立教大学　数理統計学序論 1 ・数理統計学 1

今野 良彦

大阪公立大学

2025 年 12 月 20 日

講義内容

- 前回までの記号と重要事項の復習
- 6.4 十分統計量
- 7.2 最尤(さいゆう)法
- 7.3 不偏推定量と情報不等式

記号の復習 (1)

- $(\Omega, \mathcal{A}, \mathsf{Pr})$: 確率空間.
  - $\mathcal{A}$ は σ 加法族.
  - $\mathsf{Pr} : \mathcal{A} \longrightarrow [0, 1]$ は確率測度.
- X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 $\Leftrightarrow \forall r \in \mathbb{R}$ に対して

$$X^{-1}((r, \infty)) = \{\omega \in \Omega;\ X(\omega) > r\}.$$

- 確率変数 X の分布 P^X

$$\mathsf{P}^X(B) = \mathsf{Pr}(\omega \in \Omega;\ X(\omega) \in B\}) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

  ここで, $\mathcal{B}(\mathbb{R})$ は Borel 集合族. すなわち, $\mathbb{R}$ の開集合族を含む最小の σ 加法族.
- 確率変数 X の累積分布関数 F^X: $x \in \mathbb{R}$ に対して

$$\mathsf{F}^X(x) = \mathsf{P}^X((-\infty, x])] = \mathsf{Pr}(\omega \in \Omega;\ X(\omega) \le x).$$

- 累積分布関数 F^X の性質: ① 右連続, ② 非減少, ③ $\lim_{x\to\infty} \mathsf{F}^X(x) = 1,\ \lim_{x\to-\infty} \mathsf{F}^X(x) = 0.$
  - F^X が階段関数 $\Rightarrow$ X は離散型確率変数.
  - F^X が連続関数 $\Rightarrow$ X は連続型確率変数.

記号の復習 (2)

- 確率変数 X は離散型のとき

$$\mathsf{p}^X(x) = \mathsf{F}^X(x) - \mathsf{F}^X(x-)$$

を確率関数 (p.m.f.) という.

- 確率変数 X は連続型のとき

$$\mathsf{F}^X = \int_{-\infty}^{x} \mathsf{p}^X(t)\,\mathrm{d}t$$

をみたす非負関数 p^X(存在するならば) を確率密度関数 (p.d.f.) という.

- 確率変数 X の期待値を

$$\mathsf{E}[X] = \begin{cases} \sum_x x\mathsf{p}^X(x) & (\text{離散型の場合}) \\ \int_{\mathbb{R}} x\mathsf{p}^X(x)\,\mathrm{d}x & (\text{連続型の場合}) \end{cases}$$

ただし, 和および積分の中を絶対値をとったものが有限ののとき.

記号の復習 (3)

- 確率変数 X の分散: $\mathsf{E}[X^2] < \infty$ のとき

$$\mathsf{Var}[X] = \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

- $\mathsf{F}^X(\cdot)$ が階段関数のとき, X を離散型確率変数という.
- 確率変数 X の積率母関数 (m.g.f.): 原点を含むある近傍 $(-z_0, z_0)$ があって, $\forall t \in (-z_0, z_0)$ に対して

$$\mathsf{m}^X(t) = \mathsf{E}[e^{tX}]$$

不適切逆問題のひとつの対応: 統計的モデル

- 探索の範囲を確率分布全体の集合 $\mathcal{M}$ から狭い集合に限定する!
- $\mathcal{M}$ の都合のよい部分集合を設定して, そこに制限した探索を行う. この $\mathcal{M}$ の部分集合で都合のよいように設定したものを 統計的モデル ということにする.
- 設定した統計的モデルを $\mathcal{P} \subset \mathcal{M}$ と記すことにする.

仮定 この講義では, 想定（設定）した統計的モデル $\mathcal{P}$ は真の分布 $\mathbf{P}^*$ を含んでいるとする. すなわち

$$\mathbf{P}^* \in \mathcal{P}.$$

母数化: 統計的モデル $\mathcal{P}$ の要素（分布）のレベリングをする集合 $\boldsymbol{\Theta} \subset \mathbb{R}^k$ $(\exists k \in \mathbb{N})$ を想定する.

$$\boldsymbol{\Theta} \ni \theta \longleftrightarrow \mathbf{P}_\theta \in \mathcal{P}.$$

すなわち, 統計的モデル $\mathcal{P}$ と $\mathbb{R}^k$ の部分集合間に対応関係を想定する. $\boldsymbol{\Theta}$ を統計的モデルの 母数空間 といい, $\boldsymbol{\Theta}$ の元を 母数 ということにする.

仮定

- 以下では, 対応関係

$$\boldsymbol{\Theta} \ni \theta \mapsto \mathbf{P}_\theta \in \mathcal{P} \tag{1}$$

は母数空間 $\boldsymbol{\Theta}$ から統計的モデル $\mathcal{P}$ への全単射と仮定する.

- すると $\mathbf{P}^* \in \mathcal{P}$ と仮定していたので, ある $\exists \theta^* \in \boldsymbol{\Theta}$ があって

$$\mathbf{P}_{\theta^*} = \mathbf{P}^* \Leftrightarrow \mathbf{P}_{\theta^*}(B) = \mathbf{P}(B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

となる.

- 真の分布 $\mathbf{P}^*$ に対応する母数 θ^* を「真の母数」という. この講義の仮定から, $\theta^* \in \boldsymbol{\Theta}$ となっていることに注意する.
- 写像 (1) を設定することを統計的モデルの「母数化」という.
- 母数化（写像と母数空間の設定）は一意ではないこと (便宜的である) に注意する.

復習: 統計的実験と母数モデル (3)

$\boxed{\text{記号のまとめ}}$

$\mathcal{P}$	統計的モデル
$\mathbf{\Theta}$	母数空間
$\mathrm{P}^* \in \mathcal{P}$	真の分布
$\theta^* \in \mathbf{\Theta}$	真の母数

$\boxed{\text{注意}}$ 「モデル」という言葉には注意をすること. この講義では「モデル」という言葉を単独では用いないようにする.

記号と設定

- 空でない部分集合 $\mathbb{X} \subset \mathbb{R}$ を標本空間とする.
- $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は確率空間 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 上の $\mathbb{X}$ 値確率変数列.
- $\{X_j\}_{j=1}^n$ は互いに独立で各々同一の確率分布を持つとする.
- $X = (X_1, X_2, \ldots, X_n)^\top$ とおき $B \in \mathcal{B}(\mathbb{X}^n)$ に対し

$$
\begin{aligned}
\mathrm{P}^X(B) &= \mathrm{Pr}\{X^{-1}(B)\} \\
&= \mathrm{Pr}\Big(\{\omega \in \Omega;\ X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \in B\}\Big)
\end{aligned}
$$

とおく.

記号と設定の続き

- $B \in \mathcal{B}(\mathbb{X})$ に対し

$$\mathsf{P}^{X_j}(B) = \mathsf{Pr}\Big(\{\omega \in \Omega;\ X_j(\omega) \in B\}\Big) \qquad (j = 1, 2, \ldots, n)$$

  と書いたとき

$$\mathsf{P}^X = \mathsf{P}^{X_1} \times \mathsf{P}^{X_2} \times \cdots \times \mathsf{P}^{X_n}$$

  と書けることに注意する.
- X により誘導された確率空間を $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}^X)$ と書く.

記号と設定の続き

- $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の統計的モデルを

$$\mathcal{P} = \{\mathrm{P}_\theta;\ \theta \in \Theta\}$$

と書く. ただし θ は母数で Θ は母数空間である. さらにある $\theta^* \in \Theta$ が存在して

$$\mathrm{P}^{X_1} = \mathrm{P}_{\theta^*}$$

とする.

- 統計量 $T(X)$ を

$$T : (\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$$

なる可測関数で θ に依存しないものとする. ただし $k \in \mathbb{N}$ である.

6.4 十分統計量 (4)

定義 6.12 $X \sim \mathsf{P}_\theta^{\otimes n}$ $(\theta \in \Theta)$ とする. $\forall B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 統計量 $T(X)$ を与えたときの X の条件付き確率

$$\mathsf{P}_\theta^{\otimes n}(B | T)$$

が $\theta \in \Theta$ に無関係であるとき, T は $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta\}$ に対する十分統計量 (sufficient statistic) であるという.

定理 6.13 (Fisher-Neyman の因子分解定理) 統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta\}$ とする. ただし θ は母数で Θ は母数空間とする. X は確率空間 $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}_\theta^{\otimes n})$ 上の大きさが n の標本とし, 同時 p.d.f.(または p.m.f.) $\mathsf{p}^X(x|\,\theta)$ [1] を持つとする. このとき統計量 $T:\ \mathbb{R}^n \longrightarrow \mathbb{R}^k$ が統計的モデル $\mathcal{P}$ に対する十分統計量であるための必要十分条件は, X の同時 p.d.f.(または同時 p.m.f.) $\mathsf{p}^X(x|\,\theta)$ が

$$\mathsf{p}_\theta^X(x) = g_\theta\{T(x)\}h(x) \qquad (x \in \mathbb{X}^n) \tag{2}$$

の形で表されるときである. ここで $g_\theta\{T(x)\}$ と $h(x)$ は非負値関数で $h(x)$ は θ に無関係な関数で, g_θ は T を通してのみ x に依存する.

[1] 離散型のときは同時 p.m.f. $\mathsf{p}_\theta^X(x|\,\theta)$ をもつ.

定理 6.13 の証明 (離散型確率変数の場合) 一般の場合の証明は測度論の議論が必要. 補遺 A.9 を参照のこと. 以下は cheat proof.

① 十分統計量 ⇒ (2) の証明: T の p.m.f. を $\mathsf{p}_\theta^T(t)$ ($t \in \mathbb{R}^k$) と書く. $T = t$ が与えられたときの X の条件付き p.m.f. を $\mathsf{p}^{X|T}(x|T = t)$ とする. 仮定から T は十分統計量なので $\mathsf{p}^{X|T}(x|T = t)$ は θ に依存しない. 条件付き p.m.f. の定義から

$$\mathsf{p}_\theta^X(x) = \mathsf{p}_\theta^T(t)\mathsf{p}^{X|T}(x|T = t) \tag{3}$$

となる. (3) において

$$h(x) = \mathsf{p}^{X|T}(x|T = t), \quad g_\theta\{T(x)\} = \mathsf{p}_\theta^T(t)$$

とおけば, (2) の形になる.

定理 6.13 の証明の続き ②(2) ⇒ 十分統計量 の証明: (2) と書けたときに T は十分統計量であることを示す. まず

$$
\begin{aligned}
\mathrm{p}_\theta^T(t) &= \sum_{y\in\mathbb{X}^n;\,T(y)=t} \mathrm{p}^X(y|\theta) \\
&= \sum_{y\in\mathbb{X}^n;\,T(y)=t} g_\theta\{T(y)\}h(y) \\
&= g_\theta(t) \sum_{y\in\mathbb{X}^n;\,T(y)=t} h(y)
\end{aligned}
$$

に注意する. (2) から $T=t$ を与えたときの X の条件付き p.m.f. は

$$
\begin{aligned}
\frac{\mathrm{p}^X(x|\theta)}{\mathrm{p}^T(t|\theta)} &= \frac{g_\theta(t)h(x)}{g_\theta(t)\sum_{y\in\mathbb{X}^n;\,T(y)=t} h(y)} \\
&= \frac{h(x)}{\sum_{y\in\mathbb{X}^n;\,T(y)=t} h(y)}
\end{aligned}
$$

となり, $\mathrm{p}^{X|T}$ は θ に依存しない. よって T は $\mathcal{P}=\{\mathrm{P}_\theta;\ \theta\in\Theta\}$ に対する十分統計量である.

例 6.15 $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(\theta)\ (0 \leq \theta \leq 1)$ とする. $X = (X_1, X_2, \ldots, X_n)$ の同時 p.m.f. は

$$\mathrm{p}_\theta^X(x) = \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j}\mathbb{1}_{\{0,1\}^n}(x),$$

$$x = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n,$$

$$\mathbb{1}_{\{0,1\}^n}(x) = \begin{cases} 1 & (x \in \{0, 1\}^n) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. したがって

$$T(x) = \sum_{j=1}^n x_j, \quad g_\theta\{T(x)\} = \theta^{T(x)}(1-\theta)^{n-T(x)}, \quad h(x) = \mathbb{1}_{\{0,1\}^n}(x)$$

とおけば, 定理 ?? から $T(X) = X_1 + X_2 + \cdots + X_n$ は

$$\mathcal{P} = \{\mathrm{p}_\theta(x) = \theta^x(1-\theta)^{1-x}\mathbb{1}_{\{0,1\}}(x);\ 0 \leq \theta \leq 1\}$$

の十分統計量であることがわかる. $\square$

例 6.16 $\theta \in \mathbb{R}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\theta, 1)$ とする.
$X = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathrm{p}_\theta^X(x) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^n (x_j - \theta)^2\right\} \qquad x = (x_1, x_2, \ldots, x_n)$$

である. $T(x) = n^{-1}\sum_{j=1}^n x_j$ とすると

$$\mathrm{p}_\theta^X(x) = \exp\left\{nT(x)\theta - \frac{n\theta^2}{2}\right\} \times \left(\frac{1}{2\pi}\right)^{n/2} \times \exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

となる. よって (2) で

$$g_\theta\{T(x)\} = \exp\left\{nT(x)\theta - \frac{n\theta^2}{2}\right\}, \quad h(x) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

とおけば, 定理 ?? から $T(X) = (X_1 + X_2 + \cdots + X_n)/n$ は
$\mathcal{P} = \{\mathrm{N}(\theta, 1);\ \theta \in \mathbb{R}\}$ の十分統計量であることがわかる. □

$\boxed{\text{例 6.17}}$ $n \in \mathbb{N}\,(n \geq 2)$ とし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ とする.
$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする. $X = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}_\theta^X(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (x_j - \mu)^2\right\}$$

となる. ただし $\theta = (\mu, \sigma^2)$ である. ここで

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

とおけば

$$\mathsf{p}_\theta^X(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}\left(s_n^2 + (\bar{x}_n - \mu)^2\right)\right\}$$

と書き直せる.

例 **6.17** の続き (2) において

$$T(x) = (\overline{x}_n, s_n^2), \quad g_\theta\{T(x)\} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}(s_n^2 + (\overline{x}_n - \mu)^2)\right\},$$

$h(x) = 1$

とすれば, 定理 **6.13** から

$$T(X) = \left(\frac{1}{n}\sum_{j=1}^n X_j (=: \overline{X}), \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X})^2\right)$$

は $\mathcal{P} = \{\mathrm{N}^{\otimes n}(\mu, \sigma^2); \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ の十分統計量である. ただし, $\mathrm{N}^{\otimes n}(\mu, \sigma^2)$ は $\mathrm{N}(\mu, \sigma)$ の n 個の直積分布である. □

6.4 十分統計量 (12)

注意

- 十分統計量の表現は一意でない.
- 最小十分性と完備性については資料を参照してください.

記号

- $\mathbb{X}$ を $\mathbb{R}$ の空でない部分集合, 標本空間.
- $d, n \in \mathbb{N}$.
- 統計的実験を $(\mathbb{X}^n, \{\mathrm{p}(x|\theta);\ \theta \in \Theta\})$ とする. ただし $\mathrm{p}(x|\theta)$ は p.d.f. (または p.m.f.) で $\Theta \subset \mathbb{R}^d$.
- $\theta^* \in \Theta$ とし

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

定義 7.4 $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき θ の

ゆう
尤 度関数 $\mathrm{lik}_n(\theta|x)$ を

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

で定義し, 対数尤度 $\ell_n(\theta|x)$ を

$$\ell_n(\theta|x) = \log \mathrm{lik}_n(\theta|x)$$

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

例 $\theta > 0$ とし

$$\mathrm{p}_\theta(x) := \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x) = \begin{cases} \dfrac{1}{\theta} & (0 \le x \le \theta) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. $X \sim \mathrm{p}_\theta$ とし, $X = x\ (x > 0)$ を観測したとき

$$\mathrm{lik}_1(\theta | x) = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x) = \frac{1}{\theta}\mathbb{1}_{[x,\infty)}(\theta)$$

となる.
p_θ と lik_1 のグラフを書いてみよう.

注意 7.5 尤度関数は

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

である. □

記法 関数 $g(x)$ の最大値を取る点を表す集合を

$$\operatorname*{arg\,max}_{x\in\mathbb{R}} g(x)$$

と書く. たとえば $g(x) = -(x-1)^2$ のとき

$$\operatorname*{arg\,max}_{x\in\mathbb{R}} g(x) = \{1\}$$

となる. $g(x) = \sin x$ のとき

$$\operatorname*{arg\,max}_{0\le x\le 4\pi} g(x) = \{\pi/2, 5\pi/2\}$$

となる.

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

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

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

7.2 最尤法 (5)

例の続き　$X = x$ $(x > 0)$ を観測したとき

$$\mathrm{lik}_1(\theta | x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x) = \frac{1}{\theta} \mathbb{1}_{[x,\infty)}(\theta)$$

なので, 最尤推定値は $\widehat{\theta}(x) = x$ となることに注意せよ. $\square$

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

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

である. $X_j = x_j \ (j = 1, 2, \ldots, n)$ を観測したとき

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

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

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

となる. このことから, $0 < t_n < n$ のとき

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

がわかる.

7.2 最尤法 (6)

$\boxed{\text{例 7.7 の続き}}$ したがって, $0 < t_n < n$ のとき θ^* の最尤推定値は $\widehat{\theta}_n = \dfrac{\sum_{j=1}^{n} x_j}{n}$ となり, $t_n = 0$ または $t_n = n$ のとき, 最尤推定値は存在しない. □

例 7.9 $n \geq 2\,(n \in \mathbb{N})$ とする. $\theta^* = (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty)$ とし

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

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

$$\begin{aligned}
\mathrm{lik}_n(\mu, \sigma \mid x) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_j - \mu)^2}{\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{n s_n^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{x}_n - \mu)^2}{2\sigma^2}\right\} \qquad (4)
\end{aligned}$$

となる. ただし

$$\overline{x}_n = \frac{1}{n} \sum_{j=1}^{n} x_j, \qquad s_n^2 = \frac{1}{n} \sum_{j=1}^{n} (x_j - \overline{x}_n)^2, \qquad \sum_{j=1}^{n} (x_j - \overline{x}_n)^2 \neq 0.$$

例 7.8 の続き　(4) の最後の等号は

$$\sum_{j=1}^{n}(x_j-\mu)^2 = ns_n^2 + n(\overline{x}_n-\mu)^2 \tag{5}$$

からわかる. 対数尤度は

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

となる. よって

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

を解くと

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

例 7.8 の続き　$\ell_n(\mu, \sigma | x)$ の Hessian を求める.

$$H := \begin{pmatrix} \frac{\partial \ell_n^2}{\partial \mu^2}(\overline{x}_n, s_n | x) & \frac{\partial \ell_n^2}{\partial \mu \partial \sigma}(\overline{x}_n, s_n | x) \\ \frac{\partial \ell_n^2}{\partial \partial \sigma \mu}(\overline{x}_n, s_n | x) & \frac{\partial \ell_n^2}{\partial \sigma^2}(\overline{x}_n, s_n | x) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2s_n^2} & 0 \\ 0 & -\frac{2n}{s_n^2} \end{pmatrix} \tag{6}$$

より, $-H$ は正定値行列となる. したがって, 関数

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

は $(\mu, \sigma) = (\overline{x}_n, s_n)$ で最大となる. 以上の議論から (μ^*, σ^*) の最尤推定量は

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

となる.

7.2 最尤法 (10)

注意 **7.10** の続き　$g : \mathbb{R} \times [0, \infty) \longrightarrow [0, \infty)$ は 1 対 1 とする. すると $\widehat{\theta}_n$ が θ^* の最尤推定量ならば, $g(\widehat{\theta}_n)$ は $g(\theta^*)$ の最尤推定量であることがわかる. すなわち, 最尤推定量は母数の変換に関して不変である. このことに注意すれば, 例 **7.8** の結果から $(\widehat{\mu}_n, s_n^2)$ は $(\mu^*, (\sigma^*)^2)$ の最尤推定量になることがわかる. □

定理 7.11

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

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 $\ddot{\kappa}(\eta^*) > 0\ (\eta^* \in \mathcal{E}^o)$ を仮定する. このとき η^* の最尤推定量 $\widehat{\eta}_n$ は十分大きな n に対して確率 1 で一意的に存在 [2] して

$$\widehat{\eta}_n \xrightarrow{\text{a.s.}} \eta^*,$$

$$\sqrt{n}(\widehat{\eta}_n - \eta^*) \rightsquigarrow \mathrm{N}\left(0, \frac{1}{\ddot{\kappa}(\eta^*)}\right).$$

[2] これは不明瞭な表現である. 意味するところは, $A_n = \{\widehat{\eta}_n$ は存在 $\}$ としたとき

$$\lim_{n\to\infty} \Pr\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right) = 1$$

が成立することである.

定理 7.11 の証明　① 尤度方程式の導出

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

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

となる. このことから対数尤度関数は

$$\ell_n(\eta | x) = \log(\mathrm{lik}_n(\eta | x)) = n\{\eta \overline{T}_n - \kappa(\eta)\} + (\mathrm{constant})$$

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

$$\dot{\ell}_n(\eta | x) = \frac{d\ell_n}{d\eta}(\eta | x) = n(\overline{T}_n - \dot{\kappa}) = 0 \Leftrightarrow \overline{T}_n = \dot{\kappa}(\eta) = \mathrm{E}[T(X_1)]$$

となる. 最後の等号は (6.4) からわかる.

定理 **7.11** の証明の続き ②十分統計量への大数の法則の適用 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 **??**) から

$$\overline{T}_n = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \xrightarrow{\text{a,s.}} \mathrm{E}[T(X_1)] = \dot{\kappa}(\eta^*) \quad (n \to \infty) \tag{7}$$

がわかる.

定理 7.11 の証明の続き　③ 母数 η の最尤推定量の構成 η^* の適当な近傍を $U(\eta^*)$ と書き, $\dot{\kappa}(U(\eta^*)) := \{\dot{\kappa}(\eta);\ \eta \in U(\eta^*)\}$ とおく. (7) から

$$\Pr(\lim_{n\to\infty} \overline{T}_n \in \dot{\kappa}(U(\eta^*))) = 1$$

となる. さらに, $\ddot{\kappa}(\eta^*) > 0$ なので, $\ddot{\kappa}(\eta) > 0\ (\eta \in U(\eta^*))$ となる. したがって, 関数 $\mathcal{E}^o \ni \eta \mapsto \dot{\kappa}(\eta)$ は近傍 $U(\eta^*)$ 上で狭義単調増加関数である. このことより

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

をみたす $\eta \in U(\eta^*) \subset \mathcal{E}^o$ が存在する. したがって, 陰関数の定理から, 近傍 $\kappa(U(\eta^*))$ 上で $\dot{\kappa}$ の逆写像 $\dot{\kappa}^{-1}$ が存在する. このことを踏まえて

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

とおく.

定理 7.11 の証明の続き ④ 一致性の証明

$$\overline{T}_n \xrightarrow{\text{a.s.}} \dot{\kappa}(\eta^*) \quad \text{かつ} \quad \dot{\kappa}^{-1}(t) \text{ は近傍 } \dot{\kappa}(U(\eta^*)) \text{ 上で連続}$$

なので定理 **4.14(6)** から

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

がわかる.

定理 7.11 の証明の続き ④漸近正規性の証明 一方, (6.4) から $\mathrm{E}[T(X_1)] = \dot{\kappa}(\eta^*)$, $\mathrm{Var}[T(X_1)] = \dot{\kappa}(\eta^*)$ となることに注意して, 中心極限定理 (定理 4.22) を用いると

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

がわかる. 以上のことを踏まえて, $\widehat{\eta}_n = \dot{\kappa}^{-1}(\overline{T}_n)$ に対して, デルタ法 (定理 4.25) を適用すると

$$\sqrt{n}(\widehat{\eta}_n - \eta^*) \rightsquigarrow \mathrm{N}\left(0, \frac{1}{\ddot{\kappa}(\eta^*)}\right)$$

がわかる. □

記号

- $\mathbb{X} \subset \mathbb{R}$ を空でない部分集合, 標本空間.
- $(\Omega, \mathcal{A}, \mathrm{Pr})$ を確率空間.
- $X_1, X_2, \ldots, X_n$ をこの空間上の独立同一分布に従う $\mathbb{X}$ 値確率変数列 (ランダム標本).
- $X = (X_1, X_2, \ldots, X_n)^\top$ と記し, X の値域 (標本空間) を $\mathbb{X}^n \subset \mathbb{R}^n$.
- $$\mathrm{P} = \mathrm{Pr} \circ X_1^{-1}, \quad \mathrm{P}^{\otimes n} = \mathrm{Pr} \circ X^{-1}.$$
- すると $\mathrm{P}^{\otimes n} = \underbrace{\mathrm{P} \times \mathrm{P} \times \cdots \times \mathrm{P}}_{n\text{ 個}}$ となっている.
- $\mathcal{P} = \{\mathrm{P}_\theta;\ \theta \in \Theta \subset \mathbb{R}\}$ を $\mathbb{X}$ 上の正則母数モデル.
- $\Theta \subset \mathbb{R}$ とし, P_θ に関する期待値と分散を $\mathrm{E}_\theta[\,\cdot\,]$, $\mathrm{Var}_\theta[\,\cdot\,]$ と表記.

定義 7.12 $X := (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}$ $(\theta \in \Theta \subset \mathbb{R})$ とする.
統計量 $T(X)$ が $\forall\theta \in \Theta$ に対して

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

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

例 7.13 $n \geq 2$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, 1)$ $(\theta \in \mathbb{R})$ とする.
このとき $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ は θ の不偏推定量である. □

定理 7.14 (Rao-Blackwell の定理)

$X = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}$ $(\theta \in \Theta \subset \mathbb{R})$ で $T = T(X)$ は θ の十分統計量とする. $\widehat{\theta}_n(X)$ は θ の任意の不偏推定量で, 有限の分散 $\mathrm{Var}_\theta[\widehat{\theta}_n(X)]$ をもつものとする. 推定量 $\widetilde{\theta}_n(T)$ を

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

により定めたとき, 次の (1), (2) が成立する.

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

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

7.3 情報不等式 (4)

定理 **7.14** の証明 ① $\widetilde{\theta}_n(T)$ は θ の不偏性の証明: T は十分統計量なので, $\mathrm{E}_\theta[\widehat{\theta}_n(X)|\,T]$ は θ に依存しないので $\widetilde{\theta}_n(T)$ は推定量となる. また定理 **??(2)** より

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

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

定理 **7.14** の証明の続き ② 不等式 **(2)** の証明: 定理 **2.29(4)** より

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

となり

$$\mathsf{E}_\theta[(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)] = \mathsf{E}_\theta[\mathsf{E}_\theta[(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)\big|T]] = 0 \tag{8}$$

を得る. 一方

$$\mathsf{Var}_\theta[\widehat{\theta}_n(X)] = \mathsf{E}_\theta[\{\widehat{\theta}_n(X) - \mathsf{E}_\theta[\widehat{\theta}_n(X)]\}^2]$$
$$= \mathsf{E}_\theta[\{\widehat{\theta}_n(X) - \theta\}^2]$$
$$= \mathsf{E}_\theta[\{(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T)) + (\widetilde{\theta}_n(T) - \theta)\}^2]$$
$$= \mathsf{E}_\theta[\{\widehat{\theta}_n(X) - \widetilde{\theta}_n(T)\}^2] + \mathsf{E}_\theta[\{\widetilde{\theta}_n(T) - \theta\}^2]$$
$$+ 2\mathsf{E}_\theta[(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T))(\widetilde{\theta}_n(T) - \theta)].$$

$\boxed{\text{定理 7.14 の証明の続き}}$ しかし, (8) および $\widehat{\theta}_n(X)$ と $\widetilde{\theta}_n(T)$ は θ の不偏推定量であることに注意すると

$$\begin{aligned} &\mathrm{E}_\theta[(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T))(\widetilde{\theta}_n(T) - \theta)] \\ &\quad = \underbrace{\mathrm{E}_\theta[(\widehat{\theta}_n(X) - \widetilde{\theta}_n(T))\widetilde{\theta}_n(T)]}_{=0 \quad \because (8)} - \theta\underbrace{\{\mathrm{E}_\theta[\widehat{\theta}_n(X)] - \mathrm{E}_\theta[\widetilde{\theta}_n(T)]\}}_{=0} \\ &\quad = 0 \end{aligned}$$

がわかる. 以上から

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

を得る. □

$\boxed{\text{定義 7.15 の証明の続き}}$

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

$$\mathsf{Var}_\theta[\widehat{\theta}_n(X)] \le \mathsf{Var}_\theta[\widetilde{\theta}_n(X)] \qquad (\theta \in \Theta)$$

をみたす θ の不偏推定量 $\widehat{\theta}_n(X)$ が存在するとき, $\widehat{\theta}_n(X)$ を θ の一様最小分散不偏推定量 (uniformly minimum variance unbiased estimator: UMVUE) という.

$\boxed{\text{注意 7.16}}$ 存在すれば, 一意的である. 証明の方針は, 異なるふたつ一様最小分散不偏推定量が存在すると仮定し, その平均を新たな推定量とすると分散関数の凸性を利用して, 矛盾を導く. □

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

定理 7.17 $\{\mathsf{P}_\theta;\ \theta\in\Theta\subset R\}$ は正則母数モデルとする. $X\sim\mathsf{P}_\theta^{\otimes n}$ $(\theta\in\Theta\subset\mathbb{R})$ とし X は同時 p.d.f. または p.m.f. $\mathsf{p}^X(x|\theta)$ をもつとする. $\widehat{\theta}_n=\widehat{\theta}_n(X)$ を $g(\theta)$ の任意の不偏推定量とし

$$A(\widetilde{\theta},\theta):=\mathrm{Var}_\theta\left[\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)}\right]\qquad(\widetilde{\theta},\theta\in\Theta)\tag{9}$$

$$A(\widetilde{\theta},\theta)>0\qquad(\widetilde{\theta}\in\Theta,\ \widetilde{\theta}\neq\theta)$$

とおく. このとき

$$\mathrm{Var}_\theta[\widehat{\theta}_n(X)]\geq\sup_{\widetilde{\theta}\in\Theta}\frac{\{g(\widetilde{\theta})-g(\theta)\}^2}{A(\widetilde{\theta},\theta)}$$

が成り立つ.

定理 7.17 の証明　X は連続型確率変数の場合の証明を与える. X の値域を $\mathbb{X}^n$ と書く. さらに, $\widetilde{\mathbb{X}}^n = \{x \in \mathbb{X}^n;\ \mathrm{p}^X(x|\,\theta) > 0\}$ とおく. また, 離散型の場合は, 積分記号を和の記号に変更すればよい.

① $g(\widetilde{\theta}) - g(\theta)$ の書き換え; 推定量 $\widehat{\theta}_n$ は $g(\theta)$ に対して不偏なので

$$\mathrm{E}_\theta[\widehat{\theta}_n(X)] = \int_{\mathbb{X}^n} \widehat{\theta}_n(x)\mathrm{p}^X(x|\,\theta)\,\mathrm{d}x = g(\theta) \qquad (\theta \in \Theta)$$

となる. よって

$$\begin{aligned}
\mathrm{E}_\theta\left[\widehat{\theta}_n(X)\left\{\frac{\mathrm{p}^X(X|\,\widetilde{\theta})}{\mathrm{p}^X(X|\,\theta)} - 1\right\}\right] &= \int_{\widetilde{\mathbb{X}}^n} \widehat{\theta}_n(x)\left\{\frac{\mathrm{p}^X(x|\,\widetilde{\theta})}{\mathrm{p}^X(x|\,\theta)} - 1\right\}\mathrm{p}^X(x|\,\theta)\,\mathrm{d}x \\
&= \int_{\mathbb{X}^n} \widehat{\theta}_n(x)\mathrm{p}^X(x|\,\widetilde{\theta})\,dx - \int_{\mathbb{X}^n} \widehat{\theta}_n(x)\mathrm{p}^X(x|\,\theta)\,\mathrm{d}x \\
&= g(\widetilde{\theta}) - g(\theta) \qquad (10)
\end{aligned}$$

となる.

定理 7.17 の証明の続き ② (10) の最左辺の書き換え;

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

から

$$\begin{aligned}
g(\widetilde{\theta}) - g(\theta) &= \mathsf{E}_\theta\left[\widehat{\theta}_n(X)\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)} - 1\right\}\right] \qquad (11)\\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}_n(X) - \mathsf{E}_\theta[\widehat{\theta}_n(X)] + \mathsf{E}_\theta[\widehat{\theta}_n(X)]\right\}\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)} - 1\right\}\right]\\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}_n(X) - \mathsf{E}_\theta[\widehat{\theta}_n(X)]\right\}\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)} - 1\right\}\right]\\
&\quad + \mathsf{E}_\theta[\widehat{\theta}_n(X)]\,\underbrace{\mathsf{E}_\theta\left[\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)} - 1\right]}_{=0}\\
&= \mathsf{E}_\theta\left[\{\widehat{\theta}_n(X) - \mathsf{E}_\theta[\widehat{\theta}_n(X)]\}\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(X|\theta)} - 1\right\}\right].
\end{aligned}$$

定理 7.17 の証明の続き ③ $|g(\widetilde{\theta}) - g(\theta)|$ の評価;

Cauchy-Schwarz の不等式 (定理 3.10) を上の式の最右辺に適用すると

$$
\begin{aligned}
\left|g(\widetilde{\theta}) - g(\theta)\right| &= \left|\mathsf{E}_\theta\left[\widehat{\theta}_n(X)\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(x|\theta)} - 1\right\}\right]\right| \\
&\le \sqrt{\{\mathsf{E}_\theta[\widehat{\theta}_n(X) - \mathsf{E}_\theta[\widehat{\theta}_n(X)]\}^2]}\sqrt{\mathsf{E}_\theta\left[\left\{\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(x|\theta)} - 1\right\}^2\right]} \\
&= \sqrt{\mathsf{Var}_\theta[\widehat{\theta}_n(X)]}\sqrt{\mathsf{Var}_\theta\left[\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(x|\theta)}\right]} \qquad (12)
\end{aligned}
$$

を得る. 最後の等号は (11) を用いた. したがって

$$
\mathsf{Var}_\theta[\widehat{\theta}_n(X)]\mathsf{Var}_\theta\left[\frac{\mathsf{p}^X(X|\widetilde{\theta})}{\mathsf{p}^X(x|\theta)}\right] \ge \{g(\widetilde{\theta}) - g(\theta)\}^2 \qquad (13)
$$

を得る.

7.3 情報不等式 (12)

$\boxed{\text{定理 7.17 の証明の続き}}$ したがって (9) と (13) から $\forall\theta \in \Theta$ に対して

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

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 7.18 (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) $g:\ \mathbb{R} \longrightarrow \mathbb{R}$ は連続微分可能でその導関数は $\dot{g}(\theta) \neq 0\ (\theta \in \Theta)$ をみたす. さらに

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

が存在する. ただし, $A(\widetilde{\theta}, \theta)$ は (9) で与えたものである.

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

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

をみたす.

定理 7.18 の続き　このとき $g(\theta)$ の任意の不偏推定量 $T(X)$ に対して

$$\mathrm{Var}_\theta[\widehat{\theta}_n(X)] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_X(\theta)} \qquad (\theta \in \Theta) \tag{14}$$

が成り立つ. ただし

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

である.

注意 7.19　(14) 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 7.18 の証明 ① 推定量の分散の下限の導出: X の値域を $\mathbb{X}^n$ とし, $\widetilde{\mathbb{X}}^n = \{x \in \mathbb{X}^n;\ \mathrm{p}^X(x|\,\theta) > 0\}$ とおく. 定理 ?? と仮定 (2) から

$$\mathrm{Var}_\theta[\widehat{\theta}_n(X)] \geq \lim_{\widetilde{\theta}\to\theta} \frac{\left\{\dfrac{g(\widetilde{\theta}) - g(\theta)}{\widetilde{\theta} - \theta}\right\}^2}{\dfrac{A(\widetilde{\theta},\ \theta)}{(\widetilde{\theta} - \theta)^2}} = \frac{\{\dot{g}\ (\theta)\}^2}{J(\theta)} \qquad (15)$$

を得る.

定理 7.18 の証明の続き　② $J(\theta)$ の評価: 仮定 (3) に注意して Lebesgue の優収束定理を用いると

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

となる. これと (15) を合わせると

$$
\mathrm{Var}_\theta[\widehat{\theta}_n(X)] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_X(\theta)}. \tag{16}
$$

定義 7.21 $X \sim \mathrm{P}_\theta^{\otimes n}$ $(\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\widehat{\theta}_n(X)$ は R 有効[3] であるとは

$$\mathrm{Var}_\theta[\widehat{\theta}_n(X)] = \frac{1}{\mathcal{F}_X(\theta)} \qquad (\theta \in \Theta)$$

をみたすときをいう.

[3] Rao 有効の意味であろう.

定理 7.22 $X = (X_1, X_2, \ldots, X_n) \sim \mathrm{P}_\theta^{\otimes n}$ $(\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\widehat{\theta}_n$ 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

$$\widehat{\theta}_n(X) = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \tag{17}$$

$$\log \mathrm{p}(x|\,\theta) = A(\theta)T(x) - \kappa(\theta) + g(x), \tag{18}$$

$$\int_{\mathbb{X}} T(x)\mathrm{p}(x|\,\theta)\,\mathrm{d}x = \theta \tag{19}$$

をみたすときである. ただし X_1 の p.d.f.(または p.m.f.) と値域をそれぞれ $\mathrm{p}(x|\,\theta)$ と $\mathbb{X}(\subset \mathbb{R})$ と書いた.

7.3 情報不等式 (17)

$\boxed{\text{定理 \textbf{7.22} の証明}}$ 資料を参照のこと. □

$\boxed{\text{注意 \textbf{7.23}}}$ 定理 **7.22** から, ***R*** 有効な推定量が存在する統計的モデルは指数型分布族となることがわかる. □

本日の講義のまとめ

- 6.4 十分統計量
  - Fisher-Neyman の因子分解定理.
- 7.2 最尤(さいゆう)法
  - 尤度関数と対数尤度関数
  - 最尤推定量の定義
  - 指数分布族の最尤推定量の漸近正規性
- 7.3 不偏推定量と情報不等式
  - 不偏推定量
  - 最小分散不偏推定量
  - Cramér-Rao の不等式 (情報不等式)