

第12章 大標本理論

この章では、正則母数モデルにおける最尤推定量の一致性と漸近正規性について説明する。節 12.1 では、準備として、ランダム関数の確率収束について説明する。尤度関数ないし対数尤度関数を未知母数の関数とみなして議論をするためのものである。節 12.2 では、前節の結果を利用して、最尤推定量の一致性を証明する。節 12.3 では、最尤推定量の漸近正規性を証明する。節 12.4 では、帰無仮説のもとで尤度比検定統計量の対数の -2 倍が χ^2 分布に分布収束することを主張する Wilks の定理を証明する。

12.1 ランダム関数の確率収束

$K \subset \mathbb{R}^d$ をコンパクト集合として、 $\{X_n\}_{n=1}^\infty$ を i.i.d. 確率変数列とする。さらに、 $h: K \times \mathbb{R} \rightarrow \mathbb{R}$ を各 $x \in \mathbb{R}$ に対して

$$h(\cdot, x): K \ni \mathbf{t} \mapsto h(\mathbf{t}, x) \in \mathbb{R}$$

が連続となるような関数とする。つぎに、 $n = 1, 2, \dots$ に対して

$$W_n(\mathbf{t}) := h(\mathbf{t}, X_n) \quad (\mathbf{t} \in K)$$

と定めると $\{W_n\}_{n=1}^\infty$ は K 上の i.i.d. ランダム連続関数列となる。したがって、 W_n ($n \in \mathbb{N}$) は $C(K)$ 値ランダム要素とみなすことができる。ただし、 $C(K)$ は K 上の連続関数全体が成す集合である。

集合 $C(K)$ に値を取るランダム要素はベクトルと同じように和、差とスカラー倍が定義される。このような性質を持つ空間を 線型空間 または ベクトル空間 という。さらに、線型空間 $C(K)$ に収束の概念も導入できる。 $w \in C(K)$ に対して、 w のノルム $\|w\|_\infty$ を

$$\|w\|_\infty := \sup_{\mathbf{t} \in K} |w(\mathbf{t})|$$

で定める。これを w の sup ノルム と呼ぶ。関数列 $w, w_n \in C(K)$ ($n = 1, 2, \dots$) に対して、 $\{w_n\}_{n=1}^\infty$ は w に収束するとは

$$\|w_n - w\|_\infty \xrightarrow{n \rightarrow \infty} 0$$

で定める. 線型空間 $C(K)$ はノルム $\|\cdot\|_\infty$ に関して完備¹ となることが知られている. 完備な線型ノルム空間を Banach 空間 と呼ぶ. Banach 空間 $C(K)$ の最後のよい性質は可分性である. すなわち, 線型ノルム空間 $C(K)$ は稠密² な可算部分集合を持つ.

補題 12.1. $K \subset \mathbb{R}^d$ をコンパクトな部分集合とし, W を $C(K)$ 値ランダム要素とし

$$\mu(t) = E[W(t)] \quad (t \in K)$$

と定める. このとき, 以下のことが成立する.

- (1) $E[\|W\|_\infty] < \infty$ のとき, μ は K 上の連続関数となる.
 (2) $E[\|W\|_\infty] < \infty$ とする. このとき, 任意の $\epsilon > 0$ に対して

$$\sup_{t \in K} E \left[\sup_{s \in K; |s-t|_{2,d} < \epsilon} |W(s) - W(t)| \right] \xrightarrow{\epsilon \rightarrow 0} 0$$

となる. ただし, $|\cdot|_{2,d}$ は $\mathbb{R}^d$ の Euclid ノルムである.

Proof. (1) の証明: $t_n, t \in K$ ($n = 1, 2, \dots$) とし

$$\lim_{n \rightarrow \infty} t_n = t$$

とする. ランダム関数 W は連続だから

$$W(t_n) \xrightarrow{\text{a.s.}} W(t) \quad (n \rightarrow \infty)$$

となる. さらに, $|W(t_n)| \leq \|W\|_\infty$ かつ $E[\|W\|_\infty] < \infty$ であることに注意して, 優収束定理を用いると

$$\lim_{n \rightarrow \infty} \mu(t_n) = \lim_{n \rightarrow \infty} E[W(t_n)] = E \left[\lim_{n \rightarrow \infty} W(t_n) \right] = E[W(t)] = \mu(t)$$

となる. 点列 $\{t_n\}_{n=1}^\infty$ は任意だったので, μ は t で連続であることがわかる. さらに, $t \in K$ も任意だったので, μ は K 上で連続であることがわかる.

(2) の証明: W の見本路を 1 つ固定し

$$w(t) := W(\omega)(t) \quad (\omega \in \Omega, t \in K)$$

¹ $C(K)$ の任意の Cauchy 列は $C(K)$ のある元に必ず収束することである.

² $C(K)$ の可算部分集合 $\{w_i; i \in \mathbb{N}\}$ で, 任意の $\epsilon > 0$ と $w \in C(K)$ に対して, ある $i_0 \in \mathbb{N}$ が存在して

$$\|w - w_{i_0}\|_\infty < \epsilon$$

となることをいう.

と記す. $w \in C(K)$ であり, K はコンパクトなので, w は K 上で一様連続である. $\epsilon > 0$ に対して

$$\tau_w(\epsilon) := \sup_{\mathbf{u}, \mathbf{v} \in K; |\mathbf{u} - \mathbf{v}|_2, d < \epsilon} |w(\mathbf{u}) - w(\mathbf{v})|$$

と定める. このとき, w は一様連続だから, $\tau_w(\epsilon) \downarrow 0$ ($\epsilon \downarrow 0$) となる. 一方, $M_\epsilon(\mathbf{t}) = \sup_{\mathbf{s} \in K; |\mathbf{t} - \mathbf{s}|_2, d < \epsilon} |W(\mathbf{s}) - W(\mathbf{t})|$ だったので, 固定した $\mathbf{t} \in K$ に対して

$$M_\epsilon(\mathbf{t}) \leq \tau_W(\epsilon)$$

である. したがって

$$\sup_{\mathbf{t} \in K} \mathbb{E}[M_\epsilon(\mathbf{t})] \leq \mathbb{E}[\tau_W(\epsilon)]$$

である. また

$$0 \leq \tau_W(\epsilon) \leq 2\|W\|_\infty$$

である. 仮定から $\mathbb{E}[\|W\|_\infty] < \infty$ だったので

$$0 \leq \mathbb{E}[\tau_W(\epsilon)] \leq 2\mathbb{E}[\|W\|_\infty] < \infty$$

である. さらに, 各見本路について

$$\tau_{W(\omega)}(\epsilon) \downarrow 0 \quad (\epsilon \rightarrow 0) \quad (\omega \in \Omega)$$

である. したがって, 優収束定理から

$$\mathbb{E}[\tau_W(\epsilon)] \xrightarrow{\epsilon \rightarrow 0} 0$$

がわかる. よって

$$\sup_{\mathbf{t} \in K} \mathbb{E} \left[\sup_{\mathbf{s} \in K; |\mathbf{s} - \mathbf{t}|_2, d < \epsilon} |W(\mathbf{s}) - W(\mathbf{t})| \right] \xrightarrow{\epsilon \rightarrow 0} 0$$

が証明された.

□

以下の定理は, i.i.d. ランダム関数の標本平均がその期待値に一様に収束に関する議論をする. 各点の収束は大数の弱法則から直ちにわかる. さらに, 次の定理は, コンパクト集合 K 上でランダム標本の平均がその期待値に一様に確率収束することを主張している.

定理 12.2. $K \subset \mathbb{R}^d$ をコンパクト部分集合とする. $W, W_1, W_2, \dots$ を $C(K)$ 値 i.i.d. ランダム要素とする. さらに

$$\mu(\mathbf{t}) = \mathbb{E}[W(\mathbf{t})] \quad (\mathbf{t} \in K)$$

とおき

$$\mathbb{E}[\|W\|_\infty] < \infty$$

を仮定する. ただし, $\|W\|_\infty = \sup_{\mathbf{t} \in K} |W(\mathbf{t})|$ である. $n \in \mathbb{N}$ に対して

$$\overline{W}_n := \frac{1}{n} \{W_1 + W_2 + \dots + W_n\}$$

とおいたとき

$$\|\overline{W}_n - \mu\|_\infty \xrightarrow{P} 0 \quad (n \rightarrow \infty) \quad (12.1)$$

が成り立つ.

Proof. $\epsilon > 0$ を固定する. $\delta > 0$ と $j = 1, 2, \dots, n$ に対して

$$M_{\delta,j}(\mathbf{t}) := \sup_{\mathbf{s} \in K: |\mathbf{s}-\mathbf{t}|_{2,d} < \delta} |W_j(\mathbf{s}) - W_j(\mathbf{t})|; \quad \lambda_{\delta,j}(\mathbf{t}) := \mathbb{E}[M_{\delta,j}(\mathbf{t})]$$

と定める. 補題 12.1 から δ を十分小さくすると

$$\lambda_{\delta,j}(\mathbf{t}) = \mathbb{E} \left[\sup_{\mathbf{s} \in K: |\mathbf{s}-\mathbf{t}|_{2,d} < \delta} |W_j(\mathbf{s}) - W_j(\mathbf{t})| \right] \leq \epsilon \quad (\mathbf{t} \in K)$$

とできる. ここで

$$B(\mathbf{t}; \delta) = \{\mathbf{s} \in K; |\mathbf{s} - \mathbf{t}|_{2,d} < \delta\}$$

とおく. すると K はコンパクトなので, K の開被覆 $\{B(\mathbf{t}; \delta)\}_{\mathbf{t} \in K}$ から有限被覆を取ることができる. すなわち, ある $m \in \mathbb{N}$ と $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m\} \subset K$ が存在して

$$K \subset \bigcup_{j \in \{1, 2, \dots, m\}} B(\mathbf{t}_j; \delta)$$

となる. $O_j = B(\mathbf{t}_j; \delta)$ ($j = 1, 2, \dots, m$) とおく. このとき

$$\begin{aligned} \|\overline{W}_n - \mu\|_\infty &= \max_{j \in \{1, 2, \dots, m\}} \left\{ \sup_{\mathbf{t} \in O_j} |\overline{W}_n(\mathbf{t}) - \mu(\mathbf{t})| \right\} \\ &\leq \max_{j \in \{1, 2, \dots, m\}} \left\{ \sup_{\mathbf{t} \in O_j} |\overline{W}_n(\mathbf{t}) - \overline{W}_n(\mathbf{t}_j)| + |\overline{W}_n(\mathbf{t}_j) - \mu(\mathbf{t}_j)| \right\} \end{aligned}$$

$$\begin{aligned}
& + |\mu(\mathbf{t}_j) - \mu(\mathbf{t})| \Big\} \\
& \leq \max_{j \in \{1, 2, \dots, m\}} \left\{ \sup_{\mathbf{t} \in O_j} |\overline{W}_n(\mathbf{t}) - \overline{W}_n(\mathbf{t}_j)| \right\} \\
& \quad + \max_{j \in \{1, 2, \dots, m\}} |\overline{W}_n(\mathbf{t}_j) - \mu(\mathbf{t}_j)| + \epsilon \quad (\mu \text{ の連続性})
\end{aligned}$$

となる. さらに

$$\begin{aligned}
\sup_{\mathbf{t} \in O_j} |\overline{W}_n(\mathbf{t}) - \overline{W}_n(\mathbf{t}_j)| &= \frac{1}{n} \sup_{\mathbf{t} \in O_j} \left| \sum_{k=1}^n (W_k(\mathbf{t}) - W_k(\mathbf{t}_j)) \right| \\
&\leq \frac{1}{n} \sum_{k=1}^n \sup_{\mathbf{t} \in O_j} |W_k(\mathbf{t}) - W_k(\mathbf{t}_j)| \\
&= \frac{1}{n} \sum_{k=1}^n M_{\delta, k}(\mathbf{t}_j) =: \overline{M}_{\delta, j}
\end{aligned}$$

となる. $j = 1, 2, \dots, m$ に対して, 大数の弱法則から

$$\overline{M}_{\delta, j} \xrightarrow{P} \lambda_{\delta}(\mathbf{t}_j) < \epsilon \quad (n \rightarrow \infty)$$

が成立する. よって

$$\Pr\left(\overline{M}_{\delta, j} < \epsilon\right) \xrightarrow{n \rightarrow \infty} 1 \quad (12.2)$$

がわかる. さらに, 大数の弱法則から

$$\Pr\left(\max_{i \in \{1, 2, \dots, m\}} |\overline{W}_n(\mathbf{t}_i) - \mu(\mathbf{t}_i)| < \epsilon\right) \xrightarrow{n \rightarrow \infty} 1 \quad (12.3)$$

したがって, $\bigcap_{i=1}^n \{\overline{M}_{\delta, i} < \epsilon\}$ と $\sup_{i \in \{1, 2, \dots, m\}} |\overline{W}_n(\mathbf{t}_i) - \mu(\mathbf{t}_i)| < \epsilon$ が起こったとき

$$\begin{aligned}
\|\overline{W}_n - \mu\|_{\infty} &\leq \max_{i \in \{1, 2, \dots, m\}} \left\{ \underbrace{\sup_{\mathbf{t} \in O_i} |\overline{W}_n(\mathbf{t}) - \overline{W}_n(\mathbf{t}_i)|}_{\leq \overline{M}_{\delta, i}(\mathbf{t}_i)} \right\} \\
&\quad + \max_{j \in \{1, 2, \dots, m\}} |\overline{W}_n(\mathbf{t}_j) - \mu(\mathbf{t}_j)| + \epsilon \\
&\leq 3\epsilon
\end{aligned}$$

が成り立つ. したがって, (12.2) と (12.3) と合わせると

$$\Pr\left(\|\overline{W}_n - \mu\|_{\infty} \leq 3\epsilon\right)$$

$$\geq \Pr\left(\{\overline{M}_{\delta,i} < \epsilon\} \cap \left\{\sup_{j \in \{1, 2, \dots, m\}} |\overline{W}_n(\mathbf{t}_j) - \mu(\mathbf{t}_j)| < \epsilon\right\}\right) \\ \xrightarrow{n \rightarrow \infty} 1$$

となる. よって

$$\Pr\left(\|\overline{W}_n - \mu\|_{\infty} > 3\epsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

がわかる. □

注意 12.3. 大数の強法則から

$$\|\overline{W}_n - \mu\|_{\infty} \xrightarrow{\text{a.s.}} 0 \quad (n \rightarrow \infty)$$

も同様に証明できる. □

定理 12.4. $K \subset \mathbb{R}^d$ をコンパクト部分集合とし, $\{G_n\}_{n=1}^{\infty}$ を $C(K)$ 値ランダム要素とする. ある非ランダムな関数 $g (\in C(K))$ が存在して

$$\|G_n - g\|_{\infty} \xrightarrow{\text{P}} 0 \quad (n \rightarrow \infty) \quad (12.4)$$

が成立すると仮定する. このとき, 以下が成立する.

(1) 確率ベクトル列 $\{\mathbf{t}_n\}_{n=1}^{\infty} \subset K$ はある非ランダムベクトル $\mathbf{t}^* (\in K)$ に確率収束するとする. このとき

$$G_n(\mathbf{t}_n) \xrightarrow{\text{P}} g(\mathbf{t}^*) \quad (n \rightarrow \infty)$$

が成立する.

(2) 非ランダムな関数 g は $\mathbf{t}^* (\in K)$ で唯一の最大値を取るとする. さらに, 各 n に対して, 確率ベクトル $\mathbf{t}_n (n = 1, 2, \dots)$ で G_n は最大をとるとする. すなわち

$$G_n(\mathbf{t}_n) = \sup_{\mathbf{t} \in K} G_n(\mathbf{t})$$

である. このとき

$$\mathbf{t}_n \xrightarrow{\text{P}} \mathbf{t}^* \quad (n \rightarrow \infty)$$

が成立する.

(3) $K \subset \mathbb{R}$ とし, $t^* \in K$ は方程式 $g(t) = 0$ の唯一の解とする. 各 n に対して, 確率変数 t_n は方程式 $G_n(t_n) = 0$ の解とする. このとき

$$t_n \xrightarrow{\text{P}} t^* \quad (n \rightarrow \infty)$$

が成立する.

Proof. (1) の証明: まず

$$\begin{aligned} |G_n(\mathbf{t}_n) - g(\mathbf{t}^*)| &\leq |G_n(\mathbf{t}_n) - g(\mathbf{t}_n)| + |g(\mathbf{t}_n) - g(\mathbf{t}^*)| \\ &\leq \|G_n - g\|_\infty + |g(\mathbf{t}_n) - g(\mathbf{t}^*)| \end{aligned} \quad (12.5)$$

となることに注意する. $\mathbf{t}_n \xrightarrow{P} \mathbf{t}^* (n \rightarrow \infty)$ と g の連続性から

$$g(\mathbf{t}_n) \xrightarrow{P} g(\mathbf{t}^*) \quad (n \rightarrow \infty)$$

となる. したがって, (12.5) から $\forall \epsilon > 0$ に対して

$$\begin{aligned} \Pr(|G_n(\mathbf{t}_n) - g(\mathbf{t}^*)| \geq \epsilon) &\leq \Pr\left(\|G_n - g\|_\infty + |g(\mathbf{t}_n) - g(\mathbf{t}^*)| \geq \epsilon\right) \\ &\leq \Pr\left(\|G_n - g\|_\infty \geq \frac{\epsilon}{2}\right) + \Pr\left(|g(\mathbf{t}_n) - g(\mathbf{t}^*)| \geq \frac{\epsilon}{2}\right) \\ &\xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

がわかる. よって, (1) が示された.

(2) の証明: $\epsilon > 0$ を固定して

$$K_\epsilon := K \setminus B(\mathbf{t}^*; \epsilon) = K \cap \{B(\mathbf{t}^*; \epsilon)\}^c \quad (12.6)$$

とおく. ただし, $B(\mathbf{t}^*; \epsilon) = \{\mathbf{t} \in \mathbb{R}^d; |\mathbf{t} - \mathbf{t}^*|_{2,d} < \epsilon\}$ である. K は有界なので, K_ϵ も有界である. また, K_ϵ は 2 つの閉集合の共通部分なので, 閉集合である. これらのことから K_ϵ はコンパクトであることがわかる. ここで

$$M := g(\mathbf{t}^*) = \sup_{\mathbf{t} \in K} g(\mathbf{t}); \quad M_\epsilon := \sup_{\mathbf{t} \in K_\epsilon} g(\mathbf{t})$$

とおく. K_ϵ のコンパクト性と g の連続性から, ある $\mathbf{t}_\epsilon^* \in K_\epsilon$ が存在して

$$M_\epsilon = g(\mathbf{t}_\epsilon^*)$$

とできる. g は K 上で唯一の点 $\mathbf{t}^*$ で最大値を取るので, $\mathbf{t}_\epsilon^* \neq \mathbf{t}^*$ である. よって

$$M_\epsilon < M$$

となる. ここで

$$\delta := M - M_\epsilon > 0$$

とおく. $\|G_n - g\|_\infty < \frac{\delta}{2}$ 上では

$$\sup_{\mathbf{t} \in K_\epsilon} G_n(\mathbf{t}) \leq \sup_{\mathbf{t} \in K_\epsilon} g(\mathbf{t}) + \sup_{\mathbf{t} \in K_\epsilon} |G_n(\mathbf{t}) - g(\mathbf{t})|$$

$$\begin{aligned}
&\leq \sup_{t \in K_\epsilon} g(t) + \|G_n - g\|_\infty \\
&\leq M_\epsilon + \frac{\delta}{2} = M - \frac{\delta}{2}
\end{aligned} \tag{12.7}$$

となる. 一方, $\|G_n - g\|_\infty < \frac{\delta}{2}$ 上では

$$\begin{aligned}
\sup_{t \in K} G_n(t) &\geq G_n(t^*) = g(t^*) + G_n(t^*) - g(t^*) \\
&\geq g(t^*) - \sup_{t \in K} |G_n(t) - g(t)| \\
&\geq g(t^*) - \|G_n - g\|_\infty
\end{aligned} \tag{12.8}$$

$$> g(t^*) - \frac{\delta}{2} = M - \frac{\delta}{2} \tag{12.9}$$

となる. したがって, (12.7) と (12.9) から

$$\|G_n - g\|_\infty < \frac{\delta}{2} \Rightarrow \sup_{t \in K} G_n(t) > \sup_{t \in K_\epsilon} G_n(t)$$

となる. このことと K_ϵ の定義 (12.6) から

$$\|G_n - g\|_\infty < \frac{\delta}{2} \Rightarrow t_n \in B(t^*; \epsilon)$$

がわかる. よって

$$\Pr\left(\|G_n - g\|_\infty < \frac{\delta}{2}\right) \leq \Pr(|t_n - t^*|_{2,d} < \epsilon)$$

がわかる. 上の不等式の両辺の事象の補事象を取り, $\|G_n - g\|_\infty \xrightarrow{P} 0$ に注意すると

$$\Pr(|t_n - t^*|_{2,d} \geq \epsilon) \leq \Pr\left(\|G_n - g\|_\infty \geq \frac{\delta}{2}\right) \xrightarrow{n \rightarrow \infty} 0$$

がわかる. よって, $t_n \xrightarrow{P} t^*$ が示せた.

(3) の証明: ③a 準備の補題: 任意の $\epsilon > 0$ に対して

$$\inf_{t \in K; |t - t^*| \geq \epsilon} |g(t)| > 0 \tag{12.10}$$

となることを確認する. $K_\epsilon := \{t \in K; |t - t^*| \geq \epsilon\}$ とおく. K はコンパクトなので, K_ϵ もコンパクトである. 関数 $|g|$ は K_ϵ 上でも連続なので, K_ϵ 上で最小値をもつ. しかし, $g(t) = 0$ をみたす唯一の零点 $t^* \notin K_\epsilon$ なので, (12.10) が示せた.

③ (3) の証明: 固定した $\epsilon > 0$ に対して, $c_\epsilon := \inf_{t \in K; |t-t^*| \geq \epsilon} |g(t)| > 0$ とおく. $G_n(t_n) = 0$ なので

$$|g(t_n)| = |g(t_n) - G_n(t_n)| \leq \sup_{t \in K} |g(t) - G_n(t)| = \|G_n - g\|_\infty$$

となる. よって

$$|t - t^*| \geq \epsilon \Rightarrow \|G_n - g\|_\infty \geq c_\epsilon$$

がわかる. すなわち, $\{t_n - t^*\} \subset \{\|G_n - g\|_\infty \geq c_\epsilon\}$ となる. 補題 1.7(4) と定理の仮定 (12.4) から

$$\Pr(|t_n - t^*| \geq \epsilon) \leq \Pr(\|G_n - g\|_\infty \geq c_\epsilon) \xrightarrow{n \rightarrow \infty} 0$$

がわかる. よって, (3) の主張は証明された. $\square$

12.2 最尤推定量の一致性

$\Theta \subset \mathbb{R}^d$ をコンパクト部分集合とし, 正則母数モデル $\mathcal{P} = \{\mathbf{p}_\theta; \theta \in \Theta\}$ を考える. ただし, $\mathbf{p}_\theta$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. $\theta^* \in \Theta$ を固定して

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{p}_{\theta^*}$$

とする. $\mathbf{X} = (X_1, X_2, \dots, X_n) = \mathbf{x} = (x_1, x_2, \dots, x_n)$ を観測したときの対数尤度 $\ell_n(\theta)$ を

$$\ell_n(\theta) = \log \left(\prod_{i=1}^n \mathbf{p}_\theta(x_i) \right) = \sum_{i=1}^n \log \mathbf{p}_\theta(x_i)$$

で定める. このとき, 関数 $\Theta \ni \theta \mapsto \ell_n(\theta)$ を最大にする θ の値 $\hat{\theta}_n(\mathbf{x})$ を 最尤推定値 という. 変数ベクトルに $\mathbf{X}$ を形式的に代入したものを $\hat{\theta}_n := \hat{\theta}_n(\mathbf{X})$ を 最尤推定量 という. 統計的モデルの正則性から, ほとんど至るところの $x \in \mathbb{R}$ に対して, 写像

$$\Theta \ni \theta \mapsto \mathbf{p}_\theta(x) \in [0, \infty)$$

は連続である.

定義 12.5. $\mathcal{P} = \{\mathbf{p}_\theta; \theta \in \Theta\}$ を統計的モデルとする. $\theta^* \in \Theta$ を真の母数とする. $\theta \in \Theta$ に対する Kullback-Leibler 情報量 $\text{KL}(\theta^*, \theta)$ を

$$\text{KL}(\theta^*, \theta) := \mathbb{E}_{\theta^*} \left[\log \left\{ \frac{\mathbf{p}_{\theta^*}(X)}{\mathbf{p}_\theta(X)} \right\} \right]$$

で定める.

補題 12.6. $\forall \theta \in \Theta$ に対して, $P_{\theta^*} \neq P_{\theta}$ ならば, $KL(\theta^*, \theta) > 0$ である.

Proof. Jensen の不等式 (定理 3.12) から

$$\begin{aligned}
 -KL(\theta^*, \theta) &= E_{\theta^*} \left[\log \left\{ \frac{p_{\theta}(X)}{p_{\theta^*}(X)} \right\} \right] \\
 &\leq \log \left\{ E_{\theta^*} \left[\frac{p_{\theta}(X)}{p_{\theta^*}(X)} \right] \right\} \\
 &= \log \left\{ \int_{\{x \in \mathbb{R}; p_{\theta^*}(x) > 0\}} \frac{p_{\theta}(x)}{p_{\theta^*}(x)} p_{\theta^*}(x) dx \right\} \\
 &= \log \left\{ \int_{\{x \in \mathbb{R}; p_{\theta^*}(x) > 0\}} p_{\theta}(x) dx \right\} \\
 &\leq \log \left\{ \int_{-\infty}^{\infty} p_{\theta}(x) dx \right\} \\
 &= \log(1) = 0
 \end{aligned}$$

となる. 上の不等式の等号成立条件は $\frac{p_{\theta}(x)}{p_{\theta^*}(x)} p_{\theta^*}(x)$ がほとんど至る所の $x \in \mathbb{R}$ において定数のときである. この場合, $P_{\theta} = P_{\theta^*}$ となるので

$$KL(\theta^*, \theta) > 0$$

がわかる. □

記号: $\theta^* \in \Theta$ を固定する. $X \sim p_{\theta^*}$ に対して

$$W(\theta) := \log \left[\frac{p_{\theta}(X)}{p_{\theta^*}(X)} \right] \quad (\theta \in \Theta)$$

と定める.

定理 12.7. 以下を仮定する.

- (a) $\Theta(\subset \mathbb{R}^d)$ はコンパクト部分集合.
- (b) $E_{\theta^*}[\|W\|_{\infty}] < \infty$ である. ただし, $\|W\|_{\infty} = \sup_{\theta \in \Theta} |W(\theta)|$ である.
- (c) ほとんど至ところの $x \in \mathbb{R}$ で写像

$$\Theta \ni \theta \mapsto p_{\theta}(x) \in [0, \infty)$$

は連続である.

- (d) $\forall \theta \in \Theta$ に対して

$$\theta \neq \theta^* \Rightarrow P_{\theta} \neq P_{\theta^*}$$

である.

このとき, θ^* の最尤推定量 $\hat{\theta}_n$ に対して

$$\hat{\theta}_n \xrightarrow{P} \theta^* \quad (n \rightarrow \infty)$$

が成り立つ.

Proof. $n \in \mathbb{N}$ とし, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*}$ とする. $j = 1, 2, \dots, n$ に対して

$$W_j(\theta) := \log \left[\frac{p_{\theta}(X_j)}{p_{\theta^*}(X_j)} \right] \quad (\theta \in \Theta)$$

とおく. このとき, $\{W_j(\theta)\}_{j=1}^n$ は $C(\Theta)$ 値 i.i.d. 確率要素である. さらに

$$\mu(\theta) := E[W_j(\theta)] = -\text{KL}(\theta^*, \theta) \quad (\theta \in \Theta)$$

と書く. 仮定 (d) と補題 12.6 から

$$\theta \neq \theta^* \Rightarrow \mu(\theta) < 0$$

であり, $\mu(\theta^*) = 0$ となる. よって, μ は $\theta = \theta^*$ で唯一の最大値を取る. また

$$\bar{W}_n(\theta) := \frac{1}{n} \sum_{j=1}^n W_j(\theta) = \frac{\ell_n(\theta) - \ell_n(\theta^*)}{n}$$

なので, $\bar{W}_n$ は最尤推定量 $\hat{\theta}_n$ で最大となる. よって, 定理 12.2 から

$$\|W_n - \mu\|_{\infty} \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

となる. よって, 定理 12.4(2) から

$$\hat{\theta}_n \xrightarrow{P} \theta^* \quad (n \rightarrow \infty)$$

がわかる. □

定理 12.7 では母数空間をコンパクトとした. いくつかの仮定を加えることにより, コンパクトでない母数空間をもつ統計的モデルにおいても最尤推定量の一致性は保証されることがわかる.

定理 12.8. 統計的モデル $\mathcal{P} = \{p_{\theta}; \theta \in \Theta\}$ を考える. ただし, $\Theta = \mathbb{R}^d$ もしくは有界でなくともよい. また, p_{θ} は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. さらに, p_{θ} に対応する $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ と $(\Omega, \mathcal{A})$ それぞれの上の測度を P_{θ} と Pr_{θ} と書く. $\theta^* \in \Theta$ とし

$$X, X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*}(\cdot),$$

$$W_j(\boldsymbol{\theta}) = \log \left[\frac{p_{\boldsymbol{\theta}}(X_j)}{p_{\boldsymbol{\theta}^*}(X_j)} \right] \quad (j = 1, 2, \dots, n),$$

$$W(\boldsymbol{\theta}) = \log \left[\frac{p_{\boldsymbol{\theta}}(X)}{p_{\boldsymbol{\theta}^*}(X)} \right]$$

とする. 以下を仮定する.

(a) ほとんど至ところの $x \in \mathbb{R}$ で写像

$$\Theta \ni \boldsymbol{\theta} \mapsto p_{\boldsymbol{\theta}}(x) \in [0, \infty)$$

は連続である.

(b) $\forall \boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow P_{\boldsymbol{\theta}} \neq P_{\boldsymbol{\theta}^*}$$

である.

(c) ほとんど至ところの $x \in \mathbb{R}$ に対して, $|\boldsymbol{\theta}|_{2,d} \rightarrow \infty$ のとき, $p_{\boldsymbol{\theta}}(x) \rightarrow 0$ となる. ただし, $|\cdot|_{2,d}$ は $\mathbb{R}^d$ の Euclid ノルムである.

(d) Θ の任意のコンパクト部分集合 K に対して

$$E_{\boldsymbol{\theta}^*} [\|\mathbb{1}_K W\|_{\infty}] < \infty$$

である. ただし, $g \in C(\Theta)$ に対して, $\|g\|_{\infty} = \sup_{\boldsymbol{\theta} \in \Theta} |g(\boldsymbol{\theta})|$ である.

(e) ある $M > 0$ が存在して

$$E_{\boldsymbol{\theta}^*} \left[\left\| \sup_{|\boldsymbol{\theta}|_{2,d} > M} W(\boldsymbol{\theta}) \right\| \right] < \infty$$

である.

このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\hat{\boldsymbol{\theta}}_n$ に対して

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^* \quad (n \rightarrow \infty)$$

が成り立つ.

Proof. $|\boldsymbol{\theta}|_{2,d} \rightarrow \infty$ のとき, $p_{\boldsymbol{\theta}}(x) \rightarrow 0$ なので, 固定した x と任意の $\epsilon > 0$ に対して, ある $M_{\epsilon} > 0$ が存在して

$$|\boldsymbol{\theta}|_{2,d} > M_{\epsilon} \Rightarrow p_{\boldsymbol{\theta}}(x) < \epsilon$$

となる. したがって

$$\sup_{|\boldsymbol{\theta}|_{2,d} > M} p_{\boldsymbol{\theta}}(x) \xrightarrow{M \rightarrow \infty} 0 \quad (12.11)$$

となる. $\mathbf{p}_{\boldsymbol{\theta}^*}(x) > 0$ なる $x \in \mathbb{R}$ に対して

$$\sup_{|\boldsymbol{\theta}|_2, d > M} \log \left(\frac{\mathbf{p}_{\boldsymbol{\theta}}(x)}{\mathbf{p}_{\boldsymbol{\theta}^*}(x)} \right) = \log \left(\frac{\sup_{|\boldsymbol{\theta}|_2, d > M} \mathbf{p}_{\boldsymbol{\theta}}(x)}{\mathbf{p}_{\boldsymbol{\theta}^*}(x)} \right) \\ \xrightarrow{M \rightarrow \infty} -\infty \quad (\because (12.11))$$

となる. よって

$$\Pr \left(\lim_{M \rightarrow \infty} \sup_{|\boldsymbol{\theta}|_2, d > M} W(\boldsymbol{\theta}) = -\infty \right) = 1 \quad (12.12)$$

となる. いま, 条件 (e) をみたす $M_1 > 0$ を固定し

$$Z_M := \sup_{|\boldsymbol{\theta}|_2, d > M} W(\boldsymbol{\theta}) \quad (M \geq M_1)$$

とする. $M \rightarrow \infty$ のとき, $Z_M \rightarrow -\infty$ a.s. なので, ほとんど確実に $Z_{M_1} - Z_M \uparrow +\infty$ ($M \rightarrow \infty$) となる. よって, (12.12) に注意して単調収束定理を用いると

$$\lim_{M \rightarrow \infty} \mathbf{E}_{\boldsymbol{\theta}^*} [Z_{M_1} - Z_M] = +\infty$$

となる. さらに, $\mathbf{E}_{\boldsymbol{\theta}^*} [Z_{M_1}] < \infty$ なので

$$\mathbf{E}_{\boldsymbol{\theta}^*} \left[\sup_{|\boldsymbol{\theta}|_2, d > M} W(\boldsymbol{\theta}) \right] = \mathbf{E}_{\boldsymbol{\theta}^*} [Z_M] = \mathbf{E}_{\boldsymbol{\theta}^*} [Z_{M_1}] - \mathbf{E}_{\boldsymbol{\theta}^*} [Z_{M_1} - Z_M] \xrightarrow{M \rightarrow \infty} -\infty$$

がわかる. ここで, 十分大きな $M_0 > 0$ に対して

$$\mathbf{E}_{\boldsymbol{\theta}^*} \left[\sup_{|\boldsymbol{\theta}|_2, d > M_0} W(\boldsymbol{\theta}) \right] < 0 \quad (12.13)$$

とできる. $W(\boldsymbol{\theta}^*) = 0$ なので

$$|\boldsymbol{\theta}^*|_2, d < M_0$$

である. ここで,

$$Y_j := \sup_{|\boldsymbol{\theta}|_2, d > M_0} W_j(\boldsymbol{\theta}) \quad (j = 1, 2, \dots)$$

とおく. 仮定 (e) を満たす $M_1 \leq M_0$ に対して

$$Y_j \leq \sup_{|\boldsymbol{\theta}|_2, d > M_1} W_j(\boldsymbol{\theta})$$

であるから

$$\mathbf{E}_{\theta^*}[Y_j^+] < \infty, \quad Y_j^+ := \max\{Y_j, 0\}$$

である. $L > 0$ に対して

$$Y_j^{(L)} := \max\{Y_j, -L\}$$

とおく. このとき

$$|Y_j^{(L)}| \leq Y_j^+ + L$$

であるから

$$\mathbf{E}_{\theta^*}[|Y_j^{(L)}|] < \infty$$

となる. また

$$Y_j^{(L)} \downarrow Y_j \quad (L \rightarrow \infty)$$

である. さらに,

$$\begin{aligned} Y_1^{(L)} &= \max\{Y_1, -L\} = \max\{Y_1^+ - Y_1^{-1}, -L\} \\ &= \max\{Y_1^+, -L\} + \max\{-Y_1^-, -L\} = Y_1^+ - \min\{Y_1^{-1}, L\} \end{aligned}$$

である. したがって, $L \rightarrow \infty$ のとき, 単調収束定理より

$$\begin{aligned} \mathbf{E}_{\theta^*}[Y_1^{(L)}] &= \mathbf{E}_{\theta^*}[Y_1^+] - \mathbf{E}_{\theta^*}[\min\{Y_1^{-1}, L\}] \downarrow \mathbf{E}_{\theta^*}[Y_1^+] - \mathbf{E}_{\theta^*}[Y_1^{-}] \\ &= \mathbf{E}_{\theta^*}[Y_1] \end{aligned}$$

となる. ただし, $\mathbf{E}_{\theta^*}[Y_1] = -\infty$ の場合も許していることに注意せよ. 以上から, 十分大きな $L_0 > 0$ に対して

$$\mathbf{E}_{\theta^*}[Y_1^{(L_0)}] < 0$$

とできる. このとき, $Y_1^{(L_0)}, Y_2^{(L_0)}, \dots$ は i.i.d. かつ可積分であるから, 大数の強法則より

$$\frac{1}{n} \sum_{j=1}^n Y_j^{(L_0)} \xrightarrow{\text{a.s.}} \mathbf{E}_{\theta^*}[Y_1^{(L_0)}] < 0$$

となる. 一方

$$\sup_{|\theta|_2, d > M_0} \bar{W}_n(\theta) = \sup_{|\theta|_2, d > M_0} \frac{1}{n} \sum_{j=1}^n W_j(\theta) \leq \frac{1}{n} \sum_{j=1}^n \sup_{|\theta|_2, d > M_0} W_j(\theta)$$

$$= \frac{1}{n} \sum_{j=1}^n Y_j \leq \frac{1}{n} \sum_{j=1}^n Y_j^{(L_0)}$$

となる. したがって,

$$\Pr_{\theta^*} \left(\sup_{|\theta|_{2,d} > M_0} \bar{W}_n(\theta) \geq 0 \right) \leq \Pr_{\theta^*} \left(\frac{1}{n} \sum_{j=1}^n Y_j^{(L_0)} \geq 0 \right) \xrightarrow{n \rightarrow \infty} 0$$

となるので

$$\begin{aligned} \sup_{|\theta|_{2,d} > M_0} \bar{W}_n(\theta) &\leq \frac{1}{n} \sum_{j=1}^n \sup_{|\theta|_{2,d} > M_0} W_j(\theta) \\ &\xrightarrow{P} E_{\theta^*} \left[\sup_{|\theta|_{2,d} > M_0} W(\theta) \right] < 0 \quad (n \rightarrow \infty) \end{aligned}$$

となるので

$$\Pr \left(\sup_{|\theta|_{2,d} > M_0} \bar{W}_n(\theta) \geq 0 \right) \xrightarrow{n \rightarrow \infty} 0 \quad (12.14)$$

となることがわかる. いま, K を θ^* を中心とする半径 M_0 の閉球, すなわち

$$K := \{\theta \in \Theta; |\theta - \theta^*|_{2,d} \leq M_0\}$$

と³する. さらに,

$$A_n := \left\{ \sup_{\theta \in \Theta; |\theta - \theta^*|_{2,d} > M_0} \bar{W}_n(\theta) < 0 \right\}$$

とおく. (12.25) から

$$\Pr(A_n) \longrightarrow 1 \quad (n \rightarrow \infty)$$

である. 一方

$$\bar{W}_n(\theta^*) = 0$$

であり, $\hat{\theta}_n$ は $\bar{W}_n$ の最大化点なので

$$\bar{W}_n(\hat{\theta}_n) \geq \bar{W}_n(\theta^*) = 0$$

である. したがって, 事象 A_n 上では

$$\hat{\theta}_n \in K$$

³ $\Theta = \mathbb{R}^d$ のときは, 自動的に K はコンパクトとなる. しかし, $\Theta \neq \mathbb{R}^d$ で有界でないときには, K がコンパクトになるように M_0 を注意してとる必要がある.

となる. ここで, $\tilde{\theta}_n$ を K 上で $\overline{W}_n(\theta)$ を最大にする確率ベクトルとする. ただし, A_n 上では

$$\tilde{\theta}_n = \hat{\theta}_n$$

となるように選ぶ. 定理 12.7 から

$$\tilde{\theta}_n \xrightarrow{P} \theta^* \quad (n \rightarrow \infty)$$

となる. また,

$$\Pr(\hat{\theta}_n \neq \tilde{\theta}_n) \leq \Pr(A_n^c) \xrightarrow{n \rightarrow \infty} 0$$

である. したがって, 任意の $\epsilon > 0$ に対して

$$\Pr(|\hat{\theta}_n - \theta^*|_{2,d} > \epsilon) \leq \Pr(\hat{\theta}_n \neq \tilde{\theta}_n) + \Pr(|\tilde{\theta}_n - \theta^*|_{2,d} > \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

がわかる. よって

$$\hat{\theta}_n \xrightarrow{P} \theta^* \quad (n \rightarrow \infty)$$

が示された. □

12.3 最尤推定量の漸近正規性

定理 12.9. 統計的モデル $\mathcal{P} = \{p_\theta; \theta \in \Theta\}$ を考える. ただし, $\Theta \subset \mathbb{R}$ で p_θ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. 以下を仮定する.

- (a) $X, X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*}$ とする. ただし, $\theta^* \in \Theta \subset \mathbb{R}$ である.
- (b) $A := \{x \in \mathbb{R}; p_\theta(x) > 0\}$ は $\theta \in \Theta$ に依存しない.
- (c) すべての $x \in A$ において, $p_\theta(x)$ は θ に関する 2 回連続微分可能な関数である.
- (d) $X \sim p_{\theta^*}$ に対して

$$W(\theta) = \log p_\theta(X) \quad (\theta \in \Theta)$$

と定める. 1 つの標本に基づく Fisher 情報量 $\mathcal{F}_X(\theta^*)$ は有界で

$$\mathcal{F}_X(\theta^*) = \text{Var} \left[\dot{W}(\theta^*) \right], \quad \mathcal{F}_X(\theta^*) = -\mathbb{E}_{\theta^*} \left[\ddot{W}(\theta^*) \right]$$

と表現できるとする. ただし, $\dot{W}(\theta) = \frac{dW}{d\theta}(\theta)$ と $\ddot{W}(\theta) = \frac{d^2W}{d\theta^2}(\theta)$ と定めた.

(e) Θ のすべての内点 θ^* に対して, ある $\epsilon > 0$ が存在して

$$E_{\theta^*} \left[\left\| \mathbb{1}_{[\theta^*-\epsilon, \theta^*+\epsilon]} \ddot{W} \right\|_{\infty} \right] < \infty$$

とする. ただし, Θ 上の実数値関数 g に対して, $\|g\|_{\infty} = \sup_{\theta \in \Theta} |g(\theta)|$ と定めた.

(f) θ^* の最尤推定量 $\hat{\theta}_n$ は一意性をもつ.

このとき, Θ の任意の内点 θ^* に対して

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \rightsquigarrow N\left(0, \frac{1}{\mathcal{F}_X(\theta^*)}\right) \quad (n \rightarrow \infty)$$

が成り立つ.

この定理を証明するために次の補題を利用する.

補題 12.10. $Y, \{Y_n\}_{n=1}^{\infty}$ を確率変数列とし, $Y_n \rightsquigarrow Y (n \rightarrow \infty)$ とする. $\{B_n\}_{n=1}^{\infty}$ を事象の列とし, $\Pr(B_n) \xrightarrow{n \rightarrow \infty} 1$ とする. このとき, 任意の確率変数列 $\{Z_n\}_{n=1}^{\infty}$ に対して

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}_{B_n^c} \rightsquigarrow Y \quad (n \rightarrow \infty)$$

が成り立つ. ただし, $\mathbb{1}_{B_n} = \begin{cases} 1 & (B_n \text{ が起こったとき}) \\ 0 & (\text{その他の場合}) \end{cases}$ である.

Proof. $\forall \epsilon > 0$ に対して

$$\Pr(|Z_n \mathbb{1}_{B_n^c}| > \epsilon) \leq \Pr(B_n^c) = 1 - \Pr(B_n) \xrightarrow{n \rightarrow \infty} 0$$

となる. したがって

$$Z_n \mathbb{1}_{B_n^c} \xrightarrow{P} 0 \quad (n \rightarrow \infty) \quad (12.15)$$

がわかる. さらに

$$\Pr(|\mathbb{1}_{B_n} - 1| > \epsilon) \leq \Pr(B_n^c) = 1 - \Pr(B_n) \xrightarrow{n \rightarrow \infty} 0$$

となるので

$$\mathbb{1}_{B_n} \xrightarrow{P} 1 \quad (n \rightarrow \infty) \quad (12.16)$$

がわかる. (12.15) と (12.16) に注意して Slutsky の定理 (定理 4.14(3)(4)) を用いると

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}_{B_n^c} \rightsquigarrow Y \quad (n \rightarrow \infty)$$

がわかる. □

定理 12.9 の証明: 仮定 (e) から $\epsilon > 0$ をうまく取って

$$[\theta^* - \epsilon, \theta^* + \epsilon] \subset \Theta \quad \text{かつ} \quad E_{\theta^*} \left[\left\| \mathbb{1}_{[\theta^* - \epsilon, \theta^* + \epsilon]} \ddot{W} \right\|_{\infty} \right] < \infty$$

となるようにできる. 事象 B_n を

$$B_n := \{\hat{\theta}_n \in [\theta^* - \epsilon, \theta^* + \epsilon]\}$$

と定める. $\hat{\theta}_n$ の一貫性から

$$\Pr_{\theta^*}(B_n) \xrightarrow{n \rightarrow \infty} 1 \quad (12.17)$$

となる. さらに, B_n 上では, $\hat{\theta}_n$ は $n\bar{W}_n(\theta) = \ell_n(\theta)$ ($\theta \in \Theta$) を最大にするので

$$\dot{\bar{W}}_n(\hat{\theta}_n) = \frac{d\bar{W}}{d\theta}(\theta) \Big|_{\theta=\hat{\theta}_n} = 0$$

となる. ここで, θ^* のまわりで $\dot{\bar{W}}_n(\theta^*)$ を Taylor 展開すると

$$\dot{\bar{W}}_n(\theta^*) = \dot{\bar{W}}_n(\hat{\theta}_n) + \ddot{\bar{W}}_n(\tilde{\theta}_n)(\theta^* - \hat{\theta}_n) \quad (12.18)$$

を得る. ただし, $\tilde{\theta}_n$ は θ^* と $\hat{\theta}_n$ を結ぶ直線上の確率変数である. $\dot{\bar{W}}_n(\hat{\theta}) = 0$ なので

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \frac{\sqrt{n} \dot{\bar{W}}_n(\theta^*)}{-\ddot{\bar{W}}_n(\tilde{\theta}_n)}$$

となる. 仮定 (d) から $\dot{\bar{W}}_n(\theta^*)$ は平均 0, 分散 $\mathcal{F}_X(\theta^*)$ の正規分布に従う i.i.d. 確率変数の標本平均であることに注意して, 中心極限定理 (定理 4.22) を用いると

$$\sqrt{n} \dot{\bar{W}}_n(\theta^*) \rightsquigarrow Z \sim \mathbf{N}(0, \mathcal{F}_X(\theta^*)) \quad (n \rightarrow \infty) \quad (12.19)$$

となる. B_n 上では

$$|\tilde{\theta}_n - \theta^*| \leq |\hat{\theta}_n - \theta^*|$$

であることと $\hat{\theta}_n$ の一貫性から

$$\Pr(|\tilde{\theta}_n - \theta^*| < \epsilon) \geq \Pr(|\hat{\theta}_n - \theta^*| < \epsilon) \xrightarrow{n \rightarrow \infty} 1$$

となることがわかる. よって

$$\tilde{\theta}_n \xrightarrow{P} \theta^* \quad (n \rightarrow \infty) \quad (12.20)$$

となる. 定理 12.2 から

$$\|\mathbb{1}_{[\theta^*-\epsilon, \theta^*+\epsilon]}(\ddot{\bar{W}}_n - \mu)\|_\infty \xrightarrow{P} 0 \quad (n \rightarrow \infty) \quad (12.21)$$

となる. ただし, $\mu(\theta) = E_{\theta^*}[\ddot{\bar{W}}_n(\theta)]$ ($\theta \in \Theta$) である. さらに, (12.20) と (12.21) に注意して, 定理 12.4(1) を用いると

$$\ddot{\bar{W}}_n(\tilde{\theta}_n) \xrightarrow{P} \mu(\theta^*) = -\mathcal{F}_X(\theta^*) \quad (12.22)$$

となる. B_n^c 上では $\hat{\theta}_n$ の振る舞いは関係ない. 最後に, (12.17), (12.19) および (12.22) に注意して補題 12.10 を用いると

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta^*) &= \sqrt{n}(\hat{\theta}_n - \theta^*)\mathbb{1}_{B_n} + \sqrt{n}(\hat{\theta}_n - \theta^*)\mathbb{1}_{B_n^c} \\ &= \frac{\sqrt{n} \ddot{\bar{W}}_n(\theta^*)}{\ddot{\bar{W}}_n(\tilde{\theta}_n)}\mathbb{1}_{B_n} + \sqrt{n}(\hat{\theta}_n - \theta^*)\mathbb{1}_{B_n^c} \\ &\rightsquigarrow \frac{Z}{\mathcal{F}_X(\theta^*)} \sim N\left(0, \frac{1}{\mathcal{F}_X(\theta^*)}\right) \quad (n \rightarrow \infty) \end{aligned}$$

がわかる. □

定理 12.11 (Camér の定理). $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*}$ とする. ただし, $\theta^* \in \Theta \subset \mathbb{R}^d$ とする. さらに, 以下の条件 (a) – (f) を仮定する.

(a) Θ は $\mathbb{R}^d$ の開部分集合である.

(b) $A := \{x \in \mathbb{R}; p_{\theta}(x) > 0\}$ は $\theta \in \Theta$ に依存しない.

(c) ほとんど至るところの $x \in A$ について, 関数 $\Theta \ni \theta \mapsto p_{\theta}(x)$ は 2 回連続微分可能である. さらに, その 2 回の微分演算と積分記号の交換ができることと仮定する.

(d) $\mathbb{R}$ 上の関数 K が存在して, $E_{\theta^*}[|K(X)|] < \infty$ であり

$$\ddot{\bar{W}}(\theta) := \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log p_{\theta}(X) \right)_{j, k=1, 2, \dots, d}; \quad \theta = (\theta_1, \theta_2, \dots, \theta_d)^\top$$

のどの成分の絶対値も θ^* の近傍において, 一様に K で上からおさえられる.

(e) $\mathcal{F}_X(\theta^*) = -E_{\theta^*}[\ddot{\bar{W}}(\theta^*)]$ は正定値である.

(f) $p_{\theta} = p_{\theta^*} \Leftrightarrow \theta = \theta^*$.

このとき, θ^* の最尤推定量 $\hat{\theta}_n$ は

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \rightsquigarrow N\left(\mathbf{0}_d, \{\mathcal{F}_X(\theta^*)\}^{-1}\right) \quad (n \rightarrow \infty)$$

Proof. 証明は Fergusson (2017, pp.135-136) を参照のこと.

【コメント】 定理 12.9 の条件を確認すればよい.

□

12.4 尤度比検定統計量の漸近分布

正則母数モデル $\mathcal{P} = \{p_{\theta}; \theta \in \Theta \subset \mathbb{R}^d\}$ を考える. ただし, p_{θ} は $\mathbb{R}^d$ 上の p.d.f. または p.m.f. である. いま

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta^*} \quad (\theta^* \in \Theta)$$

を仮定する. 標本 $\mathbf{X} := (X_1, X_2, \dots, X_n)^{\top}$ に基づき, 仮説

$$H_0: \theta^* \in \Theta_0 \quad \text{v.s.} \quad H_1: \theta^* \in \Theta \setminus \Theta_0$$

を検定する. ここで Θ_0 は Θ の空でない真部分集合である. 汎用的検定統計量である尤度比検定統計量

$$\lambda_n(\mathbf{X}) := \frac{\sup_{\theta \in \Theta_0} \prod_{j=1}^n p_{\theta}(X_j)}{\sup_{\theta \in \Theta} \prod_{j=1}^n p_{\theta}(X_j)} = \frac{\text{lik}_n(\hat{\theta}_{n,0})}{\text{lik}_n(\hat{\theta}_n)} \quad (12.23)$$

を考える. ここで, $\hat{\theta}_n$ と $\hat{\theta}_{n,0}$ はそれぞれ Θ と Θ_0 における最尤推定量である. $\lambda_n(\mathbf{X})$ の実現値が小さい値のとき, H_0 は棄却されるという検定方式である.

帰無仮説 H_0 のもとでの尤度比検定統計量の漸近分布を求めるために, Θ_0 を含む最小の affine 平面の次元 $\dim \Theta_0$ が $d-r$ ($1 \leq r < d$) であると仮定する. さらに, $\theta = (\theta_1, \theta_2, \dots, \theta_d)^{\top}$ と書いたとき, 帰無仮説 H_0 が

$$H_0: \theta_1 = \theta_2 = \dots = \theta_r = 0$$

と書けたと仮定する.

定理 12.12 (Wilks の定理). 定理 12.11 の仮定のもとで真の母数 θ^* が帰無仮説 H_0 を満足しているとする. このとき

$$-2 \log \lambda_n(\mathbf{X}) \rightsquigarrow \chi_r^2 \quad (n \rightarrow \infty)$$

が成立する. ただし, χ_r^2 は自由度 r の χ^2 分布を表す.

Proof. (12.23) から

$$-2 \log \lambda_n(\mathbf{X}) = 2 \{ \ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_{n,0}) \}$$

と書き直せることに注意する. ただし, $\ell_n(\boldsymbol{\theta}) = \log \text{lik}_n(\boldsymbol{\theta})$ である. $\ell_n(\hat{\boldsymbol{\theta}}_{n,0})$ を $\hat{\boldsymbol{\theta}}_n$ のまわりで 2 次の項まで Taylor 展開すると

$$\begin{aligned}\ell_n(\hat{\boldsymbol{\theta}}_{n,0}) &= \ell_n(\hat{\boldsymbol{\theta}}_n) + \dot{\ell}_n(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) \\ &\quad - n(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)^\top \tilde{\mathcal{F}}_X(\tilde{\boldsymbol{\theta}}_{n,0})(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)\end{aligned}$$

と書ける. ただし

$$\tilde{\mathcal{F}}_X(\tilde{\boldsymbol{\theta}}_{n,0}) = -\frac{1}{n} \int_0^1 \int_0^1 v \ddot{\ell}_n\{\hat{\boldsymbol{\theta}}_n + uv(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)\} du dv$$

とした. さらに

$$\tilde{\mathcal{F}}_X(\tilde{\boldsymbol{\theta}}_{n,0}) \xrightarrow{\text{a.s.}} \frac{1}{2} \mathcal{F}_X(\boldsymbol{\theta}^*) \quad (n \rightarrow \infty) \quad (12.24)$$

となることがわかる. ここで $\mathcal{F}_X(\cdot)$ は定理 12.11(e) で与えられた Fisher 情報量である. $\dot{\ell}_n(\hat{\boldsymbol{\theta}}_n) = 0$ が成り立つので, 十分大きな n に対して

$$\begin{aligned}-2 \log \lambda_n(\mathbf{X}) &= 2n(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)^\top \tilde{\mathcal{F}}_X(\tilde{\boldsymbol{\theta}}_{n,0})(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) \\ &= n(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)^\top \mathcal{F}_X(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) + o_P(1)\end{aligned} \quad (12.25)$$

が成り立つ. ただし, 確率変数列 A_n, B_n ($n \in \mathbb{N}$) に対して, $A_n = B_n + o_P(1)$ は $A_n - B_n \xrightarrow{P} 0$ ($n \rightarrow \infty$) を意味する.

帰無仮説 H_0 が単純のとき, $\hat{\boldsymbol{\theta}}_{n,0} = \boldsymbol{\theta}^*$ なので

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \rightsquigarrow N_d\left(\mathbf{0}_d, \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1}\right) \quad (n \rightarrow \infty)$$

となるので, 証明は終わりである.

帰無仮説 H_0 が複合のとき, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_{n,0})$ の漸近分布を求めるために, $\dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0})$ を $\hat{\boldsymbol{\theta}}_n$ のまわりで Taylor 展開する.

$$\begin{aligned}\frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) &= \frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) + \frac{1}{n} \left(\int_0^1 \ddot{\ell}_n\{\hat{\boldsymbol{\theta}}_n + v(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n)\} dv \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) \\ &= -\mathcal{F}_X(\boldsymbol{\theta}^*) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) + o_P(1)\end{aligned}$$

となる. ただし, $\boldsymbol{\theta} = (\Theta_1, \Theta_2, \dots, \Theta_d)^\top$ と書いたとき

$$\begin{aligned}\dot{\ell}_n(\boldsymbol{\theta}) &= \left(\frac{\partial \ell_n}{\partial \Theta_1}(\boldsymbol{\theta}), \frac{\partial \ell_n}{\partial \Theta_2}(\boldsymbol{\theta}), \dots, \frac{\partial \ell_n}{\partial \Theta_d}(\boldsymbol{\theta}) \right)^\top, \\ \ddot{\ell}_n(\boldsymbol{\theta}) &= \left(\frac{\partial^2 \ell_n}{\partial \Theta_j \partial \Theta_k}(\boldsymbol{\theta}) \right)_{j,k=1,2,\dots,d}\end{aligned}$$

と定めた. したがって

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \hat{\boldsymbol{\theta}}_n) = \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) + o_P(1)$$

を得る. このことから

$$-2 \log \lambda_n = \frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0})^\top \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) + o_P(1)$$

と表現できる. $\dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0})$ の漸近分布を求めるために $\boldsymbol{\theta}^*$ の周りでこれを Taylor 展開する.

$$\begin{aligned} & \frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) \\ &= \frac{1}{\sqrt{n}} \dot{\ell}_n(\boldsymbol{\theta}^*) + \frac{1}{n} \left(\int_0^1 \ddot{\ell}_n \{ \boldsymbol{\theta}^* + v(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \} dv \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \end{aligned} \quad (12.26)$$

を得る. ここで, $d \times d$ 行列 $\mathcal{F}_X(\boldsymbol{\theta}^*)$ を 4 つのブロック行列に分解する.

$$\mathcal{F}_X(\boldsymbol{\theta}^*) = \begin{pmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{G}_2^\top & \mathbf{G}_3 \end{pmatrix}.$$

ただし, $\mathbf{G}_1$ は $r \times r$ 行列である. さらに

$$\mathbf{H} = \begin{pmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (d-r)} \\ \mathbf{0}_{r \times r}^\top & \mathbf{G}_3^{-1} \end{pmatrix}$$

とする. $\dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0})$ の後半の $(d-r)$ 個の要素は 0 なので

$$\mathbf{H} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) = \mathbf{0}_d \quad (12.27)$$

となる. よって, (12.26) と (12.27) から

$$\begin{aligned} \mathbf{H} \frac{1}{\sqrt{n}} \dot{\ell}_n(\boldsymbol{\theta}^*) &= \mathbf{H} \left\{ \frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) - \frac{1}{n} \left(\int_0^1 \ddot{\ell}_n \{ \boldsymbol{\theta}^* + v(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \} dv \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \right\} \\ &= -\mathbf{H} \frac{1}{n} \left(\int_0^1 \ddot{\ell}_n \{ \boldsymbol{\theta}^* + v(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \} dv \right) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) \\ &= -\mathbf{H} \mathcal{F}_X(\boldsymbol{\theta}^*) \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) + o_P(1) \\ &= - \begin{pmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (d-r)} \\ \mathbf{0}_{r \times r}^\top & \mathbf{G}_3^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{G}_2^\top & \mathbf{G}_3 \end{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) + o_P(1) \\ &= - \begin{pmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (d-r)} \\ \mathbf{G}_3^{-1} \mathbf{G}_2^\top & \mathbf{I}_{d-r} \end{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) + o_P(1) \\ &= \sqrt{n}(\hat{\boldsymbol{\theta}}_{n,0} - \boldsymbol{\theta}^*) + o_P(1) \end{aligned} \quad (12.28)$$

となる. 最後の等号は $\hat{\boldsymbol{\theta}}_{n,0}$ と $\boldsymbol{\theta}^*$ の最初の r 個の成分は 0 であることからわかる. (12.28) を (12.27) に代入すると

$$\frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) = \{\mathbf{I}_d - \mathcal{F}_X(\boldsymbol{\theta}^*)\mathbf{H}\} \frac{1}{\sqrt{n}} \dot{\ell}_n(\boldsymbol{\theta}^*) + o_P(1)$$

を得る. 中心極限定理 (定理 4.24) から

$$\frac{1}{\sqrt{n}} \dot{\ell}_n(\boldsymbol{\theta}^*) \rightsquigarrow N_d(\mathbf{0}_d, \mathcal{F}_X(\boldsymbol{\theta}^*)) \quad (n \rightarrow \infty)$$

が成り立つので

$$\begin{aligned} \frac{1}{\sqrt{n}} \dot{\ell}_n(\hat{\boldsymbol{\theta}}_{n,0}) &\rightsquigarrow (\mathbf{I}_d - \mathcal{F}_X(\boldsymbol{\theta}^*)\mathbf{H})\mathbf{Y} \quad (n \rightarrow \infty), \\ \mathbf{Y} &\sim N_d(\mathbf{0}_d, \mathcal{F}_X(\boldsymbol{\theta}^*)) \end{aligned}$$

である. これより

$$\begin{aligned} -2 \log \lambda_n &= \mathbf{Y}^\top (\mathbf{I}_d - \mathcal{F}_X(\boldsymbol{\theta}^*)\mathbf{H})^\top \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} (\mathbf{I}_d - \mathcal{F}_X(\boldsymbol{\theta}^*)\mathbf{H})\mathbf{Y} + o_P(1) \\ &= \mathbf{Y}^\top \left[\{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} - \mathbf{H} \right] \mathbf{Y} + o_P(1) \\ &= \mathbf{Z}^\top \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{1/2} \left[\{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} - \mathbf{H} \right] \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{1/2} \mathbf{Z} + o_P(1) \end{aligned}$$

となる. ただし, $\mathbf{Z} \sim N_d(\mathbf{0}_d, \mathbf{I}_d)$ である. ここで, 行列

$$\mathbf{P} = \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{1/2} \left[\{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{-1} - \mathbf{H} \right] \{\mathcal{F}_X(\boldsymbol{\theta}^*)\}^{1/2}$$

は

$$\mathbf{P}^2 = \mathbf{P}; \quad \mathbf{P}^\top = \mathbf{P}; \quad \text{rank}(\mathbf{P}) = \text{tr}(\mathbf{P}) = \text{tr}(\mathbf{I}_d - \mathcal{F}_X(\boldsymbol{\theta}^*)\mathbf{H}) = r$$

をみたす. $\mathbf{P}$ はランク r の正射影となる. したがって

$$-2 \log \lambda_n(\mathbf{X}) \rightsquigarrow \mathbf{Z}^\top \mathbf{P} \mathbf{Z} \sim \chi_r^2 \quad (n \rightarrow \infty)$$

を得る. □

(12.24) の証明 □

12.5 章末注釈と参考文献

12.6 演習問題

演習問題 12.1.

演習問題 12.2.