

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。

$k, n \in \mathbb{N}$ とし、統計的実験を $(\mathbb{R}^n, \{P_\theta; \theta \in \Theta \subset \mathbb{R}^k\})$ を考え

$$X_1, X_2, \dots, X_n \sim \text{i.i.d. } P_{\theta^*} \quad (\theta^* \in \Theta)$$

とする。

$n \geq k$ とする。 $1 \leq j \leq k$ に対して

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

と定める。

定義 7.1. $n \geq k$ とし $X_1, X_2, \dots, X_n \sim \text{i.i.d. } P_{\theta^*} (\theta^* \in \Theta)$ とする。このとき θ^* のモーメント法推定量 $\hat{\theta}_n$ とは次をみたす解 (存在すれば) である。

$$\alpha_1(\hat{\theta}_n) = \hat{\alpha}_1, \quad \alpha_2(\hat{\theta}_n) = \hat{\alpha}_2, \dots, \alpha_k(\hat{\theta}_n) = \hat{\alpha}_k \quad (7.1)$$

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

$$\alpha_1 = E[X_1] = \mu^*, \quad \alpha_2 = E[X_1^2] = \mathbb{V}[X_1] + \{E[X_1]\}^2 = (\sigma^*)^2 + (\mu^*)^2$$

である。したがって $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n)$ とすれば、

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

を解くと (μ^*, σ^*) のモーメント法推定量は

$$\hat{\mu}_n = \frac{1}{n} \sum_{\ell=1}^n X_{\ell}, \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{\ell=1}^n (X_{\ell} - \hat{\mu}_n)^2}$$

となる. □

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

$$X_1, X_2, \dots, X_n \sim \text{i.i.d. } \Pr_{\theta^*} (\theta^* \in \Theta \subset \mathbb{R}^k), \quad E[|X_1|^{2k}] < \infty$$

とする. (7.1) で定義した $\alpha_j (j = 1, 2, \dots, k)$ の逆写像 α_j^{-1} が存在して, θ^* 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) $\Pr_{\theta^*} (\hat{\theta}_n \text{ は存在}) \rightarrow 0 \quad (n \rightarrow \infty).$

(2) $\forall \epsilon > 0$ に対して

$$\Pr_{\theta^*} (|\hat{\theta}_n - \theta^*|_2 > \epsilon) \rightarrow 0 \quad (n \rightarrow \infty).$$

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

(3) $\sqrt{n}(\hat{\theta}_n - \theta^*) \rightsquigarrow N_k(\mathbf{0}, \Sigma)$. ただし

$$\begin{aligned} \Sigma &= \mathbf{G}(\theta^*) E_{\theta^*} [\mathbf{Y}\mathbf{Y}^{\top}] \mathbf{G}^{\top}(\theta^*), \\ \mathbf{G}(\theta) &= (\dot{g}_1(\theta), \dot{g}_2(\theta), \dots, \dot{g}_k(\theta)), \\ \dot{g}_j(\theta) &= \left(\frac{\partial \alpha_j^{-1}}{\partial \theta_1}(\theta), \frac{\partial \alpha_j^{-1}}{\partial \theta_2}(\theta), \dots, \frac{\partial \alpha_j^{-1}}{\partial \theta_k}(\theta) \right)^{\top} \quad (j = 1, 2, \dots, k), \\ \mathbf{Y} &= (X_1, X_2^2, \dots, X_1^k)^{\top}. \end{aligned}$$

Proof. 証明は省略. □

7.2 最尤法

$k \in \mathbb{N}$ とする. 統計的実験 $(\mathbb{R}^n, \{p(x|\theta); \theta \in \Theta\})$ を考える. ただし $p(x|\theta)$ は p.d.f. もしくは p.m.f. とする. $\theta^* \in \Theta$ とし

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

とする.

定義 7.4. $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ を観測したとき θ の尤度関数 $\text{lik}_n(\theta)$ を

$$\text{lik}_n(\theta) = \prod_{j=1}^n p(x_j | \theta)$$

で定義し、対数尤度 $\ell_n(\theta)$ を

$$\ell_n(\theta) = \log \text{lik}_n(\theta)$$

で定義する.

注意 7.5. 尤度関数は

$$\text{lik}_n : \Theta \ni \theta \mapsto \text{lik}_n(\theta) \in [0, \infty)$$

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

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

となる. $g(x) = \sin x$ のとき

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

となる.

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

$$\hat{\theta}_n(x) \in \arg \max_{\theta \in \Theta} \text{lik}_n(\theta).$$

x に $\mathbf{X} = (X_1, X_2, \dots, X_n)$ を代入したものの $\hat{\theta}_n(\mathbf{X})$ を θ^* の最尤推定量 (maximum likelihood estimator=m.l.e.) という.

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

$$p(x | \theta^*) = (\theta^*)^x (1 - \theta^*)^{1-x}, \quad (x = 0, 1)$$

である. このとき

$$\text{lik}_n(\theta) = \prod_{j=1}^n p(x_j | \theta) = \prod_{j=1}^n \theta^{x_j} (1 - \theta)^{1-x_j} = \theta^s (1 - \theta)^{n-s}$$

となる. ただし $t_n = \sum_{j=1}^n x_j$ である. よって対数尤度は

$$\ell_n(\theta) = t_n \log \theta + (n - t_n) \log(1 - \theta) \quad (0 < \theta < 1)$$

となる. このことから

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

がわかる. したがって θ^* の最尤推定量は $\hat{\theta}_n = \frac{\sum_{j=1}^n X_j}{n}$ となる. □

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

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

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

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

ただし

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

最後の等号は

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

からわかる. 対数尤度は

$$\ell_n(\mu, \sigma) = -n \log \sigma - \frac{ns_n^2}{2\sigma^2} - \frac{n(\bar{x}_n - \mu)^2}{2\sigma^2}.$$

よって

$$\begin{cases} \frac{\partial \ell_n(\mu, \sigma)}{\partial \mu}(\mu, \sigma) = -\frac{n(\bar{x}_n - \mu)}{2\sigma^2} = 0 \\ \frac{\partial \ell_n(\mu, \sigma)}{\partial \sigma}(\mu, \sigma) = -\frac{n}{\sigma} + \frac{ns_n^2}{\sigma^3} + \frac{n(\bar{x}_n - \mu)^2}{\sigma^3} = 0 \end{cases}$$

を解くと

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

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

$$\mathbf{H} := \begin{pmatrix} \frac{\partial^2 \ell_n}{\partial \mu^2}(\bar{x}_n, s_n) & \frac{\partial^2 \ell_n}{\partial \mu \partial \sigma}(\bar{x}_n, s_n) \\ \frac{\partial^2 \ell_n}{\partial \sigma \partial \mu}(\bar{x}_n, s_n) & \frac{\partial^2 \ell_n}{\partial \sigma^2}(\bar{x}_n, s_n) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2\sigma^2} & 0 \\ 0 & -\frac{2n}{s_n^2} \end{pmatrix} \quad (7.3)$$

より, $-\mathbf{H}$ は正定値行列となるので, 関数

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

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

$$\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n X_j =: \bar{X}_n, \quad \hat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2}$$

となる. □

問 7.1. (7.2) と (7.3) を確認せよ.

定理 7.9.

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

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 $\ddot{\kappa}(\eta) > 0$ ($\eta \in \mathcal{E}^o$) を仮定する. このとき η^* の最尤推定量 $\hat{\eta}$ は確率 1 で一意的に存在して

$$\begin{aligned} \hat{\eta} &\xrightarrow{\text{P}} \eta^* \quad (n \rightarrow \infty) \\ \sqrt{n}(\hat{\eta} - \eta^*) &\rightsquigarrow \text{N}(0, \ddot{\kappa}^{-1}(\eta^*)) \end{aligned}$$

が成立する.

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

$$\text{lik}_n = \prod_{j=1}^n h(x_j) \exp\left[\eta \sum_{j=1}^n T(x_j) - n\kappa(\eta)\right]$$

となる. このことから対数尤度関数は

$$\ell(\eta) = \log \mathbf{p}_n = n\{\eta \bar{T} - \kappa(\eta)\} + (\text{constant})$$

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

$$\dot{\ell}(\eta) = \frac{d\ell}{d\eta} = n(\bar{T} - \dot{A}) = 0 \iff \bar{T} = \dot{\kappa}(\eta) = E[T(X_1)]$$

となる.

以後, 対数尤度関数の x_j ($j = 1, 2, \dots, n$) 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 4.8) から

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

がわかる. したがって \bar{T} は関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}(\eta)$ 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 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}(\eta)$ は狭義単調増加関数である. このことより

$$\bar{T} = \dot{\kappa}(\eta)$$

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

$$\hat{\eta} = \dot{\kappa}^{-1}(\bar{T})$$

は一意的に定まる. さらに

$$\bar{T} \xrightarrow{\text{a.s.}} \dot{\kappa}(\eta^*) \quad (n \rightarrow \infty) \quad \text{かつ} \quad \dot{\kappa}^{-1}(\eta) \text{ は } \eta^* \text{ の近傍で連続}$$

なので定理 4.7(6) から

$$\hat{\eta} = \dot{\kappa}^{-1}(\bar{T}) \xrightarrow{\text{P}} \dot{\kappa}^{-1}(\eta^*) (\dot{\kappa}(\eta^*)) = \eta^*$$

がわかる.

$E[T(X_1)] = \dot{\kappa}(\eta^*)$, $E[T(X_1)] = \ddot{\kappa}(\eta^*)$ と中心極限定理 (定理 4.11) から

$$\sqrt{n}(\bar{T} - \dot{\kappa}(\eta^*)) \rightsquigarrow N(0, \ddot{\kappa}(\eta^*))$$

がわかる. $\hat{\eta} = \dot{\kappa}^{-1}(\bar{T})$ とデルタ法 (定理 4.15) より

$$\sqrt{n}(\hat{\eta} - \eta^*) \rightsquigarrow N(0, \ddot{\kappa}^{-1}(\eta^*))$$

がわかる. □

7.3 不偏推定と情報不等式

$(\Omega, \mathcal{A}, \Pr)$ を確率空間とし, X_1, X_2, \dots, 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 $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ である. \mathbf{X} の値域 (標本空間) を $\mathbb{X} \subset R^n$ と表記する.

$$P = \Pr \circ X_1^{-1}, \quad P^{\otimes n} = \Pr \circ \mathbf{X}^{-1}$$

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

$\mathcal{P} = \{P_\theta; \theta \in \Theta \subset \mathbb{R}\}$ を \mathbb{X} 上の正則母数モデルとし, ある $\theta^* \in \Theta$ が存在し, $P_{\theta^*} = P$ となることを仮定する. すなわち $P \in \mathcal{P}$ 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 $E_\theta[\cdot], \text{Var}_\theta[\cdot]$ と表記する.

定義 7.10. $\mathbf{X} \sim P_\theta (\theta \in \Theta \subset R)$ とする. 統計量 $T(\mathbf{X})$ が $\forall \theta \in \Theta$ に対して

$$E_\theta[T(\mathbf{X})] = \theta$$

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

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

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

$$\tilde{\theta}(T) := E_\theta[\hat{\theta}(\mathbf{X})|T]$$

としたとき, 次の (1), (2) が成立する.

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

(2) $\text{Var}_\theta[\tilde{\theta}(T)] \leq \text{Var}_\theta[\hat{\theta}(\mathbf{X})] (\forall \theta \in \Theta)$ が成立する.

Proof. (1) T は十分統計量なので, $E_\theta[\hat{\theta}(\mathbf{X})|T]$ は θ に依存しないので $\tilde{\theta}(T)$ は推定量となる. また定理 2.26(2) より

$$E_\theta[\tilde{\theta}(T)] = E_\theta[E[\hat{\theta}(\mathbf{X})|T]] = E_\theta[\hat{\theta}(\mathbf{X})] = \theta, \quad (\theta \in \Theta)$$

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

(2) 定理 2.26(4) より

$$\begin{aligned} E_\theta[(\hat{\theta}(\mathbf{X}) - \tilde{\theta}(T))\tilde{\theta}(T)|T] &= \tilde{\theta}(T)E_\theta[\hat{\theta}(\mathbf{X}) - \tilde{\theta}(T)|T] \\ &= \tilde{\theta}(T)\left(E_\theta[\hat{\theta}(\mathbf{X})|T] - \tilde{\theta}(T)\right) \\ &= \tilde{\theta}(T)(\tilde{\theta}(T) - \tilde{\theta}(T)) = 0 \end{aligned}$$

となり

$$E_{\theta}[(\widehat{\theta}(\mathbf{X}) - \widetilde{\theta}(T))\widetilde{\theta}(T)] = 0 \quad (7.4)$$

を得る. 一方

$$\begin{aligned} \text{Var}_{\theta}[\widehat{\theta}(\mathbf{X})] &= E_{\theta}[\{\widehat{\theta}(\mathbf{X}) - \theta\}^2] \\ &= E_{\theta}[\{(\widehat{\theta}(\mathbf{X}) - \widetilde{\theta}(T)) + (\widetilde{\theta}(T) - \theta)\}^2] \\ &= E_{\theta}[\{\widehat{\theta}(\mathbf{X}) - \widetilde{\theta}(T)\}^2] + E_{\theta}[\{\widetilde{\theta}(T) - \theta\}^2] \\ &\quad + 2E_{\theta}[(\widehat{\theta}(\mathbf{X}) - \widetilde{\theta}(T))(\widetilde{\theta}(T) - \theta)] \end{aligned}$$

であるので, (7.4) から

$$\begin{aligned} \text{Var}_{\theta}[\widehat{\theta}(\mathbf{X})] &= E_{\theta}[\{\widehat{\theta}(\mathbf{X}) - \widetilde{\theta}(T)\}^2] + \text{Var}_{\theta}[\widetilde{\theta}(T)] \\ &\geq \text{Var}_{\theta}[\widetilde{\theta}(T)], \quad (\theta \in \Theta) \end{aligned}$$

を得る. □

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

$$\text{Var}_{\theta}[T_0(\mathbf{X})] \leq \text{Var}_{\theta}[T(\mathbf{X})], \quad (\theta \in \Theta)$$

をみたく θ の不偏推定量 $T_0(\mathbf{X})$ が存在するとき $T_0(\mathbf{X})$ 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 $g: \mathbb{R} \rightarrow \mathbb{R}$ は連続微分可能とする. $\mathbf{X} \sim P_{\theta}$ に基づき $g(\theta)$ の推定問題を考える.

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

$$A(\tilde{\theta}, \theta) := \text{Var}_{\theta} \left[\frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} \right] \quad (\tilde{\theta}, \theta \in \Theta)$$

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

$$\text{Var}_{\theta}[T(\mathbf{X})] \geq \sup_{\tilde{\theta} \in \Theta} \frac{\{g(\tilde{\theta}) - g(\theta)\}^2}{A(\tilde{\theta}, \theta)}$$

が成り立つ.

Proof. T は $g(\theta)$ の不偏推定量なので

$$E_{\theta}[T(\mathbf{X})] = \int_{\mathbf{X}} T(\mathbf{x})p(\mathbf{x}|\theta) d\mathbf{x} = g(\theta) \quad (\theta \in \Theta)$$

となる. よって

$$\begin{aligned} E_{\theta} \left[T(\mathbf{X}) \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] &= \int_{\mathbf{X}} T(\mathbf{x}) \left\{ \frac{p(\mathbf{x}|\tilde{\theta})}{p(\mathbf{x}|\theta)} - 1 \right\} p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \int_{\mathbf{X}} T(\mathbf{x})p(\mathbf{x}|\tilde{\theta}) d\mathbf{x} - \int_{\mathbf{X}} T(\mathbf{x})p(\mathbf{x}|\theta) d\mathbf{x} \\ &= g(\tilde{\theta}) - g(\theta) \end{aligned} \quad (7.5)$$

となる. 一方

$$E_{\theta} \left[\frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} \right] = \int_{\mathbf{X}} \frac{p(\mathbf{x}|\tilde{\theta})}{p(\mathbf{x}|\theta)} p(\mathbf{x}|\theta) d\mathbf{x} = \int_{\mathbf{X}} p(\mathbf{x}|\tilde{\theta}) d\mathbf{x} = 1$$

に注意すると

$$\begin{aligned} &E_{\theta} \left[T(\mathbf{X}) \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] \\ &= E_{\theta} \left[\left\{ T(\mathbf{X}) - E_{\theta}[T(\mathbf{X})] + E_{\theta}[T(\mathbf{X})] \right\} \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] \\ &= E_{\theta} \left[\left\{ T(\mathbf{X}) - E_{\theta}[T(\mathbf{X})] \right\} \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] + E_{\theta}[T(\mathbf{X})] E_{\theta} \left[\frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right] \\ &= E_{\theta} \left[\left\{ T(\mathbf{X}) - E_{\theta}[T(\mathbf{X})] \right\} \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] \end{aligned}$$

を得る. Cauchy-Schwarz の不等式を上式の最右辺に適用すると

$$\begin{aligned} &\left| E_{\theta} \left[T(\mathbf{X}) \left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\} \right] \right| \\ &\leq \sqrt{\{E_{\theta}[T(\mathbf{X}) - E_{\theta}[T(\mathbf{X})]]\}^2} \sqrt{E_{\theta} \left[\left\{ \frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} - 1 \right\}^2 \right]} \\ &= \sqrt{\text{Var}_{\theta}[T(\mathbf{X})]} \sqrt{\text{Var}_{\theta} \left[\frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} \right]} \end{aligned} \quad (7.6)$$

を得る. (7.5) と (7.6) を合わせると

$$\text{Var}_{\theta}[T(\mathbf{X})] \text{Var}_{\theta} \left[\frac{p(\mathbf{X}|\tilde{\theta})}{p(\mathbf{X}|\theta)} \right] \geq \{g(\tilde{\theta}) - g(\theta)\}^2 \quad (7.7)$$

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

$$\text{Var}_{\theta}[T(\mathbf{X})] \geq \frac{\{g(\tilde{\theta}) - g(\theta)\}^2}{A(\tilde{\theta}, \theta)}$$

となり, 定理の主張は示された. \square

定理 7.15. (Cramér-Rao の不等式) 次の条件を仮定する.

- (1) $\{P_\theta; \theta \in \Theta \subset R\}$ は正則母数モデルとする.
- (2) $g: \mathbb{R} \rightarrow \mathbb{R}$ は微分可能でその導関数を $\dot{g}(\theta)$ としたとき

$$\lim_{\tilde{\theta} \rightarrow \theta} \frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2} = J(\theta) > 0$$

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 $\tilde{\theta}: |\tilde{\theta} - \theta| < \epsilon$ に対して, ある関数 $G(\mathbf{x}|\theta)$ が存在して

$$\left| \frac{p(\mathbf{x}|\tilde{\theta}) - p(\mathbf{x}|\theta)}{(\tilde{\theta} - \theta)p(\mathbf{x}|\theta)} \right| < G(\mathbf{x}|\theta) \quad \text{かつ} \quad E_\theta[G^2(\mathbf{X}|\theta)] < \infty$$

をみたす.

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

$$\text{Var}_\theta[T(\mathbf{X})] \geq \frac{\dot{g}(\theta)}{\mathcal{F}_{\mathbf{X}}(\theta)} \quad (\theta \in \Theta) \quad (7.8)$$

が成り立つ. ただし

$$\mathcal{F}_{\mathbf{X}}(\theta) = E_\theta \left[\left\{ \frac{\partial}{\partial \theta} \log p(\mathbf{X}|\theta) \right\}^2 \right]$$

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. \square

定理 7.15 の証明 定理 7.14 と仮定 (2) から

$$\text{Var}_\theta[T(\mathbf{X})] \geq \lim_{\tilde{\theta} \rightarrow \theta} \frac{\left\{ \frac{g(\tilde{\theta}) - g(\theta)}{\tilde{\theta} - \theta} \right\}^2}{\frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2}} = \frac{\{\dot{g}(\theta)\}^2}{J(\theta)} \quad (7.9)$$

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

$$\begin{aligned}
 J(\theta) &= \lim_{\tilde{\theta} \rightarrow \theta} \frac{A(\tilde{\theta}, \theta)}{(\tilde{\theta} - \theta)^2} \\
 &= \lim_{\tilde{\theta} \rightarrow \theta} \int_{\mathbb{X}} \left\{ \frac{p(\mathbf{x}|\tilde{\theta})}{p(\mathbf{x}|\theta)} - 1 \right\}^2 \frac{p(\mathbf{x}|\theta)}{(\tilde{\theta} - \theta)^2} d\mathbf{x} \\
 &= \int_{\mathbb{X}} \left\{ \lim_{\tilde{\theta} \rightarrow \theta} \frac{p(\mathbf{x}|\tilde{\theta}) - p(\mathbf{x}|\theta)}{(\tilde{\theta} - \theta)p(\mathbf{x}|\theta)} \right\}^2 p(\mathbf{x}|\theta) d\mathbf{x} \\
 &= \int_{\mathbb{X}} \left\{ \frac{\partial}{\partial \theta} \log p(\mathbf{x}|\theta) \right\}^2 p(\mathbf{x}|\theta) d\mathbf{x} = \mathcal{F}_{\mathbf{X}}(\theta)
 \end{aligned}$$

となる. これと (7.9) を合わせると

$$\text{Var}_{\theta}[T(\mathbf{X})] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\mathbf{X}}(\theta)} \quad (7.10)$$

を得る.

□

定義 7.17. $\mathbf{X} \sim P_{\theta} (\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\hat{\theta}$ は R 有効であるとは

$$\text{Var}_{\theta}[\hat{\theta}] = [\mathcal{F}_{\mathbf{X}}(\theta)]^{-1} \quad (\theta \in \Theta)$$

をみたすときをいう.

定理 7.18. $\mathbf{X} = (X_1, X_2, \dots, X_n) \sim P_{\theta} (\theta \in \Theta \subset R)$ とする. θ の不偏推定量 $\hat{\theta}$ 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 $T(x)$, $A(\theta)$, $\kappa(\theta)$, $h(x)$ があって

$$\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^n T(X_j) \quad (7.11)$$

$$\log p_1(x|\theta) = A(\theta)T(x) - \kappa(\theta) + h(x), \quad (7.12)$$

$$\int_{\mathbb{X}} T(x)p_1(x|\theta) dx = 0^1 \quad (7.13)$$

をみたすときである. ただし X_1 の p.d.f. を $p_1(x|\theta)$ と書いた.

Proof. 表現 (7.11)–(7.13) が成立したとする. $\ell_n(\theta|\mathbf{x}) = \sum_{j=1}^n \log p_1(x_j|\theta)$ ($\mathbf{x} = (x_1, x_2, \dots, x_n)$) とおく. すると

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

となる. よって

$$E_{\theta}[\dot{\ell}_n(\theta | \mathbf{X})] = 0 \iff \dot{A}(\theta)E_{\theta}[T(X_1)] = \dot{\kappa}(\theta)$$

と (7.13) とあわせると

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

となる. さらに

$$0 = \int_{\mathbb{X}} \{T(x) - \theta\} p_1(x | \theta) dx = \int_{\mathbb{X}} \{T(x) - \theta\} \exp\{\log p_1(x | \theta)\} dx$$

を θ に関して微分すると

$$0 = - \int_{\mathbb{X}} p(x | \theta) dx + \int_{\mathbb{X}} \{T(x) - \theta\} \{\dot{A}(\theta)T(x) - \dot{\kappa}(\theta)\} p_1(x | \theta) dx$$

より

$$1 = E_{\theta}[\{T(X_1) - \theta\} \{\dot{A}(\theta)T(X_1) - \dot{\kappa}(\theta)\}]$$

を得る. この式に (7.14) を代入すると

$$\dot{A}(\theta)E_{\theta}[\{T(X_1) - \theta\}^2] = 1$$

を得る. よって

$$\text{Var}_{\theta}[T(X_1)] = \frac{1}{\dot{A}(\theta)}$$

となる. すると

$$\begin{aligned} \mathcal{F}_{\mathbf{X}}(\theta) &= \text{Var}_{\theta}[\dot{\ell}_n(\theta | \mathbf{X})] = \text{Var}_{\theta}\left[\sum_{j=1}^n \{\dot{A}(\theta)T(X_j) - \dot{\kappa}(\theta)\}\right] = \sum_{j=1}^n \{\dot{A}(\theta)\}^2 \text{Var}[T(X_j)] \\ &= n\{\dot{A}(\theta)\}^2 \text{Var}[T(X_1)] = n\dot{A}(\theta) \end{aligned}$$

を得る. $\hat{\theta}_n(\mathbf{X}) = \frac{1}{n} \sum_{j=1}^n T(X_j)$ なので

$$E_{\theta}[\hat{\theta}_n(\mathbf{X})] = \theta$$

となり $\hat{\theta}$ は θ の不偏推定量である. さらに

$$\text{Var}_{\theta}[\hat{\theta}] = \text{Var}_{\theta}\left[\frac{1}{n} \sum_{j=1}^n T(X_j)\right] = \frac{1}{n^2} \sum_{j=1}^n \text{Var}_{\theta}[T(X_j)] = \frac{1}{n\dot{A}(\theta)} = \frac{1}{\mathcal{F}_{\mathbf{X}}(\theta)}$$

となり $\hat{\theta}$ は R 有効推定量となる.

次に逆を示す. P_θ の同時 p.d.f. は $\prod_{j=1}^n p_1(x_j|\theta)$ に対して, その対数尤度関数を $\ell_n(\theta|\mathbf{x}) = \log\left(\prod_{j=1}^n p_1(x_j|\theta)\right)$ と書くことにする. すると

$$\int_{\mathbf{X}} \exp\{\ell_n(\theta|\mathbf{X})\} d\mathbf{x} = 1$$

である. この式を θ に関して微分すると

$$0 = \int_{\mathbf{X}} \dot{\ell}_n(\theta|\mathbf{x}) \exp\{\ell_n(\theta|\mathbf{x})\} d\mathbf{x} = E_\theta[\dot{\ell}_n(\mathbf{x}|\mathbf{X})]$$

となる. 同様に $E_\theta[\widehat{\theta}_n(\mathbf{X})] = \theta$ より

$$\int_{\mathbf{X}} \widehat{\theta}_n(\mathbf{x}) \exp\{\ell_n(\theta|\mathbf{x})\} d\mathbf{x} = \theta$$

となる. この式を θ に関して微分すると

$$1 = \int_{\mathbf{X}} \widehat{\theta}_n(\mathbf{x}) \dot{\ell}_n(\theta|\mathbf{x}) \exp\{\ell_n(\theta|\mathbf{x})\} d\mathbf{x} = E_\theta[\widehat{\theta}_n(\mathbf{X}) \dot{\ell}_n(\theta|\mathbf{X})] \quad (7.15)$$

を得る. (7.14) と (7.15) を合わせると

$$E_\theta[(\widehat{\theta}_n(\mathbf{X}) - \theta) \dot{\ell}_n(\theta|\mathbf{X})] = 1 \quad (7.16)$$

を得る. いま

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

とおくと

$$E_\theta[Y \dot{\ell}_n(\theta|\mathbf{X})] = 1$$

となる. これと (7.16) を合わせると

$$E_\theta[(\widehat{\theta}_n(\mathbf{X}) - \theta - Y) \dot{\ell}_n(\theta|\mathbf{X})] = 0$$

を得る. 上の式を用いると

$$\begin{aligned} E_\theta[(\widehat{\theta}_n(\mathbf{X}) - \theta)^2] &= E_\theta[(\widehat{\theta}_n(\mathbf{X}) - \theta - Y + Y)^2] \\ &= E_\theta[(\widehat{\theta}_n(\mathbf{X}) - \theta - Y)^2] + E_\theta[Y^2] \\ &\geq E_\theta[Y^2] \end{aligned}$$

となる. よって等号成立は

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

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

$$\begin{aligned} \ell_n(\bar{\theta} | \mathbf{X}) - \ell_n(\underline{\theta} | \mathbf{X}) &= \int_{\underline{\theta}}^{\bar{\theta}} \dot{\ell}_n(\mathbf{X} | \theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \mathcal{F}_{\mathbf{X}}(\theta) (\hat{\theta}_n(\mathbf{X}) - \theta) d\theta \\ &= A(\theta) \hat{\theta}_n(\mathbf{X}) - B(\theta) \end{aligned}$$

と書ける. ただし

$$A(\theta) := \int \mathcal{F}_{\mathbf{X}}(\theta) d\theta, \quad B(\theta) := \int \theta \mathcal{F}_{\mathbf{X}}(\theta) d\theta$$

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

$$\begin{aligned} \ell_1(\theta | x) &= T(x)A(\theta) - B(\theta) + h(x), \\ T(x) &= \hat{\theta}_1(x), \quad h(x) = \ell_1(\underline{\theta} | x) \end{aligned}$$

と書ける. □

7.4 章末注釈と参考文献

7.5 演習問題