第7章　推定

7.1　モーメント法

この方法で得られた推定量は最適性を持たないことがある．しかし求めるのが容易であるので，陽には式が与えられない別の推定量の値を繰り返し計算で求めるときの初期値として用いることができる．

$d,\, n \in \mathbb{N}$ とし，統計的実験を $\bigl(\mathbb{R}^n,\, \bigl\{\mathsf{P}_{\boldsymbol{\theta}} = \mathsf{P}_{1,\boldsymbol{\theta}}^{\otimes n};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\bigr\}\bigr)$ を考え

$$X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d.}\, \mathsf{P}_{1,\boldsymbol{\theta}^*} \quad (\boldsymbol{\theta}^* \in \Theta)$$

とする．ただし，$\mathsf{P}_{\boldsymbol{\theta}}$ と $\mathsf{P}_{1,\boldsymbol{\theta}}$ は，それぞれ測度空間 $\bigl(\mathbb{R}^n,\, \mathcal{B}(\mathbb{R}^n)\bigr)$ と $\bigl(\mathbb{R},\, \mathcal{B}(\mathbb{R})\bigr)$ 上の分布である．

$n \geq d$ とする．$1 \leq j \leq d$ に対して

$$\alpha_j := \alpha_j(\boldsymbol{\theta}) := \mathsf{E}\bigl[X_1^j\bigr], \quad \widehat{\alpha}_j := \frac{1}{n}\sum_{\ell=1}^{n} X_\ell^j$$

と定める．

定義 7.1. $n \geq d$ とし $X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d.}\, \mathsf{P}_{1,\boldsymbol{\theta}^*}\ \bigl(\boldsymbol{\theta}^* \in \Theta\bigr)$ とする．このとき $\boldsymbol{\theta}^*$ の**モーメント法推定量** $\widehat{\boldsymbol{\theta}}_n$ とは次をみたす解 (存在すれば) である．

$$\alpha_1(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_1,\, \alpha_2(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_2, \ldots,\, \alpha_d(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_d \tag{7.1}$$

注意 7.2. $n \geq 2$ とする．$X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d.}\, \mathsf{N}\bigl(\mu^*,\, (\sigma^*)^2\bigr)$ とする．ただし $\boldsymbol{\theta}^* := (\mu^*,\, \sigma^*) \in \mathbb{R} \times (0,\, \infty) =: \Theta$ で $(\mu^*,\, \sigma^*)$ は未知とする．このとき

$$\alpha_1 = \mathsf{E}\bigl[X_1\bigr] = \mu^*, \quad \alpha_2 = \mathsf{E}\bigl[X_1^2\bigr] = \mathsf{Var}\bigl[X_1\bigr] + \bigl\{\mathsf{E}\bigl[X_1\bigr]\bigr\}^2 = (\sigma^*)^2 + (\mu^*)^2$$

である．したがって $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n,\, \widehat{\sigma}_n)$ とすれば，

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma}_n^2 + \bigl\{\widehat{\mu}_n\bigr\}^2 = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell^2$$

を解くと (μ^*, σ^*) のモーメント法推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{\ell=1}^{n}(X_\ell - \widehat{\mu}_n)}$$

となる. □

定理 7.3. $d, n \in \mathbb{N}$ は $n \geq d$ とする.

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d.}\, \mathsf{P}_{1,\boldsymbol{\theta}^*}\, (\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d), \quad \mathsf{E}\big[\big|X_1\big|^{2d}\big] < \infty$$

とする. (7.1) で定義した $\alpha_j\, (j = 1, 2, \ldots, d)$ の逆写像 α_j^{-1} が存在して, $\boldsymbol{\theta}^*$ の近傍で全微分可能とする. このとき適当な条件のもとで以下が成立する.

(1) $\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\, \widehat{\boldsymbol{\theta}}_n$ は存在$\big) \longrightarrow 0 \qquad (n \to \infty)$.

(2) $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\big|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big|_2 > \epsilon\big) \longrightarrow 0 \qquad (n \to \infty).$$

ただし $\big|\cdot\big|_2$ は $\mathbb{R}^d$ の Euclid ノルムである.

(3) $\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}, \boldsymbol{\Sigma}\big)$. ただし

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{G}\big(\boldsymbol{\theta}^*\big)\mathsf{E}_{\boldsymbol{\theta}^*}\big[\boldsymbol{Y}\boldsymbol{Y}^\top\big]\boldsymbol{G}^\top\big(\boldsymbol{\theta}^*\big), \\
\boldsymbol{G}\big(\boldsymbol{\theta}\big) &= \big(\dot{\boldsymbol{g}}_1(\boldsymbol{\theta}), \dot{\boldsymbol{g}}_2(\boldsymbol{\theta}), \ldots, \dot{\boldsymbol{g}}_d(\boldsymbol{\theta})\big), \\
\dot{g}_j(\boldsymbol{\theta}) &= \left(\frac{\partial \alpha_j^{-1}}{\partial \theta_1}(\boldsymbol{\theta}), \frac{\partial \alpha_j^{-1}}{\partial \theta_2}(\boldsymbol{\theta}), \ldots, \frac{\partial \alpha_j^{-1}}{\partial \theta_d}(\boldsymbol{\theta})\right)^\top (j = 1, 2, \ldots, d), \\
\boldsymbol{Y} &= \big(X_1, X_1^2, \ldots, X_1^d\big)^\top.
\end{aligned}$$

Proof. 証明は省略. □

7.2 最尤法

$d, n \in \mathbb{N}$ とする. 統計的実験 $\big(\mathbb{R}^n, \big\{\prod_{j=1}^n \mathsf{p}_1(x_j|\,\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\big\}\big)$ を考える. ただし $\mathsf{p}_1(x|\,\boldsymbol{\theta})$ は p.d.f. (または p.m.f.) で $\Theta \subset \mathbb{R}^d$ である. $\boldsymbol{\theta}^* \in \Theta$ とし

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d.}\, \mathsf{p}_1(x|\,\boldsymbol{\theta}^*)$$

とする.

定義 7.4. $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したとき $\boldsymbol{\theta}$ の **尤(ゆう)度関数** $\mathsf{lik}_n(\boldsymbol{\theta}|\, \boldsymbol{x})$ を

$$\mathsf{lik}_n\bigl(\boldsymbol{\theta}|\, \boldsymbol{x}\bigr) = \prod_{j=1}^{n} \mathsf{p}(x_j|\, \boldsymbol{\theta})$$

で定義し, **対数尤度** $\ell_n(\boldsymbol{\theta}|\, \boldsymbol{x})$ を

$$\ell_n(\boldsymbol{\theta}|\, \boldsymbol{x}) = \log \mathsf{lik}_n(\boldsymbol{\theta}|\, \boldsymbol{x})$$

で定義する. ただし, $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ である.

注意 7.5. 尤度関数は

$$\mathsf{lik}_n(\,\cdot\,|\, \boldsymbol{x}) :\, \Theta \ni \boldsymbol{\theta} \mapsto \mathsf{lik}_n(\boldsymbol{\theta}|\, \boldsymbol{x}) \in [0,\, \infty)$$

である. □

関数 $g(x)$ の最大値を取る点を表す集合を

$$\arg\max_{x \in \mathbb{R}} g(x)$$

と書く. たとえば $g(x) = -(x-1)^2$ のとき

$$\arg\max_{0 \le x \le 4\pi} g(x) = \{1\}$$

となる. $g(x) = \sin x$ のとき

$$\arg\max_{0 \le x \le 4\pi} g(x) = \{\pi/2,\, 5\pi/2\}$$

となる.

定義 7.6. $X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したとき $\boldsymbol{\theta}^*$ の**最尤推定値** (maximum likelihood estimate) を $\mathsf{lik}_n(\boldsymbol{\theta}|\, \boldsymbol{x})$ を最大にする値 $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ で定義する. すなわち

$$\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x}) \in \arg\max_{\boldsymbol{\theta} \in \Theta} \mathsf{lik}_n(\boldsymbol{\theta}|\, \boldsymbol{x})$$

である. $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ に $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ を代入したもの $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ を $\boldsymbol{\theta}^*$ の**最尤推定量** (maximum likelihood estimator=m.l.e.) という.

注意 7.7. $X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d.}\, \mathsf{Ber}(\theta^*)$ とする. ただし $(0,\, 1) =: \Theta \ni \theta^*$ は未知とする. すなわち

$$\mathsf{p}(x|\, \theta^*) = \begin{cases} (\theta^*)^x \bigl(1-\theta^*\bigr)^{1-x}, & (x = 0,\, 1), \\ 0 & (\text{その他の場合}) \end{cases}$$

である．このとき

$$\mathsf{lik}_n\big(\theta|\,\boldsymbol{x}\big) = \prod_{j=1}^{n} \mathsf{p}(x_j|\,\theta) = \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j} = \theta^{t_n}(1-\theta)^{n-t_n}$$

となる．ただし, $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ と $t_n = \sum_{j=1}^{n} x_j$ である．よって対数尤度は

$$\ell_n(\theta|\,\boldsymbol{x}) = t_n \log\theta + (n-t_n)\log(1-\theta) \qquad (0<\theta<1)$$

となる．このことから, $0 < t_n < n$ のとき

$$\frac{t_n}{n} \in \arg\max_{\theta\in(0,\,1)} \ell_n(\theta|\,\boldsymbol{x})$$

がわかる．したがって, , $0 < t_n < n$ のとき θ^* の最尤推定量は $\widehat{\theta}_n = \dfrac{\sum_{j=1}^{n} X_j}{n}$ となり, $t_n = 0$ または $t_n = n$ のとき, 最尤推定量は存在しない. □

注意 7.8. $\boldsymbol{\theta}^* = (\mu^*,\, \sigma^*) \in \mathbb{R}\times(0,\,\infty)$ とし,

$$X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d.}\, \mathsf{N}\big(\mu^*,\, (\sigma^*)^2\big)$$

とする．$X_j = x_j\,(j = 1,\, 2,\, \ldots,\, n)$ を観測したとき, 尤度関数は

$$\begin{aligned}
\mathsf{lik}_n(\mu,\, \sigma|\,\boldsymbol{x}) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_j-\mu)^2}{\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{ns_n^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{x}_n-\mu)^2}{2\sigma^2}\right\}.
\end{aligned}$$

ただし

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad s_n^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\overline{x}_n)^2.$$

最後の等号は

$$\sum_{j=1}^{n}(x_j-\mu)^2 = ns_n^2 + n(\overline{x}_n-\mu)^2 \tag{7.2}$$

からわかる．対数尤度は

$$\ell_n(\mu,\, \sigma|\,\boldsymbol{x}) = -n\log\sigma - \frac{ns_n^2}{2\sigma^2} - \frac{n(\overline{x}_n-\mu)^2}{2\sigma^2} + (\text{定数項})$$

となる．よって

$$\begin{cases} \dfrac{\partial \ell_n(\mu,\, \sigma|\, \boldsymbol{x})}{\partial \mu}(\mu,\, \sigma) = -\dfrac{n(\overline{x}_n - \mu)}{2\sigma^2} = 0 \\ \dfrac{\partial \ell_n}{\partial \sigma}(\mu,\, \sigma|\, \boldsymbol{x}) = -\dfrac{n}{\sigma} + \dfrac{n s_n^2}{\sigma^3} + \dfrac{n(\overline{x}_n - \overline{x}_n)^2}{\sigma^3} = 0 \end{cases}$$

を解くと

$$\mu = \overline{x}_n, \qquad \sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\big(x_j - \overline{x}_n\big)^2}$$

となる．$\ell_n(\mu,\, \sigma)$ の Hessian を求める．

$$\boldsymbol{H} := \begin{pmatrix} \dfrac{\partial \ell_n^2}{\partial \mu^2}\big(\overline{x}_n,\, s_n|\, \boldsymbol{x}\big) & \dfrac{\partial \ell_n^2}{\partial \mu \partial \sigma}\big(\overline{x}_n,\, s_n|\, \boldsymbol{x}\big) \\ \dfrac{\partial \ell_n^2}{\partial \partial \sigma \mu}\big(\overline{x}_n,\, s_n|\, \boldsymbol{x}\big) & \dfrac{\partial \ell_n^2}{\partial \sigma^2}\big(\overline{x}_n,\, s_n|\, \boldsymbol{x}\big) \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{2s_n^2} & 0 \\ 0 & -\dfrac{2n}{s_n^2} \end{pmatrix} \tag{7.3}$$

より，$-\boldsymbol{H}$ は正定値行列となるので，関数

$$\Theta \ni (\mu,\, \sigma) \mapsto \ell_n(\mu,\, \sigma)$$

は $(\mu,\, \sigma) = (\overline{x}_n,\, s_n)$ で最大となる．よって $(\mu^*,\, \sigma^*)$ の最尤推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} X_j =: \overline{X}_n, \quad \widehat{s}_n = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\big(X_j - \overline{X}_n\big)^2}$$

となる．　□

問 7.1. (7.2) と (7.3) を確認せよ．

定理 7.9.

$$X_1,\, X_2,\, \ldots,\, X_n \sim \text{i.i.d. } \mathsf{p}(x|\, \eta) = h(x)\exp\{\eta^* T(x) - \kappa^\vee(\eta^*)\}\, (\eta^* \in \mathcal{E}^o)$$

とする．ただし $\mathcal{E}^o$ は自然母数空間 $\mathcal{E} \subset \mathbb{R}$ の内部である．さらに $\ddot{\kappa}^\vee(\eta) > 0\, (\eta \in \mathcal{E}^o)$ を仮定する．このとき η^* の最尤推定量 $\widehat{\eta}$ は確率 1 で一意的に存在して

$$\widehat{\eta} \xrightarrow{\text{a.s.}} \eta^*\, (n \to \infty)$$
$$\sqrt{n}\big(\widehat{\eta} - \eta^*\big) \rightsquigarrow \mathsf{N}\big(0,\, \big(\ddot{\kappa}^\vee\big)^{-1}(\eta^*)\big)$$

が成立する．

Proof. $X_1 = x_1,\, X_2,\, \ldots,\, X_n = x_n$ を観測したときの尤度関数 lik_n は

$$\mathsf{lik}_n(\eta|\,\boldsymbol{x}) = \prod_{j=1}^{n} h(x_j) \exp\bigl[\eta \sum_{j=1}^{n} T(x_j) - n\kappa^{\vee}(\eta)\bigr]$$

となる．このことから対数尤度関数は

$$\ell_n(\eta|\,\boldsymbol{x}) = \log\bigl(\mathsf{lik}_n(\eta|\,\boldsymbol{x})\bigr) = n\bigl\{\eta\overline{T} - \kappa^{\vee}(\eta)\bigr\} + (\mathrm{constant})$$

となる．ただし $\overline{T} = n^{-1}\sum_{j=1}^{n} T(x_j)$ である．したがって

$$\dot{\ell}(\eta|\,\boldsymbol{x}) = \frac{d\ell}{d\eta}(\eta|\,\boldsymbol{x}) = n\bigl(\overline{T} - \dot{A}\bigr) = 0 \Longleftrightarrow \overline{T} = \dot{\kappa}^{\vee}(\eta) = \mathsf{E}[T(X_1)]$$

となる．

以後, 対数尤度関数の $x_j\,(j = 1,\, 2,\, \ldots,\, n)$ のところに X_j を代入して, 確率変数にしたものを考える．$T(X_1)$ は有限の期待値をもつので, 大数の法則 (定理 4.10) から

$$\overline{T} = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \xrightarrow{\text{a,s.}} \mathsf{E}[T(X_1)] = \dot{\kappa}^{\vee}(\eta^*) \quad (n \to \infty)$$

がわかる．したがって $\overline{T}$ は関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}^{\vee}(\eta)$ の値域に含まれる．さらに $\ddot{\kappa}(\eta) > 0\,(\eta \in \mathcal{E}^o)$ なので関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}^{\vee}(\eta)$ は狭義単調増加関数である．このことより

$$\overline{T} = \dot{\kappa}^{\vee}(\eta)$$

をみたす $\eta \in \mathcal{E}^o$ が存在する．したがって

$$\widehat{\eta} = \bigl(\dot{\kappa}^{\vee}\bigr)^{-1}\bigl(\overline{T}\bigr)$$

は一意的に定まる．さらに

$$\overline{T} \xrightarrow{\text{a.s.}} \dot{\kappa}^{\vee}(\eta^*) \quad (n \to \infty) \qquad \text{かつ} \quad \bigl(\dot{\kappa}^{\vee}\bigr)^{-1}(\eta) \text{ は } \eta^* \text{ の近傍で連続}$$

なので定理 4.9(6) から

$$\widehat{\eta} = \bigl(\dot{\kappa}^{\vee}\bigr)^{-1}\bigl(\overline{T}\bigr) \xrightarrow{\text{a.s.}} \bigl(\dot{\kappa}^{\vee}\bigr)^{-1}(\eta^*)\bigl(\dot{\kappa}^{\vee}(\eta^*)\bigr) = \eta^*$$

がわかる．一方，$\mathsf{E}[T(X_1)] = \dot{\kappa}^{\vee}(\eta^*)$, $\mathsf{Var}[T(X_1)] = \ddot{\kappa}^{\vee}(\eta^*)$ と中心極限定理 (定理 4.14) から

$$\sqrt{n}\bigl(\overline{T} - \dot{\kappa}^{\vee}(\eta^*)\bigr) \rightsquigarrow \mathsf{N}\bigl(0,\, \ddot{\kappa}^{\vee}(\eta^*)\bigr)$$

がわかる．以上のことを踏まえて，$\widehat{\eta} = \bigl(\dot{\kappa}^{\vee}\bigr)^{-1}\bigl(\overline{T}\bigr)$ に対してデルタ法 (定理 4.18) を適用すると

$$\sqrt{n}\bigl(\widehat{\eta} - \eta^*\bigr) \rightsquigarrow \mathsf{N}\bigl(0,\, \bigl(\ddot{\kappa}^{\vee}\bigr)^{-1}(\eta^*)\bigr)$$

がわかる． □

7.3　不偏推定と情報不等式

$(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, $X_1, X_2, \ldots, X_n$ をこの空間上の独立同一分布に従う確率変数列 (ランダム標本) とする．ただし $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ である．$\boldsymbol{X}$ の値域 (標本空間) を $\mathbb{X} \subset \mathbb{R}^n$ と表記する．

$$\mathsf{P}_1 = \mathsf{Pr} \circ X_1^{-1}, \quad \mathsf{P} = \mathsf{Pr} \circ \boldsymbol{X}^{-1}$$

とおく．すると $\mathsf{P}^{\otimes n} = \underbrace{\mathsf{P}_1 \times \mathsf{P}_1 \times \cdots \times \mathsf{P}_1}_{n\text{ 個}}$ となっている．

$\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}\}$ を $\mathbb{X}$ 上の正則母数モデルとし, ある $\boldsymbol{\theta}^* \in \Theta$ が存在し, $\mathsf{P}_{\boldsymbol{\theta}^*} = \mathsf{P}$ となることを仮定する．すなわち $\mathsf{P} \in \mathcal{P}$ である．以下では, $\Theta \subset \mathbb{R}$ とし, P_θ に関する期待値と分散を $\mathsf{E}_\theta[\,\cdot\,]$, $\mathsf{Var}_\theta[\,\cdot\,]$ と表記する．

定義 7.10. $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta\,(\theta \in \Theta \subset \mathbb{R})$ とする．統計量 $T(\boldsymbol{X})$ が $\forall\theta \in \Theta$ に対して

$$\mathsf{E}_\theta\big[T(\boldsymbol{X})\big] = \theta$$

をみたすとき $T(\boldsymbol{X})$ は θ の**不偏推定量** (unbiased estimator) という．

例 7.11. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n \sim$ i.i.d. $\mathsf{N}(\mu, 1)\,(\mu \in \mathbb{R})$ とする．このとき $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^n X_j$ は μ の不偏推定量である．□

定理 7.12. (Rao-Blackwell の定理) $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta\,(\theta \in \Theta \subset \mathbb{R})$ で $T = T(\boldsymbol{X})$ は θ の十分統計量とする．$\widehat{\theta}(\boldsymbol{X})$ を θ の任意の不偏推定量で, 分散 $\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]$ は存在するものとする．

$$\widetilde{\theta}(T) := \mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X}) | T\big]$$

としたとき, 次の (1), (2) が成立する．

(1) $\widetilde{\theta}(T)$ は θ の不偏推定量である．

(2) $\mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big] \leq \mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]$ $(\forall\theta \in \Theta)$ が成立する．

Proof. (1) T は十分統計量なので, $\mathsf{E}_\theta\big[\widehat{\theta} | T\big]$ は θ に依存しないので $\widetilde{\theta}(T)$ は推定量となる．また定理 2.27(2) より

$$\mathsf{E}_\theta\big[\widetilde{\theta}(T)\big] = \mathsf{E}_\theta\big[\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X}) | T\big]\big] = \mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] = \theta, \qquad (\theta \in \Theta)$$

となり $\widetilde{\theta}(T)$ は θ の不偏推定量であることが示せた．

(2) 定理 2.27(4) より

$$\begin{aligned}\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big|T\big] &= \widetilde{\theta}(T)\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)|T\big] \\ &= \widetilde{\theta}(T)\Big(\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})|T\big]-\widetilde{\theta}(T)\Big) \\ &= \widetilde{\theta}(T)\big(\widetilde{\theta}(T)-\widetilde{\theta}(T)\big)=0\end{aligned}$$

となり

$$\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big]=\mathsf{E}_\theta\big[\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big|T\big]\big]=0 \tag{7.4}$$

を得る．一方

$$\begin{aligned}\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]\big\}^2\big] \\ &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\theta\big\}^2\big] \\ &= \mathsf{E}_\theta\big[\big\{\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)+\big(\widetilde{\theta}(T)-\theta\big)\big\}^2\big] \\ &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big\}^2\big]+\mathsf{E}_\theta\big[\big\{\widetilde{\theta}(T)-\theta\big\}^2\big] \\ &\quad +2\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\big(\widetilde{\theta}(T)-\theta\big)\big]\end{aligned}$$

である．しかし, (7.4) および $\widehat{\theta}(\boldsymbol{X})$ と $\widetilde{\theta}(T)$ は θ の不偏推定量であることに注意すると

$$\begin{aligned}&\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\big(\widetilde{\theta}(T)-\theta\big)\big] \\ &\quad=\mathsf{E}_\theta\big[\big(\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big)\widetilde{\theta}(T)\big]-\theta\big\{\mathsf{E}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big]-\mathsf{E}_\theta\big[\widetilde{\theta}(T)\big]\big\} \\ &\quad=0\end{aligned}$$

がわかる．以上から

$$\begin{aligned}\mathsf{Var}_\theta\big[\widehat{\theta}(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}(\boldsymbol{X})-\widetilde{\theta}(T)\big\}^2\big]+\mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big] \\ &\geq \mathsf{Var}_\theta\big[\widetilde{\theta}(T)\big], \qquad (\theta\in\Theta)\end{aligned}$$

を得る． □

定義 7.13. $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)^\top \sim \mathsf{P}_\theta\,(\theta\in\Theta\subset\mathbb{R})$ とする．θ の任意の不偏推定量 $T(\boldsymbol{X})$ に対して

$$\mathsf{Var}_\theta\big[T_0(\boldsymbol{X})\big]\leq\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big], \qquad (\theta\in\Theta)$$

をみたす θ の不偏推定量 $T_0(\boldsymbol{X})$ が存在するとき $T_0(\boldsymbol{X})$ を θ の**一様最小分散不偏推定量** (uniformly minimum variance unbiased estimator: UMVUE) という．

可測関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ は連続微分可能とする．$\boldsymbol{X} \sim \mathsf{P}_\theta$ に基づき $g(\theta)$ の推定問題を考える．

定理 7.14. $\{\mathsf{P}_\theta; \theta \in \Theta \subset R\}$ は正則母数モデルとする．$\boldsymbol{X} \sim \mathsf{P}_\theta\,(\theta \in \Theta \subset \mathbb{R})$ とし $\boldsymbol{X}$ は同時 p.d.f. または p.m.f. $\mathsf{p}(\boldsymbol{x}|\,\theta)$ をもつとする．$T = T(\boldsymbol{X})$ を $g(\theta)$ の任意の不偏推定量とし

$$A(\widetilde{\theta}, \theta) := \mathsf{Var}_\theta\left[\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)}\right] \qquad (\widetilde{\theta}, \theta \in \Theta)$$

とおく．$A(\widetilde{\theta}, \theta) > 0\,(\widetilde{\theta} \in \Theta, \widetilde{\theta} \neq \theta)$ と仮定する．このとき

$$\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \geq \sup_{\widetilde{\theta} \in \Theta} \frac{\{g(\widetilde{\theta}) - g(\theta)\}^2}{A(\widetilde{\theta}, \theta)}$$

が成り立つ．

Proof. $\boldsymbol{X}$ は連続型確率変数の場合の証明を与える．離散型の場合は, 積分記号を和の記号に変更すればよい． 推定量 T は $g(\theta)$ に対して不偏なので

$$\mathsf{E}_\theta\big[T(\boldsymbol{X})\big] = \int_{\mathbb{R}^n} T(\boldsymbol{x})\mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} = g(\theta) \qquad (\theta \in \Theta)$$

となる．よって

$$\begin{aligned}\mathsf{E}_\theta\left[T(\boldsymbol{X})\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right\}\right] &= \int_{\mathbb{R}^n} T(\boldsymbol{x})\left\{\frac{\mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)} - 1\right\}\mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} T(\boldsymbol{x})\mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta})\,d\boldsymbol{x} - \int_{\mathbb{R}^n} T(\boldsymbol{x})\mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\ &= g(\widetilde{\theta}) - g(\theta) \end{aligned} \tag{7.5}$$

となる．一方

$$\mathsf{E}_\theta\left[\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)}\right] = \int_{\mathbb{R}^n} \frac{\mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)}\mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} = \int_{\mathbb{R}^n} \mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta})\,d\boldsymbol{x} = 1$$

と (7.5) に注意すると

$$
\begin{aligned}
g(\widetilde{\theta}) - g(\theta) &= \mathsf{E}_\theta\left[T(\boldsymbol{X})\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right\}\right] \\
&= \mathsf{E}_\theta\left[\left\{T(\boldsymbol{X}) - \mathsf{E}_\theta\big[T(\boldsymbol{X})\big] + \mathsf{E}_\theta\big[T(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right\}\right] \\
&= \mathsf{E}_\theta\left[\left\{T(\boldsymbol{X}) - \mathsf{E}_\theta\big[T(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right\}\right] \\
&\qquad + \mathsf{E}_\theta\big[T(\boldsymbol{X})\big]\underbrace{\mathsf{E}_\theta\left[\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right]}_{=0} \\
&= \mathsf{E}_\theta\left[\big\{T(\boldsymbol{X}) - \mathsf{E}_\theta\big[T(\boldsymbol{X})\big]\big\}\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{X}|\,\theta)} - 1\right\}\right]
\end{aligned}
$$

を得る．Cauchy-Schwarz の不等式を上の式の最右辺に適用すると

$$
\begin{aligned}
\big|g(\widetilde{\theta}) - g(\theta)\big| &= \left|\mathsf{E}_\theta\left[T(\boldsymbol{X})\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)} - 1\right\}\right]\right| \\
&\leq \sqrt{\big\{\mathsf{E}_\theta\big[T(\boldsymbol{X}) - \mathsf{E}_\theta\big[T(\boldsymbol{X})\big]\big\}^2\big]}\sqrt{\mathsf{E}_\theta\left[\left\{\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)} - 1\right\}^2\right]} \\
&= \sqrt{\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big]}\sqrt{\mathsf{Var}_\theta\left[\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)}\right]} \qquad (7.6)
\end{aligned}
$$

を得る．したがって

$$
\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big]\mathsf{Var}_\theta\left[\frac{\mathsf{p}(\boldsymbol{X}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)}\right] \geq \{g(\widetilde{\theta}) - g(\theta)\}^2 \qquad (7.7)
$$

を得る．したがって (7.7) から $\forall\theta \in \Theta$ に対して

$$
\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \geq \frac{\{g(\widetilde{\theta}) - g(\theta)\}^2}{A\big(\widetilde{\theta},\,\theta\big)}
$$

となる．上式の $\widetilde{\theta}$ は任意だったので, 上式の左辺において $\widetilde{\theta}$ に関して sup を取ると定理の主張はわかる． □

定理 7.15. (Cramér-Rao の不等式) 次の条件を仮定する．

(1) $\{\mathsf{P}_\theta;\, \theta \in \Theta \subset R\}$ は正則母数モデルとする．

(2) $g : \mathbb{R} \longrightarrow \mathbb{R}$ は微分可能でその導関数を $\dot{g}(\theta)$ としたとき

$$
\lim_{\widetilde{\theta}\to\theta} \frac{A\big(\widetilde{\theta},\,\theta\big)}{(\widetilde{\theta} - \theta)^2} = J(\theta) > 0
$$

が存在する．

(3) $\forall\theta \in \Theta$ に対して十分小さな $\epsilon > 0$ をとると任意の $\widetilde{\theta}$: $\left|\widetilde{\theta}-\theta\right| < \epsilon$ に対して, ある関数 $G(\boldsymbol{x}|\theta)$ が存在して

$$\left|\frac{p(\boldsymbol{x}|\,\widetilde{\theta}) - p(\boldsymbol{x}|\,\theta)}{(\widetilde{\theta}-\theta)p(\boldsymbol{x}|\,\theta)}\right| < G(\boldsymbol{x}|\,\theta) \quad \text{かつ} \quad \mathsf{E}_\theta\big[G^2(\boldsymbol{X}|\,\theta)\big] < \infty$$

をみたす.

このとき $g(\theta)$ の任意の不偏推定量 $T(\boldsymbol{X})$ に対して

$$\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \geq \frac{\dot{g}(\theta)}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \qquad (\theta \in \Theta) \tag{7.8}$$

が成り立つ. ただし

$$\mathcal{F}_{\boldsymbol{X}}(\theta) = \mathsf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log \mathsf{p}(\boldsymbol{X}|\,\theta)\right\}^2\right]$$

である.

注意 7.16. (7.8) を Cramér-Rao の不等式といい, その右辺を Cramér-Rao の下限という. □

定理 7.15 **の証明**　定理 7.14 と仮定 (2) から

$$\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \geq \lim_{\widetilde{\theta}\to\theta} \frac{\left\{\dfrac{g(\widetilde{\theta})-g(\theta)}{\widetilde{\theta}-\theta}\right\}^2}{\dfrac{A\big(\widetilde{\theta},\,\theta\big)}{(\widetilde{\theta}-\theta)^2}} = \frac{\{\dot{g}(\theta)\}^2}{J(\theta)} \tag{7.9}$$

を得る. 次に仮定 (3) に注意して Lebeague の優収束定理を用いると

$$\begin{aligned}
J(\theta) &= \lim_{\widetilde{\theta}\to\theta} \frac{A\big(\widetilde{\theta},\,\theta\big)}{\big(\widetilde{\theta}-\theta\big)^2} \\
&= \lim_{\widetilde{\theta}\to\theta} \int_{\mathbb{R}^n} \left\{\frac{\mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta})}{\mathsf{p}(\boldsymbol{x}|\,\theta)} - 1\right\}^2 \frac{\mathsf{p}(\boldsymbol{x}|\,\theta)}{\big(\widetilde{\theta}-\theta\big)^2}\,d\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \left\{\lim_{\widetilde{\theta}\to\theta} \frac{\mathsf{p}(\boldsymbol{x}|\,\widetilde{\theta}) - \mathsf{p}(\boldsymbol{x}|\,\theta)}{\big(\widetilde{\theta}-\theta\big)\mathsf{p}(\boldsymbol{x}|\,\theta)}\right\}^2 \mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \left\{\frac{\partial}{\partial\theta}\log \mathsf{p}(\boldsymbol{x}|\,\theta)\right\}^2 \mathsf{p}(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} = \mathcal{F}_{\boldsymbol{X}}(\theta)
\end{aligned}$$

となる. これと (7.9) を合わせると

$$\mathsf{Var}_\theta\big[T(\boldsymbol{X})\big] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \tag{7.10}$$

を得る.

□

定義 7.17. $\boldsymbol{X} \sim \mathsf{P}_\theta\,(\theta \in \Theta \subset \mathbb{R})$ とする．θ の不偏推定量 $\widehat{\theta}$ は R 有効[1]であるとは

$$\mathsf{Var}_\theta\big[\widehat{\theta}\big] = \big[\mathcal{F}_{\boldsymbol{X}}(\theta)\big]^{-1} \qquad (\theta \in \Theta)$$

をみたすときをいう．

定理 7.18. $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n) \sim \mathsf{P}_\theta\,(\theta \in \Theta \subset R)$ とする．θ の不偏推定量 $\widehat{\theta}$ が R 有効であるための必要十分条件は $\mathbb{R}$ 上の $\mathbb{R}$ 値関数 $T(x)$, $A(\theta)$, $\kappa(\theta)$, $g(x)$ があって

$$\widehat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \tag{7.11}$$

$$\log \mathsf{p}_1(x|\,\theta) = A(\theta)T(x) - \kappa(\theta) + g(x), \tag{7.12}$$

$$\int_{\mathbb{R}^n} T(x)\mathsf{p}_1(x|\,\theta)\,dx = \theta \tag{7.13}$$

をみたすときである．ただし X_1 の p.d.f.(または p.m.f.) を $\mathsf{p}_1(x|\,\theta)$ と書いた．

Proof. 連続型分布に対する証明を与える．離散型分布に対しては, 積分記号を和の記号に替えればよい. 表現 (7.11) – (7.13) が成立したとする. $\ell_n(\theta|\,\boldsymbol{x}) = \sum_{j=1}^{n} \log \mathsf{p}_1(x_j|\theta)\,\big(\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)\big)$ とおく．すると

$$\dot{\ell}_n(\theta|\,\boldsymbol{x}) = \frac{\partial}{\partial\theta}\ell_n(\theta|\,\boldsymbol{x}) = \dot{A}(\theta)\sum_{j=1}^{n} T(x_j) - n\dot{\kappa}(\theta),$$

$$\dot{A}(\theta) = \frac{\partial A}{\partial\theta}(\theta), \quad \dot{\kappa}(\theta) = \frac{\partial B}{\partial\theta}(\theta)$$

となる．よって

$$\mathsf{E}_\theta\big[\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] = 0 \Longleftrightarrow \dot{A}(\theta)\mathsf{E}_\theta\big[T(X_1)\big] = \dot{\kappa}(\theta)$$

と (7.13) とあわせると

$$\theta\dot{A}(\theta) = \dot{\kappa}(\theta) \tag{7.14}$$

となる．さらに

$$0 = \int_{\mathbb{R}^n}\{T(x)-\theta\}\mathsf{p}_1(x|\,\theta)\,dx = \int_{\mathbb{R}^n}\{T(x)-\theta\}\exp\{\log \mathsf{p}_1(x|\,\theta)\}\,dx$$

を (7.12) と 7.14) に注意して, θ に関して微分すると

$$\begin{aligned} 0 &= -\int_{\mathbb{R}} \mathsf{p}_1(x|\,\theta)\,dx + \int_{\mathbb{R}}\{T(x)-\theta\}\big\{\dot{A}(\theta)T(x) - \underbrace{\dot{\kappa}(\theta)}_{=\theta\dot{A}(\theta)}\big\}\mathsf{p}_1(x|\,\theta)\,dx \\ &= -\int_{\mathbb{R}} \mathsf{p}_1(x|\,\theta)\,dx + \int_{\mathbb{R}}\{T(x)-\theta\}\big\{\dot{A}(\theta)T(x) - \theta\dot{A}(\theta)\big\}\mathsf{p}_1(x|\,\theta)\,dx \end{aligned}$$

[1]Rao 有効の意味であろう．

となることがわかる．よって, $\int_{\mathbb{R}} \mathsf{p}_1(x|\,\theta)\,dx = 1$ に注意して, 上式を整理すると

$$1 = \dot{A}(\theta)\mathsf{E}_\theta\big[\{T(X_1) - \theta\}^2\big]$$

を得る．さらに $\mathsf{E}_\theta[T(X_1)] = \theta$ に注意すれば

$$\mathsf{Var}_\theta\big[T(X_1)\big] = \frac{1}{\dot{A}(\theta)} \tag{7.15}$$

となる．したがって

$$\begin{aligned}
\mathcal{F}_{\boldsymbol{X}}(\theta) &= \mathsf{Var}_\theta\big[\dot{\ell}_n(\theta|\,\boldsymbol{X})\big] && (7.16)\\
&= \mathsf{Var}_\theta\Bigg[\sum_{j=1}^{n}\{\dot{A}(\theta)T(X_j) - \dot{\kappa}(\theta)\}\Bigg]\\
&= \mathsf{Var}_\theta\Bigg[\sum_{j=1}^{n}\dot{A}(\theta)\big\{T(X_j)\big\}\Bigg] \qquad (\because \text{分散は平行移動に関して不変})\\
&= \sum_{j=1}^{n}\{\dot{A}(\theta)\}^2\mathsf{Var}\big[T(X_j)\big] \qquad (\because X_1,\,X_2,\,\ldots,\,X_n \text{ は独立})\\
&= n\{\dot{A}(\theta)\}^2\mathsf{Var}\big[T(X_1)\big] && (7.17)\\
&= n\dot{A}(\theta) \qquad (\because (7.15))
\end{aligned}$$

を得る．また, $\widehat{\theta}_n(\boldsymbol{X}) = \dfrac{1}{n}\sum_{j=1}^{n} T(X_j)$ なので

$$\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \theta$$

となり $\widehat{\theta}_n$ は θ の不偏推定量である．再度, (7.15) に注意すると

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n\big] = \mathsf{Var}_\theta\Bigg[\frac{1}{n}\sum_{j=1}^{n}T(X_j)\Bigg] = \frac{1}{n^2}\sum_{j=1}^{n}\mathsf{Var}_\theta\big[T(X_j)\big] = \frac{1}{n\dot{A}(\theta)} = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

となり $\widehat{\theta}_n$ は R 有効推定量となる．

次に逆を示す．P_θ の同時 p.d.f. $\prod_{j=1}^{n}\mathsf{p}_1(x_j|\,\theta)$ に対して, その対数尤度関数を $\ell_n(\theta|\,\boldsymbol{x}) = \log\Big(\prod_{j=1}^{n}\mathsf{p}_1(x_j|\,\theta)\Big)$ と書くことにする．すると

$$\int_{\mathbb{R}^n}\exp\{\ell_n(\theta|\,\boldsymbol{X})\}\,d\boldsymbol{x} = 1$$

である．この式を θ に関して微分すると

$$0 = \int_{\mathbb{R}^n}\dot{\ell}_n(\theta|\,\boldsymbol{x})\exp\{\ell_n(\theta|\,\boldsymbol{x})\}\,d\boldsymbol{x} = \mathsf{E}_\theta\big[\dot{\ell}_n(\boldsymbol{x}|\,\boldsymbol{X})\big] \tag{7.18}$$

となる．同様に $\mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\bigr] = \theta$ より

$$\int_{\mathbb{R}^n} \widehat{\theta}_n(\boldsymbol{x}) \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\, d\boldsymbol{x} = \theta$$

となる．この式を θ に関して微分すると

$$1 = \int_{\mathbb{R}^n} \widehat{\theta}_n(\boldsymbol{x}) \dot{\ell}_n(\theta|\,\boldsymbol{x}) \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\, d\boldsymbol{x} = \mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr] \tag{7.19}$$

を得る．(7.18) と (7.19) を合わせると

$$\mathsf{E}_\theta\bigl[(\widehat{\theta}_n(\boldsymbol{X}) - \theta)\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr] = \underbrace{\mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr]}_{=1} - \theta\, \underbrace{\mathsf{E}_\theta\bigl[\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr]}_{=0} = 1 \tag{7.20}$$

を得る．いま

$$Y = \frac{\dot{\ell}_n(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

とおく．(7.16) に注意すると

$$\mathsf{E}_\theta\bigl[Y\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr] = \mathsf{E}_\theta\left[\frac{\bigl\{\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\right] = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\mathsf{Var}[\dot{\ell}_n(\theta|\,\boldsymbol{X})] = 1$$

となる．これと (7.20) を合わせると

$$\begin{aligned}
&\mathsf{E}_\theta\bigl[(\widehat{\theta}_n(\boldsymbol{X}) - \theta - Y)Y\bigr] \\
&\quad = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\bigl[(\widehat{\theta}_n(\boldsymbol{X}) - \theta)\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr]}_{=1} - \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\bigl[Y\dot{\ell}_n(\theta|\,\boldsymbol{X})\bigr]}_{=1} \\
&\quad = 0
\end{aligned}$$

を得る．上の式を用いると

$$\begin{aligned}
\mathsf{E}_\theta\Bigl[\bigl(\widehat{\theta}_n(\boldsymbol{X}) - \theta\bigr)^2\Bigr] &= \mathsf{E}_\theta\Bigl[\bigl(\widehat{\theta}_n(\boldsymbol{X}) - \theta - Y + Y\bigr)^2\Bigr] \\
&= \mathsf{E}_\theta\Bigl[\bigl(\widehat{\theta}_n(\boldsymbol{X}) - \theta - Y\bigr)^2\Bigr] + \mathsf{E}_\theta\bigl[Y^2\bigr] \\
&\geq \mathsf{E}_\theta\bigl[Y^2\bigr]
\end{aligned}$$

となる．よって等号成立は

$$\begin{aligned}
\widehat{\theta}_n(\boldsymbol{X}) - \theta - Y = 0 &\iff \widehat{\theta}_n(\boldsymbol{X}) - \theta = \frac{\dot{\ell}(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \\
&\iff \mathcal{F}_{\boldsymbol{X}}(\theta)\bigl(\widehat{\theta}_n(\boldsymbol{X}) - \theta\bigr) = \dot{\ell}_n(\theta|\,\boldsymbol{X})
\end{aligned}$$

となる．このことより, $\overline{\theta}, \underline{\theta} \in \mathbb{R}$ とすると

$$\begin{aligned}\ell_n(\overline{\theta}|\,\boldsymbol{X}) - \ell_n(\underline{\theta}|\,\boldsymbol{X}) &= \int_{\underline{\theta}}^{\overline{\theta}} \dot{\ell}_n(\boldsymbol{X}|\,\theta)\,d\theta = \int_{\underline{\theta}}^{\overline{\theta}} \mathcal{F}_{\boldsymbol{X}}(\theta)\big(\widehat{\theta}_n(\boldsymbol{X}) - \theta\big)\,d\theta \\ &= \widetilde{A}(\overline{\theta}, \underline{\theta})\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{B}(\overline{\theta}, \underline{\theta})\end{aligned}$$

と書ける．ただし

$$\widetilde{A}(\overline{\theta}, \underline{\theta}) := \int_{\underline{\theta}}^{\overline{\theta}} \mathcal{F}_{\boldsymbol{X}}(\theta)\,d\theta,\ \widetilde{B}(\overline{\theta}, \underline{\theta}) := \int_{\underline{\theta}}^{\overline{\theta}} \theta\mathcal{F}_{\boldsymbol{X}}(\theta)\,d\theta$$

である．ここで $n = 1$, $\overline{\theta} = \theta$, $\underline{\theta} = 1$, $X_1 = x$ とおくと

$$\ell_1(\theta|\,x) = T(x)A(\theta) - B(\theta) + g(x)$$

と書ける．ただし

$$T(x) = \widehat{\theta}_1(x), \quad g(x) = \ell_1(1|\,x), \quad A(\theta) = \widetilde{A}(\theta, 1), \quad B(\theta) = \widetilde{B}(\theta, 1)$$

と書ける． □

7.4 章末注釈と参考文献

7.5 演習問題