第6章　母数モデル

第 6.1 では, 統計的モデルのなかでも最も基本的な正則母数モデルとそのもとので Fisher 情報量を導入する. 第 6.2 では, 母数モデルのなかでも非常によい性質と数学的な美しさをもつ指数型分布族とその基本的な性質を説明する. 第 6.4 では, 十分統計量を定義し, 十分性の利便性の高い判定条件を述べる. 第 6.5 では, 最小十分統計量と完備十分統計量を定義し, それらに関わる重要な定理を述べる. いくつかの主張の証明で Lebeague 積分の積分の収束定理を用いた. この点について慣れていない読者は証明を読み飛ばしても全体の流れを理解するのには困難が伴わないようにした.

6.1　正則母数モデルと Fisher 情報量

定義 6.1. $d, n \in \mathbb{N}$ とし, $\mathbb{X}^n \left(\subset \mathbb{R}^n\right)$ を標本空間[1] とする. $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする. 分布 $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f.(または p.m.f.) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ と表記する. このとき母数モデルは正則であるとは次の条件をみたすときをいう.

(1) $\Theta \subset \mathbb{R}^d$ は開集合.

(2) $\mathcal{P}$ に属する分布の p.d.f.(または p.m.f.) は同じ台をもつ. すなわち, 集合 $\{x \in \mathbb{R}; \mathsf{p}_{\boldsymbol{\theta}}(x) > 0\}$ は $\boldsymbol{\theta} \in \Theta$ に依存しない.

(3) $\forall \boldsymbol{\theta} \in \Theta$ とする. $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の $\boldsymbol{\theta}$ の 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ に関して連続である.

(4) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の $\boldsymbol{\theta}$ に関する 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ の関数として可積分である.

(5) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の 1 次と 2 次の偏導関数は $\boldsymbol{\theta}$ の微分記号と x の積分記号と交換が可能である.

[1] すこし言葉を乱用して, $\mathbb{X}$ のことも標本空間と呼ぶことにする.

$X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n$ を観測したとき Θ 上の実数値関数 ℓ_n を

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j) \qquad (\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n))$$

で定義[2]する.

$\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のときこの分布に関する期待値, 分散および共分散を

$$\mathsf{E}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Var}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Cov}_{\boldsymbol{\theta}}[\,\cdot,\,\cdot\,]$$

と表記する. すなわち, 可積分な関数 $h\colon \mathbb{X}^n \to \mathbb{R}$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}}[h(\boldsymbol{X})] := \begin{cases} \sum_{\boldsymbol{x}\in\mathbb{X}^n} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j) & (\text{離散型の場合}) \\ \int_{\mathbb{X}^n} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j)\,\mathrm{d}\boldsymbol{x} & (\text{連続型の場合}) \end{cases}$$

と定める. ただし $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1\mathrm{d}x_2 \times \cdots \times \mathrm{d}x_n$ である. さらに, $\boldsymbol{X}$ を確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率ベクトルとしたとき, 任意の $B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 可測空間 $(\Omega,\, \mathcal{A})$ 上の確率測度 $\mathsf{Pr}_{\boldsymbol{\theta}}$ を

$$\mathsf{Pr}_{\boldsymbol{\theta}}\bigl(\boldsymbol{X}^{-1}(B)\bigr) := \mathsf{Pr}_{\boldsymbol{\theta}}(\boldsymbol{X} \in B) := \mathsf{E}_{\boldsymbol{\theta}}\bigl[\mathbb{1}_B(\boldsymbol{X})\bigr]$$

で定める. ただし, $\boldsymbol{X}^{-1}(B) := \{\omega \in \Omega;\, \boldsymbol{X}(\omega) \in B\} \in \mathcal{A}$ である.

さらに, $i,\, j = 1,\, 2, \ldots,\, d$ に対して

$$\dot{\ell}_{n,i} := \dot{\ell}_{n,i}\,(\boldsymbol{\theta}|\,\boldsymbol{x}) = \frac{\partial \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\theta_i}, \quad \ddot{\ell}_{n,ij} := \ddot{\ell}_{n,ij}\,(\boldsymbol{x}|\,\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\theta_j\partial\theta_k},$$

$$\dot{\boldsymbol{\ell}}_n = (\dot{\ell}_{n,1},\, \dot{\ell}_{n,2},\, \ldots,\, \dot{\ell}_{n,d})^\top, \quad \ddot{\boldsymbol{\ell}}_n = \bigl(\ddot{\ell}_{n,jk}\bigr), \quad \boldsymbol{\theta} = (\theta_1,\, \theta_2,\, \ldots,\, \theta_d)^\top$$

と定義する. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のとき Fisher 情報量 $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ は $d \times d$ 行列で $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ の $(i,\, j)$ 成分 $\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{jk}$ は

$$\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij} = \mathsf{E}_{\boldsymbol{\theta}}\bigl[\dot{\ell}_{n,i}\,(\boldsymbol{X}|\,\boldsymbol{\theta})\ \dot{\ell}_{n,j}\,(\boldsymbol{X}|\,\boldsymbol{\theta})\bigr] \qquad (1 \le i,\, j \le d)$$

で定義される.

定理 6.2. 標本空間 $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ は正則であるとする. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ としたとき以下が成立する.

(1) $\forall\boldsymbol{\theta} \in \Theta$ に対して $\mathsf{E}_{\boldsymbol{\theta}}\bigl[\dot{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\bigr] = \mathbf{0}_d$ となる.

(2) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = \mathsf{Cov}_{\boldsymbol{\theta}}\bigl[\dot{\boldsymbol{\ell}}_n\,(\boldsymbol{X}|\,\boldsymbol{\theta})\bigr] = \mathsf{E}_{\boldsymbol{\theta}}\bigl[\dot{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\,\dot{\boldsymbol{\ell}}_n^{\top}\,(\boldsymbol{\theta}|\,\boldsymbol{X})\bigr]$ が成り立つ.

(3) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = -\mathsf{E}_{\boldsymbol{\theta}}\bigl[\ddot{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\bigr]$ が成り立つ.

[2]対数尤度(ゆうど)関数と呼ぶ.

Proof. 証明は連続型分布の場合を示す．離散型の場合は積分を和の記号に置き替えればよい．$\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f. を $\mathsf{p}_{\boldsymbol{\theta}}(x)$ と表記する．さらに，$\widetilde{\mathbb{X}} = \{x \in \mathbb{X};\ \mathsf{p}_{\boldsymbol{\theta}}(x) > 0\}$ とおく．母数モデルの正則性から微分記号と積分記号の交換が保証されているので，$i = 1, 2, \ldots, d$ に対して

$$
\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\left[\dot{\ell}_{n,i}\,(\boldsymbol{\theta}|\,\boldsymbol{X})\right] &= \sum_{k=1}^{n} \int_{\mathbb{X}} \left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k)\right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n} \left\{ \int_{\widetilde{\mathbb{X}}} \left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k)\right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k \right. \\
&\qquad \left. + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k)\right) \underbrace{\mathsf{p}_{\boldsymbol{\theta}}(x_k)}_{=0}\,\mathrm{d}x_k \right\} \\
&= \sum_{k=1}^{n} \int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x_k)} \left(\frac{\partial}{\partial\theta_i} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n} \int_{\widetilde{\mathbb{X}}} \frac{\partial}{\partial\theta_i} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial\theta_i} \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial\theta_i} \left\{ \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \underbrace{\mathsf{p}_{\boldsymbol{\theta}}(x_k)}_{=0}\,\mathrm{d}x_k \right\} \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial\theta_i} \underbrace{\int_{\mathbb{X}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\,\mathrm{d}x_k}_{=1} \\
&= 0
\end{aligned}
$$

となる．よって (1) は示された．
(2) は明らか．
(3) を示すために次に注意する．$i, j = 1, 2, \ldots, d$ と $x \in \widetilde{\mathbb{X}}$ に対して

$$
\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i \partial\theta_j} = \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)} \frac{\partial^2 \mathsf{p}(x|\,\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} - \left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x)\right)\left(\frac{\partial}{\partial\theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(x)\right)
$$

となる．このとき

$$
\begin{aligned}
\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)} \left(\frac{\partial^2 \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i \partial\theta_j}\right) \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x &= \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left\{ \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \right\} \\
&= \frac{\partial^2}{\partial\theta_i \partial\theta_j} \underbrace{\int_{\mathbb{X}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x}_{=1} = 0
\end{aligned}
$$

と

$$
\begin{aligned}
\int_{\widetilde{\mathbb{X}}} &\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_{\boldsymbol{\theta}}(x)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_{\boldsymbol{\theta}}(x)\right)\mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \\
&= \mathsf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\right]
\end{aligned}
$$

となる．このことより

$$
\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(X_1)}{\partial\theta_i\partial\theta_j}\right] &= \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \\
&\qquad + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \\
&= \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j}\mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \\
&= \underbrace{\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)}\left(\frac{\partial^2 \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j}\right)\mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x}_{=0} \\
&\qquad - \mathsf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\right] \\
&= -\mathsf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\theta_i}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\left(\frac{\partial}{\partial\theta_j}\log \mathsf{p}_{\boldsymbol{\theta}}(X_1)\right)\right] \qquad (6.1)
\end{aligned}
$$

がわかる．よって

$$
\begin{aligned}
\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij} &= \mathsf{E}\left[\left(\sum_{k_1=1}^{n}\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i}\right)\left(\sum_{k_1=2}^{n}\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j}\right)\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial \log \mathsf{p}(\mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i}\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_j}\right] \\
&\qquad + \sum_{k_1\neq k_2}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i}\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j}\right] \\
&= \sum_{k=1}^{n}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i}\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_j}\right] \\
&\qquad + \sum_{k_1\neq k_2}\underbrace{\mathsf{E}\left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i}\right]}_{=0}\mathsf{E}\left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j}\right] \\
&= -\sum_{k=1}^{n}\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i\partial\theta_j}\right] \qquad (\because (6.1)) \\
&= -\mathsf{E}\left[\frac{\partial^2 \ell_n(\boldsymbol{X}\,|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right] = -\mathsf{E}\left[\ddot{\ell}_{n,ij}\,(\boldsymbol{X}\,|\,\boldsymbol{\theta})\right]
\end{aligned}
$$

がわかる． □

例 6.3. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ で $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする．このとき, 統計的モデルは $\mathcal{P} = \{\mathsf{N}(\mu, \sigma^2);\ \mu \in \mathbb{R},\ \sigma > 0\}$ で与えられる．したがって

$$\Theta = \{\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R}^2;\ \mu \in \mathbb{R},\ \sigma > 0\} = \mathbb{R} \times (0, \infty)$$

となる．このとき

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = -\sum_{j=1}^n \frac{(x_j-\mu)^2}{2\sigma^2} - n\log\sigma - n\log(\sqrt{2\pi}); \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

となる．よって

$$\begin{aligned}
\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta}|\,\boldsymbol{x}) &= \frac{\partial \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\mu} = \sum_{j=1}^n \frac{x_j-\mu}{\sigma^2}, \\
\overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta}|\,\boldsymbol{x}) &= \frac{\partial \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\sigma} = \sum_{j=1}^n \frac{(x_j-\mu)^2}{\sigma^3} - \frac{n}{\sigma}
\end{aligned}$$

となる．これらの期待値を取ると

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\big[\{\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta}|\,\boldsymbol{X})\}^2\big] &= \frac{n}{\sigma^2}, \\
\mathsf{E}_{\boldsymbol{\theta}}\big[\{\overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta}|\,\boldsymbol{X})\}^2\big] &= \frac{3n}{\sigma^2} - \frac{2n}{\sigma^2} + \frac{n}{\sigma^2} = \frac{2n}{\sigma^2}, \\
\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta}|\,\boldsymbol{X})\,\overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] &= 0
\end{aligned}$$

となるので

$$\mathcal{F}_X(\boldsymbol{\theta}) = \frac{n}{\sigma^2}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

を得る．

一方

$$\begin{aligned}
\overset{\bullet\bullet}{\ell}_{n,11} &= \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\mu^2} = \frac{\partial}{\partial\mu}\left(\sum_{j=1}^n \frac{x_j-\mu}{\sigma^2}\right) = -\frac{n}{\sigma^2}, \\
\overset{\bullet\bullet}{\ell}_{n,22} &= \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\sigma^2} = \frac{\partial}{\partial\sigma}\left(\sum_{j=1}^n \frac{(x_j-\mu)^2}{\sigma^2} - \frac{n}{\sigma}\right) = -3\sum_{j=1}^n \frac{(x_j-\mu)^2}{\sigma^2} + \frac{n}{\sigma^2}, \\
\overset{\bullet\bullet}{\ell}_{n,12} &= \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\mu\partial\sigma} = \frac{\partial}{\partial\sigma}\left[\sum_{j=1}^n \frac{x_j-\mu}{\sigma^2}\right] = -2\sum_{j=1}^n \frac{x_j-\mu}{\sigma^3}
\end{aligned}$$

となる．よって

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,11}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = -\frac{n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,22}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = -\frac{2n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,12}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = 0$$

なので

$$\begin{bmatrix} \mathsf{E}_{\boldsymbol{\theta}}\big[\ddot{\ell}_{n,11}(\boldsymbol{\theta}|\boldsymbol{X})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\ddot{\ell}_{n,12}(\boldsymbol{\theta}|\boldsymbol{X})\big] \\ \mathsf{E}_{\boldsymbol{\theta}}\big[\ddot{\ell}_{n,12}(\boldsymbol{\theta}|\boldsymbol{X})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\ddot{\ell}_{n,22}(\boldsymbol{\theta}|\boldsymbol{X})\big] \end{bmatrix} = -\mathcal{F}_X(\boldsymbol{\theta})$$

となる. □

6.2 指数型分布族

この節では, 統計的モデルでもっとも重要なものを導入し, そのモデルの性質を説明する.

6.2.1 1 変数の場合

定義 6.4. $d \in \mathbb{N}$ とし, $\mathbb{X} \subset \mathbb{R}^d$ を空でない部分集合とする. さらに $\Theta \subset \mathbb{R}$ も空でない部分集合とする. $\mathbb{X}$ 上の統計的モデル $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は 1 **母数指数型分布族**であるとは, Θ 上の実数値関数 A と $\kappa^\vee$, $\mathbb{X}$ 上の実数値関数 T と h が存在して [3], P_θ の p.d.f. または p.m.f. $\mathsf{p}_\theta^\vee(\boldsymbol{x})$ [4] が

$$\mathsf{p}_\theta^\vee(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x}) \exp\big\{A(\theta)T(\boldsymbol{x}) - \kappa^\vee(\theta)\big\} & (\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\boldsymbol{x} \notin \mathbb{X}) \end{cases} \tag{6.2}$$

の形で書けるときをいう.

注意 6.5. 関数 A, h, T の表現は一意ではない. □

例 6.6. (Poisson 分布族) $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. その p.m.f. は

$$\mathsf{p}_\theta^\vee(x) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}\exp\big\{x\log\theta - \theta\big\} \qquad (x = 0,\, 1,\, \ldots)$$

である. したがって $\{\mathsf{Po}(\theta);\, \theta \in (0,\, \infty)\}$ は

$$d = 1,\; A(\theta) = \log\theta,\; \kappa^\vee(\theta) = \theta,\; T(x) = x,\; h(x) = \frac{1}{x!}$$

によって生成される 1 母数指数型分布族である. □

[3] A と B は Greek letters 大文字の α と β である.

[4] 後に別の母数化である自然指数型分布族を導入するときに記号を簡単にするために, ここでは「$\mathsf{p}^\vee$ と $\kappa^\vee$」という少し不自然な記号を用いた.

例 6.7. (2 項分布族) $X \sim \mathsf{Bino}(n, \theta)$ とする．ただし $n \in \mathbb{N}, 0 < \theta < 1$ である．このとき X の p.m.f. は

$$
\begin{aligned}
\mathsf{p}_\theta^\vee(x) &= \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x} \\
&= \begin{pmatrix} n \\ x \end{pmatrix} \exp\left\{ x \log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta) \right\} \qquad (x = 0, 1, \ldots, n)
\end{aligned}
$$

と書ける．ただし

$$
\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x! \times (n-x)!}, \qquad 0! = 1
$$

である．したがって $\{\mathsf{Bino}(n, \theta); \theta \in (0, 1)\}$ は

$$
\begin{aligned}
&d = 1, A(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \kappa^\vee(\theta) = -n\log(1-\theta), T(x) = x, \\
&h(x) = \begin{pmatrix} n \\ x \end{pmatrix}
\end{aligned}
$$

によって生成される 1 母数指数型分布族である． □

例 6.8. $\boldsymbol{X} = (Y, Z)^\top$, $Z, W \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, 1)$ とし，$Y = Z + \theta W\ (\theta > 0)$ とする．このとき $\boldsymbol{X}$ の同時 p.d.f. は次で与えられる．$\boldsymbol{x} = (y, z)$ としたとき

$$
\begin{aligned}
\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) &= \mathsf{p}_\theta^{\boldsymbol{X}}(y, z) = \mathsf{p}^Z(z)\mathsf{p}^{Y|Z}(y|\,z) = \varphi(z)\theta^{-1}\varphi\left(\frac{y-z}{\theta}\right) \\
&= \frac{1}{2\pi\theta}\exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\} \\
&= \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}\exp\left\{-\frac{(y-z)^2}{2\theta^2} - \log\theta\right\}
\end{aligned}
$$

と書ける．ただし p_Z は Z の p.d.f. とし，$\mathsf{p}_{Y|Z}$ は $Z = z$ を与えたときの Y の条件付き p.d.f., φ は $\mathsf{N}(0, 1)$ の p.d.f. で

$$
\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad (z \in \mathbb{R})
$$

である．したがって $\boldsymbol{X}$ の分布は

$$
\begin{aligned}
&d = 2, A(\theta) = -\frac{1}{2\theta^2}, \kappa^\vee(\theta) = \log\theta, T(\boldsymbol{x}) = (y-z)^2, \\
&h(\boldsymbol{x}) = \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}
\end{aligned}
$$

で生成される 1 母数指数型分布族に属する． □

1 母数指数型分布族に属する分布からのランダム標本の分布も 1 母数指数型分布族に属する．特に, $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_\theta$ とする．ただし, $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は (6.2) で与えられる指数型分布族とする．すると $\{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}$ は $\boldsymbol{X}^{(n)} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m)$ に対する指数型分布族となる．ただし, $\mathbb{R}^{dn}$ 上の確率測度を

$$\mathsf{P}_\theta^{\otimes n} = \underbrace{\mathsf{P}_\theta \times \mathsf{P}_\theta \times \cdots \times \mathsf{P}_\theta}_{n\,\text{個}}$$

と定めてた．このとき, $\mathsf{P}_\theta^{\otimes n}$ の同時 p.d.f. $\prod_{j=1}^n \mathsf{p}_\theta^\vee(\,\cdot\,)$ は次で与えられる．すると

$$\begin{aligned}\prod_{j=1}^n \mathsf{p}_\theta^\vee(\boldsymbol{x}_j) &= \prod_{j=1}^n h(\boldsymbol{x}_j)\exp\Big\{A(\theta)T(\boldsymbol{x}_j) - \kappa^\vee(\theta)\Big\}\\ &= \left[\prod_{j=1}^n h(\boldsymbol{x}_j)\right]\exp\Big\{A(\theta)\sum_{j=1}^n T(\boldsymbol{x}_j) - n\kappa^\vee(\theta)\Big\}\end{aligned}$$

となる．したがって $\{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}$ は

$$A(\theta), \quad \sum_{j=1}^n T(\boldsymbol{x}_j), \quad n\kappa^\vee(\theta), \quad \prod_{j=1}^n h(\boldsymbol{x}_j)$$

で生成される $\mathbb{R}^{dn}$ 上の 1 母数指数型分布族となる．

統計量 $\sum_{j=1}^n T(\boldsymbol{X}_j)$ は θ の自然十分統計量 [5] となる．

6.3 指数型分布族の正準表示

(6.2) で表現された指数型分布族を θ でなく $\eta\big(= A(\theta)\big)$ で添え字付けることを考える．記号の乱用 [6] すると 1 母数指数型分布族は

$$\mathsf{p}_\eta(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x}) - \kappa(\eta)\big\} & (\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d)\\ 0 & (\boldsymbol{x} \notin \mathbb{X})\end{cases} \tag{6.3}$$

と書くことができる．ただし

$$\kappa(\eta) = \begin{cases} \log\Big(\sum_{\boldsymbol{x}\in\mathbb{X}} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x})\big\}\Big) & (\text{離散型の場合})\\ \log\Big(\int_{\mathbb{X}} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x})\big\}\,\mathrm{d}\boldsymbol{x}\Big) & (\text{連続型の場合})\end{cases}$$

[5] 十分統計量の定義は定義 6.12 を参照のこと．

[6] ここでは, 記号の乱用をしている．本来であれば, 領域を制限すると A の逆関数は存在するので, $\mathsf{p}(\boldsymbol{x}|\,\eta)$ は $\mathsf{p}(\boldsymbol{x}|\,A^{-1}(\eta))$ と書くべきであろう．

である. κ の定義より, 直ちに $\kappa(\eta)=\kappa(A(\theta))=\kappa^{\vee}(\theta)$ がわかる. さらに

$$\mathcal{E}:=\{\eta\in\mathbb{R};\,\kappa(\eta)<\infty\}$$

とおく. $\mathcal{E}$ は $\mathbb{R}$ の区間となる. すると確率分布族

$$\left\{\mathsf{p}(\boldsymbol{x}|\,\eta)=h(\boldsymbol{x})\exp\bigl\{\eta T(\boldsymbol{x})-\kappa(\eta)\bigr\};\,\eta\in\mathcal{E}\right\}$$

は確率分布族

$$\left\{\mathsf{p}_{\theta}(\boldsymbol{x})=h(\boldsymbol{x})\exp\bigl\{A(\theta)T(\boldsymbol{x})-\kappa(\theta)\bigr\};\,\theta\in\Theta\right\}$$

を含む. 確率分布族 $\{\mathsf{p}(\boldsymbol{x}|\,\eta);\,\eta\in\mathcal{E}\}$ は T, h によって生成された **正準**(せいじゅん)**指数型分布族**または**自然指数型分布族**といい, $\mathcal{E}$ を**自然母数空間**といい, T を自然十分統計量という.

例 6.9. (例 6.6 の続き) Poisson 分布族は

$$\begin{aligned}
\mathsf{p}_{\eta}(x)&=\frac{1}{x!}\exp\bigl\{\eta x-e^{\eta}\bigr\}\qquad(x=0,\,1,\,2,\ldots),\\
\eta&=\log\theta,\,h(x)=\frac{1}{x!},\,T(x)=x,\\
\exp\bigl(\kappa(\eta)\bigr)&=\sum_{x=0}^{\infty}h(x)\exp\bigl(\eta T(x)\bigr)=\sum_{x=0}^{\infty}\frac{\exp(\eta x)}{x!}=\sum_{x=0}^{\infty}\frac{\bigl(e^{\eta}\bigr)^{x}}{x!}=\exp\bigl(e^{\eta}\bigr),\\
\mathcal{E}&=\mathbb{R}
\end{aligned}$$

となる. □

補題 6.10. [7]

$$\mathcal{E}:=\left\{\eta\in\mathbb{R};\,\kappa(\eta):=\mathsf{E}[\exp\{\eta T(\boldsymbol{X})\}]<\infty\right\}$$

とする. このとき, 関数 $\kappa(\eta)$ は $\mathcal{E}^{\circ}$ 上の無限回微分可能である. ただし $\mathcal{E}^{\circ}$ は $\mathcal{E}$ の内部である. さらに積分記号と微分記号の交換は可能である.

Proof. $\eta\in\mathcal{E}^{\circ}$ なので, ある $\epsilon>0$ が存在して $[\eta-2\epsilon,\,\eta+2\epsilon]\subset\mathcal{E}^{\circ}$ とで

[7]証明には, 優収束的理を用いている.

きる．

$$
\begin{aligned}
\frac{d}{d\eta}\exp\{\kappa(\eta)\} &= \lim_{n\to\infty}\frac{\exp\left\{\kappa\left(\eta+\dfrac{\epsilon}{n}\right)\right\}-\exp\{\kappa(\eta)\}}{\epsilon/n} \\
&= \lim_{n\to\infty}\int\frac{\exp\left\{\left(\eta+\dfrac{\epsilon}{n}\right)T(\boldsymbol{x})\right\}-\exp\{\kappa(\eta)\}}{\epsilon/n}h(\boldsymbol{x})\,\mathrm{d}x \\
&= \lim_{n\to\infty}\int\frac{\exp\{\eta T(\boldsymbol{x})\}\left[\exp\left\{\dfrac{\epsilon\,T(\boldsymbol{x})}{n}\right\}-1\right]}{\epsilon/n}h(\boldsymbol{x})\,\mathrm{d}x \\
&=: \lim_{n\to\infty}\int\exp\{\eta T(\boldsymbol{x})\}f_n(\boldsymbol{x})\,\mathrm{d}x
\end{aligned}
$$

と書ける．ただし

$$
f_n(\boldsymbol{x}) = \frac{\exp\{\epsilon\,T(\boldsymbol{x})/n\}-1}{\epsilon/n}h(\boldsymbol{x})
$$

である．すると

$$
\lim_{n\to\infty}f_n(\boldsymbol{x}) =: f(\boldsymbol{x}) = T(\boldsymbol{x})h(\boldsymbol{x})
$$

となる．ここで

$$
|e^t-1| \le |t|e^{|t|};\qquad |t|\le e^{|t|}\qquad (t\in\mathbb{R})
$$

に注意する．これらの不等式を用いると

$$
\begin{aligned}
\left|\frac{\exp\{\epsilon\,T(\boldsymbol{x})/n\}-1}{\epsilon/n}\right| &\le \left\{\left|\frac{\epsilon\,T(\boldsymbol{x})}{n}\Big/\frac{\epsilon}{n}\right|\right\}\exp\left\{\left|\frac{\epsilon\,T(\boldsymbol{x})}{n}\right|\right\} \\
&\le \frac{1}{\epsilon}\big|\epsilon\,T(\boldsymbol{x})\big|\exp\left\{\left|\frac{\epsilon\,T(\boldsymbol{x})}{n}\right|\right\} \\
&\le \frac{1}{\epsilon}\exp\{|\epsilon\,T(\boldsymbol{x})|\}\ \exp\{|\epsilon\,T(\boldsymbol{x})|\} \\
&= \frac{1}{\epsilon}\exp\{|2\epsilon\,T(\boldsymbol{x})|\} \\
&\le \frac{1}{\epsilon}\Big\{\exp\big(2\epsilon T(\boldsymbol{x})\big)+\exp\big(-2\epsilon T(\boldsymbol{x})\big)\Big\}
\end{aligned}
$$

を得る．よって

$$
|f_n(\boldsymbol{x})| \le \frac{1}{\epsilon}\Big\{\exp\big(2\epsilon T(\boldsymbol{x})\big)+\exp\big(-2\epsilon T(\boldsymbol{x})\big)\Big\}h(\boldsymbol{x})
$$

となるので

$$\begin{aligned}\left|\exp\{\eta T(\boldsymbol{x})\}f_n(\boldsymbol{x})\right| &\le \frac{1}{\epsilon}\Big\{\exp\big((\eta+2\epsilon)T(\boldsymbol{x})\big)+\exp\big((\eta-2\epsilon)T(\boldsymbol{x})\big)\Big\}h(\boldsymbol{x})\\ &=: g(\boldsymbol{x})\end{aligned}$$

を得る. $[\eta-2\epsilon, \eta+2\epsilon]\subset\mathcal{E}^\circ$ から

$$\int g(\boldsymbol{x})\,\mathrm{d}x<\infty$$

となる. よって優収束定理から

$$\begin{aligned}\frac{d}{d\eta}\exp\{\kappa(\eta)\} &= \lim_{n\to\infty}\int\exp\{\eta T(\boldsymbol{x})\}f_n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &=\int\exp\{\eta T(x)\}T(\boldsymbol{x})h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

がわかる. この操作を繰り返して行けば, $\kappa(\eta)$ は無限回微分可能なことがわかる. □

定理 6.11. X は (6.3) で与えられた自然指数型分布族に属する分布 $\mathsf{p}(\boldsymbol{x}|\,\eta)$ に従うとする. η は $\mathcal{E}$ の内点としたとき, $T(\boldsymbol{X})$ の積率母関数は原点の近傍で存在し

$$M_T(s)=\exp\big[\kappa(s+\eta)-\kappa(\eta)\big]$$

で与えられる. ただし, s s は 0 のある近傍に含まれるとする. さらに

$$\mathsf{E}[T(\boldsymbol{X})]=\overset{\bullet}{\kappa}(\eta);\qquad \mathsf{Var}[T(\boldsymbol{X})]=\overset{\bullet\bullet}{\kappa}(\eta) \tag{6.4}$$

である. ただし

$$\overset{\bullet}{\kappa}(\eta)=\frac{\mathrm{d}\kappa}{\mathrm{d}\eta}(\eta),\ \overset{\bullet\bullet}{\kappa}(\eta)=\frac{\mathrm{d}^2\kappa}{\mathrm{d}\eta^2}(\eta)$$

である.

Proof. 連続型分布の場合について証明する. 離散型の場合には積分を和に変更するばよい. κ の定義に注意すると

$$\begin{aligned}M(s)&=\mathsf{E}\big[\exp(sT(\boldsymbol{X}))\big]\\ &=\int_{\mathbb{X}}h(\boldsymbol{x})\exp\big[(s+\eta)T(\boldsymbol{x})-\kappa(\eta)\big]\,\mathrm{d}\boldsymbol{x}\\ &=\exp\big[\kappa(s+\eta)-\kappa(\eta)\big]\int_{\mathbb{X}}h(x)\exp\big[(s+\eta)T(\boldsymbol{x})-\kappa(s+\eta)\big]\,\mathrm{d}\boldsymbol{x}\\ &=\exp\big[\kappa(s+\eta)-\kappa(\eta)\big]\end{aligned}$$

から 1 番目の主張はわかる. 最後の等号は $s+\eta \in \mathcal{E}^\circ$ となるように s をとると

$$\int_{\mathbb{X}} h(\boldsymbol{x}) \exp\bigl[(s+\eta)T(\boldsymbol{x}) - \kappa(s+\eta)\bigr] \,\mathrm{d}\boldsymbol{x} = 1$$

となることよりわかる. 次に, $\kappa(\cdot)$ の定義と補題 6.10 から

$$\begin{aligned}
\frac{\mathrm{d}\kappa}{\mathrm{d}\eta}(\eta) &= \frac{\mathrm{d}}{\mathrm{d}\eta} \log\left(\int_{\mathbb{X}} \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}\right) \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}\eta}\displaystyle\int_{\mathbb{X}} \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\displaystyle\int_{\mathbb{X}} \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}} \\
&= \frac{\displaystyle\int_{\mathbb{X}} \frac{\mathrm{d}}{\mathrm{d}\eta} \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\exp\bigl\{\kappa(\eta)\bigr\}} \\
&= \frac{\displaystyle\int_{\mathbb{X}} T(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x})\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\exp\bigl\{\kappa(\eta)\bigr\}} \\
&= \int_{\mathbb{X}} T(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\
&= \mathsf{E}\bigl[T(\boldsymbol{X})\bigr]
\end{aligned}$$

から (6.4) の 1 番目の等号が示せた. 同様に

$$\begin{aligned}
\frac{\mathrm{d}^2\kappa}{\mathrm{d}\eta^2}(\eta) &= \frac{\mathrm{d}}{\mathrm{d}\eta} \int_{\mathbb{X}} T(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{X}} T(\boldsymbol{x}) h(\boldsymbol{x}) \frac{\mathrm{d}}{\mathrm{d}\eta} \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{X}} T(\boldsymbol{x}) h(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} \frac{\mathrm{d}}{\mathrm{d}\eta}\Bigl(\eta T(\boldsymbol{x}) - \kappa(\eta)\Bigr) \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{X}} T^2(\boldsymbol{x}) h(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} \,\mathrm{d}\boldsymbol{x} \\
&\qquad - \underbrace{\frac{\mathrm{d}}{\mathrm{d}\eta}\kappa(\eta)}_{=\mathsf{E}[T(\boldsymbol{X}]} \int_{\mathbb{X}} T(\boldsymbol{x}) h(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \kappa(\eta)\bigr\} \,\mathrm{d}\boldsymbol{x} \\
&= \mathsf{E}\bigl[T^2(\boldsymbol{X}\bigr] - \bigl\{\mathsf{E}\bigl[T(\boldsymbol{X}\bigr]\bigr\}^2 \\
&= \mathsf{Var}\bigl[T(\boldsymbol{X}\bigr]
\end{aligned}$$

から (6.4) の 2 番目の等号が示せた. □

6.4 十分統計量

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

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

とおく. さらに $B \in \mathcal{B}(\mathbb{X})$ に対し

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

と書いたとき

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

と書ける. $\boldsymbol{X}$ により誘導された確率空間を $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}^{\boldsymbol{X}})$ と書く.

$(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta\}$ 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

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

とする.

次に統計量 $\boldsymbol{T}(\boldsymbol{X})$ を

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

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.

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

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

が $\theta \in \Theta$ に無関係であるとき, $\boldsymbol{T}$ 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. $\boldsymbol{X}$ は確率空間 $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}_\theta^{\otimes n}\big)$ 上の大きさが n の標本とし, 同時 p.d.f.(または p.m.f.) $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ [8] を持つとする. このとき統計量 $\boldsymbol{T}:\mathbb{R}^n \longrightarrow \mathbb{R}^k$ が統計的モデル

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

$\mathcal{P}$ に対する十分統計量であるための必要十分条件は, $\boldsymbol{X}$ の同時 p.d.f.(または同時 p.m.f.) $\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}|\,\theta)$ が

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = g_\theta\{\boldsymbol{T}(\boldsymbol{x})\}h(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{X}^n) \tag{6.5}$$

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

Proof. $\boldsymbol{X}$ が離散型のときのみの証明を与える. $\boldsymbol{T}$ は十分統計量とする. $\boldsymbol{T}$ の p.m.f. を $\mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})\,(\boldsymbol{t} \in \mathbb{R}^k)$ と書く. $\boldsymbol{T} = \boldsymbol{t}$ が与えられたときの $\boldsymbol{X}$ の条件付き p.m.f. を $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ とする. 仮定から $\boldsymbol{T}$ は十分統計量なので $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ は θ に依存しない. 条件付き p.m.f. の定義から

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t}) \tag{6.6}$$

となる. (6.6) において

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

とおけば, (6.5) の形になる.

次に (6.5) と書けたときに $\boldsymbol{T}$ は十分統計量であることを示す. まず

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

に注意する. (6.5) から $\boldsymbol{T} = \boldsymbol{t}$ を与えたときの $\boldsymbol{X}$ の条件付き p.m.f. は

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

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

□

注意 6.14. Fisher-Neyman の因子分解定理の測度論的な証明には Radon-Nikodym の微分に関わる知識が必要になる. これについては定理 **??** を参照のこと.

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

$$\begin{aligned} \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) &= \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j}\mathbb{1}_{\{0,\,1\}^n}(\boldsymbol{x}), \\ &\qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n, \\ &\qquad \mathbb{1}_{\{0,\,1\}^n}(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \{0, 1\}^n) \\ 0 & (\text{その他の場合}) \end{cases} \end{aligned}$$

となる. したがって

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

とおけば, 定理 6.13 から $T(\boldsymbol{X}) = X_1 + X_2 + \cdots + X_n$ は

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

の十分統計量であることがわかる. □

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

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

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

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

となる. よって (6.5) で

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

とおけば, 定理 6.13 から $T(\boldsymbol{X}) = (X_1 + X_2 + \cdots + X_n)/n$ は $\mathcal{P} = \{\mathsf{N}(\theta, 1);\, \theta \in \mathbb{R}\}$ の十分統計量であることがわかる. □

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. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

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

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

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

とおけば

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

と書き直せる. (6.5) において

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

$$h(\boldsymbol{x}) = 1$$

とすれば, 定理 6.13 から

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

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

6.5 統計量の最小十分性と完備性

6.5.1 十分統計量の最小性

与えられた統計的モデル $\mathcal{P}$ に対して, たくさんの十分統計量が存在する. $\boldsymbol{S}$ と $\boldsymbol{T}$ が $\mathsf{P} \in \mathcal{P}$ に対する十分統計量で, S の値域上で定義された可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S})$$

と表現できたとする. すると

$$\sigma(\boldsymbol{T}) \subset \sigma(\boldsymbol{S})$$

となる. このことから $\boldsymbol{T}$ は統計的モデル $\mathcal{P}$ の情報を失くことなく, データの情報を S より縮約していることがわかる. したがって T は S よりも有用な統計量となる. この観点を定式化してみよう.

定義 6.18. すべての $\mathsf{P} \in \mathcal{P}$ に対して, $\mathsf{P}(A) = 0$ となる事象 A を除いて, ある命題 M が成立しているとき

$$M \qquad \text{a.s.}\,\mathcal{P}$$

と記すことにする.

定義 6.19. (最小十分性) $\mathcal{P}$ を統計的モデルとし, 統計量 $\boldsymbol{T}$ を $\mathsf{P} \in \mathcal{P}$ に対する十分統計量とする. 十分統計量 T は $\mathsf{P} \in \mathcal{P}$ に関して最小であるとは, 任意の他の十分統計量 S に対して可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S}), \qquad \text{a.s.}\,\mathcal{P}$$

と表現できるときをいう. もちろん ψ は $\boldsymbol{S}$ ごとに定まればよい.

注意 6.20. $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を統計的モデルとする. 2 つの母数 $\theta_1, \theta_2\,(\in \Theta)$ の疑似測度を

$$\rho(\theta_1, \theta_2) = \sup_{B \in \mathcal{B}(\mathbb{X})} \left| \mathsf{P}_{\theta_1}(B) - \mathsf{P}_{\theta_2}(B) \right|$$

に関して Θ が可分[9] であるとき, $\mathcal{P}$ に対する最小十分統計量が存在することが知られている. 証明は [51, pp.78-81] を参照のこと. □

例 6.21. $\mathcal{P} = \big\{\mathsf{P}$ は開区間 $(\theta, \theta+1)$ 上の一様分布; $\theta \in \mathbb{R}\big\}$ とする. $n \geq 2\,(\in \mathbb{N})$ とし

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P} \in \mathcal{P}$$

とする. Lebesgue 測度に関する $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\begin{aligned} \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) &= \Pi_{j=1}^n \mathbb{1}_{(\theta,\, \theta+1)}(x_j) = \mathbb{1}_{(x_{(n)}-1,\, x_{(1)})}(\theta), \\ \boldsymbol{x} &= (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

となる. ただし $x_{(1)}, x_{(n)}$ は $x_1, x_2, \ldots, x_n$ の最小値と最大値である.

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}, \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

とおくと Fisher-Neyman の因子分解定理から $\boldsymbol{T} = (X_{(1)}, X_{(n)})$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量である.

次に, 十分統計量 $\boldsymbol{T}$ が最小性もみたすことを示す.

$$\begin{aligned} &\theta < x_j < \theta + 1 \quad (\forall j \in \{1, 2, \ldots, n\}) \\ &\Leftrightarrow\ \theta < x_{(1)} < x_{(n)} < \theta + 1 \end{aligned} \tag{6.7}$$

となることに注意する. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき

$$x_{(1)} = \sup\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}, \quad x_{(n)} = 1 + \inf\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}$$

となる. ここで統計量 $\boldsymbol{S}$ は $\mathsf{P} \in \mathcal{P}$ に対して十分であると仮定する. Fisher-Neyman の因子分解定理から可測関数 g_θ と h が存在して

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = g_\theta\big(\boldsymbol{S}(\boldsymbol{x})\big) h(\boldsymbol{x})$$

[9] Θ の可算無限集合 C が存在して, 任意の θ と ϵ に対して, $\theta_c \in C$ をうまく選ぶと $\rho(\theta, \theta_c) < \epsilon$ とできることである.

と書ける．$A^c := \{\boldsymbol{x} \in \mathbb{R}^n;\, h(\boldsymbol{x}) > 0\}$ 上で (6.7) が成立するので, ある可測関数 ψ が存在して A 上で

$$\boldsymbol{T} := (x_{(1)},\, x_{(n)}) = \psi\big(\boldsymbol{S}(\boldsymbol{x})\big)$$

と表現できる．さらに $\mathsf{Pr}(\boldsymbol{X} \in A) = 0$ なので $\boldsymbol{T}$ は最小十分統計量となる．ここでは, 最小十分統計量であることを直接的に確認したが, 次の定理 6.22 の判定条件を用いる簡単に最小性の確認をできる． □

定理 6.22. $\big(\mathbb{X}, \mathcal{B}\big)$ を標本空間とし, それ上の統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は測度 μ で支配されており, p.d.f. $\mathsf{p}(x|\,\theta) = g_\theta(T(x))h(x)$ を持つとする．任意の x, y に対して, 関数

$$\theta \mapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

が定数ならば, $T(x) = T(y)$ が成り立つとき, T は最小十分統計量となる．

Proof. 厳密な証明は測度論の知識を用いることになる．ここでは証明の基本的な考え方を述べるだけにする．$\widetilde{T}$ を任意の十分統計量とする．すると Fisher-Neyman の因子分解定理から

$$\mathsf{p}_\theta(x) = \widetilde{g}_\theta(\widetilde{T}(x))\widetilde{h}(x)$$

と書ける．かりに T は $\widetilde{T}$ の関数でないとすると, あるデータ x と y があって,

$$\widetilde{T}(x) = \widetilde{T}(y) \quad \text{かつ} \quad T(x) \neq T(y) \tag{6.8}$$

とできること [10] になる．しかし

$$\frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)} = \frac{\widetilde{g}_\theta(\widetilde{T}(x))\widetilde{h}(x)}{\widetilde{g}_\theta(\widetilde{T}(y))\widetilde{h}(y)} = \frac{\widetilde{h}(x)}{\widetilde{h}(y)}$$

となり, 関数 $\theta \mapsto \frac{p(x|\,\theta)}{p(y|\,\theta)}$ は定数となる．よって, T が最小十分統計量であるという仮定より $T(x) = T(y)$ となり, (6.8) と矛盾する．したがって, ある関数 f があって, $\widetilde{T} = f(\widetilde{T})$ とかける． □

[10] $\widetilde{T} = f(T)$ と書けるとき,

$$T(x) = T(y) \Longrightarrow \widetilde{T}(x) = \widetilde{T}(y)$$

なので, $\widetilde{T} = f(T)$ と書けない場合は, ある x, y があって

$$T(x) = T(y) \quad \text{かつ} \quad \widetilde{T}(x) \neq \widetilde{T}(y)$$

となる．

例 6.23. $(\mathbb{X}, \mathcal{B})$ を標本空間とし, $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を $(\mathbb{X}, \mathcal{B})$ 上の d 母数指数型分布族とする. すなわち p.d.f. $\mathsf{p}(x|\,\theta)\,(\theta \in \Theta)$ が

$$\mathsf{p}_\theta(x) = \exp\{\big(\eta(\theta)|\,T(x)\big) - \kappa(\theta)\}h(x)$$

と書ける. ただし $\kappa\colon \Theta \longrightarrow \mathbb{R}^d$ であり, $T \in \mathbb{R}^d$ で, $(\cdot|\cdot)$ は $\mathbb{R}^d$ の Euclid 内積である. Fisher-Neyman の因子分解定理から T は統計的モデル $\mathcal{P}$ に対する十分統計量となる. データ x と y は, 関数

$$\theta \mapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

が定数となるものとする. するとある θ に依存しない定数 c (x と y に依存してよい) があって

$$\big(\eta(\theta)|\,T(x)\big) = \big(\eta(\theta)|\,T(y)\big) + c$$

と書けることになる. 任意の点 $\theta_0,\, \theta_1 \in \Theta\,(\theta_0 \neq \theta_1)$ に対して

$$\begin{aligned}\big(\big[\eta(\theta_0) - \eta(\theta_1)\big]|\,T(x)\big) &= \big(\big[\eta(\theta_0) - \eta(\theta_1)\big]|\,T(y)\big)\\ \Leftrightarrow\ \big(\big[\eta(\theta_0) - \eta(\theta_1)|\,T(x) - T(y)\big) &= 0\end{aligned}$$

を得る. $\mathrm{span}\Big[\{\eta(\theta_0) - \eta(\theta_1) \in \mathbb{R}^d :\, \theta_0,\, \theta_1 \in \Theta\}\Big] = \mathbb{R}^d$ のとき, $T(x) = T(y)$ となり, T は最小十分統計量となる. □

例 6.24. $n \geq 2\,(n \in \mathbb{N})$ と $\theta \in \mathbb{R}$ とし, $X_1,\, X_2,\, \ldots,\, X_n$ は独立同一に分布する確率変数列で, 共通の p.d.f.

$$\mathsf{p}_\theta(x) = \frac{1}{2}e^{-|x-\theta|} \qquad (x \in \mathbb{R})$$

を持つとする. $\boldsymbol{X} :=(X_1,\, X_2,\, \ldots,\, X_n)$ の同時 p.d.f. $\mathsf{p}_X(\boldsymbol{x}|\,\theta)$ は

$$\mathsf{p}_X(\boldsymbol{x}|\,\theta) = \frac{1}{2^n}\exp\Big\{-\sum_{i=1}^n |x_i - \theta|\Big\} \qquad \big(\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n) \in \mathbb{R}^n\big)$$

となる. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ を順序統計量とする. Fisher-Neyman の因子分解定理から $\boldsymbol{T} = (X_{(1)},\, X_{(2)},\, \ldots,\, X_{(n)})$ は十分統計量となる. データを $x = (x_1,\, x_2,\, \ldots,\, x_n)$ と $y = (y_1,\, y_2,\, \ldots,\, y_n)$ としたとき, $\sum_{i=1}^n |x_i - \theta|$ と $\sum_{i=1}^n |y_i - \theta|$ の差は θ に依存しない. この二つの θ の関数は区分的に連続で, 不連続点は $x_{(1)},\, x_{(2)},\, \ldots,\, x_{(n)}$ と $y_{(1)},\, y_{(2)},\, \ldots,\, y_{(n)}$ である. 関数

$$\theta \mapsto \sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |y_i - \theta|$$

が θ に関して定数であるときは, ジャンプの点が一致することである. したがって, $x_{(i)} = y_{(i)}\,(i = 1,\, 2,\, \ldots,\, n)$ となるので, $\boldsymbol{T}$ は最小十分であることがわかる.

6.5.2 十分統計量の完備性

定義 6.25. $k \in \mathbb{R}$ とし, $\mathbb{T} \subset \mathbb{R}^k$ とする. $\mathcal{P}$ を統計的モデルとし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P} \in \mathcal{P}$ とする. $\boldsymbol{T}(\boldsymbol{X})\,(\in \mathbb{T})$ を $\mathsf{P} \in \mathcal{P}$ に対する十分統計量とする. ただし, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ と書いた. 統計量 $\boldsymbol{T}$ は $\mathsf{P} \in \mathcal{P}$ に対して完備であるとは, 任意の可測関数 $f : \mathbb{T} \to \mathbb{R}$ に対して

$$\mathsf{E}\big[f(\boldsymbol{T})\big] = 0\,(\forall \mathsf{P} \in \mathcal{P}) \;\Rightarrow\; f(\boldsymbol{T}) = 0,\ \text{a.s.}\,\mathcal{P}$$

が成り立つときをいう.

次の命題は, 指数型分布族の十分統計量は完備であることを主張している. 本講義録では証明していない Fubini の定理と Laplace 変換の一意性を用いてこの命題を証明することができる.

命題 6.26. $k, n \in \mathbb{N}$ かつ $n > k$ とする. $\boldsymbol{\eta} \in \mathbb{R}^k$, $\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とし

$$\mathsf{p}_{\boldsymbol{\eta}}(\boldsymbol{x}) = \exp\big\{\boldsymbol{\eta}^\top \boldsymbol{T}\big(x\big)\big\} h(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{X}^n)$$

とする. ただし

$$\mathcal{E} := \left\{\boldsymbol{\eta} \in \mathbb{R}^k;\, \kappa(\boldsymbol{\eta}) := \int_{\mathbb{X}^n} \exp\{\boldsymbol{\eta}^\top \boldsymbol{T}(\boldsymbol{x})\} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \infty\right\}$$

は $\mathbb{R}^n$ の開集合を含み

$$\boldsymbol{T} : \mathbb{X}^n \to \mathbb{R}^k, \quad h : \mathbb{X}^n \to \mathbb{R}, \quad \kappa : \mathcal{E} \to \mathbb{R}$$

は可測関数とする. $\boldsymbol{X} \sim \mathsf{P}\,(\mathsf{P} \in \mathcal{P} := \{\mathsf{p}_{\boldsymbol{\eta}}(\boldsymbol{x});\, \boldsymbol{\eta} \in \mathcal{E}\})$ としたとき, $\boldsymbol{T}(\boldsymbol{X})$ は $\mathsf{P} \in \mathcal{P}$ に対する完備かつ十分統計量である.

Proof. Fisher-Neyman の因子分解定理から $\boldsymbol{T}$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量であることは明らか. つぎに $f : \mathbb{R}^k \to \mathbb{R}$ を可測関数とし

$$\mathsf{E}\big[f(\boldsymbol{T})\big] = 0 \quad (\forall \boldsymbol{\eta} \in \mathcal{E})$$

とする. $\mathsf{E}\big[|f(\boldsymbol{T})|\big] < \infty$ なので Fubini の定理から

$$\begin{aligned}
0 &= \int_{\mathbb{X}} f\big(\boldsymbol{T}(\boldsymbol{x})\big) \exp\big\{\boldsymbol{\eta}^\top \boldsymbol{T}(\boldsymbol{x}) - \kappa(\boldsymbol{\eta})\big\} h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\boldsymbol{x} :\, \boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}} h(\boldsymbol{x}) \left\{\int_{\mathbb{T}} f(\boldsymbol{t}) \exp\big\{\boldsymbol{\eta}^\top \boldsymbol{t} - \kappa(\boldsymbol{\eta})\big\}\,\mathrm{d}\boldsymbol{t}\right\} \mathrm{dm}_{n-k}
\end{aligned}$$

と書ける．ただし $\mathbb{T} = \{\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}^k;\, \boldsymbol{x} \in \mathbb{X}\}$ で, m_{n-k} は $\mathbb{R}^{n-k}$ 上の Lebeague 測度である．よって

$$\int_{\mathcal{T}} f(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t} - \kappa(\boldsymbol{\eta})\bigr\}\,\mathrm{d}\boldsymbol{t} = 0 \qquad \text{a.s. } \mathsf{m}_{n-k}$$

となる．$\boldsymbol{\eta}_0 \in \mathcal{E}$ の内点とし, $\epsilon > 0$ を十分小さく取ると

$$\int f_+(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t}\bigr\}\,\mathrm{d}\boldsymbol{t} = \int f_-(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t}\bigr\}\,\mathrm{d}\boldsymbol{t}$$
$$\bigl(\forall \boldsymbol{\eta} \in N(\boldsymbol{\eta}_0) := \{\boldsymbol{\eta} \in \mathcal{R}^k;\, |\boldsymbol{\eta} - \boldsymbol{\eta}_0|_{2,k} < \epsilon\bigr)$$

となる．ここで

$$f_+(\boldsymbol{t}) := \max\{f(\boldsymbol{t}),\, 0\}, \quad f_-(\boldsymbol{t}) := \max\{-f(\boldsymbol{t}),\, 0\}$$

で, $|\cdot|_{2,k}$ は $\mathbb{R}^k$ 上の Euclid の距離である．特に

$$\int f_+(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\bigr\}\,\mathrm{d}\boldsymbol{t} = \int f_-(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\bigr\}\,\mathrm{d}\boldsymbol{t} =: c$$

となる．$c = 0$ のとき, $f(\boldsymbol{t}) = 0\,(\text{a.s.})$ となる．$c > 0$ のとき, Laplace 変換の一意性[11] から

$$f_+(\boldsymbol{t}) = f_-(\boldsymbol{t}), \quad \text{a.s.}$$

となる．よって $f(\boldsymbol{t}) = f_+(\boldsymbol{t}) - f_-(\boldsymbol{t}) = 0$ がわかる． □

注意 6.27. 完備十分統計量は最小十分統計量であることが知られている．しかし最小十分統計量で完備でないものの存在が知られている． □

6.6 補遺: Laplace 変換の一意性

定義 6.28. $[0, \infty)$ 上に集中した分布 P について

$$f(s) = \int_{[0,\infty)} e^{-sx}\,\mathrm{d}\mathsf{P}(x)$$

を P の Laplace **変換**という．また, 非負値確率変数 X の Laplace 変換とは

$$f(s) = \mathsf{E}\bigl[e^{-sX}\bigr] = \int_{[0,\infty)} e^{-sx}\,\mathrm{d}\mathsf{P}(x) \qquad (s \geq 0)$$

のことである．

[11]Laplace 変換の一意性の証明は次章の定理 *A.8* を参照のこと．

補題 6.29. 区間 $[0,\,1]$ に集中した分布の分布関数 G_1, G_2 に対して

$$\int_0^1 x^n\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 x^n\,\mathrm{d}\mathsf{G}_2(x) \qquad (n = 0,\,1,\,2,\ldots) \tag{6.9}$$

が成り立てば

$$\mathsf{G}_1 = \mathsf{G}_2$$

である.

Proof. Weierstrass の多項式近似定理により, $(A.15)$ が成り立てば, 任意の $[0,\,1]$ 上の連続関数 f に対して

$$\int_0^1 f(x)\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 f(x)\,\mathrm{d}\mathsf{G}_2(x)$$

が成り立つ.

いま, $0 \leq a < 1$ と $n \geq 1$ に対して

$$f_n(x) = \begin{cases} 1 & (0 \leq x \leq a), \\ n(x-a) & (a \leq x \leq a + 1/n), \\ 0 & (n + 1/n \leq x \leq 1) \end{cases}$$

と定める. すると $\mathsf{G}_1(0)\!= \mathsf{G}_2(0) = 0$ より

$$\begin{aligned} G_1(a) &= \int_0^a \mathrm{d}\mathsf{G}_1 = \lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}\mathsf{G}_1(x) = \lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}\mathsf{G}_2(x) \\ &= \int_0^a \mathrm{d}\mathsf{G}_2 = G_2(a) \end{aligned}$$

を得る. また $\mathsf{G}_1(1) = \mathsf{G}_2(1)$ だから, すてべの $a \in [0,\,1]$ に対して $\mathsf{G}_1(a) = \mathsf{G}_2(a)$ がわかる. □

定理 6.30. $[0,\,\infty)$ に集中した確率をもつ 2 つの分布関数 F_1, F_2 に対して, それぞれの Laplace 変換を f_1, f_2 とする. このとき, f_1 と f_2 が一致すれば, $\mathsf{F}_1 = \mathsf{F}_2$ である.

Proof. $\mathsf{F}\colon\,[0,\,\infty) \longrightarrow \mathbb{R}$ を確率分布関数とする. $x = e^{-t}$, $\mathsf{G}(x) = 1 - \mathsf{F}(t)$ とすると

$$\int_0^\infty e^{-\lambda t}\,\mathrm{d}\mathsf{F}(t) = \int_0^1 x^\lambda\,\mathrm{d}\mathsf{G}(x) \qquad (\lambda \geq 0)$$

となる. よって $f_1 = f_2$ ならば,

$$\int_0^1 x^n\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 x^n\,\mathrm{d}\mathsf{G}_2(x) \qquad (n = 0,\,1,\,2,\ldots)$$

となる. あとは補題 A.7 を用いれば, 定理は証明される. □

6.7　章末注釈と参考文献

第 6.1 節は $[1, \mathrm{pp}.69-71]$ を借用した. 補題 6.10 は $[17, \mathrm{pp}.28-30]$ を借用した.

6.8　演習問題

演習問題 6.1. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Po}(\theta)\,(\theta > 0)$ とする. $S := X_1 + X_2 + \cdots + X_n$ は $\mathcal{P} = \{\mathsf{Po}(\theta);\, \theta > 0\}$ の十分統計量であることを示せ.

演習問題 6.2.