と $\hat{\theta}_n \leq \theta_0 - \epsilon$ は同時に起こらない．すなわち，$\lambda_n(\theta_0 - \epsilon) < -\eta$ ならば，$\hat{\theta}_n > \theta_0 - \epsilon$ である．同様にすれば，どんな $\epsilon, \eta > 0$ に対しても

$$P(\lambda_n(\theta_0 - \epsilon) < -\eta, \lambda_n(\theta_0 + \epsilon) > \eta) \leq P(\theta_0 - \epsilon < \hat{\theta}_n < \theta_0 + \epsilon) + o(1)$$

となる．2η を $\lambda_F(\theta_0 - \epsilon)$ と $\lambda_F(\theta_0 + \epsilon)$ の小さい方と同じにとれば，

$$P(\lambda_n(\theta_0 - \epsilon) < -\eta, \lambda_n(\theta_0 + \epsilon) > \eta) \uparrow 1, \qquad (n \to \infty)$$

となるので，一致性は示せた．□

定理 5.11　：θ_0 を $\lambda_F(t) = 0$ の唯一の解とする．$\psi(x, t)$ は t の単調関数とする．$\lambda_F(t)$ は $t = \theta_0$ で微分可能とし，微係数 $\dot{\lambda}_F(\theta_0) \neq 0$ を持つとする．θ_0 のある近傍に含まれる t に対して，$\int \psi^2(x, t) dF(x) < \infty$ とし，また，$t = \theta_0$ において連続であるとする．このとき，方程式 $\lambda_n(t) = 0$ の解 T_n は

$$\sqrt{n}(T_n - \theta_0) \xrightarrow{L} N(0, \sigma^2(T, F))$$

を満足する．ただし，

$$\sigma^2(T, F) = \frac{\int \psi^2(x, \theta_0) dF(x)}{\{\dot{\lambda}_F(\theta_0)\}^2} \equiv \sigma^2$$

である．

証明：$\psi(x, t)$ は t に関して非増加ならば，λ_n も非増加となるので，

$$P(\lambda_n(t) < 0) \leq P(T_n \leq t) \leq P(\lambda_n(t) \leq 0)$$

となる．したがって，$t_{z,n} = \theta_0 + z\sigma n^{-1/2}$ （$z \in \boldsymbol{R}$）に対して，

$$\lim_{n \to \infty} P(\lambda_n(t_{z,n}) < 0) = \lim_{n \to \infty} P(\lambda_n(t_{z,n}) \leq 0) = \Phi(z)$$

を示せばよい．したがって，

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} \le \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}}\right) = \Phi(z)$$

を示せばよいこと[1] がわかる．ただし，$s_{z,n}^2 = \boldsymbol{var}_F\{\psi(X_1, t_{z,n})\}$ とし，

$$Y_{ni} = \frac{\psi(X_i, t_{z,n}) - \lambda_F(t_{z,n})}{s_{z,n}}$$

とし，$\Phi(\cdot)$ は標準正規分布の分布関数とする．仮定から，$n \uparrow \infty$ のとき，

$$\sqrt{n}\lambda_F(t_{z,n}) = z\sigma\frac{\lambda_F(\theta_0 + z\sigma/\sqrt{n}) - \lambda_F(\theta_0)}{z\sigma/\sqrt{n}} \to z\sigma\dot{\lambda}_F(\theta_0)$$

と

$$s_{z,n}^2 = \int \psi^2(X_1, \theta_0 + \frac{z\sigma}{\sqrt{n}})dF(x) = \dot{\lambda}_F^2(\theta_0)\frac{\int \psi^2(X_1, \theta_0 + z\sigma/\sqrt{n})dF(x)}{\dot{\lambda}_F^2(\theta_0)} \to \dot{\lambda}_F^2(\theta_0)\sigma^2$$

となる[2]．$\dot{\lambda}_F(\theta_0) < 0$ に注意すれば，$s_{z,n} \to -\dot{\lambda}_F(\theta_0)\sigma$ を得る．したがって，$n \uparrow \infty$ のとき，

$$\frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}} \to z$$

を得る．よって，

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} \le z\right) = \Phi(z)$$

を示せばよい．Y_{ni} は平均 0, 分散 1 の独立同一の分布に従う確率変数なので，Lindeberg 条件 (1.27) を確認すれば，よいことがわかる：すなわち，正の数 $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} \int_{|\psi(x, t_{z,n})| > \sqrt{n}\epsilon} \psi^2(x, t_{z,n})dF(x) = 0. \tag{5.119}$$

どんな $\eta > 0$ に対しても n を十分大きくとれば，

$$\psi(x, \theta_0 + \eta) \le \psi(x, t_{z,n}) \le \psi(x, \theta_0 - \eta)$$

---

[1] これは以下のことからわまる．$P(\lambda_n(t) < 0) = P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} < \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}}\right)$ となることに注意すれば

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} < \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}} - \frac{1}{m}\right) \le P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} < \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}}\right)$$

となる．また，

$$\lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} < \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}} - \frac{1}{m}\right) = \lim_{m\to\infty}\Phi\left(z - \frac{1}{m}\right) = \Phi(z)$$

となることも同様の議論をすればわかる．したがって，

$$\Phi(z) = \lim_{m\to\infty}\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} < \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}} - \frac{1}{m}\right) \le \lim_{n\to\infty}$$

$$\le \lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{ni} \le \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}}\right) = \Phi(z)$$

となることからわかる．

[2] なぜならば，すべての固定された x に対して，$\psi(x,t)$ は t の単調関数なので，単調収束定理を使えば，$\lim_{n\to\infty} \boldsymbol{E}\psi^2(X_1, \theta_0 + z\sigma/\sqrt{n}) = \boldsymbol{E}\psi^2(X_1, \theta_0)$ となることよりわかる．

より，$u(x) = \max\{|\psi(x, \theta_0 - \eta)|, |\psi(x, \theta_0 + \eta)|\}$ とおけば，

$$\int_{|\psi(x,t_{z,n})|>\sqrt{n}\epsilon} \psi^2(x, t_{z,n})dF(x) \le \int_{u(x)>\sqrt{n}\epsilon} \psi^2(x, t_{z,n})dF(x) \to 0, \qquad (n \to \infty)$$

となることが，$\int \psi^2(x, \theta_0 \pm \eta)dF(x) < \infty$ からわかり，(5.122) が示せた．

つぎに，$\psi(x,t)$ は t に関して非減少ならば，λ_n も非減少となるので，

$$P(\lambda_n(t) > 0) \le P(T_n \le t) \le P(\lambda_n(t) \ge 0)$$

となる．さらに

$$P(\lambda_n(t) > 0) = P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{ni} > \frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}}\right)$$

になることと $\dot{\lambda}_F(\theta_0) > 0$ となるので，

$$\frac{-\sqrt{n}\lambda_F(t_{z,n})}{s_{z,n}} \to -z$$

となることから

$$\lim_{n\to\infty} P(\lambda_n(t) < 0) = 1 - \Phi(-z) = \Phi(z)$$

がわかる．よって，定理は証明された．□

例 5.14　$X_1, X_2, \ldots, X_n$ は独立同一に $\mathbb{R}$ 上の累積分布関数 F に従うとし，$\psi(x, t) = x - t$ とする．

$$\sum_{i=1}^{n} \psi(X_i, T(F_n)) = \sum_{i=1}^{n} (X_i - T(F_n)) = 0$$

の解は最小 2 上推定量 $T(F_n) = \bar{X}_n$ となる．

(1)　F が正規分布 $N(\theta_0, \sigma^2)$ ならば，

$$\dot{\lambda}_F(t) = \frac{d}{dt} \int \psi(x, t)\, dF(x) = \frac{d}{dt}(\mu - t) = -1$$

と

$$\int \psi^2(x, \theta_0)\, dF(x) = \sigma^2$$

となるので，定理 (5.10) と (5.11) から

$$\sqrt{n}(\bar{X}_n - \theta_0) \xrightarrow{L} N(0, \sigma^2)$$

となる．

(2)　F が自由度 ν の t 分布

$$f(x) = \frac{c(\nu)}{(1 + (x - \theta_\theta)^2/\nu)^{(\nu+1)/2}}$$

に従う場合を考えよう．ただし，$\nu \geq 3$ で

$$c(\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}$$

である．

$$\dot{\lambda}_F(t) = \int \psi(x,\,t)\,dF(x) = \frac{d}{dt}(\mu - t) = -1$$

と

$$\begin{aligned}
\int \psi^2(x,\,\theta_0)\,dF(x) &= c(\nu)\int \frac{(x-\theta_0)^2 dF}{(1+(x-\theta_0)^2/\nu)^{(\nu+1)/2}}\,dx \\
&= c(\nu)\int \frac{x^2}{(1+x^2/\nu)^{(\nu+1)/2}}\,dx \\
&= c(\nu)\left[\frac{1}{(1+x^2/\nu)^{(\nu-1)/2}}\frac{-\nu x}{\nu-1}\right]_{-\infty}^{\infty} + c(\nu)\frac{\nu}{\nu-1}\int \frac{1}{(1+x^2/\nu)^{(\nu-1)/2}}\,dx \\
&= \frac{c(\nu)}{c(\nu-2)}\frac{\nu}{\nu-1}\sqrt{\frac{\nu}{\nu-2}}\,c(\nu-2)\int \frac{1}{(1+x^2/(\nu-2))^{(\nu-1)/2}}\,dx \\
&= \frac{c(\nu)}{c(\nu-2)}\frac{\nu}{\nu-1}\sqrt{\frac{\nu}{\nu-2}} = \frac{\nu}{\nu-2}
\end{aligned}$$

となる[3)]．定理 (5.10) と (5.11) から

$$\sqrt{n}(\bar{X}_n - \theta_0) \xrightarrow{L} N(0,\,\nu/(\nu-2))$$

となる．

(3)　$\nu = 2$ の t 分布のときには，条件をみたさないので，定理 (5.11) は適用できない．

例 5.15　　$X_1, X_2, \ldots, X_n$ は独立同一に $\mathbb{R}$ 上の確率密度関数 f に従うとする．ただし，f は θ_0 に関して対称でひと山分布ある．ψ は例 5.13 のものを用いる．

$$\lambda_F(t) = \int_{t-k}^{t+k}(x-t)f(x)\,dx + k\int_{t+k}^{\infty} f(x)\,dx - k\int_{-\infty}^{t-k} f(x)\,dx$$

となるので，$\lambda_F(\theta_0) = 0$ となる．また，

$$\dot{\lambda}_F(\theta_0) = -\int_{\theta_0-k}^{\theta_0+k} f(x)\,dx = -\mathbb{P}\{\theta_0 - k \leq X \leq \theta_0 + k\}$$

となる[4)]．さらに，

$$\int \psi^2(x,\,\theta_0)f(x)\,dx = \int_{-\theta_0-k}^{\theta_0+k}(x-\theta)^2 f(x)\,dx + 2k^2\mathbb{P}\{X \geq \theta_0 + k\}$$

[3)] 最後の等号は $\Gamma(n+1) = n\Gamma(n)$ からわかる．

[4)]

$$\begin{aligned}
\frac{d}{dt}\int_{t-k}^{t+k}(x-t)f(x)\,dx &= \frac{d}{dt}\int_{t-k}^{t+k} xf(x)\,dx - \frac{d}{dt}t\int_{t-k}^{t+k} f(x)\,dx \\
&= (t+k)f(t+k) - (t-k)f(t-k) - \int_{t-k}^{t+k} \\
&= kf(t+k) + f(t-k) - \int_{t-k}^{t+k} f(x)\,dx
\end{aligned}$$

と $\frac{d}{dt}P(X \geq t+k) = -f(t+k)$ および $\frac{d}{dt}P(X \geq t-k) = f(t-k)$ からわかる．

となる．したがって，$\hat{\theta}_n$ を方程式

$$\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,\,t)=0$$

解とすれば，

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{L}N(0,\,\sigma^2)$$

となる．ただし，

$$\sigma^2=\frac{\int_{-\theta_0-k}^{\theta_0+k}(x-\theta)^2f(x)\,dx+2k^2\mathbb{P}\{X\geq\theta_0+k\}}{(\mathbb{P}\{\theta_0-k\leq X\leq\theta_0+k\})^2}$$

定理 5.12　：T を $\psi(x,t)$ に対応する M －汎関数とする．ψ は有界で，どんな x に対しても $\psi(x,t)$ は $t=T(F)$ で連続とする．また，$\lambda_F(t)$ は $t=T(F)$ で微分可能で微係数 $\dot{\lambda}_F(T(F))\neq 0$ をもつとする．このとき，T は F において ρ_∞ － Hadamard 微分可能で影響関数

$$\phi_F(x)\equiv\dot{T}_F(\delta_x-F)=-\frac{\psi(x,F)}{\dot{\lambda}_F(T(F))}$$

をもつ．

証明:$j\uparrow\infty$ のとき，$t_j\to 0,\Delta_j\in\mathcal{D},\|\Delta_j-\Delta\|_\infty\to 0,G_j=F+t_j\Delta_j\in\mathcal{F}$ とする．$\lambda_G(T(G))=0$ より

$$|\lambda_F(T(G_j))-\lambda_F(T(F))|=|t_j\int\psi(x,T(G_j))d\Delta_j|\to 0,\qquad(t_j\to 0)$$

となることが，$\|\Delta_j-\Delta\|_\infty\to 0$ と ψ の有界性からわかる．また，$\dot{\lambda}_F(T(F))\neq 0$ より，$\lambda_F(t)$ の逆写像が $0=\lambda_F(T(F))$ の近傍で存在する．したがって，

$$T(G_j)-T(F)\to 0 \tag{5.120}$$

を得る．いま，

$$h_F(t)=\begin{cases}\dot{\lambda}_F(T(F)) & (t=T(F))\\ (\lambda_F(t)-\lambda_F(T(F)))/(t-T(F)) & (t\neq T(F))\end{cases}$$

とし，

$$R_{1j}=\int\psi(x,T(F))d\Delta_j(x)\left[\frac{1}{\dot{\lambda}_F(T(F))}-\frac{1}{h_F(T(G_j))}\right]$$

$$R_{2j}=\frac{1}{h_F(T(G_j))}\left[\int[\psi(x,T(G_j))-\psi(x,T(F))]d\Delta_j(x)\right]$$

$$L_F(\Delta)=\frac{1}{\dot{\lambda}_F(T(F))}\int\psi(x,T(F))d\Delta,\qquad D\in\mathcal{D}$$

とおけば，

$$T(G_j)-T(F)=-L_F(t_j\Delta_j)+t_j(R_{1j}-R_{2j})$$

と書ける．(5.120)，$\|\Delta_j - \Delta\|_\infty \to 0$ および ψ の有界性から，$R_{1j} \to 0$ となる．また，ψ の有界性と連続性から，$R_{2j} \to 0$ となることもわかる．したがって，

$$\frac{T(G_j) - T(F)}{t_j} \to -\frac{\psi(x,F)}{\dot{\lambda}_F(T(F))}$$

を得る． □

5.8.1 補遺：double array 確率変数列に対する中心極限定理

double arrray の確率変数列を考える．

$$\begin{array}{llll} X_{11}, & X_{12}, & \ldots, & X_{1k_1}; \\ X_{21}, & X_{22}, & \ldots, & X_{2k_2}; \\ \vdots & & & \vdots \\ X_{n1}, & X_{n2}, & \ldots, & X_{nk_n}; \end{array}$$

各 $n \geq 1$ に対し，k_n 個の確率変数 $\{X_{nj} : 1 \leq j \leq k_j\}$ があり，この k_n 個の確率変数は独立とする．$n \to \infty$ のとき $k_n \to \infty$ を仮定する．

X_{nj} の累積分布関数を F_{nj} と記し，

$$\begin{aligned} \mu_{nj} &= \mathbb{E}[X_{nj}] \\ A_n &= \mathbb{E}[\sum_{j=1}^{k_n} X_{nj}] = \sum_{j=1}^{k_n} \mu_{nj} \\ B_n^2 &= \mathbb{VAR}[\sum_{j=1}^{k_n} X_{nj}] \end{aligned}$$

とおく．

定理 5.13　$\{X_{nj} : 1 \leq j \leq k_j\}$ は独立な確率変数列（double array の行ごとに）とし，$\mathbb{E}[X_{nj}] = 0$ かつ $B_n^2 = 1$ とする．このとき，$\sum_{j=1}^{k_n} X_{nj}$ が標準正規分布に分布収束するための必要十分条件は，すべての $\epsilon > 0$ に対し

$$\sum_{j=1}^{k_n} \int_{|t|>\epsilon} t^2 \, dF_{nj}(t) \to 0, \qquad (n \to \infty) \tag{5.121}$$

である．

証明　証明は Chow and Teicher(Probability Theory, 3rd edition, p. 466) を参照せよ． □

定理 5.11 の証明の最後の部分の訂正

Y_{ni} は平均 0, 分散 1 の独立同一の分布に従う確率変数なので，$Y_{nj}/\sqrt{n}$ に対して定理 5.13 を適用する．そのために，条件 (5.121) を確認すればよいことがわかる：すなわち，正の数 $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} \frac{1}{s^2_{t_{z,n}}} \int_{|\psi(x,t_{z,n})-\lambda(t_{z,n})|>\sqrt{n}\epsilon s_{t_{z,n}}} (\psi(x,t_{z,n}) - \lambda(t_{z,n}))^2 dF(x) = 0. \tag{5.122}$$

どんな $\eta > 0$ に対しても n を十分大きくとれば，

$$\psi(x,\theta_0+\eta) - \lambda(\theta_0-\eta) \le \psi(x,t_{z,n}) - \lambda(t_{z,n}) \le \psi(x,\theta_0-\eta) - \lambda(\theta_0+\eta)$$

より，$u(x) = \max\{|\psi(x,\theta_0+\eta) - \lambda(\theta_0-\eta)|, |\psi(x,\theta_0-\eta) - \lambda(\theta_0+\eta)|\}$ とおけば，$n \to \infty$ のとき

$$\int_{|\psi(x,t_{z,n})-\lambda(t_{z,n})|>\sqrt{n}\epsilon s_{t_{z,n}}} (\psi(x,t_{z,n})-\lambda(t_{z,n}))^2 dF(x) \le \int_{u(x)>\sqrt{n}\epsilon s_{t_{z,n}}} (\psi(x,t_{z,n})-\lambda(t_{z,n}))^2 dF(x) \to 0,$$

となることが，

$$s^2_{z,n} \le \mathbb{E}[\psi^2(X,\,\theta_0-\eta)] - \min(\lambda^2_F(\theta_0-\eta),\,\lambda^2_F(\theta_0+\eta)) < \infty$$

からわかり，(5.122) が示せた．よって，定理は証明された．

・・・・・5.9・・・・・

最大経験尤度推定量

$X_1, \dots, X_n$ を $G \in \mathcal{F}$ からのランダム標本とし，P_G を G に対応する確率測度とする．$X_1 = x_1, \dots, X_n = x_n$ が与えられたとき，**ノンパラメトリック尤度関数** $l : \mathcal{F} \longrightarrow [0, \infty)$ を

$$l(G) = \Pi_{i=1}^n P_G\{x_i\} \tag{5.123}$$

で定義する．

定理 5.14　：$X_1, \dots, X_n$ を $F \in \mathcal{F}$ からのランダム標本とする．このとき，

$$l(F_n) = \sup_{G \in \mathcal{F}} l(G)$$

が成立する．

証明：$P_G\{x_i\} = 0$ ならば，$l(G) = 0$ なので，$l(G) > 0$ なる $G \in \mathcal{F}$ のみを考える．$c \in (0, 1]$ とし，

$$\mathcal{F}_c = \{G \in \mathcal{F} : p_i = P_G\{x_i\} > 0, \sum_{i=1}^n p_i = c\}$$

とする．最大値は存在するので，Lagrange multiplier 法を用いる：

$$H(p_1, \dots, p_n, \lambda) = \Pi_{i=1}^n p_i + \lambda(\sum_{i=1}^n p_i - c)$$

とおく．$p_1, \dots, p_n, \lambda$ に関して微分すれば，

$$\frac{\partial H}{\partial \lambda} = \sum_{i=1}^n p_i - c = 0, \qquad \frac{\partial H}{\partial p_i} = \frac{1}{p_i}\Pi_{i=1}^n p_i + \lambda = 0, \qquad (i = 1, 2, \dots, n)$$

を得る．これを解けば，

$$p_i = \frac{c}{n}, \qquad \lambda = -\left(\frac{c}{n}\right)^{n-1}, \qquad (i = 1, 2, \dots, n)$$

となる．

$$\max_{G \in \mathcal{F}_c} l(G) = \left(\frac{c}{n}\right)^n$$

となり，これは $c = 1$ のとき最大となる．したがって，F_n は $l(G)$ を最大にする．　□

F について補助的な情報があると仮定する．たとえば，ある既知の関数 $u : \boldsymbol{R}^d \longrightarrow \boldsymbol{R}^s$ が存在して，

$$\int u(x)dF(x) = 0$$

を満足するとする．このとき，F の推定量 $\hat{F}_n$ は

$$\int u(x)d\hat{F}_n(x)$$

を満足するのが自然であろう．しかし，経験分布 F_n は一般に

$$\int u(x)dF_n(x) \neq 0$$

である．(5.123) を修正するひとつの方法として，Owen (1988, 1990) による**経験尤度**がある．経験尤度とは (5.123) の最大化の過程で新たな制約を加えて最大化をする：すなわち，

$$p_i > 0, \qquad \sum_{i=1}^n p_i = 1, \qquad \sum_{i=1}^n p_i u(x_i) = 0$$

のもとで $l(G)$ を最大化する．ただし，$p_i = P_G\{x_i\}$ である．これを最大にする推定量を**最大経験尤度推定量**（MLME）とよぶ．

前定理の証明と同様の議論により F の MLME は

$$\hat{F}_n(t) = \sum_{i=1}^n \hat{p}_i \mathbf{1}_{(-\infty,t]}(x_i) \tag{5.124}$$

で与えられる．ただし，

$$\hat{p}_i = \frac{1}{n[1+u^T\lambda_i]}, \qquad i = 1, \ldots, n \tag{5.125}$$

で λ_i は

$$\sum_{i=1}^n \hat{p}_i u(x_i) = \sum_{i=1}^n \frac{u(x_i)}{n[1+u^T(x_i)\lambda_i]} = 0 \tag{5.126}$$

を満足する． $u = 0$ のとき，$\hat{F}_n$ は F_n に一致する．

定理 5.15 　：以下を仮定する．
(E1)

$$\boldsymbol{var}(u(X)) < \infty, \qquad \boldsymbol{E}\left|\frac{u_j(X_1)}{1+u_j(X_1)}\right| < \infty, \qquad (j = 1, 2, \ldots, s)$$

(E2) $\boldsymbol{E}|h(X_1)| < \infty$ なる関数で

$$\sup_{\gamma \in B_n(c)} \left\| \frac{u(x)u^T(x)}{(1+u^T(x)\gamma)^2} \right\| < h(x)$$

であるものが存在する．
このとき，漸近的には，(5.126) の解は存在する．

証明：$\boldsymbol{E}u(X_1) = 0$ から

$$\boldsymbol{E}\left[\frac{u(X_1)}{1+u^T(X_1)\lambda}\right] = 0$$

の解は $\lambda = 0$ となることに注意する．いま，$c > 0$ に対し，

$$B_n(c) = \{r : \|\gamma[\boldsymbol{var}(u(X_1))]^{1/2}\| \leq \frac{c}{n}\}$$

とおく．$n \uparrow \infty$ のとき，$B_n(c) \to 0$ となるので，どんな正の数 $\epsilon > 0$ に対してもある正の整数 N が存在して，$n > N$ ならば，どんな $\gamma \in \partial B_n(c)$[1)]に対しても

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\log(1+u^T(X_i)\gamma) < 0\right) \geq 1-\epsilon$$

が成立することを示せばいい．$\log(1+u^T(x)\gamma)$ を原点のまわりで展開：

$$\frac{1}{n}\sum_{i=1}^{n}\log(1+u^T(X_i)\gamma) = -\tilde{\gamma}^T\sum_{i=1}^{n}\frac{u(X_i)u^T(X_i)}{1+u^T(X_i)\gamma}\tilde{\gamma}$$

となる．ただし，$\gamma \in B_n(c)$ である．したがって，定理 2.1 をもちいれば，

$$\sup_{\gamma\in B_n(c)}\left\|\frac{1}{n}\sum_{1=1}^{n}\frac{u(X_1)u^T(X_1)}{(1+u^T(X_1)\gamma)^2} - \boldsymbol{var}(u(X_1))\right\| \xrightarrow{as} 0$$

となる．したがって，

$$\frac{1}{n}\sum_{i=1}^{n}\log(1+u^T(X_1))\gamma) = -\gamma^T[\boldsymbol{var}(u(X_1))]\gamma + o_P(1)$$

より，$n \uparrow \infty$ のとき，

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\log(1+u^T(X_i)\gamma) < 0\right) \to 1$$

を得る．よって，

$$P\left(\sum_{i=1}^{n}\frac{u(X_i)}{n[1+u(X_i)^T\lambda_n]}\right) \to 1 \tag{5.127}$$

と

$$\lambda_n \xrightarrow{P} 0$$

が成り立つことがわかる．□

定理 5.16　：$X_1, \ldots, X_n$ を $\boldsymbol{R}^d$ 上の分布関数 F からのランダム標本とし，$\int u(x)dF(x) = 0$ を満足するとする．$\hat{F}_n$ は (5.124)−−(5.126) によって与えられるとする．さらに，$U = \boldsymbol{var}(u(X_1))$ は正定置とする．このとき，固定した異なる $t_1, \ldots, t_m \in \boldsymbol{R}^d$ に対して，

$$\sqrt{n}[(\hat{F}_n(t_1), \ldots, \hat{F}_n(t_m)) - (F(t_1), \ldots, F(t_m))] \xrightarrow{L} N(0, \Sigma_u)$$

1) $\partial B_n(c)$ は $B_n(c)$ の境界．

が成り立つ．ただし，

$$\Sigma_u = \Sigma - W^T U^{-1} W, \qquad W = (W^T(t_1), \ldots, W^T(t_m)), \qquad W(t_j) = \boldsymbol{E}[u(X_i) 1_{(-\infty, t_j]}(X_j)]$$

で Σ は (5.105) で与えられる．

証明：$m = 1$ の場合のみを示す．

$$\bar{u} = \frac{1}{n} \sum_{i=1}^{n} u(X_i)$$

とおく．(5.125)，(5.127) とテーラー展開から

$$0 = \sum_{i=1}^{n} \frac{u(X_i)}{n[1 + u^T(X_i)\lambda_n]} = \frac{1}{n} \sum_{i=1}^{n} \left\{ u(X_i) - u(X_i)u^T(X_i)\lambda_n + o_P(\|\lambda_n\|) \right\}$$

を得る．これより

$$\bar{u} = \frac{1}{n} \sum_{i=1}^{n} u(X_i)u^T(X_i)\lambda_n + o_P(\|\lambda_n\|)$$

を得る．また，大数の法則と中心極限定理から

$$\frac{1}{n} \sum_{i=1}^{n} u(X_i)u^T(X_i) \xrightarrow{as} U$$

と

$$O_P(1) = \sqrt{n}\bar{u} = \frac{1}{n} \sum_{i=1}^{n} u(X_i)u^T(X_i)\sqrt{n}\lambda_n + o_P(\|\lambda_n\|)$$

を得る．これらより

$$o_P(\sqrt{n}\|\lambda_n\|) = o_P(O_P(1)) = o_P(1)$$

がわかり，

$$o_P(\|\lambda_n\|) = o_P(n^{-1/2})$$

となる．したがって，

$$U^{-1}\bar{u} = \lambda_n + o_P(n^{-1/2})$$

を得る．これを用いれば，

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,t]}(X_i)(n\hat{p}_i - 1) &= \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,t]}(X_i) \left[\frac{1}{1 + u^T(X_i)\lambda_n} - 1 \right] \\
&= -\frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,t]}(X_i)u^T(X_i)\lambda_n + o_P(n^{-1/2}) \\
&= -W^T(t)\lambda_n + o_P(n^{-1/2}) \\
&= -W^T(t)U^{-1}\lambda_n + o_P(n^{-1/2})
\end{aligned}$$

となる．したがって，

$$\begin{aligned}\hat{F}_n(t) - F(t) &= F_n(t) - F(t) + \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i)(n\hat{p}_i - 1) \\ &= \hat{F}_n(t) - F(t) - W^T(t)U^{-1}\bar{u} + o_P(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[1_{(-\infty,t]}(X_i) - F(t) - W^T(t)U^{-1}u(X_i) + o_P(n^{-1/2})\right]\end{aligned}$$

を得る．また，

$$\begin{aligned}&\boldsymbol{cov}(1_{(-\infty,t]}(X_i), W^T(t)U^{-1}u(X_i)) \\ &\quad = \boldsymbol{E}[1_{(-\infty,t]}(X_i)W^T(t)U^{-1}u(X_i)] - \boldsymbol{E}[1_{(-\infty,t]}(X_i)]\boldsymbol{E}[W^T(t)U^{-1}u(X_i)] \\ &\quad = W^T(t)U^{-1}W(t) = W^T(t)U^{-1}UU^{-1}W(t) \\ &\quad = W^T(t)U^{-1}\{\boldsymbol{var}u(X_1)\}U^{-1}W(t) = \boldsymbol{var}(W^T(t)U^{-1}u(X_1))\end{aligned}$$

となることから中心極限定理を用いれば，定理は証明される．□

·····5.10·····

一般化推定方程式

$X_1, X_2, \ldots, X_n$ は独立な確率変数[1]とし、各 $X_i (i = 1, 2, \ldots, n)$ の次元を d_i とする．母集団に関する未知の母数 θ に興味があるとする．

母数空間 Θ を $\boldsymbol{R}^k$ の部分集合とし、$\psi_i : \boldsymbol{R}^{d_i} \times \Theta \to \boldsymbol{R}^k, i = 1, 2, \ldots, n,$ をボレロ関数とし、

$$s_n(\gamma) = \sum_{i=1}^{n} \psi_i(X_i, \gamma), \qquad \gamma \in \Theta$$

とおく．θ を方程式 $s_n(\hat{\theta}) = 0$ の解 $\hat{\theta}$ で推定するとき、$\hat{\theta}$ を **GEE 推定量**（generalized estimating equation esitmator）と呼ぶことにする．

通常、GEE は

$$\mathbb{E}[s_n(\theta)] = \sum_{i=1}^{n} \mathbb{E}[\psi_i(X_i, \theta)] = 0 \tag{5.128}$$

となるようにとる．したがって、$\hat{\theta}$ は $\mathbb{E}[s_n(\theta)] = 0$ の標本版である $s_n(\hat{\theta}) =$ の解とみなすことができる．

5.10.1 一致性

定理 5.17　：$X_1, X_2, \ldots, X_n$ は独立とし、θ は一次元とする．以下を仮定する．

(G1) $\psi_i(x, \gamma)$ は実数値関数とし、すべての i に対して、γ の非増加関数とする．

(G2) ある $\delta > 0$ が存在して、θ の近傍 N_θ に含まれるすべての γ に対して、

$$\mathbb{E}|\psi_i(X_i, \gamma)|^{1+\delta} < \infty$$

が成立する[2]．

(G3) $\psi_i(x, \gamma)$ は N_θ で連続とする．

(G4) (5.128) が成立する．

(G5) 任意の $\epsilon > 0$ に対して、

$$\limsup_n \mathbb{E}[\Psi_n(\theta + \epsilon)] < 0 < \liminf_n \mathbb{E}[\Psi_n(\theta - \epsilon)]$$

が成立する．ただし、$\Psi_n(\gamma) = n^{-1} s_n(\gamma)$ である．

このとき、$\{\hat{\theta}_n\}$ を $s_n(\hat{\theta}_n) = 0$ を満たす任意の列とすれば、

$$\hat{\theta}_n \xrightarrow{P} \theta$$

---

[1] 必ずしも同一分布に従う必要はない

[2] *i.i.d.* 標本のとき、$\psi_i = \psi$ とすれば、$\mathbb{E}|\psi(X_1, \gamma)| < \infty$ となる．

が成立する．

証明：ψ_i は非増加から、$\Psi_n(\gamma)$ と $\mathbb{E}[\Psi_n(\gamma)]$ も非増加である．任意の $\epsilon > 0$ を固定し、$\theta \pm \epsilon \in N_\theta$ とする．このとき、弱大数の法則から

$$\Psi_n(\theta \pm \epsilon) - \mathbb{E}[\Psi_n(\theta \pm \epsilon)] \xrightarrow{P} 0$$

となる．また、$\Psi_n(\gamma)$ は非増加より、

$$P(\Psi_n(\theta + \epsilon) < 0 < \Psi_n(\theta - \epsilon)) < P(\theta - \epsilon < \hat{\theta} < \theta + \epsilon)$$

となる．さらに、(G5) から

$$P(\theta - \epsilon < \hat{\theta} < \theta + \epsilon) \xrightarrow{P} 0$$

となるので、一致性は示せた．□

一致性に関する別の形の定理を導出するためにつぎの補題[3)]を示す． そのために、つぎの概念を導入する． $\boldsymbol{R}^k$ から $\boldsymbol{R}^k$ への関数の列 $\{g_i\}$ が $\boldsymbol{R}^k$ の開部分集合 $\mathcal{O}$ 上で**同程度に連続**（equicontinuos）であるとは、どんな正の数 $\epsilon > 0$ に対しても適当に $\delta_\epsilon > 0$ をとって、$|t - s| < \delta_\epsilon (t, s \in \mathcal{O})$ である限り、$\sup_i |g_i(t) - g_i(s)| < \epsilon$ が成り立つことである．

補題 5.9　：Θ を $\boldsymbol{R}^k$ のコンパクトな部分集合とし、

$$h_i(X_i) = \sup_{\gamma \in \Theta} \|\psi_i(X_i, \gamma)\|, \qquad i = 1, 2, \ldots$$

とおく．このとき、以下を仮定する．
(G6) ある正の数 $\delta > 0$ が存在して、

$$\sup_i \mathbb{E}|h_i(X_i)|^{1+\delta} < \infty 0, \qquad \sup_i \mathbb{E}\|X_i\|^\delta < \infty \quad (i = 1, 2, \ldots)$$

が成立する[4)]とする．
(G7) どんな正の数 $c > 0$ に対しても、数列 $\{x_i\}$ が $|x_i| \le c$ を満足する限り、関数列 $\{g_i(\gamma) = \psi_i(X_i, \gamma)\}$ は Θ の開集合上で同程度に連続であるとする．
このとき、

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \{\psi_i(X_i, \gamma) - \mathbb{E}[\psi_i(X_i, \gamma)]\} \right\| \xrightarrow{P} 0$$

が成立する．

証明：一般性を失うことなく、ψ_i を実数値関数として証明すればよい．どんな正の数 $c > 0$ に対しても、

$$\sup_n \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} h_i(X_i) 1_{(c,\infty)}(\|X_i\|)\right] \le \sup_i \mathbb{E}[h_i(X_i) 1_{(c,\infty)}(\|X_i\|)]$$

---

3) 定理 2.1 の拡張とみなせる．

4) *i.i.d.* の場合に $\psi_i = \psi$ とすれば、$\mathbb{E}|h(X_1)| < \infty$ でよい．

である．いま、

$$c_0 = \sup_i \mathbb{E}(h_i(X_i))^{1+\delta}, \qquad c_1 = \sup_i \mathbb{E}\|X_i\|^{\delta}$$

とおく．Hölder の不等式と Markov の不等式から

$$\mathbb{E}[h_i(X_i)1_{(c,\infty)}(\|X_i\|)] \le c_0^{1/(1+\delta)} c_1^{\delta/(1+\delta)} c^{-\delta^2/(1+\delta)}$$

がすべての i について成立すること[5] がわかる．ここで、任意の正の数 ϵ と $\tilde{\epsilon}$ に対して、

$$c_0^{1/(1+\delta)} c_1^{\delta/(1+\delta)} c^{-\delta^2/(1+\delta)} < \epsilon\tilde{\epsilon}/2$$

を満足するように c を取る．このとき、任意の開集合 $\mathcal{O} \subset \Theta$ に対して、

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)\right)1_{(c,\infty)}(\|X_i\|) > \epsilon\right\} \le \tilde{\epsilon} \tag{5.129}$$

となる[6]．また、$\|x_i\| \le c$ ならば、$\{\psi_i(x_i,\gamma)\}$ は同程度連続性であることから、ある正の数 $\delta_\epsilon > 0$ が存在して、半径 δ_ϵ の開球 $\mathcal{O}_\delta$ に対して、n を十分大きく取れば、

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)\right)1_{[0,c]}(\|X_i\|) < \frac{\epsilon}{2}$$

とできる．ここで、

$$E_n = \left\{\omega : \frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)\right)1_{[0,c]}(\|X_i\|) < \frac{\epsilon}{2}\right\}$$

$$F_n = \left\{\omega : \frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \mathbb{E}[\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)]\right) > \epsilon\right\}$$

とおく．このとき、

$$\begin{aligned}
P(F_n) &\le P(E_n \cap F_n) + P(E_n^c) \le P(E_n \cap F_n) + \tilde{\epsilon} \\
&\le P\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \mathbb{E}[\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)]\right) > \frac{\epsilon}{2}\right\} + \tilde{\epsilon} \\
&\le 2\tilde{\epsilon}
\end{aligned}$$

---

[5] これは、

$$\begin{aligned}
\mathbb{E}[h_i(X_i)1_{(c,\infty)}(\|X_i\|)] &\le \left[\mathbb{E}|h(X_i)|^{1+\delta}\right]^{1/(1+\delta)}\left[\{P(1_{(c,\infty)}(\|X_i\|))\}^{1-1/(1+\delta)}\right] \\
&\le \left[\mathbb{E}|h(X_i)|^{1+\delta}\right]^{1/(1+\delta)}[\{P(\|X_i\| > c)\}]^{\delta/(1+\delta)} \\
&\le c_0^{1/(1+\delta)} c_1^{\delta/(1+\delta)} c^{-\delta^2/(1+\delta)}
\end{aligned}$$

からわかる．

[6] なぜなら、

$$\begin{aligned}
&P\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)\right)1_{(c,\infty)}(\|X_i\|) > \epsilon\right\} \\
&\le \frac{1}{\epsilon}\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)\right)1_{(c,\infty)}(\|X_i\|)\right| \\
&\le \frac{1}{\epsilon}\left\{\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)1_{(c,\infty)}(\|X_i\|)\right| + \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\right.\right. \\
&\le \frac{2}{\epsilon}\sup_n \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}h(X_i)1_{(c,\infty)}(\|X_i\|)\right] \le \tilde{\epsilon}
\end{aligned}$$

と弱大数の法則よりできることがわかる．したがって、

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)-\mathbb{E}[\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)]\right)>\epsilon\right\}\to 0 \tag{5.130}$$

となる．いま、

$$H_n(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\{\psi_i(X_i,\gamma)-\mathbb{E}[\psi_i(X_i,\gamma)]\}$$

とおく．このとき、Fatou の補題から

$$\sup_{\gamma\in\mathcal{O}_\delta}H_n(\gamma)\le\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)-\mathbb{E}[\inf_{\gamma\in\mathcal{O}}\psi_i(X_i,\gamma)]\right)$$

となるので、(5.130) から

$$P(\text{すべての } i \text{ に対して、}H_n(\gamma)>\epsilon)=P(\sup_{\gamma\in\mathcal{O}_\delta}H_n(\gamma)>\epsilon)\to 0$$

となる．同様に、

$$P(\sup_{\gamma\in\mathcal{O}_\delta}H_n(\gamma)<-\epsilon)\to 0$$

となることがわかる． Θ はコンパクトなので、ある有限の m_δ 個の開球 $\mathcal{O}_{\delta,j}$ が存在して、$\Theta\subset\cup_{j=1}^{m_\epsilon}\mathcal{O}_{\delta,j}$ とできる．これより、

$$P(\sup_{\gamma\in\Theta}|H_n(\gamma)|>\epsilon)<\sum_{j=1}^{m_\delta}P(\sup_{\gamma\in\ \mathcal{O}_{\delta,j}}|H_n(\gamma)|>\epsilon)\to 0$$

がわかり、補題は証明された． □

定理 5.18 ：(G4)、(G6)、(G7) が成立すると仮定する．さらに、関数 $\Delta_n(\gamma)=\mathbb{E}[n^{-1}s_n(\gamma)]$ はつぎの性質を持つとする．
(G8) $\lim_{n\to\infty}\Delta_n(\gamma)=0$ が成立するために必要十分条件は $\gamma=0$ である．
このとき、$\{\hat{\theta}_n\}$ は GEE 推定量の列で、$\hat{\theta}_n=O_P(1)$ ならば、

$$\hat{\theta}_n\xrightarrow{P}\theta$$

が成立する．

証明：まず、Θ がコンパクトの場合に定理が成立することを示す．補題 3.6 から

$$s_n(\hat{\theta}_n)=0,\qquad \Delta_n(\hat{\theta}_n)\xrightarrow{P}0$$

が成立する．すると、ある部分列 $\{n_j\}$ が存在し、

$$\Delta_{n_j}(\hat{\theta}_{n_j})\xrightarrow{a.s}0 \tag{5.131}$$

とできる．また、Θ はコンパクトだから、$\{\theta_{n_j}\}$ の中からある点 θ_0 に収束する部分列 $\{\theta_{m_j}$ $(\{m_j\} \subset \{n_j\})$ を選べだすことができる．一方、(5.129) から

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\left\|\sup_{\gamma\in\mathcal{O}_\epsilon}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_\epsilon}\psi_i(X_i,\gamma)\right\| 1_{(c,\infty)}(\|X_i\|)\right\} \le \tilde{\epsilon}$$

とできることと $\|x_i\| \le c$ 上では $\{\psi_i(X_i,\gamma)\}$ は同程度連続であるに注意すれば、

$$\sup_{|\gamma-\gamma'|\le\delta_\epsilon}\left\|\frac{1}{n}\sum_{i=1}^{n}\{\mathbb{E}[\psi_i(X_i,\gamma)] - \mathbb{E}[\psi_i(X_i,\gamma')]\}\right\| \xrightarrow{P} 0$$

となること[7] が、補題 3.6 よりわかる．さらなる部分列 $\{k_j\} \subset \{m_j\}$ をとれば、$\Delta_{k_j}(\gamma)$ は同程度連続になるので、

$$\Delta_{k_j}(\hat{\theta}_{k_j}) - \delta_{k_j}(\theta_0) \to 0$$

となり、(G8) から $\theta_0 = \theta$ となる．これより、

$$\hat{\theta}_n \xrightarrow{P} \theta$$

がわかる．

つぎに、Θ がコンパクトでない場合を考える．どんな正の数 $\epsilon > 0$ に対しても、ある $M_\epsilon > 0$ が存在して、

$$P(\|\hat{\theta}_n\| \le M_\epsilon) > 1 - \epsilon$$

とできる．よって、前段の証明を $\Theta \cap \{\gamma : \|\gamma\| \le M_\epsilon\}$ の閉包に適用すれば、定理は証明される． □

5.10.2 漸近正規性

定理 5.19　：$\varphi_i = \partial\psi_i(x,\gamma)/\partial\gamma$ は存在するとし、$\varphi_i = (\varphi_i^{(j)})$ とおき、N_θ を θ の近傍とする．関数列 $\{\varphi_i^{(j)}\}$ は以下を満足すると仮定する．

(G6') ある正の数 $\delta > 0$ が存在して、

$$\sup_i \mathbb{E}|h_i(X_i)|^{1+\delta} < \infty (i = 1,2,\ldots), \qquad \sup_i \mathbb{E}\|X_i\|^\delta < \infty,$$

が成立[8]する．ただし、

$$h_i(X_i) = \sup_{\gamma\in N_\theta}\|\varphi_i(X_i,\gamma)\|, \qquad i = 1,2,\ldots$$

---

[7] なぜなら、

$$\begin{aligned}
&\sup_{|\gamma-\gamma'|\le\delta_\epsilon}\left\|\frac{1}{n}\sum_{i=1}^{n}\{\mathbb{E}[\psi_i(X_i,\gamma)] - \mathbb{E}[\psi_i(X_i,\gamma')]\}\right\| \\
&\le 2\sup_{\gamma\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\{\psi_i(X_i,\gamma) - \mathbb{E}[\psi_i(X_i,\gamma)]\}\right\| \\
&+ \sup_{|\gamma-\gamma'|\le\delta_\epsilon}\left\|\frac{1}{n}\sum_{i=1}^{n}\{\psi_i(X_i,\gamma) - \psi_i(X_i,\gamma')\}\right\| \\
&+ \sum_{l=1}^{m_\epsilon}\frac{1}{n}\sum_{i=1}^{n}\left\|\sup_{\gamma\in\mathcal{O}_{l,\epsilon}}\psi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_{l,\epsilon}}\psi_i(X_i,\gamma)\right\| 1_{(c,\infty)}(\|X_i\|) \\
&\xrightarrow{P} 0
\end{aligned}$$

であることよりわかる．

[8] *i.i.d.* の場合に $\psi_i = \psi$ とすれば、$\mathbb{E}|h(X_1)| < \infty$ でよい．

とした．
(G7') どんな正の数 $c > 0$ に対しても、数列 $\{x_i\}$ が $|x_i| \leq c$ を満足する限り、関数列 $\{g_i(\gamma) = \varphi_i(X_i, \gamma)\}$ は Θ の開集合上で同程度に連続であるとする．
(G9) ある正の数 $\delta > 0$ が存在し、

$$\sup_i \mathbb{E}\|\psi_i(X, \theta)\|^{2+\delta} < \infty$$

が成立[9]する．
(G10)

$$\liminf_{n\to\infty} \mathrm{ch}_p(n^{-1}\mathbb{VAR}(s_n(\theta))) > 0, \qquad \liminf_{n\to\infty} \mathrm{ch}_p(n^{-1}M_n(\theta)) > 0,$$

が成立する．ただし、$M_n(\theta) = -\mathbb{E}[\nabla s_n(\theta)]$ とし、$\mathrm{ch}_p(A)$ は行列 A の最小固有根とする．このとき、GEE 推定量の列 $\{\hat{\theta}_n\}$ が一致性を持てば、

$$(\hat{\theta} - \theta)V_n^{-1} \xrightarrow{L} N(0, I_p) \tag{5.132}$$

が成立する．ただし、

$$V_n = [M_n(\theta)]^{-1}\mathbb{VAR}(s_n(\theta))[M_n(\theta)]^{-1}$$

である[10]．

証明：$\{\hat{\theta}\}$ の一致性から任意の正の数 $\tau > 0$ をとり、$N_\theta = \{\gamma : \|\gamma - \theta\| \leq \tau\}$ に $\hat{\theta}$ が集中しているとしてよい．(G6') から Hölder の不等式と Markov の不等式を用いて、補題 3.6 の証明中の (5.129) を得る議論を繰りす．半径 δ_δ の開球 $\mathcal{O}_\delta$ と適当に正の数 $c > 0$ を取れば、

$$P\left\{\frac{1}{n}\sum_{i=1}^n\left(\sup_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma)\right)1_{(c,\infty)}(\|X_i\|) > \epsilon\right\} \leq \tilde{\epsilon}$$

とできる．さらに、(G7') より、n を十分大きく取れば、

$$\frac{1}{n}\sum_{i=1}^n\left(\sup_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma)\right)1_{[0,c]}(\|X_i\|) < \frac{\epsilon}{2}$$

とできる．N_θ はコンパクトなので、m_ϵ 個の半径 δ_ϵ の開球 $\mathcal{O}_{\delta,j}$ を有限被覆としてとることができるので、

$$\sup_{\gamma\in N_\theta}\frac{1}{n}\|\nabla s_n(\gamma) - \nabla s_n(\theta)\| = o_P(1) \tag{5.133}$$

となること[11]がわかる．さらに、平均値の定理より、

$$-s_n(\theta) = (\hat{\theta}_n - \theta)\int_0^1 \nabla s_n(\theta + t(\hat{\theta}_n - \theta))dt$$

---

9) $i.i.d.$ の場合に $\psi_i = \psi$ とすれば、$\sup_i \mathbb{E}\|\psi_i(X, \theta)\|^2 < \infty$ でよい．

10) $n^{-1}V_n \to$ ある行列 であるので、(5.132) のオーダーは正しい．

11)

$$P\left\{\sup_{\gamma\in N_\theta}\frac{1}{n}\|\nabla s_n(\gamma) - \nabla s_n(\theta)\| > \frac{m_\epsilon + 1}{2}\right\}$$
$$< \sum_{j=1}^{m_\epsilon}\left[P\left\{\frac{1}{n}\sum_{i=1}^n\left(\sup_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_\delta}\varphi_i(X_i,\gamma)\right)1_{(c,\infty)}(\|X_i\right.\right.$$
$$\left.\left. +P\left\{\frac{1}{n}\sum_{i=1}^n\left(\sup_{\gamma\in\mathcal{O}_{\delta,j}}\varphi_i(X_i,\gamma) - \inf_{\gamma\in\mathcal{O}_{\delta,j}}\varphi_i(X_i,\gamma)\right)1_{[0,c]}(\|X_i\|)\right.\right]$$

からわかる．

とかけ、(5.133) から

$$\frac{1}{n}\left\|\int_0^1 \nabla s_n(\theta+t(\hat{\theta}_n-\theta))dt-\nabla s_n(\theta)\right\| \le \frac{1}{n}\sup_{\gamma\in N_\theta}\|\nabla s_n(\gamma)-\nabla s_n(\theta)\| = o_P(1)$$

となる．また、弱対数の法則から

$$\frac{1}{n}[\nabla s_n(\theta)+M_n(\theta)] = o_P(1)$$

となる．これより、

$$-\frac{s_n(\theta)}{n} = (\hat{\theta}_n-\theta)[\frac{1}{n}\nabla s_n(\theta)+o_P(1)] = (\hat{\theta}_n-\theta)[\frac{1}{n}M_n(\theta)+o_P(1)]$$

と $\mathrm{ch}_p(n^{-1}M_n(\theta))>0$ と $n^{-1}M_n(\theta)<\sup_i \mathbb{E}\|\varphi_i(X_i,\theta)\|<\infty$ に注意すれば、

$$s_n(\theta)[M_n(\theta)]^{-1} = (\hat{\theta}_n-\theta)(1+o_P(1))$$

となる．さらに、

$$\liminf \mathrm{ch}_p(n^{-1}\mathbb{VAR}(s_n(\theta)))>0, \qquad \sup_i \mathbb{E}\|\psi_i(X_i,\theta)\|^{2+\delta}<\infty$$

から任意の零でない横ベクトル $z\in \boldsymbol{R}^p$ に対して、

$$\frac{1}{(zV_nz')^{1+\delta/2}}\sum_{i=1}^n \mathbb{E}|\psi_i(X_i,\theta)[M_n(\theta)]^{-1}z'|^{2+\delta}\to 0$$

となること[12)] から Liapunov 条件を満たすので、

$$\frac{s_n(\theta)[M_n(\theta)]^{-1}z'}{\sqrt{zV_nz'}} \xrightarrow{L} N(0,1)$$

を得る．さらに、$V_n^{1/2}z'/\sqrt{zV_nz'}=\tilde{z}$ とおけば、

$$s_n(\theta)[M_n(\theta)]^{-1}V_n^{-1/2}\tilde{z} \xrightarrow{L} N(0,1)$$

となることから、Wald's device を用いれば、(5.132) がわかる．　□

5.10.3 推定方程式と一致性判定条件

補題 5.10　(Brouwer の不動点定理)　$D=\{\boldsymbol{x}\in\mathbb{R}^p:\|x\|\le 1\}$ とする．関数 $g: D\mapsto D$ を任意の連続写像とするとき，g の不動点 $\boldsymbol{x}_0\in D$ が存在する：

$$g(\boldsymbol{x}_0)=\boldsymbol{x}_0$$

---

12) この式の左辺は $z'=[M_n(\theta)]^{-1}w'$ とおけば、

$$\frac{1}{(w(n^{-1}\mathbb{VAR}(s_n(\theta))w')^{1+\delta/2}}\sum_{i=1}^n \mathbb{E}|\psi_i(X_i,\theta)w'|^{2+\delta}$$
$$\le \frac{1}{n^{\delta/2}}\frac{1}{(w\mathbb{VAR}(s_n(\theta))w')^{1+\delta/2}}\sup_i \mathbb{E}\|\varphi_i(X_i,\theta)\|^{2+\delta}p\|w\|^{2+\delta}$$

と評価できることよりわかる．

証明　リー群論 (杉浦光夫，共立) の p. 442 を参照. □

定理 5.20　　Θ を $\mathbb{R}^k$ の開部分集合とする．$\{\boldsymbol{s}_n(\theta)\}_{n=1}^{\infty}$ を推定方程式の列とし，つぎの条件をみたすと仮定する．
(A1)　ある事象 B 上では，すべての $n \geq 1$ に対し，$\boldsymbol{s}_n(\theta)$ は θ の連続関数である：
(A1)　十分小さな正数 $\delta > 0$ と任意の正数 $\epsilon > 0$ に対し，

$$\lim_{n\to\infty} P\left\{ \sup_{\|\theta-\theta_0\|=\delta} (\theta-\theta_0)'\boldsymbol{s}_n(\theta) < -\epsilon \,\middle|\, B \right\} = 1$$

となる．
　このとき，ある推定量列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ が存在し，

$$\lim_{n\to\infty} P\left\{ \|\hat{\theta}_n - \theta_0\| < \delta \,\middle|\, B \right\} = 1$$

と

$$\lim_{n\to\infty} P\left\{ \boldsymbol{s}_n(\hat{\theta}_n) = \boldsymbol{0} \,\middle|\, B \right\} = 1$$

が成立する．

証明　十分小さな正数 ϵ, δ に対し，

$$B_n = \{\omega \in B : \sup_{\|\theta-\theta_0\|=\delta} (\theta-\theta_0)'\boldsymbol{s}_n(\theta) < -\epsilon\}$$

とおく．まず，$\omega \in B_n$ に対し，$\{\theta : \|\theta-\theta_0\| \leq \delta\}$ 上で $\boldsymbol{s}_n(\theta_n) = \boldsymbol{0}$ の解 $\hat{\theta}_n(\omega)$ が存在することを示す．そのために，$\theta = \theta_0 + \delta\gamma$, $\|\gamma\| le 1$ とおき，すべての $\|\gamma\| \leq 1$ なる γ に対して，

$$\boldsymbol{r}_n(\gamma) = \boldsymbol{s}_n(\theta_0 + \delta\gamma) \neq \boldsymbol{0}$$

と仮定して矛盾することを示す．仮定 (A1) から

$$\boldsymbol{t}_n(\gamma) = \frac{\boldsymbol{r}_n(\gamma)}{\|\boldsymbol{r}_n(\gamma)\|}$$

は γ の連続関数となり，Brouwer の不動点定理からある γ^*, $\|\gamma^*\| \leq 1$ が存在して，$\boldsymbol{t}_n(\gamma^*) = \gamma^*$ を満足し，$\|\boldsymbol{t}_n(\gamma)\| = 1$ から $\|\gamma^*\| = 1$ となる．また，

$$(\gamma^*)'\boldsymbol{t}_n(\gamma^*) = (\gamma^*)'\gamma^* = 1$$

である．しかし，B 上では $\|\gamma\| = 1$ となる γ に対し，$\gamma'\boldsymbol{t}_n(\gamma) < 0$ となるので，矛盾する．したがって，$\boldsymbol{r}_n(\hat{\gamma}_n) = \boldsymbol{0}$ となる解 $\hat{\gamma}_n$ が $\{\gamma : \|\gamma\| \leq 1\}$ 上に存在し，(A2) から明らかに $\|\hat{\gamma}_n\| \neq 1$ であるので，$\|\gamma\| < 1$ となる．よって，$\boldsymbol{s}_n(\theta) = \boldsymbol{0}$ の解 $\hat{\theta}_n(\omega)$ で $\|\hat{\theta}_n - \theta_0\| < \delta$ なるものが存在する．つぎに $\omega \in B_n^c$ に対して，$\hat{\theta}(\omega)$ は Θ の任意の点とする．上記の議論と $\hat{\theta}$ の定義より

$$\{\omega : \|\hat{\theta}(\omega) - \theta_0\| < \delta\} \supset B_n$$

となるので，$\{\omega : \|\hat{\theta}(\omega) - \theta_0\| \geq \delta\} \supset B_n^c$ となる．したがって，

$$P\left\{\|\hat{\theta}_n - \theta_0\| < \delta \,\middle|\, B\right\} \geq P(B_n \mid B) \to 1 \qquad (n \to \infty),$$

となる．また，$\omega \in B_n$ に対し，$\boldsymbol{s}_n(\hat{\theta}_n(\omega)) = \mathbf{0}$ より

$$P\left\{\boldsymbol{s}_n(\hat{\theta}_n(\omega)) = \mathbf{0}\right\} \geq P(B_n \mid B) \to 1 \qquad (n \to \infty)$$

がわかる． □

参考文献

1. Aitchison, J. and Silvey, S.D. (1967). Maximum-likelihood estimation of parameters subject to restrains. AMS **29**, 813-823.
2. Heyde, C.C. (1997). *Quasi-Likelihood and its application*, Springer.

·····5.11·····

確率密度関数の推定

分布関数 F が確率密度関数 f を持つとき，与えられた点 x における $f(x)$ の推定を考える．ここで，f は x において連続で $f(x) > 0$ とする．

$X_1, X_2, \dots, X_n$ を F からのランダム標本とし，F_n をそれらに基づく経験分布関数とする．

$$f(x) = \lim_{h \to \infty} \frac{F(x+h) - F(x-h)}{2h}$$

から

$$f_n(x) = \lim_{h \to \infty} \frac{F_n(x+h) - F_n(x-h)}{2h}$$

は $f(x)$ のひとつの推定量となることが期待される．

$f_n(x)$ が確率密度関数であることをまず示す．そのために，

$$\int_{-\infty}^{\infty} f_n(x)\,dx = 1$$

を示せば十分である．実際，

$$f_n(x) = \frac{1}{2nh} \sum_{j=1}^{n} 1\{x - h < X_j \leq x + h\} \tag{5.134}$$

と書けることより，

$$\int_{-\infty}^{\infty} f_n(x)\,dx = \frac{1}{2nh} \sum_{j=1}^{n} \int_{X_j - h}^{X_j + h} dx = 1$$

からわかる．

また，$f_n(x)$ の基本的な性質は $n[F_n(x+h) - F_n(x-h)]$ が二項分布 $Bi(p, n)$ に従うことから容易に導くことができる．ここで，$p = F(x+h) - F(x-h)$ とおく．これより，

$$\mathbb{E}[f_n(x)] = \frac{F(x+h) - F(x-h)}{2h} = \frac{p}{2h}$$

となる．したがって，

$$\boldsymbol{bias}(f_n(x)) = \frac{F(x+h) - F(x-h)}{2h} - f(x)$$

となり，

$$h = h_n \downarrow 0, \qquad (n \uparrow \infty) \tag{5.135}$$

ならば，$\boldsymbol{bias}(f_n(x)) \downarrow 0$ となる．同様に，

$$\mathbb{VAR}(f_n(x)) = \frac{p(1-p)}{4nh^2}$$

を得る．$h_n \downarrow 0$ のとき，$p_n = F(x+h_n) - F(x-h_n) \downarrow 0$ より，

$$\mathbb{VAR}(f_n(x)) \approx \frac{p_n}{2h_n} \cdot \frac{1}{2nh_n}$$

となる．$n \uparrow \infty$ のとき，$p_n/2h_n \to f(x) > 0$ から，$h_n \downarrow 0$ のとき，

$$nh_n \to \infty \tag{5.136}$$

ならば，$\mathbb{VAR}[f_n(x)] \to 0$ が成り立つ．

定理 5.21　：$f_n(x)$ は $f(x)$ の一致推定量となるために十分条件は (5.135) と (5.136) である．

証明：定理 1.5 を用いればよい．□

$\hat{f}_n(x)$ は連続ではない．そこで，(5.134) の中の一様分布 $U(X_j - h, X_j + h)$ を連続型分布の確率密度関数

$$\frac{1}{h}K\left(\frac{x - X_j}{h}\right)$$

で置き換えると $f(x)$ の新たな推定量として，

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n}K\left(\frac{y - X_j}{h}\right)$$

を得る．これを**核 K を持つ核型推定量**（Kernel estimator）とよぶ．核 K は

$$K(x) > 0, \quad \int K(x)dx = 1, \quad K(x) = K(-x) \quad (x > 0)$$

を満たす．$\hat{f}_n(x)$ が確率密度関数であることは，f_n と同様にただちにわかる．

定理 5.22　：f は3回連続微分可能で x の近傍で3階導関数は有界とする．K は原点に対して対称な核とし

$$\int K^2(x)\,dx < \infty, \quad \int x^2K(x)\,dx < \infty, \quad \int |x|^3K(x)\,dx < \infty$$

を満たす．
(i) (5.135) を満足する任意の列 $\{h_n\}_{n=1}^{\infty}$ に対して，

$$\boldsymbol{bias}(\hat{f}_n(x)) = \frac{1}{2}h_n^2\ddot{f}(x)\int x^2K(x)\,dx + o(h_n^2) \tag{5.137}$$

が成立する．
(ii) h_n が (5.136) をさらに満たせば，

$$\mathbb{VAR}(\hat{f}_n(x)) = \frac{1}{nh_n}f(x)\int K^2(x)dx + o(\frac{1}{nh_n}) \tag{5.138}$$

が成立する．

証明：(i) $\hat{f}_n$ の定義から

$$\begin{aligned}\mathbb{E}[\hat{f}_n(x)] &= \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right] = \frac{1}{n}\sum_{j=1}^{n}\int\frac{1}{h_n}K\left(\frac{x-t}{h_n}\right)f(t)\,dt \\ &= \frac{1}{n}\sum_{j=1}^{n}\int K(z)f(x-h_nz)\,dz \end{aligned} \tag{5.139}$$

となる．一方，Taylor 展開より，

$$f(x-h_nz) = f(x) - h_nz\dot{f}(x) + \frac{1}{2}h_n^2z^2\ddot{f}(x) + \frac{1}{6}h_n^3z^3f^{(3)}(\xi)$$

となる．ただし，ξ は x と $x-h_nz$ の間にある数である．$\int K(z)\,dz = 1$ と $\int zK(z)\,dz = 0$ に注意すれば，(5.139) の最右辺は

$$f(x) + \frac{1}{2}h_n^2\ddot{f}(x)\int z^2K(z)\,dz + R_n$$

となる．ただし，$|f^{(3)}(z)| \leq M$ と書けば，

$$|R_n| \leq \frac{Mh_n^3}{6}\int|z|^3K(z)dz = o(h_n^2)$$

である．よって，(i) は示された．
(ii) (5.134) と (5.137) から

$$\begin{aligned}n\mathbb{VAR}[\hat{f}_n(x)] &= \mathbb{VAR}\left[\frac{1}{k}K\left(\frac{x-X_1}{h_n}\right)\right] \\ &= \int\frac{1}{h_n^2}K^2\left(\frac{x-t}{h_n}\right)f(t)\,dt - \left\{\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{x-X_1}{h_n}\right)\right]\right\}^2 \\ &= \frac{1}{h_n}\int K^2(z)\big[f(x) - h_nz\dot{f}(x) + \frac{1}{2}h_n^2z^2f^{(2)}(\xi)\big]\,dz + O(1)\end{aligned}$$

となる．ただし，ξ は x と $x-h_nz$ の間にある数である．(5.135) と (5.136) から，$n\uparrow\infty$ のとき，

$$n\mathbb{VAR}[\hat{f}_n(x)] = \frac{1}{h_n}f(x)\int K^2(z)dz + O(h_n) + o(1)$$

を得る．したがって，(ii) は示された[1)]． □

系 5.3　：定理の仮定のもと，

$$\mathbb{E}[\hat{f}_n(x) - f(x)]^2 \to 0, \qquad (n\to\infty)$$

が成り立つ．したがって，$\hat{f}_n(x)$ は $f(x)$ の一致推定量である．

[1)] $n\to\infty$ のとき，$nh_n\to\infty$ となるので，$h_n\to 0$ である．これらより $O(h_n) = o(1)$ と $o(n^{-1}) = o((nh_n)^{-1})$ となる．

証明：上の定理と定理 1.5 からわかる．□

定理 5.23　：$h_n \to 0$ かつ $nh_n \uparrow \infty$ ならば，

$$\frac{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]}{\sqrt{\mathbb{VAR}[\hat{f}_n(x)]}} \xrightarrow{L} N(0,1)$$

が成り立つ．

証明：定理 1.17 において，

$$Z_i = \frac{1}{h_n} K\left(\frac{x - X_i}{h_n}\right)$$

として，条件 (1.26) を確認すればよい．(5.137) より

$$\sum_{i=1}^{n} \mathbb{VAR}(Z_i) = \frac{n}{h_n} + o(\frac{n}{h_n})$$

を得る．また，

$$\begin{aligned}\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right) - \mathbb{E}\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^3 &\le \sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^3 \\ +3\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^2 \cdot \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right| &+ 4\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^3\end{aligned}$$

となる．

$$\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right| = nO(1)$$

$$\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^2 \cdot \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right| = \frac{n}{h_n}O(1)$$

$$\sum_{j=1}^{n} \mathbb{E}\left|\frac{1}{h_n}K\left(\frac{x-X_j}{h_n}\right)\right|^3 = \frac{n}{h_n^2}O(1)$$

となる．これらより

$$\frac{\{\sum_{i=1}^{n} \mathbb{E}|Z_i - \mathbb{E}(Z_i)|^3\}^2}{\{\sum_{i=1}^{n} \mathbb{VAR}(Z_i)\}^3} = \frac{(n^2/h_n^4)O(1)}{(1+o(1))(n^3/h_n^3)} = \frac{1}{nh_n}O(1) = o(1)$$

となるので定理は証明される．□

6

ブートストラップ

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を分布 $P_\theta \in \boldsymbol{P}$ をもつ母集団からの大きさ n の確率標本とする．ここで、各 $X_i (i = 1, 2, \ldots, n)$ は $\boldsymbol{S}--$ 値確率要素[2) とする．また、モデル $\boldsymbol{P}$ は「母数モデル」であっても「ノンパラメトリックモデル」または「セミパラメトリックモデル」[3)]であってもよい．

いま、母数 $\theta(P)$ の信頼領域の構成に興味があるとしよう．ここで、P を $\boldsymbol{P}$ を動かした時の値域を Θ と記すことにする．すなわち、$\Theta = \{\theta(P) : p \in \boldsymbol{P}\}$ である．ここで、$\hat{\theta}$ を $\theta(P)$ の推定量としたとき、Θ の可測集合 A に対して、

$$J_n(A, P) = \mathrm{Prob}_P\{\sqrt{n}(\hat{\theta} - \theta(P)) \in A\}$$

の推定問題を考える．

[2)] 確率要素（random element）は確率変数と同じと考えてよいが、空間 $\boldsymbol{S}$ が必ずしも Euclid 空間とは限らず一般的なものであるところがあることにだけ注意すればよい．

[3)] ここでは、モデルの定義を厳密にはしないが、母数モデルはモデルを添字付ける集合が有限次元のモデルで、ノンパラメトリックモデルはそれが（絶対連続な）確率分布の集合（無限次元）そのものである．また、セミパラメトリックモデルはノンパラメトリックモデルと母数モデルの中間であり、モデルを添字付ける集合の次元は無限次元ものであり、このタイプのモデルでは、モデルを添字付ける集合は Euclid 空間の部分集合と無限次元の集合の直積で表現できる．

·····6.1·····

Efron のノンパラメトリックブートストラップ

$\boldsymbol{S}$ と Θ を Euclid 空間の部分集合とし、P に対応する分布関数を F と記し、F からの大きさ n の確率標本 $X_1, X_2, \ldots, X_n$ に基づく経験分布関数を F_n と記すこと[1)]にする．さらに、

$$J_n(t, F) = \mathrm{Prob}_F\{\sqrt{n}(\hat{\theta} - \theta(F)) \leq t\} \tag{6.140}$$

と記し、 $J_n(t, F)$ の推定を考える．ただし、$\hat{\theta}$ は F からの大きさ n の確率標本 $X_1, X_2, \ldots, X_n$ に基づく $\theta(P)$ の推定量である．

Efron のノンパラメトリックブートストラップ推定量は (6.140) において単純に F を F_n に置き換えることによっていられる $J_n(t, F)$ の推定量である．すなわち、

$$\hat{J}_n(t) = \mathrm{Prob}_{F_n}\{\sqrt{n}(\hat{\theta} - \theta(F_n)) \leq t\}$$

である．ただし、$\hat{\theta}$ は F_n からの大きさ n の確率標本 $X_1^*, X_2^*, \ldots, X_n^*$ に基づく $\theta(F_n)$ の推定量である．

実際には、Monte–Carlo 法により、$\hat{J}_n(t)$ の近似分布をもとめる．すなわち、

$$(X_{1,i}^*, X_{2,i}^*, \ldots, X_{n,i}^*), \qquad i = 1, 2, \ldots, B$$

を F_n からの大きさ n のランダム抽出を B 回繰り返して得られたものとする． $\hat{\theta}_i^*$ を標本 $(X_{1,i}^*, X_{2,i}^*, \ldots, X_{n,i}^*)$ に基づく推定量とすれば、

$$\hat{J}_{n,B}(t) = \frac{1}{B}\sum_{i=1}^{B} 1\{\sqrt{n}(\hat{\theta}_i^* - \theta(F_n)) \leq t\}$$

によって $\hat{J}_n(t)$ を近似する．B が十分大きければ、$\hat{J}_{n,B}$ は真の標本分布 J_n のよい近似となることが期待できる．

---

1) すなわち、$F_n(x) = n^{-1}\sum_{i=1}^{n} 1\{X_i \leq x\}$ である．

·····6.2·····

ノンパラメトリックブートストラップの漸近分布

確率変数 X が分布関数 F をもつとし、$\boldsymbol{E}X^2 < \infty$ とする．このとき、$\theta(F) = \int xdF(x) = \mu(F)$ の推定を考える．$\theta(F)$ の推定量 $\hat{\theta} = \theta(F_n)$ に対し、中心極限定理から、

$$\sqrt{n}(\hat{\theta} - \theta(F)) = \sqrt{n}(\bar{X}_n - \mu(F)) \xrightarrow{L} N(0, \mathbb{VAR}(X))$$

が成立する．これに対してブートストラップ近似がどのようになるかを述べたのがつぎの定理である．

定理 6.1　：$\mathbb{E}X^2 < \infty$ ならば、ほとんどすべての列 $X_1, X_2, \ldots,$ に対して、

$$\sqrt{n}(\theta(F_n^*) - \theta(F_n)) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n) \xrightarrow{L} N(0, \mathbb{VAR}(X))$$

が成立する．

ただし、$\bar{X}_n^*$ と F_n^* は F_n からの大きさ n のランダム標本 $X_{n1}, X_{n2}, \ldots, X_{nn}$ に基づく標本平均と経験分布関数ものである．

証明：$\mathbb{E}_{F_n} X_{ni}^* = \bar{X}_n$ と

$$\mathbb{VAR}_{F_n}(X_{ni}^*) = \frac{1}{n}\sum_{i=1}^{n}(X_{ni} - \bar{X}_n)^2 \equiv S_n^2$$

に注意する．$Z_{ni}^* = n^{-1/2}(X_{ni}^* - \bar{X}_n)$ とおけば、

$$\sqrt{n}(X_n^* - \bar{X}_n) = \sum_{i=1}^{n} Z_{ni}^*, \qquad \mathbb{E}_{F_n} Z_{ni}^* = 0, \qquad \mathbb{VAR}_{F_n}(Z_{ni}^*) = \frac{1}{n}S_n^2 \equiv \sigma_n^2$$

となる．また、強大数の法則から

$$\sum_{i=1}^{n} \mathbb{VAR}_{F_n}(Z_{ni}) = S_n^2 \xrightarrow{a.s} \mathbb{VAR}(X)$$

となるので、あとは Lindeber 条件を Z_{ni}^* が満たすことを示せばよい．そのために、正の数 $\epsilon > 0$ ととる．このとき、

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{n} \mathbb{E}_{F_n}|Z_{ni}^*|^2 1\{|Z_{ni}^* > \epsilon\sigma_n\} \le 1\{\max_{1\le i\le n}|X_i - \bar{X}| > \epsilon\sqrt{n}S_n\}$$

となること[1]がわかる．したがって、$\mathbb{E}X_{ni} - \mu|^2 < \infty$ から

$$\frac{1}{\sqrt{n}}\max_{1\le i\le n}|X_{ni} - \bar{X}_n| \le \frac{1}{\sqrt{n}}\max_{1\le i\le n}|X_{ni} - \mu|\frac{1}{\sqrt{n}}|\bar{X}_n - \mu| \xrightarrow{a.s} 0$$

から Lindeber 条件 (1.27) は確認された．よって、定理は証明された． □

[1] これは、　　　　　　　　　　　　からわかる．

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{n}\mathbb{E}_{F_n}|Z_{ni}^*|^2 \mathbf{1}\{|Z_{ni}^*>\epsilon\sigma_n\} = \frac{n}{S_n^2}\sum_{i=1}^{n}\mathbb{E}_{F_n}\left|\frac{X_{ni}^* - \bar{X}_n}{\sqrt{n}}\right|^2 \mathbf{1}\{|X_{ni}^* - \bar{X}_n| > \sqrt{n}\epsilon\sigma_n\}$$

$$= \frac{1}{S_n^2}\frac{1}{n}\sum_{i=1}^{n}|X_{ni}^* - \bar{X}_n|^2 \mathbf{1}\{|X_{ni}^* - \bar{X}_n| > \sqrt{n}\epsilon\sigma_n\} \le \mathbf{1}\{\max_{1\le i\le n}|X_i - \bar{X}| > \epsilon\sqrt{n}S_n\}$$

7

U 統計量のEdgeworth 展開

·····7.1·····

Edgeworth 展開

7.1.1 記号と導入

F_0，f_0 および ϕ_0 を標準正規分布の分布関数，確率密度関数および特性関数とする．したがって，

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \phi_0(t) = e^{-t^2/2}$$

である．これらは変転公式で関係付けられる：

$$f_0(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-tx\sqrt{-1}} \phi_0(t)\, dt \tag{7.141}$$

f_0 の導関数は以下のように与えられる：

$$\begin{aligned}
&f_0'(x) = -x f_0(x), \qquad f_0''(x) = (x^2-1) f_0(x) \\
&f_0^{(3)}(x) = -(x^3-3x) f_0(x), \qquad f_0^{(4)}(x) = (x^4 - 6x^2 + 3) f_0(x) \\
&f_0^{(5)}(x) = -(x^5 - 10x^3 + 15x) f_0(x), \\
&f_0^{(6)}(x) = (x^6 - 15x^4 + 45x^2 - 15) f_0(x),
\end{aligned}$$

一般には，

$$f_0^{(k)}(x) = (-1)^k H_k f_0(x)$$

で $H_k(x)$ は**エルミート多項式**である．

(7.141) の両辺を微分すれば，

$$(-1)^k H_k(x) f_0(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-tx\sqrt{-1}} (-\sqrt{-1}t)^k \phi_0(t)\, dt$$

となり，$H_k(x) f_0(x)$ は $(t\sqrt{-1})^k \phi_0(t)$ の逆フーリエ変換となる．したがって，

$$(t\sqrt{-1})^k \phi_0(t) \quad \text{は} \quad H_k(x) f_0(x) \quad \text{のフーリエ変換} \tag{7.142}$$

となる．

いま，$X_1, X_2, \ldots X_n$ は平均ゼロで分散が σ^2 の分布 F に独立同一に従うと仮定する．さらに，この分布の特性関数を ϕ とかく．$S_n = (1/n)\sum_{k=1}^{n} X_k$ とし，その分布関数 F_n をエルミート多項式に関して展開することを考える．さらに，密度関数が存在すれば，それを f_n とかくことにする．

7.1.2 密度関数に関する Edgeworth 展開

特性関数 ϕ はある正の正数 m に対し

$$\int_{-\infty}^{\infty} |\phi(t)|^m \, dt < \infty \tag{7.143}$$

を満足するとする．$n \geq m$ に対し，密度関数 f_n の存在を保障する．ここで，X は平均がゼロで分散 σ^2　で特性関数 ϕ をもつとしたとき，

$$\gamma_1 = \mathbb{E}\left(\frac{X}{\sigma}\right)^3, \qquad \gamma_2 = \mathbb{E}\left(\frac{X}{\sigma}\right)^4 - 3$$

とおく．

定理 7.1　：(7.143) を仮定する．このとき，
(i) $\mathbb{E}|X|^3 < \infty$ のならば，$n \to \infty$ としたとき，x に関して一様に

$$|f_n(x) - f_0(x)\{1 + \frac{\gamma_1}{3!\sqrt{n}} H_3(x)\}| = o\left(\frac{1}{\sqrt{n}}\right)$$

が成り立つ．
(ii) $\mathbb{E}|X|^4 < \infty$ のならば，$n \to \infty$ としたとき，x に関して一様に

$$\begin{aligned}
&|f_n(x) - f_0(x)\{1 + \frac{\gamma_1}{3!\sqrt{n}} H_3(x) + \frac{1}{n}[\frac{\gamma_2}{4!} H_4(x) + \frac{\gamma_1^2}{2(3!)^2} H_6(x)]\}| \\
&\quad = o\left(\frac{1}{n}\right)
\end{aligned}$$

が成り立つ．

証明：まず，(i) を示す．一般性を失わず $\sigma = 1$ と仮定できる．f_n と一次近似との差を d_n とおく：

$$d_n = f_n - [1 + \frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} H_3] f_0$$

d_n のフーリエ変換はつぎで与えられる：

$$\begin{aligned}
\phi_n(t) &= \int_{-\infty}^{\infty} e^{-tx\sqrt{-1}} d_n(x) \, dx \\
&= \left(\phi(\frac{t}{\sqrt{n}})\right)^n - \phi_0(t)[1 + \frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} (t\sqrt{-1})^3] \\
&= e^{-t^2/2}\{e^{t^2/2 + n\psi(t/\sqrt{n})} - [1 + \frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} (t\sqrt{-1})^3]\} \\
&= e^{-t^2/2}[e^{z+\epsilon} - (1+z)]
\end{aligned}$$

となる．ただし，$\psi(t) = \log\phi(t)$ と[1)]し，

$$z = \frac{\gamma_1}{3!}\frac{(t\sqrt{-1})^3}{\sqrt{n}}, \qquad \epsilon = \frac{t^2}{2} + n\psi(\frac{t}{\sqrt{n}}) - z$$

とおいた．格子点分布[2)]でなければ，$t \not\emptyset$ に対して $\phi(t) < 1$ と[3)]なり，さらに，密度関数が存在することから Rieman-Lebeague の定理を使えば $t \to \infty$ のとき，$\phi(t) \to 0$ となるので，$\lambda > 0$ と $0 < \theta < 1$ を適当にとり，$|t| > \lambda$ に対して

$$|\phi(t)| < \theta$$

とできる．ここで

$$A = \{|t| > \lambda\sqrt{n}/\mathbb{E}|X|^3\}$$

おく．すると

$$\begin{aligned}\int_A e^{-t^2/2}|e^{z+\epsilon} - (1+z)|\,dt &\le \theta^{n-1}\sqrt{n}\int_{-\infty}^{\infty}|\phi(t)|\,dt + \int_A |\phi_0(t)|(1 + +\frac{\gamma_1}{3!\sqrt{n}}(t\sqrt{-1})^3)|\,dt \\ &= o(n^{-r}) \end{aligned} \tag{7.144}$$

となる．ただし，どんな $r > 0$ である．また，

$$\begin{aligned}|\phi_n(t)| &= e^{-t^2/2}|e^z(e^\epsilon - 1 + (e^z - (1+z))| \\ &\le e^{-t^2/2}[|\epsilon|e^{|\epsilon|}e^{|z|} + \frac{z^2}{2}e^{|z|}]\end{aligned}$$

となること[4)]がわかる．すべての $|t| \le \lambda\sqrt{n}/\mathbb{E}|X|^3$ に対して

$$|z| \le \frac{|t|^3}{6}\frac{\mathbb{E}|X|^2}{\sqrt{n}} \le \frac{\lambda t^2}{6}$$

となる．また，λ を十分小さくとれば，$y\downarrow$ ならば $\delta(y)\downarrow$ なる関数とある定数 c が存在し，

$$\begin{aligned}|\epsilon| &= |n\psi(\frac{t}{\sqrt{n}}) - [-\frac{t^2}{2} + \frac{\gamma_1}{6}\frac{(t\sqrt{-1})^3}{\sqrt{n}}]| \\ &\le \frac{c|t|^3\mathbb{E}|X|^3}{\sqrt{n}}\delta(\frac{\lambda}{\sqrt{n}}) \\ &\le ct^2\lambda\delta(\frac{t}{\sqrt{n}})\end{aligned} \tag{7.145}$$

となる．さらに，n を大きくとれば，

$$|z| \le \frac{t^2}{8} \qquad \text{かつ } |\epsilon| \le \frac{t^2}{8}$$

---

1) $t = re^{\theta\sqrt{-1}}$ としたとき，$M = 0, \pm1, \pm2, \dots$ に対して $\log t = \log r + \sqrt{-1}(\theta + 2m\pi)$ となるが，とくに $-\pi < \theta < \pi$ と制約して，$\log z = r + \theta\sqrt{-1}$ と書いた．

2) すなわち，ある実数 a と b が存在して $p_n = P(X = a + nb)$ で $\sum_{-\infty}^{\infty} p_n = 0$ となるように書けること．

3) Probability for Statisticnas (G.R. Shorack 著，Springer ページ 361) を参照．

4) これは，テーラー展開から

$$|e^z - \sum_{k=0}^{m-1}\frac{z^k}{k!}| = |\sum_{k=m}^{\infty}\frac{z^k}{k!}| \le |z|^m|\sum_{j=0}^{\infty}\frac{|z|^j}{j!}|\frac{j!}{(j+m)!} \le |z|^m\frac{e^{|z|}}{m!}$$

よりわかる．

とできる．したがって，

$$
\begin{aligned}
|d_n(x)| &\leq \int_{A^c} e^{-t^2/2}\{|\epsilon|e^{|\epsilon|}e^{|z|} + \frac{z^2}{2}e^{|z|}\}\,dt + o(n^{-r}) \\
&\leq c\delta(\frac{t}{\sqrt{n}})\frac{\mathbb{E}|X|^3}{\sqrt{n}}\int_{-\infty}^{\infty}|t|^3e^{-t^2/2}\,dt + \frac{(\mathbb{E}|X|^3)^2}{72n}\int_{-\infty}^{\infty}|t|^6e^{-3t^2/8}\,dt + o(n^{-r}) \quad (7.146) \\
&= o(n^{-1})
\end{aligned}
$$

となることがわかり，(i) は証明された．□

·····7.2·····

分布関数に対する展開

補題 7.1　(Essen の不等式)：G を分布関数とし，有界な密度関数 g を持ち，$\int xg(x)\,dx = 0$ とする．F を任意の分布関数とする．さらに，ϕ_G と ϕ_F を G と F の特性関数とする．このとき，任意の $a > 0$ に対し

$$\|F - G\| \le \frac{1}{\pi}\int_{-a}^{a}\left|\frac{\phi_F(t) - \phi_G(t)}{t}\right| dt + \frac{24\pi\|g\|}{\pi a}$$

が成立する．

証明：H_a を分布関数とし，その密度関数と特性関数がつぎで与えられるとする：

$$h_a(x) = \frac{1 - \cos ax}{\pi a x^2}, \qquad \phi_a(t) = \begin{cases} 1 - |t|/a & (|t| \le a), \\ 0 & (|t| > a) \end{cases}.$$

さらに，F_a と G_a を H_a との F および G との畳み込みとする．

$$F_a(z) = \int F(z-y)\,dH_a(y), \qquad G_a(z) = \int G(z-y)\,dH_a(y).$$

関係式

$$\|F - G\| \le 2\|F_a - G_a\| + \frac{24\|g\|}{\pi a} \tag{7.147}$$

が成立することが以下の議論からわかる．$\Delta = F - G$ とおけば，$\lim_{x\to\pm\infty}\Delta(x) = 0$，$\Delta(x-)$ と $\Delta(x+)$ はほとんどいたるところ存在する．したがって，

$$D \equiv \|F - G\| = |\Delta(x_0)| \text{ または } |\Delta(x_0-)|$$

と仮定する．まず，$D = |\Delta(x_0)|$ とする．一般性を失うことなく，$\Delta(x_0) > 0$ とできる．F は単調増加なので g は有界（$|g| \le M$ とする）なので，$z = x_0 + \epsilon$ と $\epsilon = D/2M$ とおけば，$|x| \le \epsilon$ に対して

$$\Delta(z - x) \ge \frac{D}{2} + Mx \tag{7.148}$$

となる．$\Delta_a = F_a - G_a$ とおけば，(7.147) と (7.148) から

$$\begin{aligned}
\|F_a - G_a\| \geq \Delta_a(z) &= \int_{-\infty}^{\infty} \Delta(z-x) h_a(x)\,dx \\
&\geq \int_{-\epsilon}^{\infty} \left[\frac{D}{2} + Mx\right] dx - D\int_{|x|\geq\epsilon} h_a(x)\,dx \\
&= \frac{D}{2}\left[1 - \int_{|x|\geq\epsilon} h_a(x)\,dx\right] - D\int_{|x|\geq\epsilon} h_a(x)\,dx \\
&= \frac{D}{2}\left[1 - \int_{|x|\geq\epsilon} h_a(x)\,dx\right] + M\int_{-\epsilon}^{\epsilon} x h_a(x)\,dx - D\int_{|x|\geq\epsilon} h_a(x)\,dx \\
&= \frac{D}{2} - \frac{3D}{2}\int_{|x|\geq\epsilon} h_a(x)\,dx \\
&\geq \frac{D}{2} - \frac{12M}{\pi a} = \frac{1}{2}\|F-G\| - \frac{12M}{\pi a}
\end{aligned}$$

となる．最後の不等式は

$$\int_{|x|\geq\epsilon} h_a(x)\,dx \leq 2\int_{\epsilon}^{\infty} \frac{2}{\pi a x^2}\,dx = \frac{4}{\pi a \epsilon} = \frac{8M}{\pi a D}$$

からわかる．よって，(7.147) が示せた．

つぎに $\|F_a - G_a\|$ を評価する．逆フーリエ変換から

$$f_a(x) - g_a(x) = \frac{1}{2\pi}\int_{-a}^{a} e^{-tx\sqrt{-1}}[\phi_F(t) - \phi_G(t)]\phi_a(t)\,dt.$$

これより

$$\Delta_a(x) = \frac{1}{2\pi}\int_{-a}^{a} e^{-tx\sqrt{-1}} \frac{\phi_F(t) - \phi_G(t)}{-t\sqrt{-1}}\phi_a(t)\,dt \tag{7.149}$$

が成立することが期待できる．$\phi_F(0) = \phi_G(0) = 1$ と $\dot{\phi}_F(0) = \dot{\phi}_G(0) = 0$ より (7.149) の右辺の被積分関数は $t = 0$ で連続で 0 となるので，右辺は定義される．また，有界収束定理を用いれば

$$\frac{d}{dx}\Delta_a(x) = f_a(x) - g_a(x)$$

となるので，(7.149) の左辺と右辺は定数を除いて等しいことがわかる．明らかに，$|x| \to \infty$ のとき，$\Delta(x) \to 0$ となる．一方，Riemann - Lebeague の補題から

$$\lim_{|x|\to\infty} e^{-tx\sqrt{-1}} \frac{\phi_F(t) - \phi_G(t)}{-t\sqrt{-1}}\phi_a(t)\,dt = 0$$

となるので，(7.149) は示された．よって，(7.149) から

$$|\Delta_a(x)| \leq \frac{1}{2\pi}\int_{-a}^{a} \left|\frac{\phi_F(t) - \phi_G(t)}{t}\right| dt$$

となるので，補題は証明された．□

ここで，

$$\lim_{|t|\to\infty} \sup |\phi(t)| < 1 \tag{7.150}$$

と仮定[1)]する．

定理 7.2　：X は格子点分布ではないとし，(7.150) を仮定する．このとき，
(i) $\mathbb{E}|X|^3 < \infty$ ならば，

$$\left\| F_n(\cdot) - F_0(\cdot) + f_0(\cdot)\frac{1}{\sqrt{n}}\frac{\gamma_1}{3|}H_2(\cdot)\right\| = o(n^{-1/2})$$

(ii) $\mathbb{E}|X|^4 < \infty$ ならば，

$$\begin{aligned}&\left\| F_n(\cdot) - \{F_0(\cdot) + f_0(\cdot)\left[\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_2(\cdot)\right.\right.\\ &\left.\left.\frac{1}{n}\left\{\frac{\gamma_2}{4!}H_3(\cdot) + \frac{\gamma_1^2}{2\cdot(3!)^2}H_5(\cdot)\right\}\right]\right\| = o^{-1}\end{aligned}$$

証明：(i) の証明.

$$D_n \equiv F_n - F_0 + f_0\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_2$$

とおき，その特性関数を

$$\phi_n(t) = [\phi(t/\sqrt{n})]^n - \phi_0[1 + \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}(t\sqrt{-1})^3]$$

とかく．Essen の不等式から

$$\|D_n\| \le \frac{1}{\pi}\int_{|t|\le a\sqrt{n}/\mathbb{E}|X|^3} \frac{|\phi_n(t)|}{|t|}\,dt + \frac{24\|f_0(1+(\gamma_1/3!)H_2/\sqrt{n}}{\pi a\sqrt{n}/\mathbb{E}|X|^3} \tag{7.151}$$

を得る．(7.151) の右辺の 2 項目のノルムは有界なので，任意の $\epsilon > 0$ に対して適当な $a \equiv a(\epsilon, FX/\epsilon)$ を取れば，$\epsilon/\sqrt{n}$ 以下にできる．つぎに (7.151) の 1 項目の積分の領域を $A =\equiv \{|t| \ge \lambda\sqrt{n}/\mathbb{E}|X|^3\}$ と $\{|t| < \lambda\sqrt{n}/\mathbb{E}|X|^3\}$ にわかて評価する[2)]．(7.144) と同様の評価をすれば，

$$\begin{aligned}\int_A \frac{|\phi_n(t)|}{|t|}\,dt &\le \theta^{n-1}\int_{-\infty}^{\infty}|\phi(t)|\,dt + \int_A \phi_0(t)|1 + \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}(t\sqrt{-1})^3|\,dt\\ &= o(n^{-1})\end{aligned}$$

となる．(7.145) と (7.146) と同様の評価をすれば，

$$\int_{A^c}\frac{|\phi_n(t)|}{|t|}\,dt \le c\delta(\frac{t}{\sqrt{n}})\frac{\mathbb{E}|X|^3}{\sqrt{n}}\int_{-\infty}^{\infty}|t|^2 e^{-t^2/2}\,dt + \frac{(\mathbb{E}|X|^3)^2}{72n}\int_{-\infty}^{\infty}|t|^5 e^{-3t^2/8}\,dt$$

となることがわかる．これらふたつの式から

$$\|D_n\| o(n^{-1/2})$$

がわかり，(i) が証明された.　□

---

1) 絶対連続な密度関数を持てば，Riemann-Lebeague の補題からこの仮定は満たされることがわかる．

2) λ は定理 7.1 の証明中で定義したものである．

·····7.3·····

U 統計量に対する Edgeworth 展開

7.3.1 形式的な展開

$\{T_n\}_{n=1}^{\infty}$ を統計量の列とし，$\mathbb{E}T_n = 0, \mathbb{E}T_n^2 = 1$ とし，j 次キュムラント $\kappa_{j,n}$ がつぎの展開を持つとする．

$$\kappa_{j,n} = \sum_{l=0}^{r-j+2} K_j^{(l)} n^{-(jL-2)/2} + o(n^{-r/2}), \qquad j \geq 3. \tag{7.152}$$

$\phi_n(t)$ を T_n の特性関数とすれば，$t \downarrow 0$ のとき，

$$\log \phi_n(t) = \sum_{j=1}^{r} \frac{\kappa_{j,n}}{j!}(t\sqrt{-1})^j + o(t^r) \tag{7.153}$$

と表現できる．(7.152) を (7.153) に代入し整理すれば，

$$\log \phi_n(t) = -\frac{t^2}{2} + P_1(t\sqrt{-1})n^{-1/2} + \ldots + P_r(t\sqrt{-1})n^{-r/2} + o(n^{-r/2})$$

となる．ただし，

$$P_k(x) = \sum_{j=3}^{k+2} \frac{K_j^{k+2-j}}{j!} x^j$$

である．さらに，$e^z = \sum_{j=0}^{\infty} z^j/j!$ を用いれば，

$$\begin{aligned}\phi_n(t) &= \exp\left[-\frac{t^2}{2} + P_1(t\sqrt{-1})n^{-1/2} + \ldots + P_r(t\sqrt{-1})n^{-r/2} + o(n^{-r/2})\right] \\ &= e^{-t^2/2}\left(1 + \sum_{j=1}^{r} Q_j(t\sqrt{-1})n{-}j/2\right) + o(n^{-r/2})\end{aligned} \tag{7.154}$$

と書き直せる．ただし， Q は P で表現される．たとえば，

$$Q_1(x) = P_1(x) \qquad Q_2(x) = P_2(x) + \frac{1}{2}P_1^2(x)$$

$$Q_3(x) = P_3(x) + \frac{1}{2}P_2^2(x) + \frac{1}{6}P_1^3(x)$$

である．$Q_j(x) = \sum_{k\geq 1} A_{jk}x^j$ とおき，これを (7.154) に代入し，各項に形式的逆フーリエ変換 (7.142) を用いれば，

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-tx\sqrt{-1}} Q_j(t\sqrt{-1})e^{-t^2/2}\,dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-tx\sqrt{-1}} \sum_{k\geq 1} A_{jk}(t\sqrt{-1})^k e^{-t^2/2}\,dt \\ &= \sum_{k\geq 1} A_{jk}H_k(x)f_0(x)\end{aligned}$$

となるので，

$$\int_{-\infty}^{x} H_k(u)f_0(u)\,du = -H_{k-1}(x)f_0(x)$$

から

$$F_n(x) = F_0(x) - f_0(x)\left[\sum_{j\geq 1}^{r}\{A_{jk}H_{k-1}(x)\}\,n^{-j/2}\right] + o(n^{-r/2})$$

と書ける．

　$r=2$ の場合

$$\begin{aligned}&\kappa_{1,n} = 0, \qquad \kappa_{2.\,n} = 0, \qquad k_{3,n} = K_3^{(0)}n^{-1/2} + K_3^{(1)}n^{-1} + o(n^{-1}) \\ &k_{4,n} = K_4^{(4)}n^{-1} + o(n^{-1})\end{aligned}$$

から

$$\begin{aligned}Q_1(x) &= \frac{1}{6}K_3^{(0)}x^3 \\ Q_2(x) &= \frac{1}{6}K_3^{(1)}x^3 + \frac{1}{24}K_4^{(4)}x^4 + \frac{1}{72}(K_3^{(0)})^2x^6\end{aligned}$$

となる．したがって，

$$\begin{aligned}&A_{13} = \frac{1}{6}K_3^{(0)}, \qquad A_{1k} = 0, \qquad k\geq 4, \\ &A_{23} = \frac{1}{6}K_3^{(1)}, \qquad A_{24} = \frac{1}{24}K_4^{(0)}, \qquad A_{25} = 0, \\ &A_{26} = \frac{1}{72}(K_3^{(0)})^2, \qquad A_{2k} = 0, \qquad k\geq 7\end{aligned}$$

となるので，

$$\begin{aligned}\sum_{k\geq 1} A_{1k}H_{k-1}(x) &= \frac{1}{6}K_3^{(0)}H_2(x) = \frac{1}{6}K_3^{(0)}(x^2-1) \\ \sum_{k\geq 1} A_{2k}H_{k-1}(x) &= \frac{1}{6}K_3^{(1)}H_2(x) + A_{24} = \frac{1}{24}K_4^{(0)}H_3(x) + \frac{1}{72}(K_3^{(0)})^2H_5(x) \\ &= \frac{1}{6}K_3^{(1)}(x^2-1) + A_{24} = \frac{1}{24}K_4^{(0)}(x^3-3x) + \frac{1}{72}(K_3^{(0)})^2(x^5-10x+15)\end{aligned}$$

となる．これより

$$F_n(x) = F_0(x) - f_0(x)\left[\frac{1}{6}K_3^{(0)}(x^2-1)n^{-1/2} + \left\{\frac{1}{6}K_3^{(1)}(x^2-1) + A_{24} = \frac{1}{24}K_4^{(0)}(x^3-3x)\right.\right.$$
$$\left.\left. + \frac{1}{72}(K_3^{(0)})^2(x^5-10x+15)\right\}\right] + o(n^{-1})$$

となる．

参考文献

この講義録を作成するにあたり参考にした文献をあげる．欧文の成書としては以下である．

[C] Chung, K.L., *A course in Probability theory*, Academic Press, 1974.

[F] Ferguson, T.S., *A course in large sample theory*, Chapman & Hall, 1996.

[K] Karr, A.F., *Probability*, Springer, 1993.

[L] Lehmann, E.L., *Elements of large – sample theory*, Springer, 1999.

[Le] Lee, A.J. *U – Statistics*, Marcel Dekker, Inc., 1990.

[Sh] Shao, J., *Mathematical Statistics*, Springer, 1999.

[Se] Serfling, R.J., *Approximation theorems of Mathematical Statistics*, Wiley, 1980.

[V] van der Vaart, A.W., *Asymptotic Statistics*, Cambridge, 1998.

和文の成書としては以下のものを適宜参考にした．

[I] 稲垣宣夫, 数理統計学, 裳華房, 1986.

[Ku] 楠岡成雄, 確率と確率過程, 岩波書店, 1992.

[Ni] 西尾真喜子, 確率論, 実教出版, 1978.

[No] 野田一雄 ·宮岡悦良, 数理統計学の基礎, 共立出版, 1992.

[Si] Sinai, Ya.G., シナイ確率論入門コース（森真訳）, Springer, 1999.