第5章　データの縮約

5.1　十分統計量

$X_1, X_2, \ldots, X_n$ を確率 (密度) 関数 $f(x|\theta)$ を持つ分布からのランダム標本とする．ただし，θ は母数空間 $\Theta \subset \mathbb{R}^k$ (k は自然数) の元とする．また，$T = T(X_1, X_2, \ldots, X_n)$ を統計量とする．統計量は n 個の確率変数の「情報」をひとつの確率変数に縮約している．縮約をするときに，どのような「情報」が失われているかを知りたい．別のいい方をすれば，推測に必要な「情報」が失われていない保障があるかど うかを知りたい．このようなことをどのように定式化するかをここでは考えていく．

$\mathcal{X}$ を標本空間とし，T を統計量とすれば，T は標本空間を分割していると考えることができる．このことを簡単な例でみてみよう．

例 5.1　X_1, X_2, X_3 はベルヌーイ試行 $Bi(1, p), 0 < p < 1$ からのランダム標本とする．したがって，標本空間は

$$\mathcal{X} = \{(0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), (1,1,1)\}$$

となる．$S(x_1, x_2, x_3) = x_1 + x_2 + x_3$ とすれば，$S(\cdot)$ の値域は $0, 1, 2, 3$ となり，$\mathcal{X}$ を分割している：

$S(\cdot)$ の値	標本点
0	(0, 0, 0)
1	(0, 0, 1), (0, 1, 0), (1, 0, 0)
2	(0, 1, 1), (1, 0, 1), (1, 1, 0)
3	(1, 1, 1)

たとえば，$S = 1$ とすれば,「1」 が 1 回出現したが，何回目に出現したかという「情報」は失われる．

つぎに，$T = T(X_1, X_2, X_3) = X_1X_2 + X_3$ という統計量を考えて，S と T をそれぞれ与えたときの (X_1, X_2, X_3) の条件付き確率分布をしらべてみよう．

標本点	T の値	$f_{X_1, X_2, X_3 \mid T}(\cdot)$	S 値	$f_{X_1, X_2, X_3 \mid S}(\cdot)$
(0, 0, 0)	0	$\frac{1-p}{1+p}$	0	1
(0, 0, 1)	1	$\frac{1-p}{1+2p}$	1	$\frac{1}{3}$
(0, 1, 0)	0	$\frac{p}{1+p}$	1	$\frac{1}{3}$
(1, 0, 0)	0	$\frac{p}{1+p}$	1	$\frac{1}{3}$
(0, 1, 1)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 0, 1)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 1, 0)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 1, 1)	2	1	3	1

表から S を与えたときの (X_1, X_2, X_3) の条件付き分布は p には依存しないが，T を与えたときの (X_1, X_2, X_3) の条件付き分布は p には依存ことがわかる．

5.1.1 十分統計量の定義

定義 5.1 $X_1, X_2, \ldots, X_n$ は確率 (密度) 関数 $f(x|\theta)$ を持つ分布からの標本の大きさ n のランダム標本とする．ただし，$\theta \in \Theta$ とする．統計量 $S = S(X_1, X_2, \ldots, X_n)$ が十分統計量であるとは，$S = s$ を与えたときの $(X_1, X_2, \ldots, X_n)$ の条件付き分布がどんな s の値に対しても θ に依存しないときをいう．

例 5.2 X_1, X_2 は正規分布 $N(\theta, 1)$ からの標本の大きさ 2 のランダム標本とする．$S = S(X_1, X_2) = X_1 + X_2$ は十分統計量であることを示そう．そのために，変換

$$\begin{cases} S = X_1 + X_2 \\ R = X_1 - X_2 \end{cases}$$

を考える．(S, R) と (X_1, X_2) は一対一対応である．また，正規分布の性質から S と R は独立となるので，

$$f_{R|S}(r|s) = \frac{f_{R,S}(r, s)}{f_S(s)} = \frac{f_R(r) f_S(s)}{f_S(s)} = f_R(r)$$

となる．しかし，

$$f_R(r) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2}} e^{-r^2/4}$$

となるので，S を与えたときの R の条件付き分布は θ に依存しない．さらに，(S, R) と (X_1, X_2) は一対一対応なので，S を与えたときの (X_1, X_2) の条件付き分布は θ に依存しないことがわかる．

統計量が複数あるときも十分性を拡張して定義できる．

定義 5.2 $X_1, X_2, \ldots, X_n$ は確率（密度）関数 $f(x|\theta), \theta \in \Theta$ を持つ分布からの標本の大きさ n のランダム標本とする．統計量 $S_1 = S_1(X_1, X_2, \ldots, X_n), S_2 = S_2(X_1, X_2, \ldots, X_n), \ldots, S_\ell = S_\ell(X_1, X_2, \ldots, X_n)$ が十分統計量であるとは，$S_1 = s_1, S_2 = s_2, \ldots, S_\ell = s_\ell$ を与えたときの $(X_1, X_2, \ldots, X_n)$ の条件付き分布がどんな $s_1, s_2, \ldots, s_\ell$ の値に対しても θ に依存しないときをいう．

例 5.3 ：$X_1, X_2, \ldots, X_n$ をベルヌーイ試行

$$f_\theta(z) = \theta^z (1-\theta)^{1-z}\, \mathbb{I}_{\{0,1\}}(z), \qquad z \in \mathbb{R}, \qquad \theta \in (0, 1)$$

からの標本の大きさ n のランダム標本とし，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とする．$T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ なる統計量は，直観から θ に対する情報をすべて含んでいると予想される．したがって，十分統計量となることが予想される．これを示すために，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とし，

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P(T = t)}$$

が $\theta \in (0, 1)$ に依存しないことを示せばよい．

$$P(T = t) = \begin{pmatrix} n \\ t \end{pmatrix} \theta^t (1-\theta)^{n-\theta}\, \mathbb{I}_{\{0,1,\ldots,n\}}(t)$$

と

$$\begin{aligned} P(\boldsymbol{X} = \boldsymbol{x}, T = t) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}\, \mathbb{I}_{\{0,1\}}(x_i) \\ &= \theta^t (1-\theta)^{n-t} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i) \end{aligned}$$

となることに注意する．$B_t = \{(x_1, x_2, \ldots, x_n) : x_i = 0, 1, \sum_{i=1}^n x_i = t\}$ とおけば，

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{1}{\begin{pmatrix} n \\ t \end{pmatrix}} \mathbb{I}_{B_t}(\boldsymbol{x})$$

となり，θ に依存しないので，$T(\boldsymbol{X})$ は $\theta \in (0, 1)$ に対する十分統計量である．

5.1.2 分解定理

定義に従って十分統計量をみつけるにはあらかじめ十分統計量と思われるものが事前にわかっていなければならない．つぎの定理を用いると比較的容易に十分統計量を見つけることができる．

定理 5.1 σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族を $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ とし，$\boldsymbol{X}$ を P_θ からのランダム標本とする．このとき，$T(\boldsymbol{X})$ が $\theta \in \Theta$ に対して十分であるための必要十分条件は $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の非負のボレロ可測関数 h と T の値域上の関数 g_θ （P に依存）が存在し，

$$\frac{dP_\theta}{d\nu}(\boldsymbol{x}) = g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{5.1}$$

と書けることである．

証明 まず，証明は離散型分布の場合は比較的簡単であるので，その場合ににについて証明を与える．σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族に対する証明は後で示す．

はじめに T が十分であると仮定する．このとき，

$$\begin{aligned} P_\theta(\boldsymbol{X} = \boldsymbol{x}) &= \sum_t P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t) \\ &= P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = T(\boldsymbol{x})) \\ &= P_\theta(T = T(\boldsymbol{x}))P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = T(\boldsymbol{x})) \end{aligned}$$

となる[(5-1)]．T が十分統計量であることから $P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = T(\boldsymbol{x}))$ は θ に依存しないので，

$$\begin{aligned} g_\theta(T(\boldsymbol{x})) &= P_\theta(T = T(\boldsymbol{x})) \\ h(\boldsymbol{x}) &= P(\boldsymbol{X} = \boldsymbol{x} | T = T(\boldsymbol{x})) \end{aligned}$$

とおけばよい．

つぎに，(5.1) が成立すると仮定する．$t = T(\boldsymbol{x})$ とおく．このとき，

$$\begin{aligned} P_\theta(T = t) &= \sum_{\boldsymbol{y}:\, T(\boldsymbol{y}) = t} P_\theta(\boldsymbol{X} = \boldsymbol{y}) \\ &= \sum_{\boldsymbol{y}:\, T(\boldsymbol{y}) = t} g_\theta(T(\boldsymbol{y}))h(\boldsymbol{y}) \\ &= g_\theta(t) \sum_{\boldsymbol{y}:\, T(\boldsymbol{y}) = t} h(\boldsymbol{y}) \end{aligned}$$

となる．従って，

$$\begin{aligned} P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} \\ &= \frac{g_\theta(t)h(\boldsymbol{x})}{g_\theta(t)\sum_{\boldsymbol{y}:\, T(\boldsymbol{y}) = t} h(\boldsymbol{y})} \\ &= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y}:\, T(\boldsymbol{y}) = t} h(\boldsymbol{y})} \end{aligned}$$

となり，θ に依存しないことがわかる． □

例 5.4 ：$X_1, X_2, \ldots, X_n$ を確率密度関数

$$f_X(x|\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x)$$

からのランダム標本とする．ただし，$\theta > 0$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \frac{1}{\theta^n}\mathbb{I}_{(0,\theta)^n}(\boldsymbol{x}) \\ &= \frac{1}{\theta^n}\mathbb{I}\{\max_{1\le i\le n} x_i \le \theta\}\mathbb{I}\{\min_{1\le i\le n} x_i \ge 0\} \\ &= g_\theta(\max(x_1, x_2, \ldots, x_n))h(\boldsymbol{x}) \end{aligned}$$

となる．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．従って，$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ は $\theta \in (0, \infty)$ の十分統計量である．

例 5.5 ：$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の分布が k 母数指数分布族に属するとする．すなわち，その同時確率関数または確率密度関数が

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{k} c_i(\theta)T_i(\boldsymbol{x}) - d(\boldsymbol{\theta}) + S(\boldsymbol{x})\right]\mathbb{I}\{\boldsymbol{x} \in A\}$$

で与えられる．このとき，$h(\boldsymbol{x}) = \exp[S(\boldsymbol{x})]\mathbb{I}\{\boldsymbol{x} \in A\}$ とみれば，因数分解定理から

$$T = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X}))$$

は $\boldsymbol{\theta}$ の十分統計量となる．

例 5.6 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\theta, 1)$ からの大きさ n のランダム標本とする．このとき，

$$\begin{aligned} \Pi_{i=1}^n f(x_i|\,\theta) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\imath}}\exp\bigl[-\frac{1}{2}(x_i-\theta)\bigr] = \frac{1}{(2\pi)^{n/2}}\exp\bigl[-\frac{1}{2}\sum_{i=1}^n(x_i-\theta)\bigr] \\ &= \frac{1}{(2\pi)^{n/2}}\exp\bigl[\theta\sum_{i=1}^n x_i - \frac{n}{2}\theta\bigr]\exp\bigl[-\frac{1}{2}\sum_{i=1}^n x_i^2\bigr] \end{aligned}$$

となる．したがって，$S(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n X_i$ として，

$$g(s|\,\theta) = \exp\bigl[\theta s - \frac{n}{2}\theta\bigr], \qquad h(x_1, x_2, \ldots, x_n) = \exp\bigl[-\frac{1}{2}\sum_{i=1}^n x_i^2\bigr]$$

とすれば，分解定理から $S(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n X_i$ は十分統計量となることがわかる．

例 5.7 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの大きさ n のランダム標本とする．ただし，$\theta = (\mu, \sigma^2)$ である．

$$\begin{aligned} \Pi_{i=1}^n f(x_i|\,\theta) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\imath}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\bigl(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\bigr)\right] \end{aligned}$$

となる．したがって，

$$S_1(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} X_i, \qquad S_2(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} X_i^2$$

とし，

$$g(s_1, s_2|\,\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(s_2 - 2\mu s_1 + n\mu^2)\right], \qquad h(x_1, x_2, \ldots, x_n) = 1$$

とおけば，$\sum_{i=1}^{n} X_i$ と $\sum_{i=1}^{n} X_i^2$ は十分統計量になる．さらに，

$$(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2) \longleftrightarrow (\bar{X}_n, \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2)$$

は一対一対応なので，$(\bar{X}_n, \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2)$ も十分統計量であることがわかる．十分統計量が表現のしかたは一通りでないことに注意する．

例 5.8　$X_1, X_2, \ldots, X_n$ を一様分布

$$f(x|\,\theta) = \frac{1}{\theta_2 - \theta_2}\,\mathbb{I}_{(\theta_1, \theta_2)}(x)$$

からの大きさ n のランダム標本とする．ただし，$\theta = (\theta_1, \theta_2), \theta_2 > \theta_2$ である．このとき，

$$\begin{aligned}
\Pi_{i=1}^{n} f(x_i|\,\theta) &= \Pi_{i=1}^{n}\frac{1}{\theta_2 - \theta_2}\,\mathbb{I}_{(\theta_1, \theta_2)}(x_i) \\
&= \frac{1}{(\theta_2 - \theta_2)^2}\Pi_{i=1}^{n}\,\mathbb{I}_{(\theta_1, \theta_2)}(x_i) \\
&= \frac{1}{(\theta_2 - \theta_2)^2}\Pi_{i=1}^{n}\,\mathbb{I}_{(\theta_1, y_n)}(y_1)\,\mathbb{I}_{(y_1, \theta_2)}(y_n)
\end{aligned}$$

となる(5-2)．ただし，$y_1 = \min[x_1, x_2, \ldots, x_n]$, $y_n = \max[x_1, x_2, \ldots, x_n]$ である．したがって，$(\min[X_1, X_2, \ldots, X_n], y_n = \max[X_1, X_2, \ldots, X_n])$ が十分統計量となる．

定理 5.2　最尤推定量は十分統計量を通してのみ標本に依存する．

証明　$S_1 = S_1(X_1, X_2, \ldots, X_n)$, $S_2 = S_2(X_1, X_2, \ldots, X_n), \ldots, S_\ell = S_\ell(X_1, X_2, \ldots, X_n)$ を十分統計量とする．尤度関数は

$$L(\theta|\,x_1, x_2, \ldots, x_n) = \Pi_{i=1}^{n} f(x_i|\,\theta) = g(s_1, s_2, \ldots, s_n|\,\theta)h(x_1, x_2, \ldots, x_n)$$

と書ける．ただし，$s_i = S_i(x_1, x_2, \ldots, x_n)$, $i = 1, 2, \ldots, \ell$ とした．$L(\theta|\,x_1, x_2, \ldots, x_n)$ の θ に関する最大化は $g(s_1, s_2, \ldots, s_n|\,\theta)$ θ に関する最大化と同値である．よって，定理は証明された．　□