2002 年度統計数学特論

今野 良彦
千葉大学大学院自然科学研究科

はじめに

この講義録は統計数学特論をまとめたものである．

目 次

1

確率分布と確率変数

·····1.1·····

確率測度

定義 1.1　標本空間 S の部分集合の集まりでつぎをみたすものを σ 集合体といい，$\mathcal{B}$ と記す.

(i) $\emptyset \in \mathcal{B}$
(ii) $A \in \mathcal{B}$ ならば $A^c \in \mathcal{B}$
(iii) $A_1, A_2, \ldots \in \mathcal{B}$ ならば $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$

定義 1.2　空でない標本空間 S とその σ 集合体 $\mathcal{B}$ があって，$\mathcal{B}$ 上の実数値集合関数 P がつぎをみたすとき，P を S 上の確率測度（または単に確率）という.

(i) すべての $A \in \mathcal{B}$ に対して，$P(A) \geq 0$
(ii) $P(S) = 1$
(iii) $A_1, A_2, \ldots \in \mathcal{B}, A_i \cap A_j = \emptyset\,(i \neq j)$ ならば $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

$(S, \mathcal{B}, P)$ を確率空間とよぶ.

<u>確率の基本性質</u>

(i) $P(\emptyset) = 0$
(ii) すべての $A \in \mathcal{B}$ に対して $P(A) \leq 1$
(iii) すべての $A \in \mathcal{B}$ に対して $P(A^c) = 1 - P(A)$
(iv) すべての $A, B \in \mathcal{B}$ に対して $P(B \cap A^c) = P(B) - P(A \cap B)$
(v) すべての $A, B \in \mathcal{B}$ に対して $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(vi) $A \subset B$ ならば $P(A) \leq P(B)$
(vii) $C_1, C_2, \ldots \in \mathcal{B}, C_i \cap C_j\,(i \neq j), S = \cup_{i=1}^{\infty} C_i$ のとき，$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$
(viii) $A_1, A_2, \ldots \in \mathcal{B}$ に対して $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$
(ix) $A_1, A_2, \ldots \in \mathcal{B}, A_i \subset A_{i+1}\,(i = 1, 2, \ldots)$ ならば

$$P(\cup_{i=1}^{\infty} A_i) = \lim_{i \to \infty} P(A_i)$$

(x) $A_1, A_2, \ldots \in \mathcal{B}, A_i \supset A_{i+1}\,(i = 1, 2, \ldots)$ ならば

$$P(\cap_{i=1}^{\infty} A_i) = \lim_{i \to \infty} P(A_i)$$

定義 1.3　S 上のボレル集合族とは，S の開集合の族を含む最小の σ 集合体である．これを $\mathcal{B}(S)$ と記す.

·····1.2·····

確率変数と累積分布関数

定義 1.4 確率空間 $(S, \mathcal{B}, P)$ 上において，写像 $X : S \to \mathbb{R}$ が任意の $x \in \mathbb{R}$ に対して

$$X^{-1}((-\infty, x]) = \{s \in S :\ X(s) \in (-\infty, x]\} \in \mathcal{B}$$

をみたすならば，X は（実）確率変数という．

定義 1.5 確率空間 $(S, \mathcal{B}, P)$ 上の確率変数 X について

$$F_X(x) = P(X \leq x) = P(s \in S :\ X(s) \in (-\infty, x]), \qquad x \in \mathbb{R}$$

によって定義される $\mathbb{R}$ 上の実数値関数 F_X を確率変数 X の累積分布関数という．

分布関数の性質

(i) $\lim_{x\to-\infty} F_X(x) = 0$ かつ $\lim_{x\to\infty} F_X(x) = 1$
(ii) $F_X(x)$ は x の非減少関数
(iii) $F_X(x)$ は右連続関数 ：すなわち，すべての x_0 に対して $\lim_{x\downarrow x_0} F_X(x) = F(x_0)$

定義 1.6 確率空間 $(S, \mathcal{B}, P)$ 上のふたつの確率変数 X, Y が同一分布に従うとは，すべての $\mathbb{R}$ 上の任意のボレル集合 $A \in \mathcal{B}(\mathbb{R})$ に対して $P(X \in A) = P(Y \in B)$ が成り立つことである．

定理 1.1 つぎのふたつは同値である．

(i) 確率変数 X, Y が同一分布に従う
(ii) すべての $x \in \mathbb{R}$ に対し，$F_X(x) = F_Y(x)$

定義 1.7 確率変数 X が連続型であるとは，X の分布関数 $F_X(x)$ が x の連続関数であることである．確率変数 X が離散型であるとは，X の分布関数 $F_X(x)$ が x の階段関数であることである．

定義 1.8 確率空間 $(S, \mathcal{B}, P)$ 上の離散型確率変数 X の確率関数を

$$f_X(x) = P(X = x), \qquad \forall x \in \mathbb{R}$$

定義 1.9 確率空間 $(S, \mathcal{B}, P)$ 上の連続型確率変数 X の確率密度関数 $f_X(x)$ は

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt \qquad \forall x \in \mathbb{R}$$

をみたすものである．

定義 1.10 確率変数 X についてある実数 m が

$$\mathbb{P}(X \leq m) \geq \frac{1}{2} \qquad \text{および} \qquad \mathbb{P}(X \geq m) \geq \frac{1}{2}$$

をみたすならば，m を X の中央値 (median) という．

<u>確率関数と確率密度関数の性質</u>

(i) ${}^{\forall}x \in \mathbb{R}$ に対し $f_X(x) \geq 0$

(ii) $\sum_x f_X(x) = 1$ (離散型) または $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (連続型)

·····1.3·····

期待値

定義 1.11 X を確率変数とし，g を実数値関数とする．このとき，$\int_{-\infty}^{\infty}|g(x)|f_X(x)\,dx < \infty$($X$ が連続型) もしくは $\sum_x |g(x)|f_X(x) < \infty$($X$ が離散型) が満たされるとき，$g(X)$ の期待値は存在するといい，$g(X)$ の期待値を

$$\mathbb{E}[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x)f_X(x)\,dx & X \text{ が連続型} \\ \sum_x g(x)f_X(x) & X \text{ が離散型} \end{cases}$$

で定義する．

期待値の基本的な性質

(i) $\mathbb{E}[ag_1(X) + a_2g_2(X)] = a_1\mathbb{E}[g_1(X)] + a_2\mathbb{E}[g_2(X)]$．ただし，$a_1$, a_2 は定数.
(ii) すべての x に対し，$g(x) \geq 0$ ならば，$\mathbb{E}[g(X)] \geq 0$
(iii) すべての x に対し，$g_1(x) \geq g_2(x)$ ならば，$\mathbb{E}[g_1(X)] \geq \mathbb{E}[g_2(X)]$
(iv) すべての x に対し，$a_1 \leq g(x) \leq a_2$ ならば，$a_1 \leq \mathbb{E}[g(X)] \leq a_2$.

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．

$$p(\boldsymbol{x}|\,\boldsymbol{\theta}) = S(\boldsymbol{x})\exp\left[\sum_{i=1}^{k}\eta_i(\theta)T_i(\boldsymbol{x}) - \tilde{A}(\theta)\right]I_B(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n \tag{2.1}$$

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

$$p(x|\theta) = \begin{pmatrix} n \\ x \end{pmatrix}\theta^x(1-\theta)^{1-x}$$

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

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

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，(2.1) において

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

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

$$f(\boldsymbol{x}|\eta) = S(x) \exp[\boldsymbol{T}(\boldsymbol{x})\eta - A(\eta)], \qquad \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n \tag{2.2}$$

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)

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

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

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

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 (2.2) 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，

$$A(\alpha\boldsymbol{\eta}_1 + (1-\alpha)\boldsymbol{\eta}_2)) \leq \alpha A(\boldsymbol{\eta}_1) + (1-\alpha)A(\boldsymbol{\eta}_2) \tag{2.3}$$

より (ii) は示せた．

また，$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathcal{E}$ ならば，(2.3) 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

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

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

□

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 (2.2) 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 (2.2) 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]：(iii) ⇒ (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) ⇒ (i) は示された．

(i) ⇒ (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) ⇔ (iii)，(i) ⇒ (ii)，(iii) ⇒ (iv) と (iii) ⇒ (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，

$$\frac{dP_\theta}{d\nu}(\boldsymbol{x}) = g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{2.4}$$

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，(2.4) 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

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

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

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

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

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

で与えられる．このとき，$h(\boldsymbol{x}) = \exp[S(\boldsymbol{x})] 1\{\boldsymbol{x} \in B\}$ 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，

$$\mathbb{VAR}(T_1, T_2) = n \begin{pmatrix} \mathbb{VAR}(X_1) & \mathbb{COV}(X_1, X_1^2) \\ \mathbb{COV}(X_1, X_1^2) & \mathbb{VAR}(X_1^2) \end{pmatrix} = 4\sigma^2 \begin{pmatrix} 1 & 2\mu \\ 2\mu & 2\sigma^2 + 4\mu^2 \end{pmatrix} \tag{2.5}$$

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

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

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

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

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 (2.1) 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，$n\ge k-1$ 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

$$\theta_k=1-\sum_{j=1}^{k-1}\theta_j \tag{2.6}$$

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．(2.6) 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 (2.6) 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，

$$\frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_j} l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_r}\left(\frac{n_j}{\theta_j} - \frac{n_k}{\theta_k}\right) = \begin{cases} -\left(\frac{n_r}{\theta_r^2} + \frac{n_k}{\theta_k^2}\right) < 0, & r = j \\ -\frac{n_k}{\theta_k^2} < 0, & r \ne j \end{cases}$$

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

$$x_{n+1} = x_n - \frac{g(x_n)}{\dot{g}(x_n)}, \qquad n = 0, 1, \ldots \tag{2.7}$$

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，(2.7) 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

$$S(\theta) = \frac{d}{d\theta} \log L_n(\theta|\,\boldsymbol{x}) = 0 \tag{2.8}$$

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，

$$H(\theta) = -\frac{d^2}{d\theta^2} \log L_n(\theta)$$

---

1) 杉浦「解析入門 I (東京大学出版会)」p.105 を参照．

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．

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 (2.8) 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

$$\widehat{\boldsymbol{\theta}}^{(k)} = \widehat{\boldsymbol{\theta}}^{(k-1)} + [H(\widehat{\boldsymbol{\theta}}^{(k-1)})]^{-1}S(\widehat{\boldsymbol{\theta}}^{(k-1)}) \tag{2.9}$$

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 (2.9) 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

$$H^*(\theta) = \frac{2n}{\pi}\int_{-\infty}^{\infty} \frac{1-(x-\theta)^2}{\{1+(x-\theta)^2\}^3}\,dx = \frac{n}{2} \tag{2.10}$$

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，(2.10) 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．

E - 段階 (E - Step)：各 $\theta \in \Theta$ に対し，条件付き期待値

$$\phi_1(\theta) = \mathbb{E}_{P_{\widehat{\theta}^{(0)}}}\{\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\} \tag{2.11}$$

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，(2.11) 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．

M - 段階 (M - Step)：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．

E - 段階 (E - Step)：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．

M - 段階 (M - Step)：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

E - 段階 (E - Step)：各 $\theta \in \Theta$ に対し，条件付き期待値

$$\phi_m(\theta) = \mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X}) = \boldsymbol{y}\} \tag{2.12}$$

を求める．

M - 段階 (M - Step)：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

$$\phi_1(\theta) = \mathbb{E}_{P_{\hat\theta^{(0)}}}[\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X}) = \boldsymbol{y}] = -\frac{1}{\theta}\mathbb{E}_{P_{\hat\theta^{(0)}}}\Bigl[\sum_{i=1}^n X_i|T(\boldsymbol{X}) = \boldsymbol{y}\Bigr] - n\log\theta$$

と

$$\begin{aligned}
\mathbb{E}_{P_{\widehat{\theta}^{(0)}}}[X_i|T(\boldsymbol{X})=\boldsymbol{y}] &= \frac{1}{\widehat{\theta}^{(0)} e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\int_y^{y+1} x e^{-x/\widehat{\theta}^{(0)}}\,dx \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{\left[-x e^{-x/\widehat{\theta}^{(0)}}\right]_y^{y+1} + \int_y^{y+1} e^{-x/\widehat{\theta}^{(0)}}\,dx\right\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\Bigg\{-(y+1)e^{-(y+1)/\widehat{\theta}^{(0)}} + y e^{-y/\widehat{\theta}^{(0)}} \\
&\quad -\widehat{\theta}^{(0)} e^{-(y+1)/\widehat{\theta}^{(0)}} + \widehat{\theta}^{(0)} e^{-y/\widehat{\theta}^{(0)}}\Bigg\} \\
&= \frac{1}{e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}})}\left\{(y+\widehat{\theta}^{(0)})e^{-y/\widehat{\theta}^{(0)}}(1-e^{-1/\widehat{\theta}^{(0)}}) - e^{-(y+1)/\widehat{\theta}^{(0)}}\right\} \\
&= y + \widehat{\theta}^{(0)} - \frac{e^{-1/\widehat{\theta}^{(0)}}}{1-e^{-1/\widehat{\theta}^{(0)}}} \\
&= y + \widehat{\theta}^{(0)} - \frac{1}{e^{1/\widehat{\theta}^{(0)}}-1}
\end{aligned}$$

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

$$\theta \mapsto \mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\} \tag{2.13}$$

を最大化する．(2.13) 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

$$\frac{\partial}{\partial\theta}\mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\}\bigg|_{\theta=\widehat{\theta}^{(\infty)}} = \mathbb{E}_{P_{\widehat{\theta}^{(\infty)}}}\left\{\frac{\partial}{\partial\theta}\log p(\boldsymbol{X}|\,\theta)|\,T(\boldsymbol{X})=\boldsymbol{y}\right\}\bigg|_{\theta=\widehat{\theta}^{(\infty)}} = 0 \tag{2.14}$$

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，(2.14) は

$$\frac{\partial}{\partial\theta}\log q(\boldsymbol{y}|\,\theta)\bigg|_{\theta=\widehat{\theta}^{(\infty)}} = 0 \tag{2.15}$$

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 (2.15) 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：

$$\log q(\boldsymbol{y}|\,\widehat{\theta}^{(m)}) > \log q(\boldsymbol{y}|\,\widehat{\theta}^{(m-1)}), \qquad m = 0,\,1,\,\ldots \tag{2.16}$$

$T(\boldsymbol{X})=\boldsymbol{y}$ 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

$$\begin{aligned}\log q(\boldsymbol{y}|\,\theta)&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log q(\boldsymbol{y}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\}\\&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)-\log k_\theta(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\\&=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\theta)|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_\theta(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\end{aligned}\tag{2.17}$$

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

$$\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\widehat{\theta}^{(m)})|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log p(\boldsymbol{X}|\,\widehat{\theta}^{(m-1)})|T(\boldsymbol{X})=\boldsymbol{y}\}\geq 0\tag{2.18}$$

がわかる．

次に，(2.17) の最右辺の 2 項目を評価する：

$$\begin{aligned}&\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}-\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\}\\&\qquad=\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\log\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|T(\boldsymbol{X})=\boldsymbol{y}\right\}\\&\qquad\leq\log\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\left\{\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\middle|T(\boldsymbol{X})=\boldsymbol{y}\right\}\end{aligned}\tag{2.19}$$

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

$$\int_B g(\boldsymbol{y})\,dQ_{\widehat{\theta}^{(m-1)}}(\boldsymbol{y})=\int_{T^{-1}(B)}\frac{k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})}{k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})}\,dP_{\widehat{\theta}^{(m-1)}}(\boldsymbol{x})\tag{2.20}$$

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，(2.20) 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 (2.19) から

$$\mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\} - \mathbb{E}_{P_{\widehat{\theta}^{(m-1)}}}\{\log k_{\widehat{\theta}^{(m-1)}}(\boldsymbol{X}|\,\boldsymbol{y})|T(\boldsymbol{X})=\boldsymbol{y}\} \le 0 \tag{2.21}$$

がわかる．よって，(2.18) と (2.21) から (2.17) は示せた．

3

大標本理論

·····3.1·····

指数分布族モデルにおける最尤推定量の漸近理論

P_η を $(\mathcal{X}, \mathcal{A})$ 上の確率測度とし，σ - 有限な測度 $\lambda(x)$ に関する確率密度関数

$$p(x|\,\boldsymbol{\eta}) = \exp\{\boldsymbol{\eta}'\boldsymbol{T}(x) - A(\boldsymbol{\eta})\}$$

を持つ[1]とする．ここでは，考える指数分布族はフルランクとし，$A(\boldsymbol{\eta})$ は 2 階連続微分可能とする．

系 2.1(i) から，スコア関数は

$$\frac{\partial}{\partial\boldsymbol{\eta}}\log p(x|\,\boldsymbol{\eta}) = \boldsymbol{T}(x) - \frac{\partial}{\partial\boldsymbol{\eta}}A(\boldsymbol{\eta}) = \boldsymbol{T}(x) - \dot{A}(\boldsymbol{\eta}) = \boldsymbol{T}(x) - \mathbb{E}_\eta[\boldsymbol{T}(X)]$$

となる．さらに，系 2.1(ii) から，スコア関数の分散は

$$I(\boldsymbol{\eta}) = \mathbb{VAR}_\eta\left\{\frac{\partial}{\partial\boldsymbol{\eta}}\log p(X|\,\boldsymbol{\eta})\right\} = \frac{\partial^2}{\partial\boldsymbol{\eta}'\partial\boldsymbol{\eta}}A(\boldsymbol{\eta}) = \ddot{A}(\boldsymbol{\eta})$$

で与えられ，これを情報行列と呼ぶことにする．

$X_1, X_2, \ldots, X_n$ を確率測度 P_{η_0} からのランダム標本とする．ただし，$\boldsymbol{\eta}_0$ は自然母数空間 $\mathcal{E}$ の内点に含まれるとする．このとき，$\mathbb{R}^n$ 上の測度 $\Pi_{i=1}^n\lambda(x_i)$ に関する $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta}_0) = \Pi_{i=1}^n p(x_i|\,\boldsymbol{\eta}_0) = \exp\Big\{\boldsymbol{\eta}_0'\sum_{i=1}^n\boldsymbol{T}(x_i) - nA(\boldsymbol{\eta}_0)\Big\}, \qquad \boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$$

となる．ここで，$A_n(\boldsymbol{\eta}) = nA(\boldsymbol{\eta})$, $n\bar{\boldsymbol{T}}_n(\boldsymbol{x}_n) = \sum_{i=1}^n\boldsymbol{T}(x_i)$ とおくと別の指数分布族ができる．$\widehat{\boldsymbol{\eta}}_n$ を $\boldsymbol{\eta}$ の最尤推定値とする．すなわち，$p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta})$ を最大にする $\boldsymbol{\eta}$ の値である．これは，関数

$$G_n(\boldsymbol{\eta}) = \frac{1}{n}\log\frac{p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta}_0)}{p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta})} = (\boldsymbol{\eta}_0 - \boldsymbol{\eta})'\bar{\boldsymbol{T}}_n + A(\boldsymbol{\eta}) - A(\boldsymbol{\eta}_0)$$

を最小にするものと同値である：

$$\widehat{\boldsymbol{\eta}}_n = \mathrm{argmin} G_n(\boldsymbol{\eta})$$

である．

以下では，$\widehat{\boldsymbol{\eta}}_n$ の漸近一致性と漸近分布を求めていく．最尤推定量の漸近分布を求める基本的な方針はつぎのふたつがある．

(1)　関数 $G_n(\boldsymbol{\eta})$ が滑らかであれば，最尤推定量は尤度方程式の解として考察する．

---

[1] (2.1) において $\lambda(x) = S(x)I_B(x)\mu(x)$ とすればよい．

(2)　最尤推定量を関数 $G_n(\boldsymbol{\eta})$ を最小にするものとして考察する．

ここでは，(2) の方針に従って議論を進めていく．

3.1.1 漸近一致性

まず，$\widehat{\boldsymbol{\eta}}_n$ において $G_n(\boldsymbol{\eta})$ が最小になることから

$$0 = G_n(\boldsymbol{\eta}_0) \geq G_n(\widehat{\boldsymbol{\eta}}_n)$$

となることに注目する．さらに，大数の法則から

$$\bar{\boldsymbol{T}}_n \xrightarrow{a.s} \mathbb{E}_{\eta_0}[\boldsymbol{T}(X)] = \dot{A}(\boldsymbol{\eta}_0) \stackrel{\text{def}}{=} \boldsymbol{\beta}_0$$

となるので，

$$G_n(\boldsymbol{\eta}) \xrightarrow{a.s} G(\boldsymbol{\eta}) \stackrel{\text{def}}{=} (\boldsymbol{\eta}_0 - \boldsymbol{\eta})'\boldsymbol{\beta}_0 + A(\boldsymbol{\eta}) - A(\boldsymbol{\eta}_0)$$

となることがわかる．この収束は，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| < \kappa, \kappa$ は正定数 $\}$ 上で一様である：

$$\sup_{|\eta-\eta_0|<\kappa} |G_n(\boldsymbol{\eta}) - G(\boldsymbol{\eta})| \leq \kappa|\bar{T}_n - \boldsymbol{\beta}_0| \xrightarrow{a.s} 0 \tag{3.1}$$

直観的に言えば，$G_n(\boldsymbol{\eta})$ は $G(\boldsymbol{\eta})$ に近いので，$G_n(\boldsymbol{\eta})$ が最小となるような点は $G(\boldsymbol{\eta})$ が最小となる点に近いと期待できる．したがって，$\widehat{\boldsymbol{\eta}}$ がどこに近づくかを観るために，$G(\boldsymbol{\eta})$ を最小とする点を求める．$G(\boldsymbol{\eta})$ は凸関数で，その導関数 $-\boldsymbol{\beta}_0 + \dot{A}(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ でゼロになる．また，2 次関数 $I(\boldsymbol{\eta}_0) = \ddot{A}(\boldsymbol{\eta}_0) > 0$ なので，$G(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ で唯一最小となる．特に，各 $\delta > 0$ に対し，正数 ϵ を

$$2\epsilon = \inf_{|\eta-\eta_0|\geq\delta} G(\boldsymbol{\eta})$$

で定める．任意の $\delta > 0$ に対し，$\widehat{\boldsymbol{\eta}}_n$ が $\boldsymbol{\eta}_0$ の δ - 近傍に高い確率で含まれることを示すためには，高い確率で

$$\inf_{|\eta-\eta_0|\geq\delta} G_n(\boldsymbol{\eta}) \geq \epsilon$$

となることを示せばよい：なぜならば，

$$\mathbb{P}_{\eta_0}\{|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| \leq \delta\} \geq \mathbb{P}_{\eta_0}\{\inf_{|\eta-\eta_0|\geq\delta} G_n(\boldsymbol{\eta}) \geq \epsilon\} \tag{3.2}$$

からわかる．したがって，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \geq \delta\}$ 上で $\boldsymbol{\eta}$ に関して $G_n(\boldsymbol{\eta})$ の最小化を考えることになる．しかし，$G_n(\boldsymbol{\eta})$ は凸関数なので，$K = \{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \delta\}$ 上のみで $G_n(\boldsymbol{\eta})$ の最小化を考えればよいことがわかる．有界集合 K 上では，(3.1) から高い確率で

$$|\inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta})| \leq \epsilon$$

となる．これから，高い確率で

$$\inf_{\eta\in K} G_n(\boldsymbol{\eta}) = \inf_{\eta\in K} G(\boldsymbol{\eta}) + \inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta}) \geq \inf_{\eta\in K} G(\boldsymbol{\eta}) - |\inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta})| \geq \epsilon$$

とできる．よって，(3.2) から $\widehat{\boldsymbol{\eta}}_n$ の漸近一致性が示せた.

3.1.2 収束の速さ

中心極限定理より

$$\sqrt{n}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) \xrightarrow{L} N(0, I(\boldsymbol{\eta}_0))$$

となる．ただし，$I(\boldsymbol{\eta}_0) = \mathbb{VAR}_{\eta_0}(\boldsymbol{T}(X)) = \ddot{A}(\boldsymbol{\eta}_0)$ である．これより

$$M_n = |\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0| = O_p(1/\sqrt{n})$$

となる．$\bar{\boldsymbol{T}}_n$ の収束の早さから

$$|G_n(\boldsymbol{\eta}_0) - G(\boldsymbol{\eta})| \leq M_n|\boldsymbol{\eta}_0 - \boldsymbol{\eta}| \tag{3.3}$$

を得る．次に，$G(\boldsymbol{\eta})$ を $\boldsymbol{\eta}_0$ のまわりでテーラー展開する：

$$\begin{aligned} G(\boldsymbol{\eta}) &= G(\boldsymbol{\eta}_0) + (\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\dot{G}(\boldsymbol{\eta}_0) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\ddot{G}(\boldsymbol{\eta}^*)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \\ &= \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o(|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2) \end{aligned}$$

となる．ただし，$\boldsymbol{\eta}^*$ は $\boldsymbol{\eta}$ と $\boldsymbol{\eta}_0$ を直線で結んだ線分上の点である．$I(\boldsymbol{\eta}_0)$ は正定値なので，ある正定数 C が存在して

$$\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \geq 2C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2$$

とできる．$\delta > 0$ を十分小さくとって，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \delta\}$ 上では

$$G(\boldsymbol{\eta}) \geq C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2 \tag{3.4}$$

とできる．さらに，n が十分大きいとき，高い確率で $M_n \leq \delta$ となるので，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = 2M_n/C\}$ 上では，(3.3) と (3.4) より

$$G_n(\boldsymbol{\eta}) \geq G(\boldsymbol{\eta}) - M_n|\boldsymbol{\eta} - \boldsymbol{\eta}_0| \geq C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2 - M_n|\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \frac{2M_n^2}{C}$$

となる．これより，高い確率で

$$\inf_{|\eta - \eta_0| \geq M_n/C} |G_n(\boldsymbol{\eta})| \geq \frac{2M_n^2}{C}$$

となる．したがって，一致性の議論と同様にすれば，高い確率で $\widehat{\boldsymbol{\eta}}_n$ は $\boldsymbol{\eta}_0$ の M_n/C - 近傍に入ることがわかる．よって

$$|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| \leq \frac{2M_n}{C} = O_p(1/\sqrt{n})$$

となる．上の式は $M_n \leq \delta$ の仮定のもとの議論だったので，もうすこし吟味が必要である．十分大きな正数 K に対し，$D_n = \{\sqrt{n}M_n \leq K\}$, $E_n = \{M_n \leq \delta\}$, $F_n = \{\sqrt{n}|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}| \leq 2K/C\}$ とおく．上の議論で $D_n \cap E_n$ が起こったならば，F_n が起こることが言えた．すなわち，

$$\mathbb{P}(D_n \cap E_n) \leq \mathbb{P}(F_n)$$

である．また，$n \to \infty$ のとき，$\mathbb{P}(D_n) \to 0$ と $\mathbb{P}(E_n) \to 0$ に注意すれば

$$\mathbb{P}(D_n \cap E_n) \geq 1 - \mathbb{P}(D_n^c) - \mathbb{P}(E_n^c) = \mathbb{P}(D_n) + \mathbb{P}(E_n) + 1 \to 1$$

となるので，$\mathbb{P}(F_n) \to 1\,(n \to \infty)$ がわかる．

3.1.3 漸近正規性

$\delta_n = O(1/\sqrt{n})$ とおく．$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \leq \delta_n\}$ 上で

$$\begin{aligned} G_n(\boldsymbol{\eta}) &= -(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)' I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o(1/n) \\ &= \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) - I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \end{aligned}$$

となる．これは，$o(|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2) = o(O(1/n)) = o(1/n)$ に注意すればわかる．ただし，$I_0^{1/2}$ は正定値行列で $I(\boldsymbol{\eta}_0) = I_0^{1/2} I_0^{1/2}$ をみたすものである．

いま，

$$Q_n(\boldsymbol{\eta}) = \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) - I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2$$

とおく．$Q_n(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}_n = \boldsymbol{\eta}_0 + I^{-1}(\boldsymbol{\eta}_0)(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)$ で最小となる．中心極限定理より

$$\sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = I^{-1}(\boldsymbol{\eta}_0)\sqrt{n}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0},\, I^{-1}(\boldsymbol{\eta}_0))$$

となる．よって

$$|\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| = O_p(1/\sqrt{n})$$

となる．したがって

$$\begin{aligned} G_n(\boldsymbol{\eta}) &= \frac{1}{2}|I_0^{1/2}\boldsymbol{\eta} - I_0^{1/2}(\boldsymbol{\eta}_0 + I(\boldsymbol{\eta}_0)^{-1}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0))|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \\ &= \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \end{aligned}$$

となる．$|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}| = O_p(1/\sqrt{n})$ から $G_n(\boldsymbol{\eta})$ に $\hat{\boldsymbol{\eta}}_n$ と $\tilde{\boldsymbol{\eta}}$ を代入すれば，上の式の最右辺の $o(1/n)$ の項は $o_P(1/n)$ になる．$\hat{\boldsymbol{\eta}}_n$ の定義から $G_n(\hat{\boldsymbol{\eta}}_n) \leq G_n(\tilde{\boldsymbol{\eta}}_n)$ となるので，

$$\frac{1}{2}|I_0^{1/2}(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n)|^2 \leq o_P(1/n)$$

を得る．よって，

$$|\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n| = o_P(1/\sqrt{n})$$

となり，Slutzky の補題を用いれば，

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = \sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + o_P(1) \xrightarrow{L} N(\mathbf{0},\, I^{-1}(\boldsymbol{\eta}_0))$$

を得る．

·····3.2·····

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,

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) \xrightarrow{as} \mu(\theta) \qquad (n \to \infty) \tag{3.5}$$

が成立する．しかし,

$$\sup_{\theta \in \Theta}\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta)\right| \xrightarrow{as} 0 \qquad (n \to \infty) \tag{3.6}$$

が成立すると便利である．

(3.6) 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, (3.6) 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,

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\theta_0)\right| &\leq \left|\frac{1}{n}\sum_{i=1}^{n} U(X_i, \hat{\theta}_n) - \mu(\hat{\theta}_n)\right| + \left|\mu(\hat{\theta}_n) - \mu(\theta_0)\right| \\
&\leq \sup_{\theta \in \Theta}\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i, \theta) - \mu(\theta)\right| + \left|\mu(\hat{\theta}_n) - \mu(\theta_0)\right| \\
&\xrightarrow{as} 0, \qquad (n \to \infty)
\end{aligned}$$

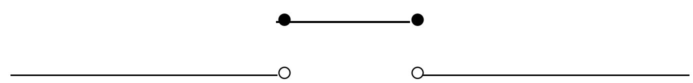
からわかる．

以下, (3.6) 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,

$$\limsup_{n\to\infty} g(\theta_n) \leq g(\theta)$$

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,

$$|U(x,\theta)| \leq K(x)$$

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,

$$P\left\{\limsup_{n\to\infty}\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta) \leq \sup_{\theta\in\Theta}\mu(\theta)\right\} = 1$$

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, $\varphi(x,\theta,\rho) \leq K(x)$ 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

$$U(x, \theta) \le \varphi(x, \theta_i, \rho_{\theta_i})$$

である．したがって,

$$\frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \le \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

となることから

$$\sup_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \le \sup_{1\le i\le m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

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,

$$P\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i}) \le \mu(\theta_i) + \epsilon\right\} = 1$$

が成り立つ．よって,

$$P\left\{\lim_{n\to\infty} \sup_{1\le i\le m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i}) \le \sup_{1\le i\le m} \mu(\theta_i) + \epsilon\right\} = 1$$

となる．さらに,

$$\sup_{1\le i\le m} \mu(\theta_i) \le \sup_{\theta\in\Theta} \mu(\theta)$$

と

$$\lim_{n\to\infty}\sup \sup_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \le \lim_{n\to\infty} \sup_{1\le i\le m} \frac{1}{n}\sum_{j=1}^n \varphi(X_j, \theta_i, \rho_{\theta_i})$$

に注意すれば,

$$P\left\{\lim_{n\to\infty}\sup \sup_{\theta\in\Theta} \frac{1}{n}\sum_{j=1}^n U(X_j, \theta) \le \sup_{\theta\in\Theta} \mu(\theta) + \epsilon\right\} = 1$$

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． □

·····3.3·····

最尤推定量の強一致性と漸近正規性

$f(x|\theta)$ を σ – 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．

3.3.1 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 σ – 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 σ – 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, σ – 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

$$\sup_{\theta \in S} \frac{1}{n} \sum_{j=1}^{n} U(X_j, \theta) \le \frac{\delta}{2} < 0 \tag{3.7}$$

とほとんど確実にできる．しかし, $\hat{\theta}_n$ の定義から

$$\frac{1}{n} \sum_{j=1}^{n} U(X_j, \hat{\theta}_n) = \sup_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} U(X_j, \theta) \ge 0 \tag{3.8}$$

であり[2), (3.7) 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．

3.3.2 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) (3.8) 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,

$$\begin{aligned} |B_n + I(\theta_0)| &\le \int_0^1 \left| \frac{1}{n} \sum_{j=1}^{n} \dot{\Psi}(X_j, \theta_0 + t(\hat{\theta}_n - \theta_0)) + I(\theta_0) \right| dt \\ &\le \int_0^1 \left[\sup_{\theta \in S_\rho} \left| \frac{1}{n} \sum_{j=1}^{n} \dot{\Psi}(X_j, \theta) - \boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta) \right| + |\boldsymbol{E}_{\theta_0} \dot{\Psi}(X, \theta) + I(\theta_0)| \right] dt \\ &\le 2\epsilon \end{aligned}$$

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．

·····3.4·····

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 3.2　（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 3.4　：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

$$\begin{aligned}\liminf_{n\to\infty} \boldsymbol{E}\{\sqrt{n}(T_n - \theta)\}^2 &= \liminf_{n\to\infty} \int_0^\infty P\big[\{\sqrt{n}(T_n - \theta)\}^2 > t\big]dt \\ &\geq \int_0^\infty \liminf_{n\to\infty} P\big[\{\sqrt{n}(T_n - \theta)\}^2 > t\big]dt\end{aligned}$$

となる．さらに、仮定より、

$$\liminf_{n\to\infty} P\big[\{\sqrt{n}(T_n - \theta)\}^2 > t\big] = 2\big\{1 - \Phi(\frac{\sqrt{t}}{\sigma(\theta)})\big\}$$

となる[1]. これを積分すれば、

$$\begin{aligned}2\int_0^\infty \big[1 - \Phi(\frac{\sqrt{t}}{\sigma(\theta)})\big]dt &= 4\sigma^2(\theta)\int_0^\infty u\{1 - \Phi(u)\}du \\ &= 2\sigma^2(\theta)\int_0^\infty u^2\phi(u)du = \sigma^2(\theta)\end{aligned}$$

より[2]、定理は示せた．　□

例 3.1　：$X_1, \ldots, X_n$ 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

$$Y_n = \begin{cases} 1 & (\text{確率}\quad \frac{1}{n^2}) \\ 0 & (\text{確率}\quad 1 - \frac{1}{n^2}) \end{cases}$$

---

[1] $P\big[\sqrt{n}(T_n - \theta) > \sqrt{t}\big] = P\big[\{\sqrt{n}(T_n - \theta)\}^2 > t\big] + P\big[\sqrt{n}(T_n - \theta) < -\sqrt{t}\big]$ と $\liminf_{n\to\infty} P\big[\{\sqrt{n}(T_n - \theta)\}^2 > t\big] > t\big] \to 1 - \Phi(\sqrt{t}/\sigma(\theta))$ および $\liminf_{n\to\infty} P\big[\sqrt{n}(T_n - \theta) < -\sqrt{t}\big] > t\big] \to \Phi(-\sqrt{t}/\sigma(\theta))$ 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 3.5　：$P_{\theta+h/\sqrt{n}}$ 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、

$$\begin{aligned} P_{\theta+h/\sqrt{n}}(B_n) &\le P_{\theta+h/\sqrt{n}}\{\log \Lambda_n(h) > K\} + \boldsymbol{E}_\theta[\exp(\log \Lambda_n(h))]1\{\log \Lambda_n(h) \le K\}1\{B_n\}] \\ &\le \epsilon + \exp(K) P_\theta(B_n) \to 0 \end{aligned} \tag{3.9}$$

となり、補題は示された．□

定理 3.6　：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、

$$\sqrt{n}(\hat{\theta}_n - \theta - \frac{h}{\sqrt{n}}) = h^{-1}I^{-1}(\theta)\log \Lambda_n(h) - \frac{1}{2}h + o_P(1) \tag{3.10}$$

を得る．$P_{\theta+h/\sqrt{n}}$ は P_θ に接触しているので、(3.10) 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 3.7　：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、

$$\sigma^2(\theta) \ge I^{-1}(\theta)$$

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

$$\lim_{n\to\infty} P_{\theta+h/\sqrt{n}}\Big\{\sqrt{n}(T_n-\theta-\frac{h}{\sqrt{n}})\leq x\Big\}=\Phi(\frac{x}{\sigma(\theta)}) \tag{3.11}$$

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

$$\lim_{n\to\infty}P_{\theta+h/\sqrt{n}}(W_n^*)\geq\lim_{n\to\infty}P_{\theta+h/\sqrt{n}}(W_n) \tag{3.12}$$

となる．しかし、補題 2.3 から、(3.12) 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、(3.12) 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、

$$\Phi(h\sqrt{I(\theta)})\geq\Phi(\frac{h}{\sigma(\theta)})$$

となり[6]、

$$I(\theta)\geq\frac{1}{\sigma^2(\theta)}$$

を得る. よって、定理は示された．□

[5] これは (3.11) 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.

……3.5……

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 3.8　：$\tilde{\theta}_n$ 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、

$$\begin{aligned} \tilde{\theta}_n^{(1)} &= \tilde{\theta}_n - \{\ddot{l}_n(\tilde{\theta})\}^{-1}\dot{l}(\tilde{\theta}_n) \\ \theta_n^* &= \tilde{\theta}_n + I^{-1}(\tilde{\theta}_n)\frac{1}{n}\dot{l}(\tilde{\theta}_n) \end{aligned}$$

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、

$$\sqrt{n}(\tilde{\theta}_n^{(1)} - \hat{\theta}_n) = \big[\mathrm{id} - \{\ddot{l}_n(\tilde{\theta}_n)\}^{-1}\int_0^1 \ddot{l}_n(\hat{\theta}_n + t(\tilde{\theta}_n - \hat{\theta}_n))dt\big]\sqrt{n}(\tilde{\theta}_n - \hat{\theta})$$

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． □

4

線形回帰モデルと一般化線形回帰モデル

·····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

$$Y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{4.1}$$

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

$$\boldsymbol{Y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n \tag{4.2}$$

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 (4.2) 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

$$\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n = (\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta})'(\boldsymbol{Y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) \tag{4.3}$$

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．(4.3) は $\boldsymbol{\beta}$ の 2 次関数なので，(4.1) を最小にする $\boldsymbol{\beta}$ は方程式

$$\boldsymbol{X}_n'\boldsymbol{X}_n\boldsymbol{\beta} = \boldsymbol{X}_n'\boldsymbol{Y}_n \tag{4.4}$$

または

$$\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2\boldsymbol{x}_i = \mathbf{0} \tag{4.5}$$

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

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\boldsymbol{Y}_n \tag{4.6}$$

で与えられる．

また，単回帰線形モデル (4.1) 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 (4.2) 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 (4.1) 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)　$\boldsymbol{c}\,(\neq \mathbf{0})$ 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．

[1)] これは (4.2) 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，(4.1) を得ることからわかる．

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

$$\boldsymbol{c}'\boldsymbol{\beta} = \mathbb{E}[\boldsymbol{a}'\boldsymbol{Y}_n] = \boldsymbol{a}'\mathbb{E}[\boldsymbol{X}_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n] = \boldsymbol{a}'\boldsymbol{X}_n\boldsymbol{\beta}$$

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

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}\log L_n(\boldsymbol{\beta},\, \sigma^2 |\, \boldsymbol{y}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i = \boldsymbol{0}\\ \frac{\partial}{\partial\sigma}\log L_n(\boldsymbol{\beta},\, \sigma^2 |\, \boldsymbol{y}) &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 = 0\end{aligned}$$

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，(4.6) 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 $\boldsymbol{H}\boldsymbol{w} \in \text{span}\,(\boldsymbol{X}_n)$ をみたし，任意のベクトル $\boldsymbol{z} \in \text{span}\,(\boldsymbol{X}_n)$ 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 $\text{tr}\,(\boldsymbol{H})$ 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，

$$\begin{aligned}\mathbb{COV}(\widehat{\boldsymbol{\beta}}_n, \boldsymbol{Y}-\boldsymbol{X}_n\boldsymbol{\beta}) &= \mathbb{COV}(\boldsymbol{A}_n\boldsymbol{Y}, (I_n-\boldsymbol{H}\boldsymbol{Y}) = \boldsymbol{A}_n\mathbb{COV}(\boldsymbol{Y},\boldsymbol{Y})(I_n-\boldsymbol{H})' = \sigma^2\boldsymbol{A}_n(I_n-\boldsymbol{H})\\ &= \sigma^2\{(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n' - (\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'\} = \boldsymbol{0}\end{aligned}$$

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 (4.2) 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 (4.2) 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

$$\sqrt{t_n}(\hat{\beta}_n - \beta) = \sum_{i=1}^{n} \frac{x_i - \bar{x}_n}{\sqrt{t_n}} \epsilon_i = \sum_{i=1}^{n} c_{ni}\epsilon_i$$

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

$$\begin{aligned}
\sqrt{n}(\hat{\alpha}_n - \alpha) &= \sqrt{n}\bar{\epsilon}_n - \sqrt{n}(\hat{\beta}_n - \beta)\bar{x}_n \\
&= \sqrt{n}\bar{\epsilon}_n - \sqrt{\frac{n}{t_n}}\bar{x}_n\sqrt{t_n}(\hat{\beta}_n - \beta) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i - \sqrt{\frac{n}{t_n}}\bar{x}_n\sum_{i=1}^{n} c_{ni}\epsilon_i \\
&= \sum_{i=1}^{n} d_{ni}\epsilon_i
\end{aligned}$$

[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．

[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．

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，

$$\sum_{i=1}^n d_{ni} = \sqrt{n}, \qquad \sum_{i=1}^n d_{ni}^2 = 1 + \frac{n\bar{x}_n^2}{t_n} \equiv r_n^2$$

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，

$$\frac{1}{B_n^2}\sum_{j=1}^{n}\mathbb{E}\left[X_{nj}^2 I\{|X_{nj}| \geq \epsilon B_n\}\right] \to 0, \qquad (n \to \infty) \tag{4.7}$$

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，(4.7) 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

$$f(y_i|\,\theta_i,\,\phi_i) = c(y_i,\,\phi_i)\exp\left[\frac{y_i\theta_i - b(\theta_i)}{\phi_i}\right] \tag{4.8}$$

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 (4.8) を持つ分布を指数分布族という．(4.8) 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

$$\int c(y,\,\phi_i)\left\{\frac{y}{\phi_i}\right\}\exp\left[\frac{y_i\theta_i}{\phi_i}\right]\,d\nu(y) = \frac{\dot{b}(\theta_i)}{\phi_i}\exp\left[\frac{b(\theta_i)}{\phi_i}\right] \tag{4.9}$$

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，(4.9) 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，

$$\mathbb{E}[Y_i] = \mu_i \equiv \mu(\theta_i) = \dot{b}(\theta_i), \quad \mathbb{VAR}[Y_i] = \phi_i\ddot{b}(\theta_i) \tag{4.10}$$

となる．さらに，$\theta_i = \dot{b}^{-1}(\mu_i)$ より

$$\mathbb{VAR}[Y_i] = \phi_i \ddot{b}(\dot{b}^{-1}(\mu_i)) \equiv \phi_i V(\mu_i) \tag{4.11}$$

と書くこと[1]にする．(4.11) 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

$$g(\mu(\theta_i)) = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{4.12}$$

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，

$$\mathcal{B} = \{\boldsymbol{\gamma} \in \mathbb{R} : \text{すべての } \boldsymbol{x} \in \mathcal{X} \text{ に対し}, (g \circ \mu)^{-1}(\boldsymbol{\gamma}'\boldsymbol{x}_i) \in \Theta^0\} \tag{4.13}$$

とする．このとき，$\boldsymbol{\beta} \in \mathcal{B}$ とする．多くのモデルでは

$$\phi_i = \frac{\phi}{t_i}, \qquad i = 1, 2, \ldots, n \tag{4.14}$$

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 (4.12) と (4.14) 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．(4.10) 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 (4.11) 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

$$\dot{l}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{\{y_i - g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})\}t_i}{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))}\boldsymbol{x}_i = \boldsymbol{0} \tag{4.15}$$

となる．

いま

$$M_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 \ddot{b}(h(\boldsymbol{x}_i'\boldsymbol{\beta})) t_i \boldsymbol{x}_i \boldsymbol{x}_i' \tag{4.16}$$

と

$$R_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} [Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))] \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta}) t_i \boldsymbol{x}_i \boldsymbol{x}_i' \tag{4.17}$$

とおく．このとき

$$\begin{aligned} \mathbb{VAR}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log L_n(\boldsymbol{\beta},\, \phi)\right] &= \frac{1}{\phi^2} \sum_{i=1}^{n} \mathbb{VAR}[\{Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))\} \dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta}) t_i \boldsymbol{x}_i] \\ &= \frac{1}{\phi^2} \sum_{i=1}^{n} \mathbb{VAR}[Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))] [\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 t_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \\ &= \frac{1}{\phi^2} \sum_{i=1}^{n} \frac{\phi}{t_i} \ddot{b}(h(\boldsymbol{x}_i'\boldsymbol{\beta})) [\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2 t_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \\ &= \frac{1}{\phi} M_n(\boldsymbol{\beta}) \end{aligned} \tag{4.18}$$

となる．また

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \log L_n(\boldsymbol{\beta},\, \phi) = \frac{1}{\phi} \ddot{l}_n(\boldsymbol{\beta}) = \frac{R_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\beta})}{\phi} \tag{4.19}$$

となる．

定理 4.6　(4.8), (4.12), (4.13), (4.14) 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)　(4.13) 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

$$0 < \inf_{i=1,\,2,\,\dots\,,\,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta}) \leq \sup_{i=1,\,2,\,\dots\,,\,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta}) < \infty$$

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

$$\max_{i=1,\,2,\,\dots\,,\,n} \boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \to 0, \qquad \mathrm{ch}_{\min}(\boldsymbol{X}_n'\boldsymbol{X}_n) \to 0$$

とする．ただし，$\boldsymbol{X}_n' = (\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \dots\,,\, \boldsymbol{x}_n)$ 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)　(4.10) 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．

一致性　このとき

$$\mathbb{P}(\dot{l}_n(\hat{\boldsymbol{\beta}}_n) = 0) \to 1, \qquad \hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta} \tag{4.20}$$

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
<u>一致性</u>　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 (4.20) 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

$$\begin{aligned}\mathbb{P}\{\ &\text{すべての } \boldsymbol{\gamma} \in \partial B_n(a)\,(n \ge n_0) \text{ とすべての } \phi > 0 \text{ に対し} \\ &\log L_n(\boldsymbol{\gamma},\, \phi) - \log L_n(\boldsymbol{\beta},\, \phi) < 0\} \ge 1 - \epsilon\end{aligned} \tag{4.21}$$

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，

$$\max_{\boldsymbol{\gamma} \in B_n(a)}\left\|[I_n(\boldsymbol{\beta})]^{-1/2}\frac{\ddot{l}_n(\boldsymbol{\gamma}^*)}{\phi}[I_n(\boldsymbol{\beta})]^{-1/2}\right\| \xrightarrow{P} 0 \tag{4.22}$$

を示すことにする．(4.16) と (4.17) に注意すれば，(4.22) を示すためには

$$\max_{\boldsymbol{\gamma} \in \partial B_n(a)}\|[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\gamma}) - M_n(\boldsymbol{\beta})][M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.23}$$

と

$$\max_{\boldsymbol{\gamma} \in \partial B_n(a)}\|[M_n(\boldsymbol{\beta})]^{-1/2}R_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.24}$$

を示せばよいことがわかる．まず，(4.23) 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

$$\begin{aligned}
&\mathrm{tr}\{[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\gamma}) - M_n(\boldsymbol{\beta})][M_n(\boldsymbol{\beta})]^{-1/2}\} \\
&\quad = \mathrm{tr}\{(\boldsymbol{X}_n'\Phi(\boldsymbol{\beta})\boldsymbol{X}_n)^{-1}[\boldsymbol{X}_n'\Phi(\boldsymbol{\gamma})\boldsymbol{X}_n - \boldsymbol{X}_n'\Phi(\boldsymbol{\beta})\boldsymbol{X}_n]\}^2 \\
&\quad = \mathrm{tr}\{(G'H'\Phi(\boldsymbol{\beta})HG)^{-1}[G'H'(\Phi(\boldsymbol{\gamma}) - \Phi(\boldsymbol{\beta}))HG]\}^2 \\
&\quad \leq q\frac{t_\infty^2 \max_{\boldsymbol{\gamma}\in B_n(a)} \sup_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{t_0^2 \min_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|^2}
\end{aligned}$$

となるので，(4.23) の上限は

$$\sqrt{q}\frac{t_\infty}{t_0}\frac{\max_{\boldsymbol{\gamma}\in B_n(a)} \sup_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})|}{\inf_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|} \to 0, \qquad (n \to \infty)$$

となる．なぜならば，$\boldsymbol{\gamma} \in B_n(a)$ に対し，$n \to \infty$ のとき

$$\begin{aligned}
|\boldsymbol{x}_i'\boldsymbol{\gamma} - \boldsymbol{x}_i'\boldsymbol{\beta}|^2 &\leq |\boldsymbol{x}_i'[I_n(\boldsymbol{\beta})]^{-1/2}[I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})|^2 \\
&\leq \|[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{x}_i\|^2\|[I_n(\boldsymbol{\beta})]^{1/2}(\boldsymbol{\gamma} - \boldsymbol{\beta})\|^2 \\
&\leq a^2 \max_{i=1,2,\dots,n} \boldsymbol{x}_i'[I_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i \qquad (4.25) \\
&\leq a^2 \max_{i=1,2,\dots,n} \boldsymbol{x}_i'[M_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i \\
&\leq a^2 \frac{1}{t_0 \min_{i=1,2,\dots,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})} \max_{i=1,2,\dots,n} \boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \to 0, \qquad (4.26)
\end{aligned}$$

となることと φ の連続性よりわかる．このことより，(4.23) は示せた．

つぎに，(4.24) 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

$$\max_{\boldsymbol{\gamma}\in B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}U_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \qquad (4.27)$$

がわかる．これは

$$\mathrm{tr}\{[M_n(\boldsymbol{\beta})]^{-1/2}U_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\}^2 \leq q\frac{t_\infty^2 \max_{\boldsymbol{\gamma}\in B_n(a)} \sup_{i=1,2,\dots,n} |\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma}) - \ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{t_0^2 \inf_{i=1,2,\dots,n} |\varphi(\boldsymbol{x}_i\boldsymbol{\beta})|^2}$$

になることに注意すれば，(4.26) と $\ddot{h}$ の連続性よりわかる．つぎに

$$\max_{\boldsymbol{\gamma}\in B_n(a)} \|[M_n(\boldsymbol{\beta})]^{-1/2}V_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.28}$$

を示す．まず

$$\begin{aligned}\mathbb{E}\left[\left\{\sum_{i=1}^n e_i[\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})-\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]t_i\boldsymbol{x}_i\boldsymbol{x}_i'\right\}^2\right] &= \sum_{i=1}^n[\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})-\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2t_i^2(\boldsymbol{x}_i\boldsymbol{x}_i')^2\mathbb{E}[e_i^2]\\ &= \sum_{i=1}^n\phi_i t_i^2[\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})-\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2t_i^2(\boldsymbol{x}_i\boldsymbol{x}_i')^2\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})\end{aligned}$$

となることに注意する．これから，$n\to\infty$ の

$$\begin{aligned}&\mathbb{E}[\mathrm{tr}\{M_n^{-1}(\boldsymbol{\beta})V_n(\boldsymbol{\gamma})M_n^{-1}(\boldsymbol{\beta})V_n(\boldsymbol{\gamma})\}] \le \mathbb{E}\left[\frac{q^2\mathrm{tr}\{V_n^2(\boldsymbol{\gamma})\}}{\{\mathrm{tr}M_n(\boldsymbol{\beta})\}^2}\right]\\ &\quad\le q^2\phi t_\infty\frac{\sum_{i=1}^n\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'\boldsymbol{x}_i)^2\max_{\boldsymbol{\gamma}\in B_n(a)}|\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})-\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})|^2}{\{\sum_{i=1}^n[\dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})]^2\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})t_i\boldsymbol{x}_i'\boldsymbol{x}_i\}^2}\\ &\quad\le\frac{q^2\phi t_\infty}{t_0^2}\sup_{i=1,2,\ldots,n}\max_{\boldsymbol{\gamma}\in B_n(a)}\left|\frac{\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\gamma})}{\ddot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})}-1\right|^2\frac{\sup_{i=1,2,\ldots,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}{\inf_{i=1,2,\ldots,n}\varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}\frac{\sum_{i=1}^n(\boldsymbol{x}_i'\boldsymbol{x}_i)^2}{(\sum_{i=1}^n\boldsymbol{x}_i'\boldsymbol{x}_i)^2}\\ &\quad\to 0\end{aligned}\tag{4.29}$$

となることが $\ddot{h}$ の連続性と (4.26) よりわかる．(4.29) と Markov の不等式より (4.28) は示せた．つぎに

$$\|[M_n(\boldsymbol{\beta})]^{-1/2}W_n[M_n(\boldsymbol{\beta})]^{-1/2}\| \xrightarrow{P} 0 \tag{4.30}$$

を示す．

$$\begin{aligned}&\mathbb{E}[\|[M_n(\boldsymbol{\beta})]^{-1/2}W_n[M_n(\boldsymbol{\beta})]^{-1/2}\|^2]\\ &\quad\le\inf_{i=1,2,\ldots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbb{E}[\{\sum_{i=1}^n h(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i\}^2]\\ &\quad=\inf_{i=1,2,\ldots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n h^2(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2\mathbb{E}[e_i]\\ &\quad=\inf_{i=1,2,\ldots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n h^2(\boldsymbol{x}_i'\boldsymbol{\beta})(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2\frac{\phi}{t_i}\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})\\ &\quad\le\frac{\phi}{t_0}\inf_{i=1,2,\ldots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sup_{i=1,2,\ldots,n}\varphi^2(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{i=1}^n(\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i)^2\\ &\quad\le\frac{q\phi}{t_0}\inf_{i=1,2,\ldots,n}\varphi^{-2}(\boldsymbol{x}_i'\boldsymbol{\beta})\sup_{i=1,2,\ldots,n}\varphi^2(\boldsymbol{x}_i'\boldsymbol{\beta})\max_{i=1,2,\ldots,n}\boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i\to 0\end{aligned}$$

となる．したがって，Markov の不等式を用いれば，(4.30) は示せた．よって，(4.27), (4.28), (4.30) より (4.24) は示せた．したがって，(4.23) と (4.24) が示せたので，(4.22) 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

$$\max_{\boldsymbol{\lambda}\in\partial B_n(a)} |\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{\lambda}|^2 \leq \max_{\boldsymbol{\lambda}\in\partial B_n(a)} \|\boldsymbol{\lambda}\|^2 \dot{l}_n'(\boldsymbol{\beta})[I_n(\boldsymbol{\beta})]^{-1}\dot{l}_n(\boldsymbol{\beta})$$

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

$$\mathbb{P}\{\|\boldsymbol{\lambda}'[I_n(\boldsymbol{\beta}]^{-1/2}\boldsymbol{\lambda}\frac{\dot{l}_n(\boldsymbol{\beta})}{\phi}\| \leq \frac{a}{4}\} \geq 1 - \left(\frac{4}{a\phi}\right)^2 \mathbb{E}[\dot{l}_n'(\boldsymbol{\beta})[I_n(\boldsymbol{\beta})]^{-1}\dot{l}_n(\boldsymbol{\beta})] = 1 - \frac{16q}{(a\phi)^2} \geq 1 - \epsilon$$

となる．したがって，(4.22) 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．(4.22) 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

$$A_\epsilon = \{\boldsymbol{\gamma} \in \mathcal{B} :\ \|\boldsymbol{\gamma} - \boldsymbol{\beta}\| \leq \epsilon\}$$

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

$$\begin{aligned}
&\left\|[M_n(\boldsymbol{\beta})]^{-1/2} \int_0^1 \ddot{l}_n(\boldsymbol{\beta} + t(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}))\, dt[M_n(\boldsymbol{\beta})]^{-1/2} + I_q\right\| \\
&\quad \leq \max_{\boldsymbol{\gamma}\in A_\epsilon} \left\|[M_n(\boldsymbol{\beta})]^{-1/2}[R_n(\boldsymbol{\gamma}) + M_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\gamma})][M_n(\boldsymbol{\beta})]^{-1/2}\right\| \\
&\quad \leq \max_{\boldsymbol{\gamma}\in A_\epsilon} \left\|[M_n(\boldsymbol{\beta})]^{-1/2}[M_n(\boldsymbol{\beta}) - M_n(\boldsymbol{\gamma})][M_n(\boldsymbol{\beta})]^{-1/2}\right\| \\
&\quad + \max_{\boldsymbol{\gamma}\in A_\epsilon} \left\|[M_n(\boldsymbol{\beta})]^{-1/2} R_n(\boldsymbol{\gamma})[M_n(\boldsymbol{\beta})]^{-1/2}\right\| \\
&\quad \xrightarrow{P} 0
\end{aligned}$$

となることが (4.23) と (4.24) 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，

$$\left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2} \frac{\dot{l}_n(\boldsymbol{\beta})}{\phi} = \left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{1/2} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + o_P(1) \tag{4.31}$$

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

$$\boldsymbol{\alpha}' \left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2} \frac{\ddot{l}_n(\boldsymbol{\beta})}{\phi} = \sum_{i=1}^n \boldsymbol{\alpha}' \left[\frac{M_n(\boldsymbol{\beta})}{\phi}\right]^{-1/2} \frac{t_i \dot{h}(\boldsymbol{x}_i'\boldsymbol{\beta})\sqrt{\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}}{\phi} \boldsymbol{x}_i \frac{Y_i - \mu(h(\boldsymbol{x}_i'\boldsymbol{\beta}))}{\sqrt{\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}}$$

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

$$\begin{aligned}
&\max_{i=1,2,\ldots,n} \left\{ \frac{t_i^2 \dot{h}^2(\boldsymbol{x}_i'\boldsymbol{\beta})\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}{\phi} [\boldsymbol{\alpha}'[M_n(\boldsymbol{\beta})]^{-1}\boldsymbol{x}_i]^2 \right\} \Bigg/ \sum_{i=1}^n \left\{ \frac{t_i^2 \dot{h}^2(\boldsymbol{x}_i'\boldsymbol{\beta})\ddot{b}(\boldsymbol{x}_i'\boldsymbol{\beta})}{\phi} [\boldsymbol{\alpha}'[M_n(\boldsymbol{\beta})]^{-1/2}\boldsymbol{x}_i]^2 \right\} \\
&\quad \leq \max_{i=1,2,\ldots,q} a_i^2 \left\{ \frac{t_\infty^2 \sup_{i=1,2,\ldots,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})}{t_0 \inf_{i=1,2,\ldots,n} \varphi(\boldsymbol{x}_i'\boldsymbol{\beta})} \right\}^2 \max_{i=1,2,\ldots,n} \boldsymbol{x}_i'(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{x}_i \Big/ \boldsymbol{\alpha}'\boldsymbol{\alpha}\phi I_q \\
&\quad \to 0
\end{aligned}$$

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，(4.31) 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．　□

<u>最尤推定値の計算アルゴリズム</u>

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

$$S(\boldsymbol{\beta}) \equiv \frac{\partial}{\partial \boldsymbol{\beta}} \log L_n(\boldsymbol{\beta},\, \phi) = \frac{1}{\phi} \sum_{i=1}^n \frac{\{y_i - g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})\} t_i}{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))} \boldsymbol{x}_i = \mathbf{0}$$

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．(4.11) と (4.12)

から

$$
\begin{aligned}
H(\boldsymbol{\beta}) \equiv \mathbb{VAR}[S(\boldsymbol{\beta})] &= \frac{1}{\phi^2}\sum_{i=1}^{n}\left\{\frac{t_i^2}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2}\mathbb{VAR}[Y_i - g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})]\right\}\boldsymbol{x}_i\boldsymbol{x}_i' \\
&= \frac{1}{\phi^2}\sum_{i=1}^{n}\left\{\frac{t_i^2}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2}\frac{\phi}{t_i}V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\right\}\boldsymbol{x}_i\boldsymbol{x}_i' \\
&= \frac{1}{\phi}\sum_{i=1}^{n}\frac{t_i}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))\}^2 V(g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}))}\boldsymbol{x}_i\boldsymbol{x}_i'
\end{aligned}
$$

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

$$
\widehat{\boldsymbol{\beta}}_n^{(k+1)} = \widehat{\boldsymbol{\beta}}_n^{(k)} + H^{-1}(\widehat{\boldsymbol{\beta}}_n^{(k)})S(\widehat{\boldsymbol{\beta}}_n^{(k)}) \tag{4.32}
$$

となる．(4.32) 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，

$$
\begin{aligned}
w_i^{(k)} &= \frac{t_i}{\{\dot{g}(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))\}^2 V(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))} \\
z_i^{(k)} &= \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)} + \dot{g}(g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))(y_i - g^{-1}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n^{(k)}))
\end{aligned}
$$

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

いろいろな話題

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，$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{L} N(\mathbf{0}, \Sigma)$ ならば，

$$\sqrt{n}(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{L} N(\mathbf{0}, \dot{g}(\boldsymbol{\mu})\,\Sigma\,\dot{g}(\boldsymbol{\mu})')$$

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，$\boldsymbol{X}_n - \boldsymbol{\mu} \xrightarrow{P} \mathbf{0}$ から
$$h(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{P} h(\mathbf{0}) = \int_0^1 \dot{g}(\boldsymbol{\mu} + t\mathbf{0})\,dt = \dot{g}(\boldsymbol{\mu})$$
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

$$\left[\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots \frac{S_n}{S_{n+1}}\right] \tag{5.1}$$

は独立で，S_{n+1} を与えたときの (5.1) 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，

$$g(x_1,\, x_2,\, x_3) = \frac{1}{x_1+x_2+x_3}\begin{bmatrix} x_1 \\ x_1+x_2 \end{bmatrix} \tag{5.2}$$

とおく．このとき，

$$g(\frac{S_{k_1}}{n+1}, \frac{S_{k_2}-S_{k_1}}{n+1}, \frac{S_{n+1}-S_{k_2}}{n+1}) = \frac{1}{S_{n+1}}\begin{bmatrix} S_{k_1} \\ S_{k_2} \end{bmatrix}$$

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，

$$\Sigma = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2-p_1 & 0 \\ 0 & 0 & 1-p_1 \end{bmatrix}$$

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，(5.2) 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あっているので，$\sigma^2 < \pi\sigma^2/2$ 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，$\dot{g}(x_1,\, x_2) = [-1/2,\, 1/2]$ となり，

$$\dot{g}(x_1,\, x_2)\frac{\pi^2\sigma^2}{4}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\dot{g}'(x_1,\, x_2) = \frac{\pi^2\sigma^2}{4}$$

よりわかる．

A

補遺

·····A.1·····

基本的な確率不等式

定理 A.1　非負確率変数 $X \geq 0,\ a.s.$ と正数 $p > 0$ に対して,

$$\mathbb{E}X^p = \int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt$$

が成立する.

証明　Fubini の定理を用いて変形すれば,

$$\begin{aligned}\int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt &= \int_0^\infty pt^{p-1}\mathbb{E}1_{(t,\infty)}(X)\,dt = \mathbb{E}\int_0^\infty pt^{p-1}1_{(t,\infty)}(x)\,dt \\ &= \mathbb{E}\int_0^X pt^{p-1}\,dt = \mathbb{E}X^p\end{aligned}$$

から命題は証明される.　□

例 A.1　X を非負値離散型確率変数で $x_1, x_2, \ldots,$ 上に値を取り

$$\mathbb{P}(X = x_i) = p_i, \qquad x_i > 0, \qquad i = 1, 2, \ldots, \qquad \sum_{i=1}^\infty p_i = 1$$

とする．ただし,

$$\sum_{i=1}^\infty p_i x_i < \infty \tag{1.1}$$

とする．さらに,

$$I_{[x_i,\infty)}(x) = \begin{cases} 1 & x \geq x_i, \\ 0 & \text{（その他）} \end{cases}$$

とおく．このとき, X の分布関数は

$$F_X(x) = \sum_{i=1}^\infty p_i I_{[x_i,\infty)}(x)$$

となることに注意すれば,

$$\mathbb{P}(X > t) = 1 - \sum_{i=1}^\infty p_i I_{[x_i,\infty)}(x) = \sum_{i=1}^\infty p_i I_{(-\infty,x_i)}(x)$$

となる．したがって，

$$\int_0^\infty \mathbb{P}(X > t)\,dt = \sum_{i=1}^\infty p_i \int_0^\infty I_{(-\infty,\,x_i)}(t)\,dt$$

となることがわかる．積分記号と和の記号の交換は (1.1) から成立することがわかる．しかし

$$\int_0^\infty I_{(-\infty,\,x_i)}(t) = \lim_{s\uparrow x_i} \int_0^s I_{(-\infty,\,x_i)}(t)\,dt + \lim_{s\downarrow x_i} \int_s^\infty I_{(-\infty,\,x_i)}(t)\,dt = x_i$$

となる．これより

$$\int_0^\infty \mathbb{P}(X > t)\,dt = \sum_{i=1}^\infty p_i x_i$$

となり，離散型確率変数の期待値と一致することが直接確認できる．

定理 A.2　（Markov の不等式）　$X \geq 0$, $a.s.$ とする．任意の $a > 0$ に対して，

$$\mathbb{P}(X \geq a) = \frac{\mathbb{E}X}{a}$$

が成立する．

証明　$X \in [a, \infty)$ のとき，$X/a \geq 1$ に注意する：

$$\mathbb{P}\{X \geq a\} = \mathbb{E}[\mathbf{1}_{[a,\,\infty)}(X)] \leq \mathbb{E}[\frac{X}{a}\mathbf{1}_{[a,\,\infty)}(X)] \leq \frac{\mathbb{E}X}{a}.$$

□

注意 A.1　Y を確率変数とする．$X = (Y - \mathbb{E}Y)^2$ とおけば，Chebyshev の不等式

$$\mathbb{P}\{|Y - \mathbb{E}Y| \geq a\} = \mathbb{P}\{(Y - \mathbb{E}Y)^2 \geq a^2\} \leq \frac{\mathbb{VAR}(Y)}{a^2}$$

を得る．

定理 A.3　（Jensen の不等式）　g は下に凸 (convex) とし，X と $g(X)$ は可積分とする．このとき，

$$g(\mathbb{E}X) \leq \mathbb{E}g(X)$$

が成立する．

証明　任意の $x_0 \in \mathbb{R}$ とある定数 c に対し，$g(x) \geq g(x_0) + c(x - x_0)$ が成立する．$x = X(\omega)$ として，期待値をとれば，$\mathbb{E}g(x) \geq g(x_0) + c(\mathbb{E}X - x_0)$ を得る．さらに，$x_0 = \mathbb{E}X$ とすれば，命題は証明される．□

定理 A.4　(Hölder の不等式) X, Y を確率変数で $\mathbb{E}[|X|^p] < \infty, \mathbb{E}[|Y|^q] < \infty$ をみたすものとする．ただし，$p > 1$ かつ $1/p + 1/q = 1$ である．このとき，

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq \{\mathbb{E}[|X|^p]\}^{1/p}\{\mathbb{E}[|Y|^q]\}^{1/q}$$

が成立する．

証明　一般性を失わず，$X \geq 0, Y \geq 0$ を仮定してよい．さらに，$\mathbb{E}[X^p] = 0$ ならば，$X = 0, a.s.$ なり，この場合には，Hólder の不等式は明らかに成立するので，$\mathbb{E}[X^p] > 0$ を仮定する．

いま，$C = \mathbb{E}[X^p]$ とし，確率測度 Q を

$$Q(A) = \frac{1}{C}\mathbb{E}[1_A X^p], \qquad A \in \mathcal{B}$$

で定義する．つぎに，

$$Z = \frac{Y}{X^{p-1}} 1_{\{X>0\}}$$

とおく．$g(x) = |x|^p$ は凸関数なので，Jensen の不等式を用いれば，

$$\{\mathbb{E}_Q[Z]\}^p \leq \mathbb{E}_Q[Z^p]$$

となる．これより，

$$\begin{aligned}
\frac{1}{C^q}\{\mathbb{E}[XY]\}^q &= \frac{1}{C^q}\left\{\mathbb{E}\left[\frac{Y}{X^{p-1}}X^p\right]\right\}^q \\
&= \{\mathbb{E}_Q[Z]\}^q \\
&\leq \mathbb{E}_Q[Z^q] \\
&= \frac{1}{C}\mathbb{E}\left[\left\{\frac{Y}{X^{p-1}}\right\}^q X^p\right] \\
&= \frac{1}{C}\mathbb{E}\left[\left\{Y^q\frac{1}{X^{(p-1)q}}\right\}^q X^p\right] \\
&= \frac{1}{C}\mathbb{E}\left[\left\{Y^q\frac{1}{X^p}\right\}^q X^p\right] \\
&= \frac{1}{C}\mathbb{E}[Y^q]
\end{aligned}$$

となる．最後から 2 番目の等号は $1/p + 1/q = 1$ より $(p-1)q = p$ となることよりわかる．よって，

$$\{\mathbb{E}[XY]^q \leq C^{q-1}\mathbb{E}[Y^q]$$

となるので，

$$\mathbb{E}[XY] \leq C^{(q-1)/q}\{\mathbb{E}[Y^q]\}^{1/q}$$

を得る．最後に，$(q-1)/q = 1/p$ に注意すれば，定理は証明される．　□

·····A.2·····

収束のモード

定義 A.1　$\{X_n\}_{n=1}^n$ は確率変数列とし，X を確率変数とする．$n \to \infty$ のとき，$\{X_n\}_{n=1}^n$ が X に分布収束するとは，

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x) = F_X(x)$$

が成立することである．ただし，x は $F_X(x)$ の連続点とする．これを $X_n \xrightarrow{d} X$ または $X_n \xrightarrow{L} X$ と記すことにする．

定義 A.2　$\{X_n\}_{n=1}^n$ は確率変数列とし，X を確率変数とする．$n \to \infty$ のとき，$\{X_n\}_{n=1}^n$ が X に確率収束するとは，任意の $\epsilon\,(\epsilon > 0)$ に対して，

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

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

定義 A.3　$\{X_n\}_{n=1}^n$ は確率変数列，X を確率変数とする．$n \to \infty$ のとき，$\{X_n\}_{n=1}^n$ が X に概収束するとは，

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

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

例 A.2　$X_1, X_2, \ldots, X_n$ は独立同一に一様分布 $f_X(x) = I_{(0,1)}(x)$ に従うとし，

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

とおく．

まず，$M_n \xrightarrow{P} 1$ を示そう．M_n の累積分布関数は

$$F_{M_n}(x) = P(M_n \leq x) = x^n I_{(0,1)}(x) + I_{[1,\infty)}(x)$$

となることに注意する．これより，$0 < \epsilon < 1$ に対して，$P(M_n > 1 + \epsilon) = 0$ に注意すれば，$n \to \infty$ のとき

$$P(|M_n - 1| > \epsilon) = P(\epsilon + 1 < M_n \quad \text{または} \quad M_n < 1 - \epsilon) = P(M_n < 1 - \epsilon) = (1-\epsilon)^n \to 0$$

となる．したがって，$M_n \xrightarrow{P} 1$ がわかる．

つぎに，$n(1-M_n)$ の分布収束先を求めよう．$x>0$ に対して，$n\to\infty$ のとき

$$P(n(1-M_n)\le x)=P(M_n\ge 1-x/n)=1-\left(1-\frac{x}{n}\right)^n\to 1-e^{-x}$$

となる．$M_n<1$ より，明らかに $x\le 0$ に対しては $P(n(1-M_n)\le x)=0$ となる．したがって，$n(1-M_n)$ は指数分布 $f_X(x)=e^{-x}I_{(0,\infty)}(x)$ に分布収束する．

定理 A.5　$X_n \xrightarrow{P} c$ となるための十分条件は

$$\mathbb{E}(X_n-c)^2\to 0,\qquad (n\to\infty)\tag{1.2}$$

である．

証明　与えられたどんな正の数 $\epsilon>0$ に対しても

$$P(|X_n-c|\ge\epsilon)\le\frac{1}{\epsilon^2}\mathbb{E}(X_n-c)^2$$

となり，$n\uparrow\infty$ とすれば，

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

を得る．□

注意 A.2　(1.2) が成立するとき，X_n は c に**2次平均収束**するという．

例 A.3　$\epsilon_1,\ldots,\epsilon_n$ 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，

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

となるので，

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

となることがわかる．

定理 A.6　$X_1,\ldots,X_n$ は独立同一の分布に従い，その平均と分散は $\mathbb{E}(X_1)=\xi$ と $\mathbb{VAR}(X_1)=\sigma^2<\infty$ 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，

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

となることから，定理 A.5 を使えば，定理は証明される．□

以下に確率変数列の収束に関する重要な結果を証明なしで述べる．

定理 A.7

(i) $X_n \xrightarrow{a.s.} X$ ならば，$X_n \xrightarrow{P} X$
(ii) $X_n \xrightarrow{P} X$ ならば，$X_n \xrightarrow{d} X$

定理 A.8　$g(x)$ を実数値連続関数とする．このとき，

(i) $X_n \xrightarrow{P} X$ ならば，$g(X_n) \xrightarrow{P} g(X)$
(ii) $X_n \xrightarrow{a.s.} X$ ならば，$g(X_n) \xrightarrow{a.s.} g(X)$
(iii) $X_n \xrightarrow{d} X$ ならば，$g(X_n) \xrightarrow{d} g(X)$

定理 A.9　確率変数列 $\{X_n\}_{n=1}^{\infty}$ と $\{Y_n\}_{n=1}^{\infty}$ および定数 c は定数に対して，$X_n \xrightarrow{d} X$ と $Y_n \xrightarrow{P} c$ が成立するとする．このとき，

(i) $X_n + Y_n \xrightarrow{d} X + c$ が成立する．
(ii) $X_n Y_n \xrightarrow{d} cX$ が成立する．

定理 A.10　確率変数列 $\{X_n\}_{n=1}^{\infty}$，数列，確率変数 Z，確率変数列 $\{a_n\}_{n=1}^{\infty}$ および定数 c に対して

$$a_n(X_n - c) \xrightarrow{d} Z$$

が成立するとする．このとき，$g(x)$ は $x = c$ で微分 $\dot{g}(c)$ を持つ[1)]ならば，

$$a_n(g(X_n) - g(c)) \xrightarrow{d} \dot{g}(c)Z$$

が成立する．

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

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

[1)] $g(x)$ が $x = c$ で連続微分可能でなくともこの定理　　は成立することに注意する．

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．

⋯⋯A.3⋯⋯

大数の法則と中心極限定理

ここでは，大数の法則と中心極限定理を証明なしで述べる．さらに，対称分布からのランダム標本の基づくメディアンは分布の中心に確率収束し，漸近正規性をもつことをこの二つの定理を用いて示す．

定理 A.11　$X_1, X_2, \ldots, X_n$ は独立に同一の分布[1)] に従うとする．また，

$$\mathbb{E}|X_1| < \infty, \qquad \mathbb{E}[X_1] = \xi$$

であるとする．このとき，

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

となる．

証明　略．

定理 A.12　$\{X_n\}_{n\geq 1}$ は独立同一の分布に従う確率変数の列とし，

$$\mathbb{E}(X_1) = \xi, \qquad \mathbb{VAR}(X_1) = \sigma^2$$

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，

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

である．また，

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

---

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．また，

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

より

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

となり，定理 1.15 からわかる．□

定理 A.13　$X_{n1},\,X_{n2},\,\ldots\,,\,X_{nn}$ は独立同一分布に従い，各 $X_{ni},\,i=1,\,2,\,\ldots\,,\,n,$ は平均 ξ_n と分散 σ_n^2 なる分布関数 $F_n(x)$ をもち，さらに 3 次の積率も有限とする．このとき，

$$\frac{\mathbb{E}_n\{|X_{n1}-\xi_n|^3\}}{\sigma_n^3}=o(\sqrt{n}) \tag{1.3}$$

ならば，

$$\frac{\sqrt{n}(\bar{X}_n-\xi_n)}{\sigma_n}\xrightarrow{d}N(0,\,1)$$

となる．特に，(1.3) が有限ならば，上の結果は成立する．ただし，$\mathbb{E}_n$ は F_n に関する期待値で，$\bar{X}_n=(1/n)(X_{n1}+X_{n2}+\cdots+X_{nn})$ である．

証明　これは Berry - Essenn の定理

$$\mathbb{P}\left[\frac{\sqrt{n}(\bar{X}_n-\xi_n)}{\sigma_n}\leq\right]-\Phi(x)\leq\frac{C}{\sqrt{n}}\frac{\mathbb{E}_n\{|X_{n1}-\xi_n|^3\}}{\sigma_n^3}$$

よりわかる．ただし，$\Phi(x)$ は標準正規分布の分布関数である．□

例 A.4　$X_1,X_2,\ldots,X_n$ を $F_\theta(x)=P_\theta(X_1\leq x)=F(x-\theta)\,(x\in\mathbb{R})$ からのランダム標本とする．ただし，θ は未知の定数とし，$F(x)$ は狭義単調増加で連続とし，$F(0)=1/2$ とする．$X_1,X_2,\ldots,X_n$ に基づいて，メディアン θ の推定を考える．

いま，

$$m_n=\begin{cases}\frac{n+1}{2}, & (n\text{ が奇数のとき})\\ \frac{n}{2}, & (n\text{ が偶数のとき})\end{cases}$$

とおく．順序統計量を $X_{(1)}\leq\cdots\leq X_{(n)}$ としたとき，

$$Z_n=X_{(m_n)}$$

を**標本メディアン**という．これを用いて，メディアン θ を推定することを考える．標本メディアン Z_n の漸近的な性質を調べてみよ．

まず，$Z_n\xrightarrow{P}\theta$ を示そう．任意の $\epsilon>0$ に対して，標本メディアンの定義から

$$P(Z_n>\theta+\epsilon)=P\left(\frac{1}{n}\sum_{i=1}^n I\{X_i>\theta+\epsilon\}\geq\frac{n-m_n+1}{n}\right)$$

$$P(Z_n<\theta-\epsilon)=P\left(\frac{1}{n}\sum_{i=1}^n I\{X_i>\theta-\epsilon\}\leq\frac{n-m_n}{n}\right)$$

となる．いま，

$$2\delta = \min\left\{F_\theta(\theta+\epsilon) - \frac{1}{2}, \frac{1}{2} - F_\theta(\theta-\epsilon)\right\}$$

とおけば，$F_\theta(\theta) = 1/2$ に注意すれば，$\delta > 0$ となることがわかる．

$$A_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} - (1 - F_\theta(\theta+\epsilon))\right| \le \delta\right\}$$

$$B_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} - (1 - F_\theta(\theta-\epsilon))\right| \le \delta\right\}$$

とおけば，大数の法則から $n \to \infty$ のとき

$$P(A_n) \to 1, \qquad P(B_n) \to 1 \tag{1.4}$$

となる．また，$\frac{1}{n} < \delta$ ならば，A_n が起これば，

$$\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \le \delta + 1 - F_\theta(\theta+\epsilon) = \delta + \frac{1}{2} + \frac{1}{2} - F_\theta(\theta+\delta) \le \frac{1}{2} - 2\delta + \delta$$

$$\le \delta + \frac{1}{2} - 2\delta = \frac{1}{2} - \delta < \frac{n - m_n + 1}{n}$$

となることが $2\delta \le F_\theta(\theta+\epsilon) - (1/2)$ に注意すればわかる．これより，事象

$$\left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \ge \frac{n - m_n + 1}{n}\right\}$$

と A_n は排反になることがわかる．したがって，

$$C_n = A_n \cap \left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \ge \frac{n - m_n + 1}{2}\right\} = \emptyset$$

となることと (1.4) から，$n \to \infty$ のとき，

$$P(Z_n > \theta+\epsilon) \le P(C_n) + P(A_n^c) \to 0$$

となる．同様に，$\frac{1}{n} < \delta$ ならば，B_n が起これば，

$$\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} \ge -\delta + 1 - F_\theta(\theta-\epsilon) = -\delta + \frac{1}{2} + \frac{1}{2} - F(\theta-\delta) \ge -\delta + \frac{1}{2} + 2\delta$$

$$= \frac{1}{2} + \delta > \frac{n - m_n}{n}$$

となることが $2\delta \le (1/2) - F_\theta(\theta-\epsilon)$ に注意すればわかる．これより，事象

$$\left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} \le \frac{n - m_n}{n}\right\}$$

と B_n は排反になることがわかる．したがって，

$$D_n = B_n \cap \left\{ \frac{1}{n} \sum_{i=1}^{n} I\{X_i > \theta - \epsilon\} \le \frac{n - m_n}{n} \right\} = \emptyset$$

となることがわかる．さらに，(1.4) から，$n \to \infty$ のとき，

$$P(Z_n < \theta - \epsilon) \le P(D_n) + P(B_n^c) \to 0$$

となる．これらから，$n \to \infty$ のとき

$$P(|Z_n - \theta| > \epsilon) \le P(Z_n > \theta + \epsilon) + P(Z_n < \theta - \epsilon) \to 0$$

となり，$Z_n \xrightarrow{P} \theta$ がわかる．

つぎに，$f = dF/dx$ とし，P_0 を $\theta = 0$ のときに対応する確率測度とする．すなわち，$P(X - \theta \le x) = P_0(X \le x)$ である．$f(0) > 0$ の仮定もとで，Z_n の極限分布を求める：

$$P[\sqrt{n}(Z_n - \theta) \le t] = P_0[\sqrt{n} Z_n \le t] = P_0 \left[X_{(m_n)} \le \frac{t}{\sqrt{n}} \right]$$

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

$$X_{(m_n)} \le \frac{t}{\sqrt{n}} \Leftrightarrow S_n \le n - m_n$$

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，

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

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

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

となることに注意する．いま，$Y_{ni} = I\{X_i \ge t/\sqrt{n}\}$, $i = 1, 2, \dots, n$, とおけば，$Y_{n1}, Y_{n2}, \dots, Y_{nn}$ は独立同一分布に従い，$P_0(Y_{ni} = 1) = p_n$ かつ $P_0(Y_{ni} = 0) = 1 - p_n$ となる．また，$S_n = \sum_{i=1}^{n} Y_{ni}$ となる．定理 $A.13$ を用いて

$$\frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}} \xrightarrow{d} N(0, 1)$$

を示す．そのために，(1.3) を確認する：$n \to \infty$ のとき，$p_n \to (1/2)$ になることに注意すれば，

$$\mathbb{E}|Y_1 - p_n|^3 = p_n(1 - p_n)^3 + (1 - p_n)p_n^3 = p_n(1 - p_n)\{p_n^2 + (1 - p_n)^2\} \to \frac{1}{8}$$

と $\sigma_n = \sqrt{p_n(1-p_n)} \to (1/2)$ となるので，

$$\frac{\mathbb{E}|Y_1 - p_n|^3}{\sigma_n^3} \to 1$$

がわかる．よって，(1.3) を確認は確認された．したがって，

$$P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \leq 2tf(0)\right] \to \Phi(2f(0)t)$$

となる．ただし，$\Phi(\cdot)$ は標準正規分布の分布関数である．さらに，Slutsky の補題（定理 $A.9$ ）を用いれば

$$\begin{aligned}
&P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \leq \frac{(n-m_n) - np_n}{\sqrt{np_n(1-p_n)}}\right] \\
&= P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} + \left\{(2f(0)t) - \frac{(n-m_n) - np_n}{\sqrt{np_n(1-p_n)}}\right\} \leq 2f(0)t\right] \\
&\to \Phi(2f(0)t)
\end{aligned}$$

となることがわかる．したがって

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

となる．したがって，

$$\sqrt{n}(Z_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{4f^2(0)}\right)$$

をえる．

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.