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. 節 12.5 では, 指数型分布族の仮定のもとで不完全データに基づく最尤推定値を求めるアルゴリズムである EM アルゴリズムを説明する.

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, $h\colon K \times \mathbb{R} \to \mathbb{R}$ を各 $x \in \mathbb{R}$ に対して

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

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

$$W_n(\boldsymbol{t}) := h(\boldsymbol{t}, X_n) \qquad (\boldsymbol{t} \in K)$$

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 (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

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

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

$$\|w_n - w\|_\infty \overset{n\to\infty}{\longrightarrow} 0$$

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

$$\mu(\boldsymbol{t}) = \mathsf{E}[W(\boldsymbol{t})] \qquad (\boldsymbol{t} \in K)$$

と定める. このとき, 以下のことが成立する.
(1) $\mathsf{E}[\|W\|_\infty] < \infty$ のとき, μ は K 上の連続関数となる.
(2) $\epsilon > 0$ を任意とし, $\mathsf{E}[\|W\|_\infty] < \infty$ とする. このとき

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

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: $\boldsymbol{t}_n, \boldsymbol{t} \in K\, (n = 1, 2, \ldots)$ とし

$$\lim_{n\to\infty} \boldsymbol{t}_n = \boldsymbol{t}$$

とする. ランダム関数 W は連続だから

$$W(\boldsymbol{t}_n) \xrightarrow{\text{a.s.}} W(\boldsymbol{t}) \qquad (n \to \infty)$$

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

$$\lim_{n\to\infty} \mu(\boldsymbol{t}_n) = \lim_{n\to\infty} \mathsf{E}[W(\boldsymbol{t}_n)] = \mathsf{E}\big[\lim_{n\to\infty} W(\boldsymbol{t}_n)\big] = \mathsf{E}[W(\boldsymbol{t})] = \mu(\boldsymbol{t})$$

となる. 点列 $\{\boldsymbol{t}_n\}_{n=1}^\infty$ は任意だったので, μ は $\boldsymbol{t}$ で連続であることがわかる. さらに, $\boldsymbol{t} \in K$ 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: 任意の $\epsilon > 0$ に対して

$$M_\epsilon(\boldsymbol{t}) := \sup_{\boldsymbol{s}\in K;\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon} |W(\boldsymbol{s}) - W(\boldsymbol{t})|,$$
$$\lambda_\epsilon(\boldsymbol{t}) := \mathsf{E}\big[M_\epsilon(\boldsymbol{t})\big] \qquad (\boldsymbol{t} \in K)$$

[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.

と定める. W は $\boldsymbol{t}(\in K)$ で連続関数だったので, M_ϵ も $\boldsymbol{t}(\in K)$ で連続となる. また, $|M_\epsilon(\boldsymbol{t})| \le \|M_\epsilon\|_\infty$ かつ

$$\begin{aligned}
\mathsf{E}\big[\|M_\epsilon\|_\infty\big] &= \mathsf{E}\left[\sup_{\boldsymbol{t}\in K}|M_\epsilon(\boldsymbol{t})|\right] \\
&= \mathsf{E}\left[\sup_{\boldsymbol{t}\in K}\left\{\sup_{\boldsymbol{s}\in K;\,|\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon}|W(\boldsymbol{s})-W(\boldsymbol{t})|\right\}\right] \\
&\le \mathsf{E}\left[\sup_{\boldsymbol{t}\in K}\left\{\sup_{\boldsymbol{s}\in K;\,|\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon}\big\{|W(\boldsymbol{s})|+|W(\boldsymbol{t})|\big\}\right\}\right] \\
&\le 2\mathsf{E}\big[\|W\|_\infty\big] < \infty
\end{aligned}$$

となることに注意する. 再度, $\boldsymbol{t}\in K$ に収束する任意の点列 $\{\boldsymbol{t}_n\}_{n=1}^\infty$ を取り, 優収束定理を用いると

$$\begin{aligned}
\lim_{n\to\infty}\lambda(\boldsymbol{t}_n) &= \lim_{n\to\infty}\mathsf{E}\big[M_\epsilon(\boldsymbol{t}_n)\big] = \mathsf{E}\big[\lim_{n\to\infty}M_\epsilon(\boldsymbol{t}_n)\big] \qquad (\because \text{優収束定理}) \\
&= \mathsf{E}\big[M_\epsilon\big(\lim_{n\to\infty}\boldsymbol{t}_n\big)\big] \qquad (\because M_\epsilon \text{ の連続性}) \\
&= \mathsf{E}\big[M_\epsilon(\boldsymbol{t})\big]
\end{aligned}$$

となるので, λ_ϵ は $\boldsymbol{t}(\in K)$ で連続である. M_ϵ の定義から $\epsilon\to 0$ のとき, $\lambda_\epsilon \downarrow 0$ となる. したがって, Dini の定理から

$$\sup_{\boldsymbol{t}\in K}\lambda_\epsilon(\boldsymbol{t}) \downarrow 0 \qquad (\epsilon\to 0)$$

がわかる. □

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. $W, W_1, W_2, \ldots$ を $C(K)$ 値 i.i.d. ランダム要素とする. さらに

$$\mu(\boldsymbol{t}) = \mathsf{E}\big[W(\boldsymbol{t})\big] \qquad (\boldsymbol{t}\in K)$$

とおき

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

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

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

とおいたとき

$$\left\|\overline{W}_n - \mu\right\|_\infty \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{12.1}$$

が成り立つ.

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

$$M_{\delta, j}(\boldsymbol{t}) := \sup_{\boldsymbol{s}\in K:\, |\boldsymbol{s}-\boldsymbol{t}|_{2,d}<\delta} \left|W_j(\boldsymbol{s}) - W_j(\boldsymbol{t})\right|; \qquad \lambda_\delta(\boldsymbol{t}) := \mathsf{E}\left[M_{\delta, i}(\boldsymbol{t})\right]$$

と定める. 補題 12.1 から δ を十分小さくすると

$$\lambda_\delta(\boldsymbol{t}) = \mathsf{E}\left[\sup_{\boldsymbol{s}\in K:\, |\boldsymbol{s}-\boldsymbol{t}|_{2,d}<\delta} \left|W_i(\boldsymbol{s}) - W_i(\boldsymbol{t})\right|\right] \le \epsilon \qquad (\boldsymbol{t} \in K)$$

とできる. ここで

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

とおく. すると K はコンパクトなので, K の開被覆 $\left\{B(\boldsymbol{t};\, \delta)\right\}_{\boldsymbol{t}\in K}$ 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 $\{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_m\} \subset K$ が存在して

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

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

$$\begin{aligned}
\left\|\overline{W}_n - \mu\right\|_\infty &= \max_{j\in\{1,2,\ldots,m\}} \left\{\sup_{\boldsymbol{t}\in O_j} \left|\overline{W}_n(\boldsymbol{t}) - \mu(\boldsymbol{t})\right|\right\} \\
&\le \max_{j\in\{1,2,\ldots,m\}} \left\{\sup_{\boldsymbol{t}\in O_j} \left|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_j)\right| + \left|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\right|\right. \\
&\qquad\qquad \left. + \left|\mu(\boldsymbol{t}_j) - \mu(\boldsymbol{t})\right|\right\} \\
&\le \max_{j\in\{1,2,\ldots,m\}} \left\{\sup_{\boldsymbol{t}\in O_j} \left|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_j)\right|\right\} \\
&\qquad\qquad + \max_{j\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_j)\right| + \epsilon \qquad (\mu \text{ の連続性})
\end{aligned}$$

となる. さらに

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

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

$$\overline{M}_{\delta,j} \xrightarrow{\mathrm{P}} \lambda_\delta(\boldsymbol{t}_j) < \epsilon \qquad (n \to \infty)$$

が成立する. よって

$$\mathsf{Pr}\Big(\overline{M}_{\delta,j} < \epsilon\Big) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.2}$$

がわかる. さらに, 大数の弱法則から

$$\mathsf{Pr}\Big(\max_{i\in\{1,2,\ldots,m\}} \big|\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)\big| < \epsilon\Big) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.3}$$

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

$$\begin{aligned}\big\|\overline{W}_n - \mu\big\|_\infty &\le \max_{i\in\{1,2,\ldots,m\}} \Big\{\underbrace{\sup_{\boldsymbol{t}\in O_i} \big|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)\big|}_{\le \overline{M}_{\delta,i}(\boldsymbol{t}_i)}\Big\} \\ &\qquad + \max_{j\in\{1,2,\ldots,m\}} \big|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\big| + \epsilon \\ &\le 3\epsilon\end{aligned}$$

が成り立つ. したがって, (12.2) と (12.3) と合わせると

$$\begin{aligned}&\mathsf{Pr}\Big(\big\|\overline{W}_n - \mu\big\|_\infty \le 3\epsilon\Big) \\ &\quad \ge \mathsf{Pr}\Big(\{\overline{M}_{\delta,i} < \epsilon\} \bigcap \Big\{\sup_{j\in\{1,2,\ldots,m\}} \big|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\big| < \epsilon\Big\}\Big) \\ &\quad \overset{n\to\infty}{\longrightarrow} 1\end{aligned}$$

となる. よって

$$\mathsf{Pr}\Big(\big\|\overline{W}_n - \mu\big\|_\infty > 3\epsilon\Big) \overset{n\to\infty}{\longrightarrow} 0$$

がわかる. □

注意 12.3. 大数の強法則から

$$\big\|\overline{W}_n - \mu\big\|_\infty \overset{\text{a.s.}}{\to} 0 \qquad (n \to \infty)$$

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

$$\left\|G_n - g\right\|_{\infty} \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

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

$$G_n(\boldsymbol{t}_n) \xrightarrow{\mathrm{P}} g(\boldsymbol{t}^*) \qquad (n \to \infty)$$

が成立する.
(2) 非ランダムな関数 g は $\boldsymbol{t}^*\,(\in K)$ 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 $\boldsymbol{t}_n\,(n = 1, 2, \ldots)$ で G_n は最大を取るとする. すなわち

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

である. このとき

$$\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t} \qquad (n \to \infty)$$

が成立する.
(3) $K \subset \mathbb{R}$ とし, $t^* \in K$ は方程式 $g(t) = 0$ の唯一の解とし, 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

$$t_n \xrightarrow{\mathrm{P}} t^* \qquad (n \to \infty)$$

が成立する.

Proof. (1) の証明: まず

$$\begin{aligned}\left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| &\le \left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}_n)\right| + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \\ &\le \left\|G_n - g\right\|_{\infty} + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \end{aligned} \tag{12.4}$$

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

$$g(\boldsymbol{t}_n) \xrightarrow{\mathrm{P}} g(\boldsymbol{t}^*) \qquad (n \to \infty)$$

となる. したがって, (12.4) から $\forall \epsilon > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\Big(\left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \ge \epsilon\Big) &\le \mathsf{Pr}\left(\left\|G_n - g\right\|_{\infty} + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \ge \epsilon\right) \\ &\le \mathsf{Pr}\left(\left\|G_n - g\right\|_{\infty} \ge \frac{\epsilon}{2}\right) + \mathsf{Pr}\left(\left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \ge \frac{\epsilon}{2}\right) \\ &\xrightarrow{n\to\infty} 0\end{aligned}$$

がわかる．よって, (1) が示された．
(2) の証明: $\epsilon > 0$ を固定して

$$K_\epsilon := K \setminus B(\boldsymbol{t}^*;\,\epsilon) = K \bigcap \bigl\{B(\boldsymbol{t}^*;\,\epsilon)\bigr\}^c \tag{12.5}$$

とおく．ただし, $B(\boldsymbol{t}^*;\,\epsilon) = \bigl\{\boldsymbol{t} \in \mathbb{R}^d;\ \bigl|\boldsymbol{t} - \boldsymbol{t}^*\bigr|_{2,\,d} < \epsilon\bigr\}$ 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

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

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

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

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

$$M_\epsilon < M$$

となる．ここで

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

とおく．$\bigl\|G_n - g\bigr\|_\infty < \dfrac{\epsilon}{2}$ 上では

$$\begin{aligned}
\sup_{\boldsymbol{t}\in K_\epsilon} G_n(\boldsymbol{t}) &= \sup_{\boldsymbol{t}\in K_\epsilon}\bigl\{G_n(\boldsymbol{t}) - g(\boldsymbol{t})\bigr\} \\
&\le \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t}) + \sup_{\boldsymbol{t}\in K_\epsilon}\bigl|g(\boldsymbol{t}) + G_n(\boldsymbol{t}) - g(\boldsymbol{t})\bigr| \\
&\le \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t}) + \bigl\|G_n - g\bigr\|_\infty \\
&\le M_\epsilon + \frac{\delta}{2} = M - \frac{\delta}{2}
\end{aligned} \tag{12.6}$$

となる．一方, $\bigl\|G_n - g\bigr\|_\infty < \dfrac{\epsilon}{2}$ 上では

$$\begin{aligned}
\sup_{\boldsymbol{t}\in K} G_n(\boldsymbol{t}) &\ge G_n(\boldsymbol{t}^*) = g(\boldsymbol{t}^*) + G_n(\boldsymbol{t}^*) - g(\boldsymbol{t}^*) \\
&\ge g(\boldsymbol{t}^*) - \sup_{\boldsymbol{t}\in K}\bigl|G_n(\boldsymbol{t}) - g(\boldsymbol{t})\bigr| \\
&\ge g(\boldsymbol{t}^*) - \bigl\|G_n - g\bigr\|_\infty > g(\boldsymbol{t}^*) - \frac{\delta}{2}
\end{aligned} \tag{12.7}$$

となる. したがって, (12.6) と (12.7) から

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

となる. このことと K_ϵ の定義 (12.5) から

$$\left\|G_n - g\right\|_\infty < \frac{\epsilon}{2} \Rightarrow \boldsymbol{t}_n \in B(\boldsymbol{t}^*;\, \epsilon)$$

がわかる. よって

$$\mathsf{Pr}\left(\left\|G_n - g\right\|_\infty < \frac{\epsilon}{2}\right) \le \mathsf{Pr}\left(\left|\boldsymbol{t}_n - \boldsymbol{t}^*\right|_{2,\,d} < \epsilon\right)$$

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

$$\mathsf{Pr}\left(\left|\boldsymbol{t}_n - \boldsymbol{t}^*\right|_{2,\,d} \ge \epsilon\right) \le \mathsf{Pr}\left(\left\|G_n - g\right\|_\infty \ge \frac{\epsilon}{2}\right) \xrightarrow{n\to\infty} 0$$

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

<u>(3) の証明</u>: 一般性を失わず, G_n は t_n の近傍で非減少とし, g は t^* の近傍で非減少と仮定できる. g の近傍での非減少性から, $\epsilon > 0$ と $\delta > 0$ をうまく取り

$$g(t^* - \epsilon) < -2\delta \quad \text{かつ} \quad g(t^* + \epsilon) > 2\delta$$

となるようにできる. さらに, G_n の近傍での非減少性から

$$G_n(t^* - \epsilon) < -\delta \quad \text{かつ} \quad t_n \le t^* - \epsilon \Rightarrow G_n(t_n) < -\delta$$

となる. 同様に

$$G_n(t^* + \epsilon) > \delta \quad \text{かつ} \quad t_n \ge t^* + \epsilon \Rightarrow G_n(t_n) > \delta$$

である. また

$$\begin{aligned}
&\mathsf{Pr}\left(G_n(t^* - \epsilon) > \delta,\, G_n(t^* + \epsilon) > \delta\right) \\
&\le \mathsf{Pr}\left(t^* - \epsilon < t_n < t^* + \epsilon\right) + \mathsf{Pr}\left(\left\{G_n(t^* - \epsilon) < -\delta\right\} \bigcap \left\{t_n \le t^* - \epsilon\right\}\right) \\
&\quad + \mathsf{Pr}\left(\left\{G_n(t^* - \epsilon) > \delta\right\} \bigcap \left\{t_n \ge t^* + \epsilon\right\}\right) \\
&\le \mathsf{Pr}\left(t^* - \epsilon < t_n < t^* + \epsilon\right) + \mathsf{Pr}\left(G_n(t_n) < -\delta\right) + \mathsf{Pr}\left(G_n(t_n) > \delta\right)
\end{aligned} \tag{12.8}$$

となる. しかし, $\|G_n - g\|_\infty < \delta$ 上では

$$\begin{aligned} G_n(t^* - \epsilon) &\le g(t^* - \epsilon) + \bigl|G_n(t^* - \epsilon) - g(t^* - \epsilon)\bigr| \\ &\le -2\delta + \|G_n - g\|_\infty = -\delta \end{aligned}$$

となる. よって

$$\|G_n - g\|_\infty < \delta \Rightarrow G_n(t^* + \epsilon) > \delta$$

なので

$$\mathsf{Pr}\bigl(\|G_n - g\|_\infty < \delta\bigr) \le \mathsf{Pr}\bigl(G_n(t^* - \epsilon) < -\delta\bigr)$$

がわかる. また, $\|G_n - g\|_\infty < \delta$ 上では

$$\begin{aligned} G_n(t^* + \epsilon) &\ge g(t^* + \epsilon) - \bigl|G_n(t^* + \epsilon) - g(t^* + \epsilon)\bigr| \\ &\ge 2\delta - \|G_n - g\|_\infty = -\delta \end{aligned}$$

となる. よって

$$\|G_n - g\|_\infty < \delta \Rightarrow G_n(t^* + \epsilon) > \delta$$

なので

$$\mathsf{Pr}\bigl(\|G_n - g\|_\infty < \delta\bigr) \le \mathsf{Pr}\bigl(G_n(t^* - \epsilon) > \delta\bigr)$$

となる. $\mathsf{Pr}\bigl(\|G_n - g\|_\infty < \delta\bigr) \overset{n\to\infty}{\longrightarrow} 1$ に注意すると

$$\mathsf{Pr}\bigl(G_n(t^* - \epsilon) < -\delta\bigr) \overset{n\to\infty}{\longrightarrow} 1 \quad \text{かつ} \quad \mathsf{Pr}\bigl(G_n(t^* - \epsilon) > \delta\bigr) \overset{n\to\infty}{\longrightarrow} 1$$

がわかる. 一方

$$\mathsf{Pr}\bigl(G_n(t_n) < -\delta\bigr) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{かつ} \quad \mathsf{Pr}\bigl(G_n(t_n) > \delta\bigr) \overset{n\to\infty}{\longrightarrow} 0$$

であることに注意すると (12.8) から

$$\mathsf{Pr}\bigl(t^* - \epsilon < t_n < t^* + \epsilon\bigr) \overset{n\to\infty}{\longrightarrow} 1$$

がわかる. □

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, 正則統計的モデル $\mathcal{P} = \{\mathsf{p}_{\boldsymbol{\theta}}(\cdot);\ \boldsymbol{\theta} \in \Theta\}$ を考える. ただし, $\mathsf{p}_{\boldsymbol{\theta}}(\cdot)$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. $\boldsymbol{\theta}^{\in}\Theta$ を固定して

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\cdot)$$

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

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

で定める. このとき, 関数 $\Theta \ni \boldsymbol{\theta} \mapsto \ell_n(\boldsymbol{\theta})$ を最大にする $\boldsymbol{\theta}$ の値を $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ を最尤推定値という. 変数ベクトルに $\boldsymbol{X}$ を形式的に代入したものを $\widehat{\boldsymbol{\theta}} := \widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ 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

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

は連続である.

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

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

で定める.

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

Proof. Jensen の不等式 (定理 3.12) から

$$\begin{aligned}
-\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) &= \mathsf{E}_{\boldsymbol{\theta}^*}\left[\log\left\{\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right\}\right] \\
&\leq \log\left\{\mathsf{E}_{\boldsymbol{\theta}^*}\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right]\right\} \\
&= \log\left\{\int_{\{x\in\mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)>0\}} \frac{\mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)\,\mathrm{d}x\right\} \\
&= \log\left\{\int_{\{x\in\mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)>0\}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x\right\} \\
&\leq \log\left\{\int_{-\infty}^{\infty} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x\right\} \\
&= \log(1) = 0
\end{aligned}$$

となる. 上の不等式の等号成立条件は $\dfrac{\mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)$ 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, $\mathsf{P}_{\boldsymbol{\theta}} = \mathsf{Pr}_{\boldsymbol{\theta}^*}$ となるので

$$\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > 0$$

がわかる. □

記号: $\boldsymbol{\theta} \in \Theta$ と $X \sim \mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta}^*)$ に対して

$$W(\boldsymbol{\theta}) := \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right]$$

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) $\mathsf{E}_{\boldsymbol{\theta}^*}\big[\|W\|_\infty\big] < \infty$ である. ただし, $\|W\|_\infty = \sup_{\boldsymbol{\theta}\in\Theta}\big|W(\boldsymbol{\theta})\big|$ である.
(c) ほとんど至ところの $x \in \mathbb{R}$ で写像

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

は連続である.
(d) $\forall\boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \;\Rightarrow\; \mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$$

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

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

が成り立つ.

Proof. $n \in \mathbb{N}$ とし, $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\,\cdot\,)$ とする. $j = 1,\, 2,\, \ldots,\, n$ に対して

$$W_j(\boldsymbol{\theta}) := \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X_j)}\right] \qquad (\boldsymbol{\theta} \in \Theta)$$

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

$$\mu(\boldsymbol{\theta}) := \mathsf{E}\big[W_j(\boldsymbol{\theta})\big] = -\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) \qquad (\boldsymbol{\theta} \in \Theta)$$

と書く. 仮定 (d) と補題 12.6 から

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \;\Rightarrow\; \mu(\boldsymbol{\theta}) < 0$$

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

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

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

$$\left\|W_n - \mu\right\|_\infty \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

となる. よって, 定理 **??**(2) から

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

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 $\mathcal{P} = \{\mathsf{p}(\,\cdot\,|\,\boldsymbol{\theta});\, \boldsymbol{\theta} \in \Theta\}$ 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, $\mathsf{p}(\,\cdot\,|\boldsymbol{\theta})$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. さらに, $\mathsf{p}(\,\cdot\,|\boldsymbol{\theta})$ 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 $\mathsf{P}_{\boldsymbol{\theta}}$ と $\mathsf{Pr}_{\boldsymbol{\theta}}$ と書く. $\boldsymbol{\theta}^* \in \Theta$ とし

$$\begin{aligned}
&X,\, X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\,\cdot\,),\\
&W_j(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X_j)}\right] \qquad (j = 1,\, 2,\, \ldots,\, n),\\
&W(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right]
\end{aligned}$$

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

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

は連続である.
(b) $\forall \boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow \mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$$

である.
(c) ほとんど至るところの $x \in \mathbb{R}$ に対して, $\left|\boldsymbol{\theta}\right|_{2,\,d} \to 0$ のとき, $\mathsf{p}(x|\,\boldsymbol{\theta}) \to 0$ 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

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\|\mathbb{1}_K W\|_\infty\right] < \infty$$

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

(e) ある $M>0$ が存在して

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

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

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n\to\infty)$$

が成り立つ.

Proof. $|\boldsymbol{\theta}|_{2,\,d}\to 0$ のとき, $\mathsf{p}(x|\,\boldsymbol{\theta})\to 0$ なので, $\mathsf{p}(X|\,\boldsymbol{\theta}^*)>0$ のとき

$$\lim_{M\to\infty}\sup_{|\boldsymbol{\theta}|_{2,\,d}>M} W(\boldsymbol{\theta})=-\infty$$

となる. 上式と仮定 (e) に注意して優収束定理を用いると

$$\lim_{M\to\infty}\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M} W(\boldsymbol{\theta})\right]=-\infty$$

がわかる. よって, 十分大きな $M_0>0$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W(\boldsymbol{\theta})\right]<0$$

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

$$|\boldsymbol{\theta}^*|_{2,\,d}<M_0$$

である. さらに, 大数の強法則から

$$\begin{aligned}\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0}\overline{W}_n(\boldsymbol{\theta}) &\le \frac{1}{n}\sum_{j=1}^{n}\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W_j(\boldsymbol{\theta})\\ &\xrightarrow{\mathrm{P}} \mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W(\boldsymbol{\theta})\right]<0 \qquad (n\to\infty)\end{aligned}$$

となるので

$$\mathsf{Pr}\left(\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0}\overline{W}_n(\boldsymbol{\theta})\ge 0\right)\xrightarrow{n\to\infty} 0 \qquad (n\to\infty)$$

となることがわかる. いま, K を $\boldsymbol{\theta}^*$ を中心とした半径 M_0 の閉球とし, $\widetilde{\boldsymbol{\theta}}_n$ を K 上で $\overline{W}_n(\boldsymbol{\theta})$ を最大にする確率ベクトルとする. 定理 12.7 から

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n\to\infty)$$

となる. さらに

$$\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} \overline{W}_n(\boldsymbol{\theta}) < \overline{W}_n(\widehat{\boldsymbol{\theta}}_n) = 0$$

なので, $\widehat{\boldsymbol{\theta}}_n \in K$ となる. よって

$$\mathsf{Pr}\big(\widehat{\boldsymbol{\theta}}_n = \widetilde{\boldsymbol{\theta}}_n\big) \overset{n\to\infty}{\longrightarrow} 1$$

となる. よって

$$\widehat{\boldsymbol{\theta}}_n \overset{\mathrm{P}}{\to} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

がわかる. □

12.3 最尤推定量の漸近正規性

定理 12.9. 統計的モデル $\mathcal{P} = \{\mathsf{p}_\theta(\,\cdot\,);\, \theta \in \Theta\}$ を考える. ただし, $\Theta \subset \mathbb{R}$ で $\mathsf{p}_\theta(\,\cdot\,)$ 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) $X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\theta^*}(\,\cdot\,)$ とする. ただし, $\theta^* \in \Theta \subset \mathbb{R}$ である.
(b) $A := \{x \in \mathbb{R};\, \mathsf{p}_\theta(x) > 0\}$ は　$\theta(\in \Theta)$ に依存しない.
(c) すべての $x \in A$ において, $\mathsf{p}_\theta(x)$ は θ に関する 2 回連続微分可能な関数である.
(d) $X \sim \mathsf{p}_{\theta^*}(\,\cdot\,)$ に対して

$$W(\theta) = \log \mathsf{p}_\theta(X) \qquad (\theta \in \Theta)$$

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

$$\mathcal{F}_X(\theta^*) = \mathsf{E}\Big[\overset{\bullet}{W}(\theta^*)\Big], \qquad \mathcal{F}_X(\theta^*) = \mathsf{E}_{\theta^*}\Big[\overset{\bullet\bullet}{W}(\theta^*)\Big]$$

と表現できるとする. ただし, $\overset{\bullet}{W}(\theta) = \dfrac{\mathrm{d}W}{\mathrm{d}\theta}(\theta)$ と $\overset{\bullet\bullet}{W}(\theta) = \dfrac{\mathrm{d}^2 W}{\mathrm{d}\theta^2}(\theta)$ 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

$$\mathsf{E}_{\theta^*}\Big[\big\|\mathbb{1}_{[\theta^*-\epsilon,\,\theta+\epsilon]}\overset{\bullet\bullet}{W}\big\|_\infty\Big] < \infty$$

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 $\widehat{\theta}_n$ は一意性をもつ.
このとき, Θ の任意の内点 θ^* に対して

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) \rightsquigarrow \mathsf{N}\bigg(0,\, \frac{1}{\mathcal{F}_X(\theta^*)}\bigg) \qquad (n \to \infty)$$

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

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}(Z_n) \rightsquigarrow Y \qquad (n \to \infty)$$

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

Proof. $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}\big(|Z_n \mathbb{1}_{B_n^c}| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となる. したがって

$$Z_n \mathbb{1}_{B_n^c} \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty) \tag{12.9}$$

がわかる. さらに

$$\mathsf{Pr}\big(|\mathbb{1}_{B_n} - 1| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となるので

$$\mathbb{1}_{B_n} \overset{\mathrm{P}}{\to} 1 \qquad (n \to \infty) \tag{12.10}$$

がわかる. (12.9) と (12.10) に注意して Slusky の定理 (定理 4.14(3)(4)) を用いると

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

がわかる. □

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

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

となるようにできる. 事象 B_n を

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

と定める. $\widehat{\theta}_n$ の一致性から

$$\mathsf{Pr}_{\theta^*}(B_n) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.11}$$

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

$$\dot{\overline{W}}_n(\widehat{\theta}_n) = \frac{\mathrm{d}\overline{W}}{\mathrm{d}\theta}(\theta)\bigg|_{\theta=\widehat{\theta}_n} = 0$$

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

$$\dot{\overline{W}}_n(\theta^*) = \dot{\overline{W}}_n(\widehat{\theta}_n) + \ddot{\overline{W}}_n(\widetilde{\theta}_n)(\theta^* - \widehat{\theta}_n) \tag{12.12}$$

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

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) = \frac{\sqrt{n}\,\dot{\overline{W}}_n(\theta^*)}{-\ddot{\overline{W}}_n(\widetilde{\theta}_n)}$$

となる．仮定 (d) から $\dot{\overline{W}}_n(\theta^*)$ 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

$$\sqrt{n}\,\dot{\overline{W}}_n(\theta^*) \rightsquigarrow Z \sim \mathsf{N}\big(0, \mathcal{F}_X(\theta^*)\big) \qquad (n \to \infty) \tag{12.13}$$

となる．B_n 上では

$$|\widetilde{\theta}_n - \theta^*| \le |\widehat{\theta}_n - \theta^*|$$

であることと $\widehat{\theta}_n$ の一致性から

$$\mathsf{Pr}\big(|\widetilde{\theta}_n - \theta^*| < \epsilon\big) \ge \mathsf{Pr}\big(|\widehat{\theta}_n - \theta^*| < \epsilon\big) \overset{n\to\infty}{\longrightarrow} 1$$

となることがわかる．よって

$$\widetilde{\theta}_n \overset{\mathrm{P}}{\to} \theta^* \qquad (n \to \infty) \tag{12.14}$$

となる．定理 12.2 から

$$\big\|\mathbb{1}_{[\theta^*-\epsilon,\,\theta^*+\epsilon]}\big(\ddot{\overline{W}}_n - \mu\big)\big\|_\infty \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty) \tag{12.15}$$

となる．ただし, $\mu(\theta) = \mathsf{E}_{\theta^*}\big[\ddot{\overline{W}}_n(\theta)\big]\,(\theta \in \Theta)$ である．さらに, (12.14) と (12.15) に注意して, 定理 12.4(1) を用いると

$$\ddot{\overline{W}}_n(\widetilde{\theta}_n) \overset{\mathrm{P}}{\to} \mu(\theta^*) = -\mathcal{F}_X(\theta^*) \tag{12.16}$$

となる. B_n^c 上では $\widehat{\theta}_n$ の振る舞いは関係ない. 最後に, (12.11), (12.13) および (12.16 に注意して補題 **??** を用いると

$$\begin{aligned}\sqrt{n}\big(\widehat{\theta}_n-\theta^*\big) &= \sqrt{n}\big(\widehat{\theta}_n-\theta^*\big)\mathbb{1}_{B_n}+\sqrt{n}\big(\widehat{\theta}_n-\theta^*\big)\mathbb{1}_{B_n^c} \\ &= \frac{\sqrt{n}\ \overline{\dot{W}}_N\ (\theta^*)}{-\ \overline{\ddot{W}}_n\ (\widetilde{\theta}_n)}\mathbb{1}_{B_n}+\sqrt{n}\big(\widehat{\theta}_n-\theta^*\big)\mathbb{1}_{B_n^c} \\ &\rightsquigarrow \frac{Z}{\mathcal{F}_X(\theta^*)}\sim \mathsf{N}\left(0,\ \frac{1}{\mathcal{F}_X(\theta^*)}\right) \qquad (n\to\infty)\end{aligned}$$

がわかる. □

定理 12.11 (Camér の定理)**.** $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\cdot)$ とする. ただし, $\boldsymbol{\theta}^*\in\Theta\subset\mathbb{R}^d$ 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) $A:=\{x\in\mathbb{R};\ \mathsf{p}_{\boldsymbol{\theta}}(x)$ は $\boldsymbol{\theta}(\in\Theta)$ 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 $\Theta\ni\boldsymbol{\theta}\mapsto \mathsf{p}_{\boldsymbol{\theta}}(x)$ 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と仮定する.
(d) $\mathbb{R}$ 上の関数 K が存在して, $\mathsf{E}_{\boldsymbol{\theta}^*}\big[|K(X)|\big]<\infty$ であり

$$\dot{\boldsymbol{W}}\ (\boldsymbol{\theta}):=\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log \mathsf{p}_{\boldsymbol{\theta}}(X)\right)_{j,\,k=1,\,2,\ldots,\,d}; \qquad \boldsymbol{\theta}=(\theta_1,\,\theta_2,\,\ldots,\,\theta_d)^\top$$

のどの成分の絶対値も $\boldsymbol{\theta}^*$ の近傍において, 一様に K で上からおさえられる.
(e) $\boldsymbol{\mathcal{F}}_X(\boldsymbol{\theta}^*)=-\mathsf{E}_{\boldsymbol{\theta}^*}\big[\ddot{\boldsymbol{W}}\ (\boldsymbol{\theta}^*)\big]$ は正定値である.
(f) $\mathsf{p}_{\boldsymbol{\theta}}(\cdot)=\mathsf{p}_{\boldsymbol{\theta}^*}(\cdot) \Leftrightarrow \boldsymbol{\theta}=\boldsymbol{\theta}^*$.
このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ は

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^*\big)\rightsquigarrow \mathsf{N}\left(\mathbf{0}_d,\ \big\{\boldsymbol{\mathcal{F}}_X(\boldsymbol{\theta}^*)\big\}^{-1}\right) \qquad (n\to\infty)$$

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

12.4 尤度比検定統計量の漸近分布

統計的モデル $\mathcal{P}=\{\mathsf{p}_{\boldsymbol{\theta}}(\cdot);\ \boldsymbol{\theta}\in\Theta\subset\mathbb{R}^d\}$ を考える. ただし, $\mathsf{p}_{\boldsymbol{\theta}}(\cdot)$ は $\mathbb{R}^d$ 上の p.d.f. または p.m.f. である. いま

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\cdot) \qquad (\boldsymbol{\theta}^*\in\Theta)$$

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

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

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

$$\lambda_n(\boldsymbol{X}) := \frac{\sup_{\boldsymbol{\theta}\in\Theta_0} \prod_{j=1}^n \mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\sup_{\boldsymbol{\theta}\in\Theta} \prod_{j=1}^n \mathsf{p}_{\boldsymbol{\theta}}(X_j)} = \frac{\mathsf{lik}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)}{\mathsf{lik}_n\big(\widehat{\boldsymbol{\theta}}_n\big)} \tag{12.17}$$

を考える. ここで, $\widehat{\boldsymbol{\theta}}_n$ と $\widehat{\boldsymbol{\theta}}_{n,0}$ はそれぞれ Θ と Θ_0 における最尤推定量である. $\lambda_n(\boldsymbol{X})$ 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 $d - r\,(1 \le r < d)$ であると仮定する. さらに, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ と書いたとき, 帰無仮説 H_0 が

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

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 $\boldsymbol{\theta}^*$ が帰無仮説 H_0 を満足しているとする. このとき

$$-2\log\lambda_n(\boldsymbol{X}) \rightsquigarrow \chi_r^2 \qquad (n \to \infty)$$

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. (12.17) から

$$-2\log\lambda_n(\boldsymbol{X}) = 2\big\{\ell_n\big(\widehat{\boldsymbol{\theta}}_n\big) - \ell_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)\big\}$$

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

$$\begin{aligned}\ell_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big) = \ell_n\big(\widehat{\boldsymbol{\theta}}_n\big) &+ \dot{\ell}_n\,\big(\widehat{\boldsymbol{\theta}}_n\big)\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) \\ &- n\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big)^\top \widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\big(\widetilde{\boldsymbol{\theta}}_{n,0}\big)\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big)\end{aligned}$$

と書ける. ただし

$$\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\big(\widetilde{\boldsymbol{\theta}}_{n,0}\big) = -\frac{1}{n}\int_0^1\int_0^1 v\,\ddot{\ell}_n\,\big\{\widehat{\boldsymbol{\theta}}_n + uv\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big)\big\}\,\mathrm{d}u\,\mathrm{d}v$$

とした. さらに

$$\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\big(\widetilde{\boldsymbol{\theta}}_{n,0}\big) \overset{\text{a.s.}}{\to} \frac{1}{2}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*) \qquad (n \to \infty) \tag{12.18}$$

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

$$\begin{aligned}
-2\log\lambda_n(\boldsymbol{X}) &= 2n\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big)^\top \widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\big(\widetilde{\boldsymbol{\theta}}_{n,0}\big)\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) \\
&= n\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big)^\top \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) + o_P(1)
\end{aligned} \tag{12.19}$$

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 $A_n - B_n \overset{\text{P}}{\to} 0\,(n \to \infty)$ を意味する.

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

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}_d\left(\boldsymbol{0}_d,\, \big\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)\big\}^{-1}\right) \qquad (n \to \infty)$$

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, $\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{n,0}\big)$ の漸近分布を求めるために, $\dot{\ell}_n\,\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)$ を $\widehat{\boldsymbol{\theta}}_n$ のまわりで Taylor 展開する.

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\,\dot{\ell}_n\,\big(\widehat{\boldsymbol{\theta}}_{n,0}\big) \\
&\quad= \frac{1}{\sqrt{n}}\,\dot{\ell}_n\,\big(\widehat{\boldsymbol{\theta}}_{n,0}\big) + \frac{1}{n}\left(\int_0^1 \ddot{\ell}_n\,\big\{\widehat{\boldsymbol{\theta}}_n + v\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big\}\,\mathrm{d}v\right)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) \\
&\quad= -\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) + o_P(1)
\end{aligned}$$

となる. ただし, $\boldsymbol{\theta} = (\theta_1,\, \theta_2,\, \ldots,\, \theta_d)^\top$ と書いたとき

$$\begin{aligned}
\dot{\ell}_n\,(\boldsymbol{\theta}) &= \left(\frac{\partial \ell_n}{\partial \theta_1}(\boldsymbol{\theta}),\, \frac{\partial \ell_n}{\partial \theta_2}(\boldsymbol{\theta}),\, \ldots,\, \frac{\partial \ell_n}{\partial \theta_d}(\boldsymbol{\theta})\right)^\top, \\
\ddot{\ell}_n\,(\boldsymbol{\theta}) &= \left(\frac{\partial^2 \ell_n}{\partial \theta_j \partial \theta_k}(\boldsymbol{\theta})\right)_{j,\,k=1,\,2,\ldots,\,n}
\end{aligned}$$

と定めた. したがって

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\big) \rightsquigarrow -\big\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)\big\}^{-1}\,\dot{\ell}_n\,\big(\widehat{\boldsymbol{\theta}}_{n,0}\big) + o_P(1)$$

を得る. このことから

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

と表現できる. $\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)$ 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.

$$\begin{aligned}&\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)\\&\quad=\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right)+\frac{1}{n}\left(\int_0^1\overset{\bullet\bullet}{\ell}_n\left\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\right\}\mathrm{d}v\right)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\end{aligned}\tag{12.20}$$

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

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

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. $\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)$ の後半の $(d-r)$ 個の要素は 0 なので

$$\mathbf{H}\,\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)=\mathbf{0}_d\tag{12.21}$$

となる. よって, (12.20) と (12.21) から

$$\begin{aligned}\mathbf{H}\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right)&=\mathbf{H}\left\{\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)-\frac{1}{n}\left(\int_0^1\overset{\bullet\bullet}{\ell}_n\left\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\right\}\mathrm{d}v\right)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\right\}\\&=-\mathbf{H}\frac{1}{n}\left(\int_0^1\overset{\bullet\bullet}{\ell}_n\left\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\right\}\mathrm{d}v\right)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\\&=-\mathbf{H}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=-\begin{pmatrix}\mathbf{0}_{r\times r}&\mathbf{0}_{r\times(d-r)}\\\mathbf{0}_{r\times r}^\top&\mathbf{G}_3^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{G}_1&\mathbf{G}_2\\\mathbf{G}_2^\top&\mathbf{G}_3\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=-\begin{pmatrix}\mathbf{0}_{r\times r}&\mathbf{0}_{r\times(d-r)}\\\mathbf{G}_3^{-1}\mathbf{G}_2^\top&\mathbf{I}_{d-r}\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\end{aligned}\tag{12.22}$$

となる. 最後の等号は $\widehat{\boldsymbol{\theta}}_{n,0}$ と $\boldsymbol{\theta}^*$ の最初の r 個の成分は 0 であることからわかる. (12.22) を (12.21) に代入すると

$$\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right)=\left\{\mathbf{I}_d-\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\mathbf{H}\right\}\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right)+o_P(1)$$

を得る. 中心極限定理 (定理 4.24) から

$$\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right)\rightsquigarrow\mathsf{N}_d\Big(\mathbf{0}_d,\,\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\Big)\qquad(n\to\infty)$$

が成り立つので

$$\frac{1}{\sqrt{n}} \overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right) \rightsquigarrow \left(\boldsymbol{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathrm{H}\right)\boldsymbol{Y} \qquad (n \to \infty),$$

$$\boldsymbol{Y} \sim \mathsf{N}_d\left(\mathbf{0}_d, \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right)$$

である. これより

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

となる. ただし, $\boldsymbol{Z} \sim \mathsf{N}_d\left(\mathbf{0}_d, \mathbf{I}_d\right)$ である. ここで, 行列

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

は

$$\mathbf{P}^2 = \mathbf{P}; \quad \mathbf{P}^\top = \mathbf{P}; \quad \operatorname{rank}(\mathbf{P}) = \operatorname{tr}(\mathbf{P}) = \operatorname{tr}\left(\mathbf{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right) = r$$

をみたす. $\mathbf{P}$ はランク r の正射影となる. しがたって

$$-2\log\lambda(\boldsymbol{X}) \rightsquigarrow \boldsymbol{Z}^\top\mathbf{P}\boldsymbol{Z} \sim \chi_r^2 \qquad (n \to \infty)$$

を得る. □

(12.18) の証明 □

12.5 EM アルゴリズム: 不完全データに基づく最尤推定値の計算

EM アルゴリズムは不完全データから最尤推定値を求めるための再帰的繰り返し計算法である. この節では, 実際にはすべてを観測されない完全データ $\boldsymbol{X}$ は指数型分布族に含まれる分布から生成されたものと仮定して, 議論を進めていく. $\boldsymbol{X}$ は未観測であるが, その代わり $\boldsymbol{X}$ の関数である $\boldsymbol{Y}$ を観測したと仮定する. すなわち, 単射ではない関数 g があって, $\boldsymbol{Y} = g(\boldsymbol{X})$ と表現できると仮定する. 以下では, $\boldsymbol{X}$ は測度 μ に関する p.d.f.

$$\mathsf{p}(\boldsymbol{x}) = h(\boldsymbol{x})\exp\left\{\eta T(\boldsymbol{x}) - \kappa^\vee(\eta)\right\}$$

を持つとする．ここで, $\mathbb{X}$ と E をそれぞれ $\boldsymbol{X}$ の標本空間と母数空間とし, $T:\mathbb{X}\to\mathbb{R}$, $\kappa^{\vee}:E\to\mathbb{R}$ とする．

EM アルゴリズムは, 観測データが不完全なとき, 有効である．たとえば, $X_1, X_2, \ldots, X_n$ は i.i.d. ランダム標本とし, $Y_j\,(j=1, 2, \ldots, n)$ は X_j を小数第 1 位を四捨五入したもの等である．また, EM アルゴリズムは, 欠損データに対してもうまく機能することが知られている．たとえば, ある調査で設問が 2 つあり, おのおの設問では, いくつかの選択肢から回答を選ぶものとする．$\boldsymbol{X}$ は 2 つの設問に対する回答としたとき, Y は 1 つ目の質問をスキップして, 2 つ目だけ回答をしたデータ等である．

いま, $\mathbb{X}$, $\mathbb{Y}$ をそれぞれ $\boldsymbol{X}$ と $\boldsymbol{Y}$ の標本空間とする．$\forall\boldsymbol{y}\in\mathbb{Y}$ に対して, 切断面 $\mathbb{X}(\boldsymbol{y})$ を

$$\mathbb{X}(\boldsymbol{y})=\bigl\{\boldsymbol{x}\in\mathbb{X};\, g(\boldsymbol{x})=\boldsymbol{y}\bigr\}$$

で定める．このとき

$$\boldsymbol{Y}=\boldsymbol{y}\ \Leftrightarrow\ \boldsymbol{X}\in\mathbb{X}(\boldsymbol{y})$$

である．

命題 12.13. μ を標本空間 $\mathbb{X}$ 上の測度とし, ν を誘導された標本空間 $\mathbb{Y}$ 上の計数測度とする．X は測度 μ に関する p.d.f.

$$\mathsf{p}(x)=h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}$$

をもち, $Y=g(X)$ は離散確率変数とする．ただし, $T:\mathbb{X}\to\mathbb{R}$, $h:\mathbb{X}\to\mathbb{R}$, $\kappa:E\to\mathbb{R}$ である．このとき, $\mu\otimes\nu$ に関する (X, Y) の同時 p.d.f. は

$$\mathsf{p}^{(X,Y)}(x, y)=\mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}$$

で与えられる．

Proof. $f:\mathbb{X}\times\mathbb{Y}\to[0, \infty)$ を任意の非負値関数とする．このとき

$$f(X, Y)=\sum_{y\in\mathbb{Y}}f(X, y)\mathbb{1}\{Y=y\}$$

と書き直す．期待値作用素の線型性と $\mathbb{1}\{g(x)=y\}=\mathbb{1}_{\mathbb{X}(y)}(x)$ であることから

$$\begin{aligned}
\mathsf{E}\bigl[f(X, Y)\bigr]&=\sum_{y\in\mathbb{Y}}\mathsf{E}\bigl[f(X, y)\mathbb{1}\{g(X)=y\}\bigr]\\
&=\sum_{y\in\mathbb{Y}}\int_{\mathbb{X}}f(x, y)\mathbb{1}\{g(x)=y\}h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)\\
&=\sum_{y\in\mathbb{Y}}\int_{\mathbb{X}}f(x, y)\mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\bigl\{\eta T(x)-\kappa^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)
\end{aligned}$$

となる. このことから命題の主張は示せた. $\square$

アルゴリズムを導入するために, 完全データに基づく最尤推定量は $\psi(T)$ であることに注意する. ただし, ψ は $\overset{\bullet}{\kappa}{}^{\vee} = \dfrac{\mathrm{d}\kappa^{\vee}}{\mathrm{d}\eta}$ の逆関数である. さらに

$$e(y,\, \eta) := \mathsf{E}_{\eta}\bigl[T(T) \,|\, Y = y\bigr]$$

とおく. これは $Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|Y}(\,\cdot\,|\, y)$ に関する積分で計算できる. 命題 12.13 から

$$\begin{aligned}
\mathsf{p}^{X|Y}(x|\, y) &= \frac{\mathsf{p}^{(X,\,Y)}(x,\, y)}{\mathsf{p}^{Y}(y|\, \eta)} \\
&= \frac{\mathbb{1}_{\mathbb{X}(y)}(x) h(x) \exp\bigl\{\eta T(x) - \kappa^{\vee}(\eta)\bigr\}}{\mathsf{p}^{Y}(y|\, \eta)}
\end{aligned}$$

である. ただし

$$\begin{aligned}
\mathsf{p}^{Y}(y|\, \eta) &= \mathsf{Pr}(Y = y) = \mathsf{Pr}\bigl(X \in \mathbb{X}(y)\bigr) \\
&= \int_{\mathbb{X}(y)} h(x) \exp\bigl\{\eta T(x) - \kappa^{\vee}(\eta)\bigr\}\, \mathrm{d}\mu(x) \qquad (12.23)
\end{aligned}$$

である.

アルゴリズムは, 最尤推定値 $\widehat{\eta}$ に対する初期値 $\widehat{\eta}_0$ から始める. この初期値とデータ Y を用いて, $T_1(X)$ の値を

$$T_1 = e(Y,\, \widehat{\eta}_0)$$

と定める. これを **E-Step** と呼ぶ. 更新された値を $\widehat{\eta}_1 = \psi(T_1)$ で定める. これを M-**Step** という. E-Step と M-Step を交互に繰り返し行き, 更新される値が収束するまでこれらの 2 つの Steps を交互に繰り返す.

指数型分布族が正準形ではなく

$$h(x) \exp\bigl\{\eta(\theta) T(x) - \kappa(\theta)\bigr\} \qquad \bigl(\theta \in \Theta = \overset{\bullet}{\kappa}{}^{\vee}(E)\bigr)$$

で表現されるとする. すると E-Step は正準形のときと同じで,

$$T_{k+1} = \mathsf{E}_{\widehat{\theta}_k}\bigl[T(X) \,|\, Y = y\bigr]$$

となり, M-Step は関数

$$\Theta \ni \theta \;\mapsto\; \eta(\theta) T_{k+1} - \kappa(\theta)$$

を最大にする値 $\widehat{\theta}_{k+1}\, (k = 1,\, 2,\, \ldots)$ である.

EM アルゴリズムの収束を調べる. そのために, EM アルゴリズムが $\widetilde{\eta}$ に収束したとき, $\widetilde{\eta}$ は関係式

$$\widetilde{\eta} = \psi\bigl(e(Y, \widetilde{\eta})\bigr) \tag{12.24}$$

をみたす. なぜならば, (12.23) から, $\mathsf{p}^Y(y|\,\eta) > 0$ なる y に対して,

$$\begin{aligned}
\frac{\partial}{\partial\eta}\log \mathsf{p}^Y(y|\,\eta) &= \frac{\dfrac{\partial}{\partial\eta}\displaystyle\int_{\mathbb{X}(y)} h(x)\exp\bigl\{\eta T(x) - \kappa^\vee(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(y)} \frac{\partial}{\partial\eta} h(x)\exp\bigl\{\eta T(x) - \kappa^\vee(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(y)} \bigl\{T(x) - \overset{\bullet}{\kappa}{}^\vee(\eta)\bigr\} h(x)\exp\bigl\{\eta T(x) - \kappa^\vee(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \int_{\mathbb{X}(y)} T(x)\,\underbrace{\frac{h(x)\exp\bigl\{\eta T(x) - \kappa^\vee(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)}}_{=\,\mathsf{p}^{X|\,Y}(x|\,y)} \\
&\qquad - \overset{\bullet}{\kappa}{}^\vee(\eta)\frac{\displaystyle\int_{\mathbb{X}(y)} h(x)\exp\bigl\{\eta T(x) - \kappa^\vee(\eta)\bigr\}\,\mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \mathsf{E}_\eta\bigl[T(X)|\,Y = y\bigr] - \overset{\bullet}{\kappa}{}^\vee(\eta) \\
&= e(y,\,\eta) - \overset{\bullet}{\kappa}{}^\vee(\eta)
\end{aligned}$$

となることからわかる.

さらに, ψ は $\overset{\bullet}{\kappa}{}^\vee$ の逆写像であることに注意すると (12.24) は

$$\overset{\bullet}{\kappa}{}^\vee(\widetilde{\eta}) = e\bigl(Y,\,\widetilde{\eta}\bigr) \tag{12.25}$$

と同値になる.

例 12.14. $X_1,\,X_2,\,\ldots,\,Xn \overset{\text{i.i.d.}}{\sim} \mathsf{p}(\,\cdot\,|\,\eta)$ とする. ただし

$$\mathsf{p}_\eta(x) = \begin{cases} \eta e^{-\eta x} & (x > 0) \\ 0 & (x \le 0) \end{cases}$$

である. いま

$$Y_j = \lfloor X_j \rfloor \qquad (j = 1,\,2,\,\ldots,\,n)$$

とする. すなわち, Y_j は X_j を越えない最大の整数である. $\boldsymbol{X} = (X_1,\,X_2,\,\ldots,\,X_n)^\top$ の同時 p.d.f. は正準母数 η と十分統計量 $T := T(\boldsymbol{X}) = -(X_1 + X_2 +$

$\cdots + X_n)$ の指数型分布族を作る．$\kappa^{\vee}(\eta) = -n\log\eta$ であることに注意すると，$\psi(x) = -\dfrac{n}{x}$ となる．よって，$\boldsymbol{X}$ に基づく η の最尤推定量は $\psi(T) = -\dfrac{n}{T}$ である．$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^{\top}$ と $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\top}$ とおく．命題 12.13 から $j = 1, 2, \ldots, n$ に対して

$$
\begin{aligned}
\mathsf{E}_{\eta}\big[X_j \big| \boldsymbol{Y} = \boldsymbol{y}\big] &= \mathsf{E}_{\eta}\big[X_j \big| y_j \le X_j < y_j + 1\big] \\
&= \frac{\displaystyle\int_{y_j}^{y_j+1} x\eta e^{-\eta x}\,\mathrm{d}x}{\displaystyle\int_{y_j}^{y_j+1} \eta e^{-\eta x}\,\mathrm{d}x} \\
&= \frac{\Big[-xe^{-\eta x}\Big]_{y_j}^{y_j+1} + \displaystyle\int_{y_j}^{y_j+1} \eta e^{-\eta x}\,\mathrm{d}x}{\Big[-e^{-\eta x}\Big]_{y_j}^{y_j+1}} \\
&= \frac{e^{-\eta y_j}\big\{y_j - (1+y_j)e^{-\eta} - \dfrac{e^{-\eta y_j}}{\eta}\big\{e^{-\eta} - 1\big\}}{e^{-\eta y_j}\big\{1 - e^{-\eta}\big\}} \\
&= \frac{y_j\big\{1 - e^{-\eta}\big\} - e^{-\eta} - \dfrac{e^{-\eta} - 1}{\eta}\big\{e^{-\eta} - 1\big\}}{1 - e^{-\eta}} \\
&= y_j + \frac{1 - e^{-\eta} - \eta e^{-\eta}}{\eta\big\{1 - e^{-\eta}\big\}} \\
&= y_j + \frac{e^{\eta} - 1 - \eta}{\eta\big\{e^{\eta} - 1\big\}}
\end{aligned}
$$

となる．さらに，独立性から

$$
\mathsf{E}_{\eta}\big[X_j \big| \boldsymbol{Y} = \boldsymbol{y}\big] = \mathsf{E}_{\eta}\big[X_j \big| Y_j = y_j\big]
$$

となるので

$$
\begin{aligned}
e(\boldsymbol{y}, \eta) &= \mathsf{E}_{\eta}\big[T(\boldsymbol{X}) \big| \boldsymbol{Y} = \boldsymbol{y}\big] \\
&= -\sum_{j=1}^{n} \mathsf{E}_{\eta}\big[X_j \big| Y_j = y_j\big] \\
&= -\sum_{j=1}^{n} \left\{ y_j + \frac{e^{\eta} - 1 - \eta}{\eta\big\{e^{\eta} - 1\big\}} \right\} \\
&= -n\left\{ \overline{y}_n + \frac{e^{\eta} - 1 - \eta}{\eta\big\{e^{\eta} - 1\big\}} \right\}
\end{aligned}
$$

となる．ただし，$\overline{y}_n = \dfrac{1}{n}\sum_{j=1}^n y_j$ である．よって，k-Step($k = 1, 2, \ldots$) の推定値を $\widehat{\eta}_k$ と書いたとき，EM アルゴリズムは

$$\widehat{\eta}_{k+1} = -\frac{n}{T_{k+1}} \quad \text{かつ} \quad T_{k+1} = -n\left\{\overline{y}_n + \frac{e^{\widehat{\eta}_k} - 1 - \widehat{\eta}_k}{\widehat{\eta}_k\left\{e^{\widehat{\eta}_k} - 1\right\}}\right\} \tag{12.26}$$

である． □

注意 12.15. 例 12.14 では，Y_j の p.m.f. を陽に計算できる．実際，$y = 0, 1, 2, \ldots$ に対して

$$\begin{aligned}
\mathsf{Pr}_\eta\big(Y_j = y\big) &= \mathsf{Pr}_\eta\big(y \le X_j < y_j + 1\big) \\
&= \int_y^{y+1} \eta e^{-\eta x}\,\mathrm{d}x \\
&= \Big[-e^{-etax}\Big]_y^{y+1} = -e^{-\eta(y+1)} + e^{-\eta y} \\
&= \big(1 - e^{-\eta}\big)\big(e^{-\eta}\big)^y
\end{aligned}$$

となる．したがって，$Y_1, Y_2, \ldots, Y_n$ は母数 $\pi := 1 - e^{-\eta}$ の幾何分布からのランダム標本とみなせる．よって，π の最尤推定量は

$$\widehat{\pi} = \frac{1}{1 + \overline{Y}_n}; \qquad \overline{Y}_n = \frac{1}{n}\sum_{j=1}^n Y_j$$

で与えられる．$\eta = -\log(1 - \pi)$ なので，η の最尤推定量は

$$\widehat{\eta} = -\log\big(1 - \widehat{\pi}\big) = \log\left(1 + \frac{1}{\overline{Y}_n}\right)$$

である．

EM アルゴリズムの繰り返し $\{\widehat{\eta}_k\}_{k=1}^\infty$ の $\widehat{\eta}$ への収束のスピードを調べてみよう．そのために，$\widehat{\eta}_k = \widehat{\eta} + \epsilon$ と書く．すると $\widehat{\eta}$ は繰り返し計算の固定点なので

$$-\frac{n}{\widehat{\eta}} = -n\left\{\overline{y}_n + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}\right\}$$

をみたしていること[3]に注意する. ここで, 次の Taylor 展開を考える.

$$\frac{1}{(\widehat{\eta}+\epsilon)+1}=\frac{1}{\widehat{\eta}}-\frac{\epsilon}{\widehat{\eta}^2}+O(\epsilon^2),$$

$$\frac{1}{e^{\widehat{\eta}+\epsilon}-1}=\frac{1}{e^{\widehat{\eta}}-1}-\frac{e^{\widehat{\eta}}\epsilon}{\left(e^{\widehat{\eta}}-1\right)^2}+O(\epsilon^2).$$

これらより

$$\begin{aligned}
T_{k+1}&=-n\left\{\overline{y}_n+\frac{e^{\widehat{\eta}_k}-1-\widehat{\eta}_k}{\widehat{\eta}_k\left(e^{\widehat{\eta}_k}-1\right)}\right\}\\
&=-n\left\{\overline{y}_n+\frac{1}{\widehat{\eta}_k}-\frac{1}{e^{\widehat{\eta}_k}-1}\right\}\\
&=-n\left\{\overline{y}_n+\frac{1}{\widehat{\eta}_k}-\frac{1}{\widehat{\eta}_k\left(e^{\widehat{\eta}_k}-1\right)}\right\}\\
&=-n\left\{\underbrace{\overline{y}_n+\frac{1}{\widehat{\eta}}-\frac{1}{e^{\widehat{\eta}}-1}}_{=\frac{1}{\widehat{\eta}}}-\frac{\epsilon}{\widehat{\eta}^2}+\frac{\epsilon e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}+O(\epsilon^2)\right\}\\
&=-\frac{n}{\widehat{\eta}}\left\{1-\left(\frac{\epsilon}{\widehat{\eta}}-\frac{\epsilon\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right)+O(\epsilon^2)\right\}
\end{aligned}$$

を得る. この式から

$$\begin{aligned}
\widehat{\eta}_{k+1}=-\frac{n}{T_{k+1}}&=\frac{\widehat{\eta}}{1-\epsilon\left(\frac{1}{\widehat{\eta}}-\frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right)+O(\epsilon^2)}\\
&=\widehat{\eta}\left\{1+\epsilon\left(\frac{1}{\widehat{\eta}}-\frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right)+O(\epsilon^2)\right\}\\
&=\widehat{\eta}+\epsilon\left\{1-\frac{\widehat{\eta}^2e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right\}+O(\epsilon^2)
\end{aligned}$$

を得る. 特に, $\widehat{\eta}_k=\widehat{\eta}$ のとき, $\epsilon=0$ なので, $\widehat{\eta}_{k+1}=\widehat{\eta}$ となる. したがって, $\widehat{\eta}$ は繰り返しの固定点であることがわかる. □

[3] $\lim_{k\to\infty}T_k=T_\infty$ とおくと $\widehat{\eta}=-\frac{n}{T_\infty}$ となる. したがって, (12.26) から

$$\begin{aligned}
-\frac{n}{\widehat{\eta}}=\lim_{k\to\infty}T_k&=-n\left\{\overline{y}_n+\frac{\exp\left\{\lim_{k\to\infty}\widehat{\eta}_k\right\}-1-\lim_{k\to\infty}\widehat{\eta}_k}{\lim_{k\to\infty}\widehat{\eta}_k\left\{\exp\left\{\lim_{k\to\infty}\widehat{\eta}_k\right\}-1\right\}}\right\}=-n\left\{\overline{y}_n+\frac{e^{\widehat{\eta}}-1-\widehat{\eta}}{\widehat{\eta}\left\{e^{\widehat{\eta}}-1\right\}}\right\}\\
&=-n\left\{\overline{y}_n+\frac{1}{\widehat{\eta}}-\frac{1}{e^{\widehat{\eta}}-1}\right\}
\end{aligned}$$

からわかる.

注意 12.16. 一般に EM アルゴリズムは安定しており, 信頼される. このアルゴリズムの良い性質は, 繰り返しをするごとに尤度関数を増加させることである. すなわち, EM アルゴリズムの $k - Step$ の推定値を $\widehat{\eta}_k\,(k = 1, 2, \ldots)$ としたとき

$$\prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_{k+1}) \geq \prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_k)$$

となることである. これは, EM アルゴリズムが MM アルゴリズムの特別な場合であることからわかる. この点については Lange (2014) を参照のこと. しかし, 収束は一般に保証されない. たとえば, 尤度関数のモードが複数ある場合には, EM アルゴリズムは収束しないことがある. EM アルゴリズムが収束するための十分条件は Wu (1983) を参照のこと. EM アルゴリズムは安定しているが, 収束のスピードは遅くことが知られている. □

12.6 章末注釈と参考文献

12.7 演習問題

演習問題 12.1.

演習問題 12.2.