数理統計

箱

2025 年 4 月 27 日

概要

数理統計の基礎を解説する．

目次

記号と用語

- 集合 $\mathcal{X}$ を考えるとき，その部分集合 $\mathcal{A}$ の特性関数を，$1_{\mathcal{A}}$ と書く．
- T を集合 $\mathcal{X}$ から可測空間 $\mathcal{Y}$ への写像とするとき，T を可測にする $\mathcal{X}$ 上の可測構造の中で最小のものを，$\sigma[T]$ と書く．

1 統計モデル

1.1 統計モデル

定義 1.1（統計モデル） 可測空間 $\mathcal{X}$ とその上の確率測度の族 $(P_\theta)_{\theta\in\Theta}$ との組 $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を，**統計モデル**（statistical model）という．$\mathcal{X}$ をこの統計モデルの**標本空間**（sample space）といい，Θ をこの統計モデルの**パラメータ空間**（parameter space）という．

$(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．このとき，$\mathcal{X}$ から集合 $\mathcal{Y}$ への写像 $\phi\colon \mathcal{X} \to \mathcal{Y}$ を，しばしば**統計量**（statistic）という．また，各 $\theta \in \Theta$ に対して，確率空間 $(\mathcal{X}, P_\theta)$ 上の確率変数としての期待値，条件付き期

待値，分散，共分散を，それぞれ $E_\theta[\phi]$, $E_\theta[\phi|\mathfrak{F}]$, $\mathrm{Var}_\theta[\phi]$, $\mathrm{Cov}_\theta[\phi, \psi]$ などと書く（ϕ と ψ は有限次元実線型空間に値をとる可測統計量であり，$\mathfrak{F}$ は $\mathcal{X}$ の可測構造の部分 σ-代数である）．

1.2 十分性

定義 1.2（十分性） $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．

(1) $\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ がこの統計モデルに対して**十分**（sufficient）であるとは，任意の可測集合 $\mathcal{A} \subseteq \mathcal{X}$ に対して，$\mathfrak{F}$-可測関数 $1_{\mathcal{A},\mathfrak{F}}\colon \mathcal{X} \to [0,1]$ であって，任意の $\theta \in \Theta$ に対して条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ の代表元であるものがとれることをいう．

(2) $\mathcal{Y}$ を可測空間とし，$T\colon \mathcal{X} \to \mathcal{Y}$ を可測統計量とする．T がこの統計モデルに対して**十分**であるとは，σ-代数 $\sigma[T]$ がこの統計モデルに対して十分であることをいう．

命題 1.3 $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．$\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ に対して，次の条件は同値である．

(a) $\mathfrak{F}$ は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分である．

(b) 任意の有限次元実線型空間 $\mathcal{V}$ と $(P_\theta)_{\theta\in\Theta}$-可積分な可測写像 $\phi\colon \mathcal{X} \to \mathcal{V}$ に対して，$\mathfrak{F}$-可測写像 $1_{\mathcal{A},\mathfrak{F}}\colon \mathcal{X} \to \mathcal{V}$ であって，任意の $\theta \in \Theta$ に対して条件付き期待値 $E_\theta[\phi|\mathfrak{F}]$ の代表元であるものがとれる．

証明 (b) $\Longrightarrow$ (a) 条件 (b) が成り立つとする．このとき，任意の可測集合 $\mathcal{A} \subseteq \mathcal{X}$ に対して，条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元 $1_{\mathcal{A},\mathfrak{F}}\colon \mathcal{X} \to \mathbb{R}$ がとれる．任意の $\theta \in \Theta$ に対して，条件付き期待値の順序保存性より，P_θ-ほとんど確実に $0 \le 1_{\mathcal{A},\mathfrak{F}} \le 1$ である．そこで，$1_{\mathcal{A},\mathfrak{F}}$ の 0 以下の値は 0 に，1 以上の値は 1 に修正して得られる関数を改めて $1_{\mathcal{A},\mathfrak{F}}$ と書くと，これも条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元である．よって，$\mathfrak{F}$ は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分である．

(a) $\Longrightarrow$ (b) $\mathfrak{F}$ が統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分であるとする．任意の有限次元実線型空間 $\mathcal{V}$ と $(P_\theta)_{\theta\in\Theta}$-可積分な可測写像 $\phi\colon \mathcal{X} \to \mathcal{V}$ に対して，条件付き期待値 $E_\theta[\phi|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元がとれることを示したい．$\mathcal{V}$ の基底を一つ固定して成分ごとに考えることにより，一般性を失わず，$\mathcal{V} = \mathbb{R}$ であると仮定する．さらに，正の部分と負の部分への分解を考えることにより，一般性を失わず，$\phi \ge 0$ であると仮定する．

$\mathcal{X}$ 上の可測単関数の増加列 $(\phi_n)_{n\in\mathbb{N}}$ であって，ϕ に各点収束するものをとる．$\mathfrak{F}$ の十分性より，各 $n \in \mathbb{N}$ に対して，条件付き期待値 $E_\theta[\phi_n|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元 $\phi_{n,\mathfrak{F}}\colon \mathcal{X} \to \mathbb{R}_{\ge 0}$ がとれる．条件付き期待値に対する Lebesgue の収束定理より，任意の $\theta \in \Theta$ に対して，$(E_\theta[\phi_n|\mathfrak{F}])_{n\in\mathbb{N}}$ は $E_\theta[\phi|\mathfrak{F}]$ に P_θ-概収束する．そこで，関数 $\phi_\mathfrak{F}\colon \mathcal{X} \to \mathbb{R}_{\ge 0}$ を

$$\phi_\mathfrak{F}(x) = \begin{cases} \limsup_{n\to\infty} \phi_{n,\mathfrak{F}}(x) & (\limsup_{n\to\infty} \phi_{n,\mathfrak{F}}(x) < \infty) \\ 0 & (\limsup_{n\to\infty} \phi_{n,\mathfrak{F}}(x) = \infty) \end{cases}$$

と定めると，これは $E_\theta[\phi|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元である．これで，主張が示された． □

$\mathfrak{F}$ が統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分であるとき，$(P_\theta)_{\theta\in\Theta}$-可積分な可測写像 $\phi\colon \mathcal{X} \to \mathcal{V}$（$\mathcal{V}$ は有限次元実線型空間）に対して，条件付き期待値 $E_\theta[\phi|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元を，単に $E[\phi|\mathfrak{F}]$ と書く．これは，$(P_\theta)_{\theta\in\Theta}$-ほとんど確実に一意に定まる．

補題 1.4　$(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．μ を $\mathcal{X}$ 上の σ-有限測度とし，各 P_θ は μ-絶対連続であるとする．このとき，パラメータの列 $(\theta_n)_{n\in\mathbb{N}_{>0}}$ を適当に選んで $Q = \sum_{n=1}^{\infty} 2^{-n} P_{\theta_n}$ と置けば，各 P_θ は Q-絶対連続となる．

証明　$\mu(\mathcal{X}) = \infty$ ならば，$\mathcal{X}$ の分割 $(\mathcal{X}_i)_{i\in\mathbb{N}_{>0}}$ であって任意の $i \in \mathbb{N}_{>0}$ に対して $0 < \mu(\mathcal{X}_i) < \infty$ を満たすものをとり，可測集合 $\mathcal{A} \subseteq \mathcal{X}$ に対して

$$\mu'(\mathcal{A}) = \sum_{i=1}^{\infty} 2^{-i} \mu(\mathcal{X}_i)^{-1} \mu(\mathcal{A} \cap \mathcal{X}_i)$$

と定めることにより，μ と同値な有限測度 μ' が得られる．そこで，一般性を失わず，μ は有限であると仮定する．

各 $\theta \in \Theta$ に対して，Radon–Nikodym 微分 $dP_\theta/d\mu$ の代表元 f_θ を一つ固定し，$\mathcal{S}_\theta = \{x \in \mathcal{X} \mid f_\theta(x) > 0\}$ と置く．μ は有限だから，パラメータの列 $(\theta_n)_{n\in\mathbb{N}_{>0}}$ を，

$$\mu\left(\bigcup_{n=1}^{\infty} S_{\theta_n}\right) = \sup\left\{\mu\left(\bigcup_{\theta\in\Theta'} S_\theta\right) \,\middle|\, \Theta' \text{ は } \Theta \text{ の可算部分集合}\right\}$$

を満たすようにとれる．$Q = \sum_{n=1}^{\infty} 2^{-n} P_{\theta_n}$ と置き, 各 P_θ が Q-絶対連続であることを示す．可測集合 $\mathcal{A} \subseteq \mathcal{X}$ であって $Q(\mathcal{A}) = 0$ を満たすものを任意にとる．Q の定義より，任意の $n \in \mathbb{N}_{>0}$ に対して，$P_{\theta_n}(\mathcal{A}) = 0$ だから，$\mu(\mathcal{A} \cap \mathcal{S}_{\theta_n}) = 0$ である．また，$(\theta_n)_{n\in\mathbb{N}_{>0}}$ のとり方より，$\mu(\mathcal{S}_\theta \setminus \bigcup_{n=1}^{\infty} \mathcal{S}_{\theta_n}) = 0$ である．したがって，

$$\mu(\mathcal{A} \cap \mathcal{S}_\theta) \leq \sum_{n=1}^{\infty} \mu(\mathcal{A} \cap \mathcal{S}_{\theta_n}) + \mu\left(\mathcal{S}_\theta \setminus \bigcup_{n=1}^{\infty} \mathcal{S}_{\theta_n}\right) = 0$$

だから，$P_\theta(\mathcal{A}) = 0$ である．よって，P_θ は Q-絶対連続である．　□

定理 1.5（因子分解定理）　$(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．μ を $\mathcal{X}$ 上の σ-有限測度とし，各 P_θ は μ-絶対連続であるとする．このとき，$\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ に対して，次の条件は同値である．

(a) $\mathfrak{F}$ は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分である．
(b) 可測関数 $g\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ と $\mathfrak{F}$-可測関数の族 $(h_\theta\colon \mathcal{X} \to \mathbb{R}_{\geq 0})_{\theta\in\Theta}$ が存在して，任意の $\theta \in \Theta$ に対して，μ-ほとんどいたるところで $dP_\theta/d\mu = gh_\theta$ が成り立つ．

証明　補題 1.4 より，パラメータの列 $(\theta_n)_{n\in\mathbb{N}_{>0}}$ を適当に選んで $Q = \sum_{n=1}^{\infty} 2^{-n} P_{\theta_n}$ と置けば，各 P_θ は Q-絶対連続となる．以下，条件 (a) と (b) が，ともに次の条件 (c) と同値であることを示す．

(c) 任意の $\theta \in \Theta$ に対して，Radon–Nikodym 微分 dP_θ/dQ の代表元として，$\mathfrak{F}$-可測であるものがとれる．

(a) $\Longrightarrow$ (c)　$\mathfrak{F}$ が統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分であるとする．可測集合 $\mathcal{A} \subseteq \mathcal{X}$ に対して，条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元 $1_{\mathcal{A},\mathfrak{F}}\colon \mathcal{X} \to [0, 1]$ をとる．$\mathcal{B} \in \mathfrak{F}$ とすると，任意の $\theta \in \Theta$ に対して

$$\int_\mathcal{B} 1_{\mathcal{A},\mathfrak{F}} \, dP_\theta = \int_\mathcal{B} 1_\mathcal{A} \, dP_\theta$$

だから，上式で $\theta = \theta_n$ として両辺に 2^{-n} を掛けたものの $n \in \mathbb{N}_{>0}$ にわたる和をとれば，

$$\int_\mathcal{B} 1_{\mathcal{A},\mathfrak{F}} \, dQ = \int_\mathcal{B} 1_\mathcal{A} \, dQ$$

を得る．したがって，$1_{\mathcal{A},\mathfrak{F}}$ は確率測度 Q に関する条件付き期待値 $E_Q[1_\mathcal{A}|\mathfrak{F}]$ の代表元でもある．

$\theta \in \Theta$ とし，可測空間 $(\mathcal{X}, \mathfrak{F})$ 上の確率測度 $P_\theta|_\mathfrak{F}$ と $Q|_\mathfrak{F}$ を考える．P_θ が Q-絶対連続であることより $P_\theta|_\mathfrak{F}$ は $Q|_\mathfrak{F}$-絶対連続だから，Radon–Nikodym 微分 $dP_\theta|_\mathfrak{F}/dQ|_\mathfrak{F}$ が定まる．$\mathfrak{F}$-可測関数 $dP_\theta|_\mathfrak{F}/dQ|_\mathfrak{F}$（の一つの代表元）が dP_θ/dQ の代表元であることを示す．$\mathcal{A} \subseteq \mathcal{X}$ を可測集合とすると，$1_{\mathcal{A},\mathfrak{F}}$ が条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ や $E_Q[1_\mathcal{A}|\mathfrak{F}]$ の代表元であることより，

$$\begin{aligned}
\int_\mathcal{A} \frac{dP_\theta|_\mathfrak{F}}{dQ|_\mathfrak{F}}\,dQ &= \int_\mathcal{X} 1_\mathcal{A} \frac{dP_\theta|_\mathfrak{F}}{dQ|_\mathfrak{F}}\,dQ \\
&= \int_\mathcal{X} 1_{\mathcal{A},\mathfrak{F}} \frac{dP_\theta|_\mathfrak{F}}{dQ|_\mathfrak{F}}\,dQ \\
&= \int_\mathcal{X} 1_{\mathcal{A},\mathfrak{F}}\,dP_\theta \\
&= \int_\mathcal{X} 1_\mathcal{A}\,dP_\theta \\
&= P_\theta(\mathcal{A})
\end{aligned}$$

が成り立つ．よって，$(dP_\theta|_\mathfrak{F}/dQ|_\mathfrak{F}) \cdot Q = P_\theta$ だから，$dP_\theta|_\mathfrak{F}/dQ|_\mathfrak{F}$ は dP_θ/dQ の代表元である．

(c) $\Longrightarrow$ (a)　条件 (c) が成り立つとして，dP_θ/dQ を $\mathfrak{F}$-可測関数とみなす．$\mathcal{A} \subseteq \mathfrak{X}$ を可測集合とすると，任意の $\theta \in \Theta$ と $\mathcal{B} \in \mathfrak{F}$ に対して

$$\int_\mathcal{B} E_Q[1_\mathcal{A}|\mathfrak{F}]\,dP_\theta = \int_\mathcal{B} E_Q[1_\mathcal{A}|\mathfrak{F}]\frac{dP_\theta}{dQ}\,dQ = \int_\mathcal{B} 1_\mathcal{A}\frac{dP_\theta}{dQ}\,dQ = \int_\mathcal{B} 1_\mathcal{A}\,dP_\theta$$

だから，$E_Q[1_\mathcal{A}|\mathfrak{F}]$ は条件付き期待値 $E_\theta[1_\mathcal{A}|\mathfrak{F}]$ の $\theta \in \Theta$ によらない代表元である．よって，$\mathfrak{F}$ は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分である．

(b) $\Longrightarrow$ (c)　条件 (b) を満たす可測関数 $g\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ と $\mathfrak{F}$-可測関数の族 $(h_\theta\colon \mathcal{X} \to \mathbb{R}_{\geq 0})_{\theta\in\Theta}$ がとれたとする．このとき，$k_1 = \sum_{n=1}^\infty 2^{-n} g h_{\theta_n}$ は $\overline{\mathbb{R}}_{\geq 0}$ に値をとる $\mathfrak{F}$-可測関数であり，μ-ほとんどいたるところで $dQ/d\mu = gk_1$ が成り立つ．gk_1 は μ-ほとんどいたるところで有限だから，k_1 の値 ∞ を 0 に修正して得られる $\mathfrak{F}$-可測関数を $k\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ と置くと，これも μ-ほとんどいたるところで $dQ/d\mu = gk$ を満たす．

各 $\theta \in \Theta$ に対して，可測関数 $\phi_\theta\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ を，

$$\phi_\theta(x) = \begin{cases} h_\theta(x)/k(x) & (k(x) > 0) \\ 0 & (k(x) = 0) \end{cases}$$

と定める．すると，

$$\phi_\theta \cdot Q = \phi_\theta gk \cdot \mu = 1_{\{k>0\}} g h_\theta \cdot \mu = 1_{\{k>0\}} \cdot P_\theta$$

が成り立つ．さらに，μ-ほとんどいたるところで $dQ/d\mu = gk$ であることより $\{k = 0\}$ は Q-無視可能だから，P_θ が Q-絶対連続であることより P_θ-無視可能であり，したがって，$1_{\{k>0\}} \cdot P_\theta = P_\theta$ である．よって，dP_θ/dQ の代表元として，$\mathfrak{F}$-可測関数 ϕ_θ がとれる．

(c) $\Longrightarrow$ (b)　条件 (c) が成り立つとして，dP_θ/dQ を $\mathfrak{F}$-可測関数とみなす．任意の $\theta \in \Theta$ に対して，$P_\theta = (dQ/d\mu)(dP_\theta/dQ) \cdot \mu$ だから，$g = dQ/d\mu$，$h_\theta = dP_\theta/dQ$ と置けばよい．□

1.3　完備性

定義 1.6（完備性）　$(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとする．

(1) $\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ がこの統計モデルに対して**完備**（complete）であるとは，任意の $\mathfrak{F}$-可測統計量 $\phi\colon \mathcal{X} \to \mathbb{R}$ に対して，

$$\text{任意の } \theta \in \Theta \text{ に対して } \phi \text{ が } P_\theta\text{-可積分かつ } E_\theta[\phi] = 0 \implies (P_\theta)_{\theta\in\Theta}\text{-ほとんど確実に } \phi = 0$$

が成り立つことをいう．

(2) $\mathcal{Y}$ を可測空間とし，$T\colon \mathcal{X} \to \mathcal{Y}$ を可測統計量とする．T がこの統計モデルに対して**完備**であるとは，σ-代数 $\sigma[T]$ がこの統計モデルに対して完備であることをいう．

1.4 指数型分布族

定義 1.7（指数型分布族） $\mathcal{X}$ を可測空間とする．$\mathcal{X}$ 上の測度 μ，可測写像 $h\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ と $T\colon \mathcal{X} \to \mathcal{V}$（$\mathcal{V}$ は有限次元実線型空間），写像 $c\colon \Theta \to \mathcal{V}^*$ と $d\colon \Theta \to \mathbb{R}$（$\Theta$ は集合）を用いて

$$P_\theta = f_\theta \cdot \mu, \qquad f_\theta(x) = h(x)\exp(\langle c(\theta), T(x)\rangle - d(\theta))$$

と表せる確率測度の族 $(P_\theta)_{\theta\in\Theta}$ を，$\mathcal{X}$ 上の**指数型分布族**（exponential family）という．

注意 1.8 定義 1.7 の状況を考える．

(1) 任意の $\theta \in \Theta$ に対して，P_θ が $\mathcal{X}$ 上の確率測度であることより，$\mathcal{X}$ 上の関数 $x \mapsto h(x)\exp(\langle c(\theta), T(x)\rangle)$ は μ-可積分であり，

$$d(\theta) = \log\left(\int_{\mathcal{X}} h(x)\exp(\langle c(\theta), T(x)\rangle)\, d\mu(x)\right)$$

が成り立つ．

(2) $h \cdot \mu$ を改めて μ と置くことで，$h = 1$ であると仮定できる．

命題 1.9 $\mathcal{X}$ を可測空間とする．$\mathcal{X}$ 上の σ-有限測度 μ，可測写像 $h\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ と $T\colon \mathcal{X} \to \mathcal{V}$（$\mathcal{V}$ は有限次元実線型空間），写像 $c\colon \Theta \to \mathcal{V}^*$ と $d\colon \Theta \to \mathbb{R}$（$\Theta$ は集合）を用いて

$$P_\theta = f_\theta \cdot \mu, \qquad f_\theta(x) = h(x)\exp(\langle c(\theta), T(x)\rangle - d(\theta))$$

と表せる指数型分布族 $(P_\theta)_{\theta\in\Theta}$ を考える．

(1) T は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して十分である．

(2) $c(\Theta)$ が $\mathcal{V}$ において内点をもつとする．このとき，T は統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ に対して完備である．

証明 (1) 因子分解定理（定理 1.5）から従う．

(2) 一般性を失わず，$h = 1$ であると仮定する（注意 1.8 (2)）．

Doob–Dynkin の補題より，$\mathcal{X}$ から $\mathbb{R}$ への任意の $\sigma[T]$-可測統計量は，可測写像 $T\colon \mathcal{V} \to \mathbb{R}$ を用いて $\phi \circ T$ と表せる．任意の $\theta \in \Theta$ に対して，$\phi \circ T$ は P_θ-可積分で $E_\theta[\phi \circ T] = 0$ を満たすと仮定する．任意の $\theta \in \Theta$ に対して，

$$\begin{aligned} E_\theta[\phi \circ T] &= \int_{\mathcal{X}} \phi(T(x))\exp(\langle c(\theta), T(x)\rangle - d(\theta))\, d\mu(x) \\ &= e^{-d(\theta)} \int_{\mathcal{V}} \phi(v)\exp(\langle c(\theta), v\rangle)\, dT_*\mu(v) \end{aligned}$$

だから，上記の仮定は，任意の $\alpha \in c(\Theta)$ に対して

$$\int_{\mathcal{V}} \phi(v) \exp(\langle \alpha, v \rangle)\, dT_*\mu(v) = 0$$

であることを意味する．さらに，$\beta \in \mathcal{V}^*$ とすると，関数 $v \mapsto \phi(v) \exp(\langle \alpha, v \rangle)$ が $T_*\mu$-可積分であることからこれと絶対値が等しい関数 $v \mapsto \phi(v) \exp(\langle \alpha - i\beta, v \rangle)$ も $T_*\mu$-可積分であり，積分記号下の微分に関する定理を用いて確かめられるように，$c(\Theta)^\circ + i\mathcal{V}^*$ 上の関数 $\alpha - i\beta \mapsto \int_{\mathcal{V}} \phi(v) \exp(\langle \alpha - i\beta, v \rangle)\, dT_*\mu(v)$ は正則である．ところが，$\beta = 0$ のときはこの積分は 0 だから，一致の定理より，任意の $\alpha - i\beta \in c(\Theta)^\circ + i\mathcal{V}^*$ に対して

$$\int_{\mathcal{V}} \phi(v) \exp(\langle \alpha - i\beta, v \rangle)\, dT_*\mu(v) = 0$$

が成り立つ．$\alpha \in c(\Theta)^\circ$ を固定すると，上式の左辺を $\beta \in \mathcal{V}^*$ の関数とみなしたものは，$\mathcal{V}$ 上の有限 Borel 測度 $\phi \exp(\langle \alpha, - \rangle) \cdot T_*\mu(v)$ の Fourier 変換である．したがって，Fourier 変換の単射性より

$$\phi \exp(\langle \alpha, - \rangle) \cdot T_*\mu(v) = 0$$

だから，$T_*\mu$-ほとんどいたるところで $\phi = 0$ である．すなわち，μ-ほとんどいたるところで $\phi \circ T = 0$ である．特に，$(P_\theta)_{\theta \in \Theta}$-ほとんど確実に $\phi \circ T = 0$ である．以上より，T は統計モデル $(\mathcal{X}, (P_\theta)_{\theta \in \Theta})$ に対して完備である．$\square$

2 推定

2.1 不偏推定量

定義 2.1（不偏推定量） $(\mathcal{X}, (P_\theta)_{\theta \in \Theta})$ を統計モデルとし，$\mathcal{V}$ を有限次元実線型空間，$g\colon \Theta \to \mathcal{V}$ を写像とする．可測統計量 $\delta\colon \mathcal{X} \to \mathcal{V}$ が $g(\theta)$ の**不偏推定量**（unbiased estimator）であるとは，任意の $\theta \in \Theta$ に対して，δ が P_θ-可積分かつ $E_\theta[\delta] = g(\theta)$ を満たすことをいう．

定義 2.2（一様最小分散不偏推定量） $(\mathcal{X}, (P_\theta)_{\theta \in \Theta})$ を統計モデルとし，$\mathcal{V}$ を有限次元実内積空間，$g\colon \Theta \to \mathcal{V}$ を写像とする．$g(\theta)$ の**一様最小分散不偏推定量**（uniformly minimum-variance unbiased estimator, UMVUE）とは，$g(\theta)$ の不偏推定量 $\delta_0\colon \mathcal{X} \to \mathcal{V}$ であって，$g(\theta)$ の任意の不偏推定量 $\delta\colon \mathcal{X} \to \mathcal{V}$ と $\theta \in \Theta$ に対して

$$E_\theta[\|\delta_0 - g(\theta)\|^2] \leq E_\theta[\|\delta - g(\theta)\|^2]$$

を満たすものをいう．

定理 2.3（Rao–Blackwell の定理） $(\mathcal{X}, (P_\theta)_{\theta \in \Theta})$ を統計モデルとし，$\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ はこれに対して十分であるとする．$\mathcal{V}$ を有限次元実線型空間，$g\colon \Theta \to \mathcal{V}$ を写像とし，$\delta\colon \mathcal{X} \to \mathcal{V}$ を $g(\theta)$ の不偏推定量とする．

(1) $E[\delta|\mathfrak{F}]$ は $g(\theta)$ の不偏推定量である．
(2) $w\colon \Theta \times \mathcal{V} \to \overline{\mathbb{R}}_{\geq 0}$ は関数であり，任意の $\theta \in \Theta$ に対して $w(\theta, -)$ は凸であるとする．このとき，任意の $\theta \in \Theta$ に対して，
$$E_\theta[w(\theta, E[\delta|\mathfrak{F}])] \leq E_\theta[w(\theta, \delta)]$$

が成り立つ．特に，$\mathcal{V}$ が内積空間ならば，$g(\theta)$ の任意の不偏推定量 $\delta'\colon \mathcal{X} \to \mathcal{V}$ と $\theta \in \Theta$ に対して，

$$E_\theta[\|E[\delta|\mathfrak{F}] - g(\theta)\|^2] \le E_\theta[\|\delta - g(\theta)\|^2]$$

が成り立つ．

証明 (1) 条件付き期待値の性質より，任意の $\theta \in \Theta$ に対して，$E[\delta|\mathfrak{F}]$ は P_θ-可積分であり $E_\theta[E[\delta|\mathfrak{F}]] = E_\theta[\delta] = g(\theta)$ が成り立つ．よって，$E[\delta|\mathfrak{F}]$ は $g(\theta)$ の不偏推定量である．

(2) 前半の主張は，条件付き期待値に対する Jensen の不等式から従う．前半の主張において $w(\theta, v) = \|v - g(\theta)\|^2$ とすれば，後半の主張が従う． □

定理 2.4（Lehmann–Scheffé の定理） $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとし，$\mathcal{X}$ の可測構造の部分 σ-代数 $\mathfrak{F}$ はこれに対して完備かつ十分であるとする．$\mathcal{V}$ を有限次元実線型空間とし，$g\colon \Theta \to \mathcal{V}$ を写像とする．

(1) $g(\theta)$ の不偏推定量が存在するとする．このとき，$\mathfrak{F}$-可測な $g(\theta)$ の不偏推定量が，$(P_\theta)_{\theta\in\Theta}$-ほとんど確実に一意に存在する．

(2) δ_0, $\delta\colon \mathcal{X} \to \mathcal{V}$ を $g(\theta)$ の不偏推定量とし，δ_0 は $\mathfrak{F}$-可測であるとする．$w\colon \Theta \times \mathcal{V} \to \overline{\mathbb{R}}_{\ge 0}$ は関数であり，任意の $\theta \in \Theta$ に対して $w(\theta, -)$ は凸であるとする．このとき，任意の $\theta \in \Theta$ に対して，

$$E_\theta[w(\theta, \delta_0)] \le E_\theta[w(\theta, \delta)]$$

が成り立つ．特に，$\mathcal{V}$ が有限次元実内積空間ならば，δ_0 は $g(\theta)$ の一様最小分散不偏推定量である．

証明 (1) 存在 $\delta\colon \mathcal{X} \to \mathcal{V}$ を $g(\theta)$ の不偏推定量とすると，Rao–Blackwell の定理（定理 2.3 (1)）より，$E[\delta|\mathfrak{F}]$ は $\mathfrak{F}$-可測な $g(\theta)$ の不偏推定量である．

一意性 δ_0, $\delta_0'\colon \mathcal{X} \to \mathcal{V}$ がともに $\mathfrak{F}$-可測な $g(\theta)$ の不偏推定量であるとする．このとき，任意の $\theta \in \Theta$ に対して $E_\theta[\delta_0 - \delta_0'] = g(\theta) - g(\theta) = 0$ だから，$\mathfrak{F}$ の完備性より，$(P_\theta)_{\theta\in\Theta}$-ほとんど確実に $\delta_0 - \delta_0' = 0$ が成り立つ．

(2) Rao–Blackwell の定理（定理 2.3 (1)）より，$E[\delta|\mathfrak{F}]$ は $\mathfrak{F}$-可測な $g(\theta)$ の不偏推定量だから，(1) の一意性より，$(P_\theta)_{\theta\in\Theta}$-ほとんど確実に $\delta_0 = E[\delta|\mathfrak{F}]$ である．よって，主張は，Rao–Blackwell の定理（定理 2.3 (2)）から従う． □

2.2 Fisher 情報量

定義 2.5（Fisher 情報量） $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとし，パラメータ空間 Θ は有限次元実線型空間 $\mathcal{V}$ の開集合であるとする．μ を $\mathcal{X}$ 上の測度とし，各 P_θ は可測関数 $f_\theta\colon \mathcal{X} \to \mathbb{R}_{\ge 0}$ を用いて $f_\theta \cdot \mu$ と表されているとする．$\theta_0 \in \Theta$ とし，次の条件が満たされるとする．

(FI1) μ-ほとんどすべての $x \in \mathcal{X}$ に対して，関数 $\theta \mapsto f_\theta(x)$ は，θ_0 のある近傍において正であり，θ_0 において微分可能である（したがって，微分 $(D_\theta \log f_\theta(x))|_{\theta=\theta_0}$ が定義される）．

(FI2) $\mathcal{X}$ 上 μ-ほとんどいたるところで定義され $\mathcal{V}^*$ に値をとる写像 $x \mapsto (D_\theta \log f_\theta(x))|_{\theta=\theta_0}$ は，P_{θ_0}-2 乗可積分である．

このとき，

$$I(\theta_0) = E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes (D_\theta \log f_\theta)|_{\theta=\theta_0}]$$

と定め，これを統計モデル $(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ の θ_0 における **Fisher 情報量**（Fisher information）という．

定義 2.5 の状況で，Fisher 情報量 $I(\theta_0)$ は，正値対称テンソルである．すなわち，$\mathcal{V}$ 上の双線型形式 $(v, w) \mapsto \langle I(\theta_0), v\otimes w\rangle$ は対称であり，任意の $v \in \mathcal{V}$ に対して $\langle I(\theta_0), v\otimes v\rangle \geq 0$ である．

注意 2.6　定義 2.5 の状況を考える．パラメータ θ_0 の下での $(D_\theta \log f_\theta)|_{\theta=\theta_0}$ の期待値（いま，$(D_\theta \log f_\theta)|_{\theta=\theta_0}$ は P_{θ_0}-2 乗可積分であると仮定しているから，この期待値が定義される）は，

$$\begin{aligned} E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}] &= \int_{\mathcal{X}} (D_\theta \log f_\theta(x))|_{\theta=\theta_0} f_{\theta_0}(x)\, d\mu(x) \\ &= \int_{\mathcal{X}} (D_\theta f_\theta(x))|_{\theta=\theta_0}\, d\mu(x) \end{aligned}$$

と表せる．ここで，微分と積分の順序交換ができると仮定すると，

$$E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}] = \left(D_\theta \int_{\mathcal{X}} f_\theta(x)\, d\mu(x) \right)\Bigg|_{\theta=\theta_0} = (D_\theta 1)|_{\theta=\theta_0} = 0$$

となる．これが成り立つとき，Fisher 情報量は，

$$I(\theta_0) = \mathrm{Var}_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}]$$

とも書ける．

積分記号下の微分に関する定理より，μ-可積分関数 $h\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$，$\mu$-無視可能な集合 $\mathcal{N} \subseteq \mathcal{X}$，$\theta_0$ の近傍 $\Xi \subseteq \Theta$ が存在して次の条件を満たす場合には，前段で述べた微分と積分の順序交換を正当化できる．

(i) 任意の $x \in \mathcal{X} \setminus \mathcal{N}$ に対して，関数 $\theta \mapsto f_\theta(x)$ は Ξ 上で微分可能である．
(ii) 任意の $\theta_1 \in \Xi$ に対して，μ-ほとんどすべての $x \in \mathcal{X} \setminus \mathcal{N}$ に対して $\|(D_\theta f_\theta(x))|_{\theta=\theta_1}\|_{\mathcal{V}} \leq h(x)$ が成り立つ．

ここで，$\mathcal{V}$ 上のノルム $\|\text{–}\|_{\mathcal{V}}$ を一つ固定した（有限次元実線型空間上のノルムはすべて同値だから，上記の条件を成否は，このノルムのとり方には依存しない）．

$\mathcal{V}$ を（可換体上の）有限次元線型空間，$T \in \mathcal{V}^* \otimes \mathcal{V}^*$ を非退化対称テンソルとするとき，T は線型同型写像 $\Phi\colon \mathcal{V} \to \mathcal{V}^*$ を定める．T を $\Phi^{-1}\otimes\Phi^{-1}$ で移して得られる非退化対称テンソル $T^\vee = (\Phi^{-1}\otimes\Phi^{-1})(T) \in \mathcal{V}\otimes\mathcal{V}$ を，T の**逆形式**（inverse form）という．

定理 2.7（Cramér–Rao の不等式）　$(\mathcal{X}, (P_\theta)_{\theta\in\Theta})$ を統計モデルとし，パラメータ空間 Θ は有限次元実線型空間 $\mathcal{V}$ の開集合であるとする．μ を $\mathcal{X}$ 上の測度とし，各 P_θ は可測関数 $f_\theta\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ を用いて $f_\theta \cdot \mu$ と表されているとする．$\mathcal{W}$ を有限次元実線型空間，$g\colon \Theta \to \mathcal{W}$ を写像とし，$\delta\colon \mathcal{X} \to \mathcal{W}$ を可測統計量とする．$\theta_0 \in \Theta$ とし，次の条件が満たされるとする（条件 (FI1) と (FI2) は，定義 2.5 のものと同一である）．

(FI1) μ-ほとんどすべての $x \in \mathcal{X}$ に対して，関数 $\theta \mapsto f_\theta(x)$ は，θ_0 のある近傍において正であり，θ_0 において微分可能である（したがって，微分 $(D_\theta \log f_\theta(x))|_{\theta=\theta_0}$ が定義される）．
(FI2) $\mathcal{X}$ 上 μ-ほとんどいたるところで定義され $\mathcal{V}^*$ に値をとる写像 $x \mapsto (D_\theta \log f_\theta(x))|_{\theta=\theta_0}$ は，P_{θ_0}-2 乗可積分である．
(FI3) Fisher 情報量 $I(\theta_0)$ は非退化である（したがって，逆形式 $I(\theta)^\vee \in \mathcal{V}\otimes\mathcal{V}$ が定義される）．

(FI4) g は θ_0 において微分可能であり（したがって，微分 $Dg(\theta_0)$ が定義される），δ は P_{θ_0}-2 乗可積分であり，

$$E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}] = 0,$$
$$E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes \delta] = Dg(\theta_0)$$

が成り立つ（いま，$(D_\theta \log f_\theta)|_{\theta=\theta_0}$ と δ は 2 乗可積分であると仮定しているから，上式の左辺の期待値が定義される）．

このとき，

$$\mathrm{Var}_{\theta_0}[\delta] \geq (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee)$$

が成り立つ．

証明 $I(\theta_0)$ が定める $\mathcal{V}$ から $\mathcal{V}^*$ への線型同型写像を $\Phi\colon \mathcal{V} \to \mathcal{V}^*$ と書き，$u(x) = \Phi^{-1}((D_\theta \log f_\theta(x))|_{\theta=\theta_0}) \in \mathcal{V}$ と置く．$(D_\theta \log f_\theta(x))|_{\theta=\theta_0}$ は μ-ほとんどすべての $x \in \mathcal{X}$ に対して定義され x の関数として P_{θ_0}-2 乗可積分だから，$u(x)$ も同様である．また，条件 (ii) より，

$$E_{\theta_0}[u] = \Phi^{-1}(E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}]) = 0, \tag{$*$}$$
$$E_{\theta_0}[u \otimes \delta] = (\Phi^{-1} \otimes \mathrm{id}_{\mathcal{W}})(E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes \delta]) = (\Phi^{-1} \otimes \mathrm{id}_{\mathcal{W}})(Dg(\theta_0)) \tag{$**$}$$

かつ

$$\begin{aligned} \mathrm{Var}_{\theta_0}[u] &= (\Phi^{-1} \otimes \Phi^{-1})(\mathrm{Var}_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}]) \\ &= (\Phi^{-1} \otimes \Phi^{-1})(I(\theta_0)) \\ &= I(\theta_0)^\vee \end{aligned} \tag{$***$}$$

である．

以下，

$$\begin{aligned} &\mathrm{Var}_{\theta_0}[\delta - Dg(\theta_0) \circ u] \\ &= \mathrm{Var}_{\theta_0}[\delta] - \mathrm{Cov}_{\theta_0}[\delta, Dg(\theta_0) \circ u] - \mathrm{Cov}_{\theta_0}[Dg(\theta_0) \circ u, \delta] + \mathrm{Var}_{\theta_0}[Dg(\theta_0) \circ u] \end{aligned} \tag{$****$}$$

の左辺の各項を計算する．まず，($***$) より，

$$\begin{aligned} \mathrm{Var}_{\theta_0}[Dg(\theta_0) \circ u] &= (Dg(\theta_0) \otimes Dg(\theta_0))(\mathrm{Var}_{\theta_0}[u]) \\ &= (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee) \end{aligned}$$

である．次に，($*$) と ($**$) より，

$$\begin{aligned} \mathrm{Cov}_{\theta_0}[Dg(\theta_0) \circ u, \delta] &= (Dg(\theta_0) \otimes \mathrm{id}_{\mathcal{W}})(\mathrm{Cov}_{\theta_0}[u, \delta]) \\ &= (Dg(\theta_0) \otimes \mathrm{id}_{\mathcal{W}})(E_{\theta_0}[(u - E_{\theta_0}[u]) \otimes (\delta - E_{\theta_0}[\delta])]) \\ &= (Dg(\theta_0) \otimes \mathrm{id}_{\mathcal{W}})(E_{\theta_0}[u \otimes \delta]) \\ &= (Dg(\theta_0)\Phi^{-1} \otimes \mathrm{id}_{\mathcal{W}})(Dg(\theta_0)) \\ &= (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee) \end{aligned}$$

である（最後の等号は，両辺とも $I(\theta_0)^\vee \in \mathcal{V} \otimes \mathcal{V}$ と二つの $Dg(\theta_0) \in \mathcal{V}^* \otimes \mathcal{W}$ の縮約であることから成り立つ）．これらを $(****)$ に代入すると

$$
\begin{aligned}
&\mathrm{Var}_{\theta_0}[\delta - Dg(\theta_0) \circ u] \\
&= \mathrm{Var}_{\theta_0}[\delta] - (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee) - (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee) + (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee) \\
&= \mathrm{Var}_{\theta_0}[\delta] - (Dg(\theta_0) \otimes Dg(\theta_0))(I(\theta_0)^\vee)
\end{aligned}
$$

となり，$\mathrm{Var}_{\theta_0}[\delta - Dg(\theta_0) \circ u] \geq 0$ であることと合わせて，主張の不等式を得る． □

注意 2.8 定理 2.7 の状況で，条件 (FI1) と (FI2) が成り立ち，さらに，μ-可積分関数 $h\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$，$\mu$-無視可能な集合 $\mathcal{N} \subseteq \mathcal{X}$，$\theta_0$ の近傍 $\mathit{\Xi} \subseteq \mathit{\Theta}$ が存在して次の条件を満たすとする．

(i) 任意の $x \in \mathcal{X} \setminus \mathcal{N}$ に対して，関数 $\theta \mapsto f_\theta(x)$ は $\mathit{\Xi}$ 上で微分可能である．
(ii) 任意の $\theta_1 \in \mathit{\Xi}$ に対して，μ-ほとんどすべての $x \in \mathcal{X} \setminus \mathcal{N}$ に対して $\|(D_\theta f_\theta(x))|_{\theta=\theta_1}\|_{\mathcal{V}}$，$\|(D_\theta(f_\theta(x)\delta(x)))|_{\theta=\theta_1}\|_{\mathrm{Hom}(\mathcal{V},\mathcal{W})} \leq h(x)$ が成り立つ．
(iii) δ は P_{θ_0}-2 乗可積分な $g(\theta)$ の不偏推定量である．

ここで，$\mathcal{V}$ 上のノルム $\|{-}\|_{\mathcal{V}}$ と $\mathrm{Hom}(\mathcal{V},\mathcal{W})$ 上のノルム $\|{-}\|_{\mathrm{Hom}(\mathcal{V},\mathcal{W})}$ を一つずつ固定した（有限次元実線型空間上のノルムはすべて同値だから，上記の条件を成否は，これらのノルムのとり方には依存しない）．このとき，条件 (FI4) が成り立つことを示そう．

δ が P_{θ_0}-2 乗可積分であることは仮定 (iii) に含まれており，仮定 (i) と (ii) より $E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0}] = 0$ が成り立つ（注意 2.6）．次に，$E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes \delta]$ について考える．この期待値は，

$$
\begin{aligned}
E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes \delta] &= \int_{\mathcal{X}} (D_\theta \log f_\theta(x))|_{\theta=\theta_0} f_{\theta_0}(x) \otimes \delta(x)\, d\mu(x) \\
&= \int_{\mathcal{X}} (D_\theta f_\theta(x))|_{\theta=\theta_0} \otimes \delta(x)\, d\mu(x) \\
&= \int_{\mathcal{X}} (D_\theta(f_\theta(x)\delta(x)))|_{\theta=\theta_0}\, d\mu(x) \qquad (*)
\end{aligned}
$$

と表せる．一方で，δ が $g(\theta)$ の不偏推定量であること（仮定 (iii)）より

$$
g(\theta) = E_\theta[\delta] = \int_{\mathcal{X}} f_\theta(x)\delta(x)\, d\mu(x)
$$

だから，仮定 (i), (ii) と積分記号下の微分に関する定理より，g は θ_0 において微分可能であり，

$$
Dg(\theta_0) = \int_{\mathcal{X}} (D_\theta(f_\theta(x)\delta(x)))|_{\theta=\theta_0}\, d\mu(x) \qquad (**)
$$

が成り立つ．$(*)$ と $(**)$ を比較して，$E_{\theta_0}[(D_\theta \log f_\theta)|_{\theta=\theta_0} \otimes \delta] = Dg(\theta_0)$ を得る．これで，主張が示された．

2.3 最尤推定量

定義 2.9（最尤推定量） $(\mathcal{X}, (P_\theta)_{\theta\in\mathit{\Theta}})$ を統計モデルとする．μ を $\mathcal{X}$ 上の測度とし，各 P_θ は可測関数 $f_\theta\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ を用いて $f_\theta \cdot \mu$ と表されているとする．写像 $\delta\colon \mathcal{X} \to \mathit{\Theta}$ が θ の**最尤推定量**（maximum likelihood estimator，MLE）であるとは，任意の $x \in \mathcal{X}$ に対して，$\mathit{\Theta}$ 上の関数 $\theta \mapsto f_\theta(x)$ が $\theta = \delta(x)$ において最大値をとることをいう．

参考文献

[1] 野田一雄，宮岡悦良,『入門・演習 数理統計』，共立出版，1990.
[2] 吉田朋広,『数理統計学』，朝倉書店，2006.