

第6章 母数モデル

6.1 正則母数モデルと Fisher 情報量

定義 6.1. $d \in \mathbb{N}$ とし, $\mathbb{X} (\subset \mathbb{R}^d)$ を標本空間とする. \mathbb{X} 上の母数モデル $\mathcal{P} = \{P_\theta; \theta \in \Theta \subset \mathbb{R}^k\}$ が与えられているとする. P_θ の同時 p.d.f. $p(\mathbf{x}|\theta)$ と表記¹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) $\Theta \subset \mathbb{R}^k$ は開集合.
- (2) \mathcal{P} に属する分布の p.d.f. は同じ台をもつ. すなわち $\{\mathbf{x} \in \mathbb{R}^n; p(\mathbf{x}|\theta) > 0\}$ は $\theta \in \Theta$ に依存しない.
- (3) $\forall \theta \in \Theta$ とする. $p(\mathbf{x}|\theta)$ の θ の 1 次と 2 次の偏導関数は \mathbf{x} に関して連続である.
- (4) $p(\mathbf{x}|\theta)$ の θ に関する 1 次と 2 次の偏導関数は \mathbf{x} の関数として可積分である.
- (5) $p(\mathbf{x}|\theta)$ の 1 次と 2 次の偏導関数は θ の微分記号と \mathbf{x} の積分記号と交換が可能である.

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

$$\ell(\theta) = \log p(\mathbf{x}|\theta) \quad (\mathbf{x} = (x_1, x_2, \dots, x_n))$$

で定義²する.

$X \sim P_\theta$ のときこの分布に関する期待値, 分散および共分散を $E_\theta[\cdot], \text{Var}_\theta[\cdot], \text{Cov}_\theta[\cdot, \cdot]$ と表記する.

さらに

$$S_j := S_j(\mathbf{x}|\theta) = \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_j}, \quad \dot{S}_{jk} := \dot{S}_{jk}(\mathbf{x}|\theta) = \frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta_j \partial \theta_k}, \quad (j, k = 1, 2, \dots, d)$$

$$\mathbf{S} = (S_1, S_2, \dots, S_d)^\top, \quad \dot{\mathbf{S}} = (\dot{S}_{jk}), \quad \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^\top$$

¹ここでは, 離散型確率分布のときでも p.d.f. と記すことにする. さらに, 積分は和の記号と読み替える.

²対数 尤度 関数 と呼ぶ.

と定義する. $\mathbf{X} \sim P_\theta$ のとき Fisher 情報量 $\mathcal{F}_X(\theta)$ は $d \times d$ 行列で $\mathcal{F}_X(\theta)$ の (j, k) 成分 $\{\mathcal{F}_X(\theta)\}_{jk}$ は

$$\{\mathcal{F}_X(\theta)\}_{jk} = E_\theta[S_j(\mathbf{X}|\theta)S_k(\mathbf{X}|\theta)], \quad (1 \leq j, k \leq d)$$

で定義される.

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

- (1) $\forall \theta \in \Theta$ と $j = 1, 2, \dots, d$ に対して $E_\theta[S_j(\mathbf{X}|\theta)] = 0$ となる.
- (2) $\mathcal{F}_X(\theta) = \text{Cov}_\theta[S(\mathbf{X}|\theta)] = E_\theta[S(\mathbf{X}|\theta)S^\top(\mathbf{X}|\theta)]$ が成り立つ.
- (3) $\mathcal{F}_X(\theta) = E_\theta[\dot{S}(\mathbf{X}|\theta)]$ が成り立つ.

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換えればよい. P_θ の同時 p.d.f. を $p(\mathbf{x}|\theta)$ と表記する. 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

$$\begin{aligned} E_\theta[S_j(\mathbf{X}|\theta)] &= \int_{\mathbb{X}} \left(\frac{\partial}{\partial \theta_j} \log p(\mathbf{x}|\theta) \right) p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \int_{\mathbb{X}} \frac{1}{p(\mathbf{x}|\theta)} \left(\frac{\partial}{\partial \theta_j} p(\mathbf{x}|\theta) \right) p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \int_{\mathbb{X}} \left(\frac{\partial}{\partial \theta_j} p(\mathbf{x}|\theta) \right) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta_j} \int_{\mathbb{X}} p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta_j} 1 = 0 \end{aligned}$$

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.

$$\frac{\partial^2 \log p(\mathbf{x}|\theta)}{\partial \theta_j \partial \theta_k} = \frac{1}{p(\mathbf{x}|\theta)} \frac{\partial^2 p(\mathbf{x}|\theta)}{\partial \theta_j \partial \theta_k} - \left(\frac{\partial}{\partial \theta_j} \log p(\mathbf{x}|\theta) \right) \left(\frac{\partial}{\partial \theta_k} \log p(\mathbf{x}|\theta) \right).$$

このとき

$$\begin{aligned} E_\theta \left[\frac{1}{p(\mathbf{X}|\theta)} \frac{\partial^2 p(\mathbf{X}|\theta)}{\partial \theta_j \partial \theta_k} \right] &= \int_{\mathbb{X}} \frac{1}{p(\mathbf{x}|\theta)} \left(\frac{\partial^2 p(\mathbf{x}|\theta)}{\partial \theta_j \partial \theta_k} \right) p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \frac{\partial^2}{\partial \theta_j \partial \theta_k} \int_{\mathbb{X}} p(\mathbf{x}|\theta) d\mathbf{x} \\ &= \frac{\partial^2}{\partial \theta_j \partial \theta_k} 1 = 0 \end{aligned}$$

となる. このことより

$$\begin{aligned} E_{\theta} \left[\frac{\partial^2 \log p(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right] &= \int_{\mathbf{X}} \frac{1}{p(\mathbf{x}|\boldsymbol{\theta})} \left(\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right) p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} \\ &\quad - E_{\theta} \left[\left(\frac{\partial}{\partial \theta_j} \log p(\mathbf{X}|\boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta_k} \log p(\mathbf{X}|\boldsymbol{\theta}) \right) \right] \\ &= -E_{\theta} \left[\left(\frac{\partial}{\partial \theta_j} \log p(\mathbf{X}|\boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta_k} \log p(\mathbf{X}|\boldsymbol{\theta}) \right) \right] \\ &= -\{\mathcal{F}_{\mathbf{X}}(\boldsymbol{\theta})\}_{jk} \end{aligned}$$

がわかる. □

例 6.3. $\mathcal{P} = \{N(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\}$ とする. 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

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = -\sum_{j=1}^n \frac{(x_j - \mu)^2}{2\sigma^2} - n \log \sigma - n \log(\sqrt{2\pi}) \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

となる. よって

$$S_1(\mathbf{x}|\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \mu} = \sum_{j=1}^n \frac{x_j - \mu}{\sigma^2}, \quad S_2(\mathbf{x}|\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \sigma} = \sum_{j=1}^n \frac{(x_j - \mu)^2}{\sigma^3} - \frac{n}{\sigma}$$

となる. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ で $X_1, X_2, \dots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$ のとき

$$\begin{aligned} E_{\theta} [S_1^2(\mathbf{X}|\boldsymbol{\theta})] &= \frac{n}{\sigma^2}, \\ E_{\theta} [S_2^2(\mathbf{X}|\boldsymbol{\theta})] &= \frac{3n}{\sigma^2} - \frac{2n}{\sigma^2} + \frac{n}{\sigma^2} = \frac{2n}{\sigma^2} \\ E_{\theta} [S_1(\mathbf{X}|\boldsymbol{\theta})S_2(\mathbf{X}|\boldsymbol{\theta})] &= 0 \end{aligned}$$

となる. よって

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

を得る.

一方

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

となる. よって

$$E_{\theta}[\dot{S}_{11}(X|\theta)] = -\frac{n}{\sigma^2}, E_{\theta}[\dot{S}_{22}(X|\theta)] = -\frac{2n}{\sigma^2}, E_{\theta}[\dot{S}_{12}(X|\theta)] = 0$$

なので

$$\begin{bmatrix} E_{\theta}[\dot{S}_{11}(X|\theta)] & E_{\theta}[\dot{S}_{12}(X|\theta)] \\ E_{\theta}[\dot{S}_{12}(X|\theta)] & E_{\theta}[\dot{S}_{22}(X|\theta)] \end{bmatrix} = -\mathcal{F}_X(\theta)$$

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 $\{P_{\theta}; \theta \in \Theta\}$ は 1 母数指数型分布族であるとは, Θ 上の実数値関数 $A(\theta)$ と $\kappa(\theta)$, \mathbb{R}^d 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. $p(\mathbf{x}|\theta)$ が

$$p(\mathbf{x}|\theta) = h(\mathbf{x}) \exp\{A(\theta)T(\mathbf{x}) - \kappa(\theta)\}, \quad \mathbf{x} \in \mathbb{X} \subset \mathbb{R}^d. \quad (6.1)$$

の形でかけるときをいう.

注意 6.5. 関数 A, B, T の表現は一意ではない. □

$X \sim p(\mathbf{x}|\theta)$ のとき統計量 $T(X)$ は θ の十分統計量になる. T をこの分布族に対する自然十分統計量という. なお十分統計量については節 6.3 で定義する.

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

$$p(x|\theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!} \exp\{x \log \theta - \theta\} \quad (x = 0, 1, \dots).$$

したがって $\{\text{Pois}(\theta); \theta \in (0, \infty)\}$ は

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

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.

例 6.7. (2 項分布族) $X \sim \text{Bino}(n, \theta)$ 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

$$\begin{aligned} p(x|\theta) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \binom{n}{x} \exp\left\{x \log\left(\frac{\theta}{1-\theta}\right) + n \log(1-\theta)\right\} \quad (x = 0, 1, \dots, n) \end{aligned}$$

と書ける. ただし

$$\binom{n}{x} = \frac{n!}{x! \times (n-x)!}, \quad 0! = 1$$

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

$$d = 1, A(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \kappa(\theta) = -n \log(1-\theta), T(x) = x, h(x) = \binom{n}{x}$$

によって生成される 1 母数指数型分布族である. \square

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

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

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

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

である. したがって \mathbf{X} の分布は

$$d = 2, A(\theta) = -\frac{1}{2\theta^2}, \kappa(\theta) = \log \theta, T(\mathbf{x}) = (y-z)^2, h(\mathbf{x}) = \frac{1}{2\pi} \exp\left\{-\frac{z^2}{2}\right\}$$

で生成される 1 母数指数型分布族に属する. \square

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 $X_1, X_2, \dots, X_n \sim \text{i.i.d. } P_\theta$ とし, $\{P_\theta; \theta \in \Theta\}$ は (6.1) で与えられる指数型分布族とする. $\{P_\theta; \theta \in \Theta\}$ は $\mathbf{X} = (X_1, X_2, \dots, X_n)$ に対する指数型分布族となる. ただし

$$P_\theta^{\otimes n} = \underbrace{P_\theta \times P_\theta \times \dots \times P_\theta}_{n \text{ 個}}$$

で \mathbb{R}^{dn} 上の確率測度である. P_θ の同時 p.d.f. は次で与えられる. $\mathbf{x} = (x_1, x_2, \dots, x_n)$ としたとき

$$\begin{aligned} p(\mathbf{x}|\theta) &= \prod_{j=1}^n h(x_j) \exp\left\{A(\theta)T(x_j) - \kappa(\theta)\right\} \\ &= \left[\prod_{j=1}^n h(x_j)\right] \exp\left\{A(\theta) \sum_{j=1}^n T(x_j) - n\kappa(\theta)\right\} \end{aligned}$$

となる. したがって $\{P_\theta; \theta \in \Theta\}$ は

$$A(\theta), \sum_{j=1}^n T(x_j), n\kappa(\theta), \prod_{j=1}^n h(x_j)$$

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 $\sum_{j=1}^n T(X_j)$ は θ の自然十分統計量となる.

指数型分布族の正準表示: (6.1) で表現された指数型分布族を θ でなく η で添え字付けることを考える. すると 1 母数指数型分布族は

$$q(x|\eta) = h(x) \exp\{\eta T(x) - \kappa(\eta)\} \quad (x \in \mathcal{X} \subset \mathbb{R}^d) \quad (6.2)$$

と書くことができる. ただし

$$\kappa(\eta) = \begin{cases} \log\left(\sum_x h(x) \exp\{\eta T(x)\}\right) & \text{(離散型の場合)} \\ \log\left(\int \dots \int h(x) \exp\{\eta T(x)\} dx\right) & \text{(連続型の場合)} \end{cases}$$

である⁴ $\theta \in \Theta$ に対して $K(\eta)$ は有界である.

逆に

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

とおくと, \mathcal{E} は \mathbb{R} の区間となる. すると確率分布族 $\{q(x|\eta); \eta \in \mathcal{E}\}$ は確率分布族 $\{p(x|\theta); \theta \in \Theta\}$ を含む. 確率分布族 $\{q(x|\eta); \eta \in A(\eta)\}$ は T, h によって生成された正準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.

⁴ K は Greek letter 大文字の κ である.

例 6.9. (例 6.6 の続き) Poisson 分布族は

$$\begin{aligned}
 q(x|\eta) &= \frac{1}{x!} \exp\{\eta x - e^\eta\} \quad (x = 0, 1, 2, \dots), \\
 \eta &= \log \theta, \quad h(x) = \frac{1}{x!}, \quad T(x) = x, \\
 \exp(K(\eta)) &= \sum_{x=0}^{\infty} h(x) \exp(\eta T(x)) = \sum_{x=0}^{\infty} \frac{\exp(\eta x)}{x!} = \sum_{x=0}^{\infty} \frac{(e^\eta)^x}{x!} = \exp(e^\eta), \\
 \mathcal{E} &= \mathbb{R}
 \end{aligned}$$

となる.

□

補題 6.10.

$$\mathcal{E} := \left\{ \eta \in \mathbb{R}; \int \exp[\eta T(x)] h(x) dx < \infty \right\}$$

とする. このとき, 関数

$$\kappa(\eta) = \log \int \exp\{\eta T(x)\} h(x) dx \quad (\eta \in \mathcal{E}^\circ)$$

は無有限回微分可能である. ただし \mathcal{E}° 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 $[\eta - 2\epsilon, \eta + 2\epsilon] \subset \mathcal{E}^\circ$ とできる.

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

と書ける. ただし

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

である. すると

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

となる. ここで

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

に注意する. これらの不等式を用いると

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

を得る. よって

$$|f_n(x)| \leq \frac{1}{\epsilon} \left\{ \exp(2\epsilon T(x)) + \exp(-2\epsilon T(x)) \right\} h(x)$$

となるので

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

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

$$\int g(x) dx < \infty$$

となる. よって優収束定理から

$$\frac{d}{d\eta} \exp\{\kappa(\eta)\} = \lim_{n \rightarrow \infty} \int \exp\{\eta T(x)\} f_n(x) dx = \int \exp\{\eta T(x)\} T(x)h(x)$$

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. \square

定理 6.11. X は (6.2) で与えられた指数型分布族に属する分布 $q(x|\eta)$ に従うとする. η は \mathcal{E} の内点としたとき, $T(X)$ の積率母関数は原点の近傍で存在し

$$M_T(s) = \exp[\kappa(s + \eta) - \kappa(\eta)] \quad (s \text{ は } 0 \text{ のある近傍に含まれる})$$

で与えられる. さらに

$$E[T(X)] = \dot{\kappa}(\eta), \quad \text{Var}[T(X)] = \ddot{\kappa}(\eta)$$

である. ただし

$$\dot{\kappa}(\eta) = \frac{d\kappa}{d\eta}(\eta), \quad \ddot{\kappa}(\eta) = \frac{d^2\kappa}{d\eta^2}(\eta)$$

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. K の定義に注意すると

$$\begin{aligned} M(s) &= E[\exp(sT(X))] \\ &= \int_{\mathbb{R}^p} h(x) \exp[(s + \eta)T(x) - \kappa(\eta)] dx \\ &= \exp[\kappa(s + \eta) - \kappa(\eta)] \int_{\mathbb{R}^p} h(x) \exp[(s + \eta)T(x) - \kappa(s + \eta)] dx \\ &= \exp[\kappa(s + \eta) - \kappa(\eta)] \end{aligned}$$

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

$$\int_{\mathbb{R}^p} h(x) \exp[(s + \eta)T(x) - \kappa(s + \eta)] dx = 1$$

となることよりわかる. 次に, 補題 6.10 から

$$\int T(x) \exp\{\eta T(x)\} h(x) dx = \frac{d}{d\eta} \exp\{\kappa(\eta)\} = \dot{\kappa}(\eta) \exp\{\kappa(\eta)\}$$

となるので

$$\dot{\kappa}(\eta) = \int T(x) \exp\{\eta T(x) - \kappa(\eta)\} h(x) dx = E[T(X)]$$

がわかる. □

6.3 十分統計量

$n \in \mathbb{N}$ とし, X_1, X_2, \dots, X_n は確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数列で, 互いに独立で各々同一の確率分布を持つとする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ とおき $B \in \mathcal{B}(\mathbb{R}^n)$ に対し

$$\begin{aligned} P_{\mathbf{X}}(B) &= \Pr\{\mathbf{X}^{-1}(B)\} \\ &= \Pr\left(\{\omega \in \Omega; \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \in B\}\right) \end{aligned}$$

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

$$P_{X_j}(B) = \Pr\left(\{\omega \in \Omega; X_j(\omega) \in B\}\right) \quad (j = 1, 2, \dots, n)$$

と書いたとき

$$P_{\mathbf{X}} = P_{X_1} \times P_{X_2} \times \dots \times P_{X_n}$$

と書ける. \mathbf{X} により誘導された確率空間を $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{\mathbf{X}})$ とかく.

$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の統計的モデルを $\mathcal{P} = \{P_{\theta}; \theta \in \Theta\}$ 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

$$P_{\mathbf{X}} = P_{\theta^*}$$

とする.

次に統計量 $T(\mathbf{X})$ を

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

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. $\mathbf{X} \sim P_{\theta} (\theta \in \Theta)$ とする. $\forall B \in \mathcal{B}(\mathbb{R}^n)$ に対し統計量 $T(\mathbf{X})$ を与えたときの条件付き確率

$$P_{\theta}(B|T)$$

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

定理 6.13. (因子分解定理) 統計的モデルを $\mathcal{P} = \{P_{\theta}; \theta \in \Theta\}$ とする. ただし θ は母数で Θ は母数空間とする. \mathbf{X} は確率空間 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{\theta})$ 上の大きさが n の標本とし, 同時 p.d.f. $p(\mathbf{x}|\theta)$ ⁵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, \mathbf{X} の同時 p.d.f.(または同時 p.m.f.) $p(\mathbf{x}|\theta)$ が

$$p(\mathbf{x}|\theta) = g_{\theta}\{T(\mathbf{x})\}h(\mathbf{x}) \quad (6.3)$$

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

⁵離散型のときは同時 p.m.f. $p(\mathbf{x}|\theta)$ をもつ.

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

$$p_{X|T}(x|\theta) = p^T(t|\theta)p_{X|T}(x|T = t). \quad (6.4)$$

(6.4) において

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

とおけば, (6.3) の形になる.

次に (6.3) と書けたときに T は十分統計量であることを示す. まず

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

に注意する. $T(x) = t$ とする. (6.3) から $T = t$ を与えたときの X の条件付き p.m.f. は

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

となり, $p_{X|T}$ は θ に依存しない. よって T は $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ 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 C.70 を参照のこと.

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

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

したがって

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

とおけば, 定理 6.13 から $T(\mathbf{X}) = X_1 + X_2 + \cdots + X_n$ は $\mathcal{P} = \{p(\mathbf{x}|\theta); 0 \leq \theta \leq 1\}$ 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, $X_1, X_2, \dots, \sim X_n$ i.i.d. $\mathbf{N}(\theta, 1)$ とする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

$$p(\mathbf{x}|\theta) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{j=1}^n (x_j - \theta)^2\right\} \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

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

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

となる. よって (6.3) で

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

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

例 6.17. $\nu \in \mathbb{R}, 0 < \sigma < \infty$ とする. $X_1, X_2, \dots, X_n \sim$ i.i.d. $\mathbf{N}(\mu, \sigma^2)$ とする. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の同時 p.d.f. は

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

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

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

とおけば

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

と書き直せる. (6.3) において

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

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

とすれば, 定理 6.13 から

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

は $\mathcal{P} = \{N(\mu, \sigma^2); \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ の十分統計量である. \square

6.4 統計量の最小十分性と完備性

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. S と T が $P \in \mathcal{P}$ 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

$$T = \psi(S)$$

と表現できたとする. すると

$$\sigma(T) \subset \sigma(S)$$

となる. このことから 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 $P \in \mathcal{P}$ に対して, $P(A) = 0$ となる事象 A を除いて, ある命題 M が成立しているとき

$$M \quad \text{a.s. } \mathcal{P}$$

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 T を $P \in \mathcal{P}$ に対する十分統計量とする. 十分統計量 T は $P \in \mathcal{P}$ 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

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

と表現できるときをいう. もちろん ψ は S ごとに定まればよい.

例 6.20. $\mathcal{P} = \{P \text{ は開区間 } (\theta, \theta + 1) \text{ 上の一様分布; } \theta \in \mathbb{R}\}$ とする.
 $n \geq 2 (\in \mathbb{N})$ とし

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

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

$$p(\mathbf{x}|\theta) = \prod_{j=1}^n \mathbb{1}_{(\theta, \theta+1)}(x_j) = \mathbb{1}_{(x_{(n)}-1, x_{(1)})}(\theta), \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

となる. ただし $x_{(1)}, x_{(n)}$ は x_1, x_2, \dots, x_n の最小値と最大値である.
 Fisher-Neyman の因子分解定理から $T = (X_{(1)}, X_{(n)})$ は $P \in \mathcal{P}$ に対する十分統計量⁶である.

$$\theta < x_j < \theta + 1 \quad (\forall j \in \{1, 2, \dots, n\}) \iff \theta < x_{(1)} < x_{(n)} < \theta + 1 \quad (6.5)$$

より, $\mathbf{X} = \mathbf{x}$ を観測したとき

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

である. ここで統計量 S は $P \in \mathcal{P}$ 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_P と h が存在して

$$p(\mathbf{x}|\theta) = g_P(S(\mathbf{x}))h(\mathbf{x})$$

と書ける. $\{\mathbf{x} \in \mathbb{R}^n; h(\mathbf{x}) > 0\}$ 上で (6.5) が成立するので, ある可測関数 ψ が存在して A 上で

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

と表現できる. さらに $\Pr(\mathbf{X} \in A) = 1$ なので T は最小十分統計量となる. \square

定義 6.21. \mathcal{P} を統計的モデルとし, $X \sim P \in \mathcal{P}$ とする. 統計量 $T(X)$ を $P \in \mathcal{P}$ に対する十分統計量とする. T は $P \in \mathcal{P}$ に対して完備であるとは, 任意の可測関数 f に対して

$$E[f(T)] = 0 (\forall P \in \mathcal{P}) \implies f(T) = 0, \text{ a.s. } \mathcal{P}$$

が成り立つときをいう.

⁶厳密には $\mathcal{P} = \{p(\mathbf{x}|\theta) = \prod_{j=1}^n \mathbb{1}_{(\theta, \theta+1)}(x_j); \theta\}$ とすべきであるが, 誤解がないので $\mathcal{P} = \{\mathbb{1}_{(\theta, \theta+1)}(x); \theta \in \mathbb{R}\}$ のまま表記している. また

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

のことである.

命題 6.22. $k, n \in \mathbb{N}$ かつ $n > k$ とする. $\boldsymbol{\eta} \in \mathbb{R}^k, \mathbb{X} \subset \mathbb{R}^n$ とし,

$$p(\boldsymbol{x}|\boldsymbol{\eta}) = \exp\{\boldsymbol{\eta}^\top T(\boldsymbol{x})\}h(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{X}$$

とする. ただし

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

は \mathbb{R}^n の開集合を含み,

$$T: \mathbb{R}^n \longrightarrow \mathbb{R}, \quad h: \mathbb{R}^k \longrightarrow \mathbb{R}, \quad K: \mathcal{E} \longrightarrow \mathbb{R}$$

は可測関数とする. $\boldsymbol{X} \sim P \in \mathcal{P} := \{p(\boldsymbol{x}|\boldsymbol{\eta}); \boldsymbol{\eta} \in \mathcal{E}\}$ としたとき, $T(\boldsymbol{X})$ は $P \in \mathcal{P}$ 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 T は $P \in \mathcal{P}$ に対する十分統計量であることは明らか. つぎに $f: \mathbb{R}^k \longrightarrow \mathbb{R}$ を可測関数とし

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

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

$$\begin{aligned} 0 &= \int_{\mathbb{X}} f(T(\boldsymbol{x})) \exp\{\boldsymbol{\eta}^\top T(\boldsymbol{x}) - K(\boldsymbol{\eta})\}h(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}: T(\boldsymbol{x})=\boldsymbol{t}} h(\boldsymbol{x}) \left\{ \int_{\mathcal{T}} f(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t} - K(\boldsymbol{\eta})\} d\boldsymbol{t} \right\} d\lambda \end{aligned}$$

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

$$\int_{\mathcal{T}} f(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t} - K(\boldsymbol{\eta})\} d\boldsymbol{t} = 0 \quad \text{a.s. } \lambda$$

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

$$\begin{aligned} \int f_+(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t}\} d\boldsymbol{t} &= \int f_-(\boldsymbol{t}) \exp\{\boldsymbol{\eta}^\top \boldsymbol{t}\} d\boldsymbol{t} \\ (\forall \boldsymbol{\eta} \in N(\boldsymbol{\eta}_0) := \{\boldsymbol{\eta} \in \mathcal{R}^k; |\boldsymbol{\eta} - \boldsymbol{\eta}_0|_2 < \epsilon\}) \end{aligned}$$

となる. ここで

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

で $|\cdot|_2$ は \mathbb{R}^k 上の Euclid の距離である. 特に

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

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

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

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

注意 6.23. 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. \square

6.5 章末注釈と参考文献

6.6 演習問題