·····3.1·····

指数分布族モデルにおける最尤推定量の漸近理論

P_η を $(\mathcal{X}, \mathcal{A})$ 上の確率測度とし，σ - 有限な測度 $\lambda(x)$ に関する確率密度関数

$$p(x|\,\boldsymbol{\eta}) = \exp\{\boldsymbol{\eta}'\boldsymbol{T}(x) - A(\boldsymbol{\eta})\}$$

を持つ[1)]とする．ここでは，考える指数分布族はフルランクとし，$A(\boldsymbol{\eta})$ は 2 階連続微分可能とする．

系 2.1(i) から，スコア関数は

$$\frac{\partial}{\partial\boldsymbol{\eta}}\log p(x|\,\boldsymbol{\eta}) = \boldsymbol{T}(x) - \frac{\partial}{\partial\boldsymbol{\eta}}A(\boldsymbol{\eta}) = \boldsymbol{T}(x) - \dot{A}(\boldsymbol{\eta}) = \boldsymbol{T}(x) - \mathbb{E}_\eta[\boldsymbol{T}(X)]$$

となる．さらに，系 2.1(ii) から，スコア関数の分散は

$$I(\boldsymbol{\eta}) = \mathbb{VAR}_\eta\left\{\frac{\partial}{\partial\boldsymbol{\eta}}\log p(X|\,\boldsymbol{\eta})\right\} = \frac{\partial^2}{\partial\boldsymbol{\eta}'\partial\boldsymbol{\eta}}A(\boldsymbol{\eta}) = \ddot{A}(\boldsymbol{\eta})$$

で与えられ，これを情報行列と呼ぶことにする．

$X_1, X_2, \ldots, X_n$ を確率測度 P_{η_0} からのランダム標本とする．ただし，$\boldsymbol{\eta}_0$ は自然母数空間 $\mathcal{E}$ の内点に含まれるとする．このとき，$\mathbb{R}^n$ 上の測度 $\Pi_{i=1}^n\lambda(x_i)$ に関する $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta}_0) = \Pi_{i=1}^n p(x_i|\,\boldsymbol{\eta}_0) = \exp\Big\{\boldsymbol{\eta}_0'\sum_{i=1}^n\boldsymbol{T}(x_i) - nA(\boldsymbol{\eta}_0)\Big\}, \qquad \boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$$

となる．ここで，$A_n(\boldsymbol{\eta}) = nA(\boldsymbol{\eta}),\ n\bar{\boldsymbol{T}}_n(\boldsymbol{x}_n) = \sum_{i=1}^n\boldsymbol{T}(x_i)$ とおくと別の指数分布族ができる．$\widehat{\boldsymbol{\eta}}_n$ を $\boldsymbol{\eta}$ の最尤推定値とする．すなわち，$p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta})$ を最大にする $\boldsymbol{\eta}$ の値である．これは，関数

$$G_n(\boldsymbol{\eta}) = \frac{1}{n}\log\frac{p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta}_0)}{p_n(\boldsymbol{x}_n|\,\boldsymbol{\eta})} = (\boldsymbol{\eta}_0 - \boldsymbol{\eta})'\bar{\boldsymbol{T}}_n + A(\boldsymbol{\eta}) - A(\boldsymbol{\eta}_0)$$

を最小にするものと同値である：

$$\widehat{\boldsymbol{\eta}}_n = \mathrm{argmin}G_n(\boldsymbol{\eta})$$

である．

以下では，$\widehat{\boldsymbol{\eta}}_n$ の漸近一致性と漸近分布を求めていく．最尤推定量の漸近分布を求める基本的な方針はつぎのふたつがある．

(1) 関数 $G_n(\boldsymbol{\eta})$ が滑らかであれば，最尤推定量は尤度方程式の解として考察する．

---

1) (2.1) において $\lambda(x) = S(x)I_B(x)\mu(x)$ とすればよい．

(2)　最尤推定量を関数 $G_n(\boldsymbol{\eta})$ を最小にするものとして考察する．

ここでは，(2) の方針に従って議論を進めていく．

3.1.1 漸近一致性

まず，$\widehat{\boldsymbol{\eta}}_n$ において $G_n(\boldsymbol{\eta})$ が最小になることから

$$0 = G_n(\boldsymbol{\eta}_0) \geq G_n(\widehat{\boldsymbol{\eta}}_n)$$

となることに注目する．さらに，大数の法則から

$$\bar{\boldsymbol{T}}_n \xrightarrow{a.s} \mathbb{E}_{\eta_0}[\boldsymbol{T}(X)] = \dot{A}(\boldsymbol{\eta}_0) \stackrel{\text{def}}{=} \boldsymbol{\beta}_0$$

となるので，

$$G_n(\boldsymbol{\eta}) \xrightarrow{a.s} G(\boldsymbol{\eta}) \stackrel{\text{def}}{=} (\boldsymbol{\eta}_0 - \boldsymbol{\eta})'\boldsymbol{\beta}_0 + A(\boldsymbol{\eta}) - A(\boldsymbol{\eta}_0)$$

となることがわかる．この収束は，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| < \kappa, \kappa$ は正定数 $\}$ 上で一様である：

$$\sup_{|\eta-\eta_0|<\kappa} |G_n(\boldsymbol{\eta}) - G(\boldsymbol{\eta})| \leq \kappa |\bar{T}_n - \boldsymbol{\beta}_0| \xrightarrow{a.s} 0 \tag{3.22}$$

直観的に言えば，$G_n(\boldsymbol{\eta})$ は $G(\boldsymbol{\eta})$ に近いので，$G_n(\boldsymbol{\eta})$ が最小となるような点は $G(\boldsymbol{\eta})$ が最小となる点に近いと期待できる．したがって，$\widehat{\boldsymbol{\eta}}$ がどこに近づくかを観るために，$G(\boldsymbol{\eta})$ を最小とする点を求める．$G(\boldsymbol{\eta})$ は凸関数で，その導関数 $-\boldsymbol{\beta}_0 + \dot{A}(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ でゼロになる．また，2 次関数 $I(\boldsymbol{\eta}_0) = \ddot{A}(\boldsymbol{\eta}_0) > 0$ なので，$G(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ で唯一最小となる．特に，各 $\delta > 0$ に対し，正数 ϵ を

$$2\epsilon = \inf_{|\eta-\eta_0|\geq\delta} G(\boldsymbol{\eta})$$

で定める．任意の $\delta > 0$ に対し，$\widehat{\boldsymbol{\eta}}_n$ が $\boldsymbol{\eta}_0$ の δ - 近傍に高い確率で含まれることを示すためには，高い確率で

$$\inf_{|\eta-\eta_0|\geq\delta} G_n(\boldsymbol{\eta}) \geq \epsilon$$

となることを示せばよい：なぜならば，

$$\mathbb{P}_{\eta_0}\{|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| \leq \delta\} \geq \mathbb{P}_{\eta_0}\{\inf_{|\eta-\eta_0|\geq\delta} G_n(\boldsymbol{\eta}) \geq \epsilon\} \tag{3.23}$$

からわかる．したがって，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \geq \delta\}$ 上で $\boldsymbol{\eta}$ に関して $G_n(\boldsymbol{\eta})$ の最小化を考えることになる．しかし，$G_n(\boldsymbol{\eta})$ は凸関数なので，$K = \{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \delta\}$ 上のみで $G_n(\boldsymbol{\eta})$ の最小化を考えればよいことがわかる．有界集合 K 上では，(3.22) から高い確率で

$$|\inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta})| \leq \epsilon$$

となる．これから，高い確率で

$$\inf_{\eta\in K} G_n(\boldsymbol{\eta}) = \inf_{\eta\in K} G(\boldsymbol{\eta}) + \inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta}) \geq \inf_{\eta\in K} G(\boldsymbol{\eta}) - |\inf_{\eta\in K} G_n(\boldsymbol{\eta}) - \inf_{\eta\in K} G(\boldsymbol{\eta})| \geq \epsilon$$

とできる．よって，(3.23) から $\widehat{\boldsymbol{\eta}}_n$ の漸近一致性が示せた．

3.1.2 収束の速さ

中心極限定理より

$$\sqrt{n}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) \xrightarrow{L} N(0,\, I(\boldsymbol{\eta}_0))$$

となる．ただし，$I(\boldsymbol{\eta}_0) = \mathbb{VAR}_{\eta_0}(\boldsymbol{T}(X)) = \ddot{A}(\boldsymbol{\eta}_0)$ である．これより

$$M_n = |\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0| = O_p(1/\sqrt{n})$$

となる．$\bar{\boldsymbol{T}}_n$ の収束の早さから

$$|G_n(\boldsymbol{\eta}_0) - G(\boldsymbol{\eta})| \le M_n|\boldsymbol{\eta}_0 - \boldsymbol{\eta}| \tag{3.24}$$

を得る．次に，$G(\boldsymbol{\eta})$ を $\boldsymbol{\eta}_0$ のまわりでテーラー展開する：

$$\begin{aligned} G(\boldsymbol{\eta}) &= G(\boldsymbol{\eta}_0) + (\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\dot{G}(\boldsymbol{\eta}_0) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'\ddot{G}(\boldsymbol{\eta}^*)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \\ &= \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o(|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2) \end{aligned}$$

となる．ただし，$\boldsymbol{\eta}^*$ は $\boldsymbol{\eta}$ と $\boldsymbol{\eta}_0$ を直線で結んだ線分上の点である．$I(\boldsymbol{\eta}_0)$ は正定値なので，ある正定数 C が存在して

$$\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \ge 2C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2$$

とできる．$\delta > 0$ を十分小さくとって，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \delta\}$ 上では

$$G(\boldsymbol{\eta}) \ge C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2 \tag{3.25}$$

とできる．さらに，n が十分大きいとき，高い確率で $M_n \le \delta$ となるので，$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| = 2M_n/C\}$ 上では，(3.24) と (3.25) より

$$G_n(\boldsymbol{\eta}) \ge G(\boldsymbol{\eta}) - M_n|\boldsymbol{\eta} - \boldsymbol{\eta}_0| \ge C|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2 - M_n|\boldsymbol{\eta} - \boldsymbol{\eta}_0| = \frac{2M_n^2}{C}$$

となる．これより，高い確率で

$$\inf_{|\eta - \eta_0| \ge M_n/C} |G_n(\boldsymbol{\eta})| \ge \frac{2M_n^2}{C}$$

となる．したがって，一致性の議論と同様にすれば，高い確率で $\widehat{\boldsymbol{\eta}}_n$ は $\boldsymbol{\eta}_0$ の M_n/C - 近傍に入ることがわかる．よって

$$|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| \le \frac{2M_n}{C} = O_p(1/\sqrt{n})$$

となる．上の式は $M_n \le \delta$ の仮定のもとの議論だったので，もうすこし吟味が必要である．十分大きな正数 K に対し，$D_n = \{\sqrt{n}M_n \le K\}$, $E_n = \{M_n \le \delta\}$, $F_n = \{\sqrt{n}|\widehat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}| \le 2K/C\}$ とおく．上の議論で $D_n \cap E_n$ が起こったならば，F_n が起こることが言えた．すなわち，

$$\mathbb{P}(D_n \cap E_n) \le \mathbb{P}(F_n)$$

である．また，$n \to \infty$ のとき，$\mathbb{P}(D_n) \to 0$ と $\mathbb{P}(E_n) \to 0$ に注意すれば

$$\mathbb{P}(D_n \cap E_n) \geq 1 - \mathbb{P}(D_n^c) - \mathbb{P}(E_n^c) = \mathbb{P}(D_n) + \mathbb{P}(E_n) + 1 \to 1$$

となるので，$\mathbb{P}(F_n) \to 1\,(n \to \infty)$ がわかる．

3.1.3 漸近正規性

$\delta_n = O(1/\sqrt{n})$ とおく．$\{\boldsymbol{\eta} : |\boldsymbol{\eta} - \boldsymbol{\eta}_0| \leq \delta_n\}$ 上で

$$\begin{aligned} G_n(\boldsymbol{\eta}) &= -(\boldsymbol{\eta} - \boldsymbol{\eta}_0)'(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0)' I(\boldsymbol{\eta}_0)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o(1/n) \\ &= \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) - I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \end{aligned}$$

となる．これは，$o(|\boldsymbol{\eta} - \boldsymbol{\eta}_0|^2) = o(O(1/n)) = o(1/n)$ に注意すればわかる．ただし，$I_0^{1/2}$ は正定値行列で $I(\boldsymbol{\eta}_0) = I_0^{1/2} I_0^{1/2}$ をみたすものである．

いま，

$$Q_n(\boldsymbol{\eta}) = \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) - I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2$$

とおく．$Q_n(\boldsymbol{\eta})$ は $\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}_n = \boldsymbol{\eta}_0 + I^{-1}(\boldsymbol{\eta}_0)(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)$ で最小となる．中心極限定理より

$$\sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = I^{-1}(\boldsymbol{\eta}_0)\sqrt{n}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0},\, I^{-1}(\boldsymbol{\eta}_0))$$

となる．よって

$$|\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0| = O_p(1/\sqrt{n})$$

となる．したがって

$$\begin{aligned} G_n(\boldsymbol{\eta}) &= \frac{1}{2}|I_0^{1/2}\boldsymbol{\eta} - I_0^{1/2}(\boldsymbol{\eta}_0 + I(\boldsymbol{\eta}_0)^{-1}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0))|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \\ &= \frac{1}{2}|I_0^{1/2}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})|^2 - \frac{1}{2}|I_0^{-1/2}(\bar{\boldsymbol{T}}_n - \boldsymbol{\beta}_0)|^2 + o(1/n) \end{aligned}$$

となる．$|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}| = O_p(1/\sqrt{n})$ から $G_n(\boldsymbol{\eta})$ に $\hat{\boldsymbol{\eta}}_n$ と $\tilde{\boldsymbol{\eta}}$ を代入すれば，上の式の最右辺の $o(1/n)$ の項は $o_P(1/n)$ になる．$\hat{\boldsymbol{\eta}}_n$ の定義から $G_n(\hat{\boldsymbol{\eta}}_n) \leq G_n(\tilde{\boldsymbol{\eta}}_n)$ となるので，

$$\frac{1}{2}|I_0^{1/2}(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n)|^2 \leq o_P(1/n)$$

を得る．よって，

$$|\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n| = o_P(1/\sqrt{n})$$

となり，Slutzky の補題を用いれば，

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = \sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + o_P(1) \xrightarrow{L} N(\mathbf{0},\, I^{-1}(\boldsymbol{\eta}_0))$$

を得る．

·····3.2·····

一様強一致性

$X_1, X_2, \ldots, X_n$ を独立な確率変数列とし, 同一の分布 $F(x)$ をもつとする．Θ を母数空間とし, $U(x,\theta)$ を x の実数値関数とする．推定や検定において,

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta)$$

の評価が重要である．θ にたいし,

$$\mu(\theta) = \boldsymbol{E}U(X_1,\theta) = \int U(x,\theta)dF(x) < \infty$$

とく．大数の法則から, 各 θ に対し,

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta) \xrightarrow{as} \mu(\theta) \qquad (n \to \infty) \tag{3.26}$$

が成立する．しかし,

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta) - \mu(\theta)\right| \xrightarrow{as} 0 \qquad (n \to \infty) \tag{3.27}$$

が成立すると便利である．

(3.27) が成立すると仮定し, 統計量の列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ が $\hat{\theta}_n \xrightarrow{as} \theta_0$ を満足するとする． さらに, $\mu(\theta)$ は θ の連続関数であるとする．すると, (3.27) は

$$\frac{1}{n}\sum_{i=1}^{n} U(X_i,\hat{\theta}_n) \xrightarrow{as} \mu(\theta_0) \qquad (n \to \infty)$$

を保障することがわかる．実際,

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i,\hat{\theta}_n) - \mu(\theta_0)\right| &\le \left|\frac{1}{n}\sum_{i=1}^{n} U(X_i,\hat{\theta}_n) - \mu(\hat{\theta}_n)\right| + \left|\mu(\hat{\theta}_n) - \mu(\theta_0)\right| \\
&\le \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} U(X_i,\theta) - \mu(\theta)\right| + \left|\mu(\hat{\theta}_n) - \mu(\theta_0)\right| \\
&\xrightarrow{as} 0, \qquad (n \to \infty)
\end{aligned}$$

からわかる．

以下, (3.27) が成立するための十分条件を求めることにする．