・・・・・2.3・・・・・

十分統計量

定義 2.2　：$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ を確率分布族とし，θ が未知の確率分布 P_θ からの標本を $\boldsymbol{X}$ とする．統計量 $T(\boldsymbol{X})$ が $\theta \in \Theta$ に対し**十分**であるとは，T が与えられたときにの $\boldsymbol{X}$ の条件付き分布が θ に依存せず既知となることである．

例 2.8　：$X_1, X_2, \ldots, X_n$ を二項分布

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

から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．

$$P(T = t) = \binom{n}{t}\theta^t(1-\theta)^{n-t}1_{\{0,1,\ldots,n\}}(t)$$

と

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

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，

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{1}{\binom{n}{t}}1_{B_t}(\boldsymbol{x})$$

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．

定義に従って十分統計量をみつけるには，あらかじめそれ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 2.3　：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，

$$\frac{dP_\theta}{d\nu}(\boldsymbol{x}) = g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x}) \tag{2.4}$$

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[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，(2.4) 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，

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

[1)]一番目の等式は

$$\{\boldsymbol{X} = \boldsymbol{x}\} = \{\boldsymbol{X} = \boldsymbol{x}\} \cap (\cup_t \{T =\}) = \cup_t(\{\boldsymbol{X} = \boldsymbol{x}\} \cap \{T = t\})$$

からわかる．また，二番目の $T(\boldsymbol{x}) = t$ を満たさない $\boldsymbol{x}$ との積事象は空事象なので，$T(\boldsymbol{x}) = t$ 以外の和の t に関する項の事象は空事象となるので，和はなくなることがわかる．

となり，θ に依存しないことがわかる．□

例 2.9　：$X_1, X_2, \ldots, X_n$ を確率密度関数

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

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

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

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 2.10　：$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 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

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{k}\eta_i(\theta)T_i(\boldsymbol{x}) - \tilde{A}(\theta) + S(\boldsymbol{x})\right]1\{\boldsymbol{x} \in B\}$$

で与えられる．このとき，$h(\boldsymbol{x}) = \exp[S(\boldsymbol{x})]1\{\boldsymbol{x} \in B\}$ とみれば，因数分解定理から

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

は $\boldsymbol{\theta}$ の十分統計量となる．

与えられた分布族に対して十分統計量は必ず存在[2)]する．また，複数存在する．たとえば，例 2.8 においては，$m(1 \le m \le n)$ を固定された自然数とすれば，$(\sum_{i=1}^{m} X_i, \sum_{i=m+1}^{n} X_i)$ は θ に対する十分統計量になる．もし，ある可測関数とある統計量を用いて十分統計量 T が $T = h(S)$ と書ければ，S も十分統計量になることが因数分解定理より直ちにわかる．このとき，$\sigma(T) \subset \sigma(S)$ なので，T の方が S より有用である．データの情報を最大限に縮約する統計量とはどのようなものであろうか？そのために以下の記号と概念を導入する．

すべての $P \in \mathcal{P}$ に対し，$P(A) = 0$ を満足する事象 A を除いてあ，る命題が成立するとき，その命題は $\mathcal{P}$ **に関してほとんどいたるところ成立する**といい，$a.e.\mathcal{P}$ とかく．

定義 2.3　：T は $P \in \mathcal{P}$ の十分統計量とする．T が**最小十分統計量**であるとは，$P \in \mathcal{P}$ の任意の十分統計量 U に対し，ある可測関数 h が存在し，

$$T = h(U), \quad a.e. \quad \mathcal{P}$$

と書けることをいう．

[2)] すなわち，与えられた統計量自体である．

もし，T と U がともに最小十分統計量ならば，一対一写像 h が存在し，

$$T = h(U), \quad a.e. \quad \mathcal{P}$$

となる．

定理 2.4　：(i) $\mathcal{P}$ をある分布族とし，その部分分布族を $\mathcal{P}_0$ とする．$\mathcal{P}_0$ に関してほとんどいたるところ成立するならば，$\mathcal{P}$ に関してもほとんどいたるところ成立すると仮定する．このとき，統計量 T が $P \in \mathcal{P}$ に関して十分で，$\mathcal{P}_0$ に関して最小十分ならば，T は $\mathcal{P}$ に関しても最小十分である．
(ii) $\mathcal{P}$ を σ-有限な測度に関する $(k+1)$ 個の密度関数 $f_0, f_1, \ldots, f_k$ からなる分布族とする．さらに，$f_i, i = 1, 2, \ldots, k$ の台は

$$\{\boldsymbol{x} : f_i(\boldsymbol{x}) > 0\} \subset \{\boldsymbol{x} : f_0(\boldsymbol{x}) > 0\}$$

を満たし，$f_0(\boldsymbol{x}) > 0$ 上で

$$T_i(\boldsymbol{X}) = \frac{f_i(\boldsymbol{X})}{f_0(\boldsymbol{X})}, \quad i = 1, 2, \ldots, k$$

なる統計量 $(T_1, T_2, \ldots, T_k)$ は $P \in \mathcal{P}$ に関して最小十分である．

証明：(i) S を $P \in \mathcal{P}$ に関して十分統計量とすれば，明らかに $P \in \mathcal{P}_0$ に関しても十分である．T が $P \in \mathcal{P}_0$ に関して最小十分であることから，ある可測関数 h が存在し，$T = h(S), a.s., \mathcal{P}_0$ と表現できる．さらに，仮定から，これは $T = h(S), a.s., \mathcal{P}$ を意味するので，(i) は示された．
(ii) $f_0 > 0, a.s.\mathcal{P}$ であることに注意する．いま，

$$g_0(T) = 1, \quad , g_i(T) = T_i, \quad i = 1, 2, \ldots, k$$

とおく．このとき，T_i の定義から

$$f_i(\boldsymbol{x}) = g_i(T(\boldsymbol{x}))f_0(\boldsymbol{x}). \quad a.s.\mathcal{P}$$

となる．因数分解定理から，T は $P \in \mathcal{P}$ に関して十分であることがわかる．T の最小性を示すために，S を他の十分統計量とする．因数分解定理を再度用いれば，ある可測関数 h と $\tilde{g}_i$ が存在し，

$$f_i(\boldsymbol{x}) = \tilde{g}_i(S(\boldsymbol{x}))h(\boldsymbol{x}), \quad i = 0, 1, \ldots, k$$

と表現できる．また，T_i の定義から

$$T_i(\boldsymbol{x}) = \frac{\tilde{g}_i(S(\boldsymbol{x}))}{\tilde{g}_0(S(\boldsymbol{x}))}$$

と $\{\boldsymbol{x} : f_0(\boldsymbol{x}) > 0\}$ 上で表現できる．したがって，最小性の定義から T は $P \in \mathcal{P}$ に関して最小十分であることがわかる．□

例 2.11　（例 2.10 の続き）：$\Theta_0 = \{\theta_0, \theta_1, \ldots, \theta_k\} \subset \Theta$ とする．$\boldsymbol{c} = (c_1(\theta), \ldots, c_k(\theta))' \subset \mathbb{R}^k$ に対し，

$$\tilde{\boldsymbol{c}}_i(\theta) = \boldsymbol{c}(\theta_i) - \boldsymbol{c}(\theta_0), \quad i = 1, 2, \ldots, k$$

としたとき，$\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立になるように Θ_0 をさだめることができると仮定[3)]する．これは，k 母数指数分布族の階数が k であるならば，この仮定をみたす．例 2.10 から T は $\theta \in \Theta$ に関して十分であるので，最小性をつぎに示す．$\mathcal{P}_0 = \{f_\theta : \theta \in \Theta_0\}$ とおけば，定理 2.4 (ii) より

$$S(\boldsymbol{x}) = (\exp\{T(\boldsymbol{x})\tilde{\boldsymbol{c}}_1'(\theta) - \tilde{d}_1\}, \ldots, \exp\{T(\boldsymbol{x})\tilde{\boldsymbol{c}}_k'(\theta) - \tilde{d}_k\})$$

は $\theta \in \Theta_0$ に関して最小十分である．ただし，$\tilde{d}_i = d(\theta_i) - d(\theta_0), i = 1, 2, \ldots, k$ である．$\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立なので，1 対 1 の可測関数 $h : \mathbb{R}^k \to \mathbb{R}^k$ が存在して，$T(\boldsymbol{x}) = h(S(\boldsymbol{x})), a.s., \mathcal{P}_0$ と書ける．したがって，$\theta \in \Theta_0$ に関して最小十分である．さらに，これは $\mathcal{P}$ においてもほとんどいたるところ成立することが $\tilde{\boldsymbol{c}}_1(\theta), \ldots, \tilde{\boldsymbol{c}}_k(\theta)$ は $\mathbb{R}^k$ において一次独立からわかる．したがって，定理 2.4 (i) から T は $\theta \in \Theta$ に関して最小十分であることが示せた．

例 2.12　$X_1, X_2, \ldots, X_n\,(n \geq 2)$ を正規分布 $N(\mu, \sigma^2)$ からのランダム標本とする．$\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ は未知とする．このとき，

$$\begin{aligned} p(\boldsymbol{x}|\,(\mu, \sigma^2)) &= \frac{1}{(\sqrt{2\pi}\sigma)^2} \exp\bigl[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\bigr] \\ &= \exp\bigl[\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 - \frac{n}{2}\bigl(\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2)\bigr)\bigr] \end{aligned}$$

となる．したがって，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ として，$T_1(\boldsymbol{X}) = \sum_{i=1}^n X_i, T_2(\boldsymbol{X}) = \sum_{i=1}^n X_i^2$ とした 2 母数指数分布族となる．よって，$(T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))$ は (μ, σ^2) の十分統計量になることがわかる．さらに，

$$\mathbb{COV}(T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) = \mathbb{COV}(\sum_{k=1}^n X_k, \sum_{l=1}^n X_l^2) = \sum_{k=1}^n \mathbb{COV}(X_k, X_k^2)$$

となる．したがって，

$$\mathbb{VAR}(T_1, T_2) = n\begin{pmatrix} \mathbb{VAR}(X_1) & \mathbb{COV}(X_1, X_1^2) \\ \mathbb{COV}(X_1, X_1^2) & \mathbb{VAR}(X_1^2) \end{pmatrix} = 4\sigma^2 \begin{pmatrix} 1 & 2\mu \\ 2\mu & 2\sigma^2 + 4\mu^2 \end{pmatrix} \tag{2.5}$$

となり，正定値であることがわかる．したがって，$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ は (μ, σ^2) の最小十分統計量となる．さらに，$\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ としたとき，$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ と $(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i - \bar{X}_n)^2)$ は 1 対 1 対応するので，$(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i - \bar{X}_n)^2)$ も (μ, σ^2) の最小十分統計量となる．

[3)] これは $\{\boldsymbol{c}(\theta) : \theta \in \Theta\}$ が $\mathbb{R}^k$ の開集合ならば可能である．

最後に，(2.5) を確認する．X_1 の積率母関数が $m_X(t) = \exp(\mu t + \sigma^2 t^2/2)$ で与えられることに注意する．これより

$$\begin{aligned}
\mathbb{E}[X_1] &= \big[(\mu + \sigma^2 t) m_X(t)\big]_{t=0} = \mu, \\
\mathbb{E}[X_1^2] &= \big[\sigma^2 m_X(t) + (\mu + \sigma^2 t)^2 m_X(t)\big]_{t=0} = \mu^2 + \sigma^2, \\
\mathbb{E}[X_1^3] &= \big[\sigma^2(\mu + \sigma^2 t) m_X(t) + 2\sigma^2(\mu + \sigma^2 t) m_X(t) + (\mu + \sigma^2 t)^3 m_X(t)\big]_{t=0} = 3\mu\sigma^2 + \mu^3, \\
\mathbb{E}[X_1^4] &= \big[3\sigma^4 m_X(t) + 6\sigma^2(\mu + \sigma^2 t)^2 m_X(t) + (\mu + \sigma^2 t)^4 m_X(t)\big]_{t=0} = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4
\end{aligned}$$

となる．したがって，

$$\begin{aligned}
\mathbb{COV}(X_1, X_1^2) &= \mathbb{E}[(X_1 - \mu)(X_1^2 - \mu^2 - \sigma^2)] = 2\mu\sigma^2, \\
\mathbb{VAR}[X_1^2] &= \mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2 = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 - (\mu^2 + \sigma^2)^2 = 2\sigma^4 + 4\mu^2\sigma^2
\end{aligned}$$

からわかる．

2.3.1 完備統計量

Θ を母数空間とする統計モデル $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ を考えよう．

定義 2.4　：確率変数 X は分布 P_θ にしたがうとする．統計量 $V(X)$ が**補助統計量**（ancillary statistics）であるとは，$V(X)$ の分布が θ に依存しないときをいう．さらに，$V(X)$ が**一次のオーダーで補助統計量**（first-order ancillary statistics）であるとは，$\mathbb{E}[V(X)]$ が θ に依存しないことをいう．

自明な補助統計量は $V(X)$ が定数であるような統計量である．しかし，もし $V(X)$ が自明でない補助統計量であれば，$\sigma(X)$ の部分 σ - 集合族 $\sigma(V(X))$ は P_θ の情報を含まない自明でない σ - 集合族となる．したがって，十分統計量 $T(X)$ のすべての定数ではない関数が補助統計量でないのならば，$T(X)$ は最もうまく情報縮約できたことになる．

定義 2.5　：統計量 $T(X)$ が $\theta \in \Theta$ に対し**完備**（*complete*）であるとはつぎの条件を満たすことである：
任意のボレル関数 f に対し，すべての $P \in \mathcal{P}$ について

$$\mathbb{E}[f(T)] = 0 \quad \text{ならば} \quad f = 0, \quad a.s., \mathcal{P}$$

が成立する．
$T(X)$ が有界な関数に対して完備であるとは，有界ボレロ関数に対して，上の条件が成立することである．

つぎの定理は指数分布族においては比較的容易に完備十分統計量をみつけることができることを示している．

定理 2.5　確率分布 P の密度関数が (2.1) で与えられる指数分布族で，自然母数空間がフルランクであるとする．このとき，$T(X)$ は $c \in \Xi$ に対して完備十分統計量となる．

証明：例 2.10 から十分性はわかる．完備性については Lehmann の TSH の 4.2 節を参照せよ．
□

つぎに完備統計量と補助統計量の独立性についての定理をあげる．

定理 2.6　（Basu の定理）：確率変数 X は $P \in \mathcal{P}$ にしたがい，$V(X)$ を補助統計量とし，$T(X)$ を有界関数にたいして完備な十分統計量とする．このとき，V と T は独立となる

証明：B を V の値域の事象とする．V は補助統計量なので，$P \in \mathcal{P}$ を動かしても $P(V^{-1}(B))$ は定数となる．また，T は十分統計量なので，$\mathbb{E}[1_B|T]$ は T のみの関数である．条件付き確率の性質から

$$0 = \mathbb{E}[1_B(V)] - P(V^{-1}(B)) = \mathbb{E}\{\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B))\}$$

となり，上の式の右辺の期待値の中 $\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B))$ は T の有界ボレロ関数となるので，完備性から

$$\mathbb{E}(1_B(V)|T)\} - P(V^{-1}(B)) = 0$$

となる．さらに，A を T の値域の事象とすれば，

$$\begin{aligned} P\{T^{-1}(A) \cap V^{-1}(B)\} &= \mathbb{E}\{\mathbb{E}(1_A(T)1_B(V)\ T)\} = \mathbb{E}\{1_A(T)\mathbb{E}(1_B(V)|T)\} \\ &= \mathbb{E}\{1_A(T)P(V^{-1}(B)\} = P(T^{-1}(A))\mathbb{E}\{1_A(T)\} \\ &= P(T^{-1}(A))P(V^{-1}(B)) \end{aligned}$$

となり，独立性が示せた．□