数理解析研究所講究録
第197巻 1974年 26-32

nuisance parameter をもつ最尤推定の empirical Bayes 的接近.

九大 理　大隈彰道

§1. 序.

最尤推定の漸近的性質については種々の結果が得られているが,ここではパラメーターの一部について,その先験分布が仮定される場合の最尤推定量の漸近的正規性を考える。即ち,確率変数 X_k $k=1,2,\cdots$ が密度関数 $f_k(x|\theta,\alpha)$ をもち, α の先験分布が G であるとする。そのときの n 個の観測値 $x_1,\cdots x_n$ に基づく θ の最尤推定量 $\tilde{\theta}_n$ の漸近正規性について考える。もちろん $\tilde{\theta}_n$ は先験分布 G に depend する。

§3. では,各 α に対し $f_k(x|\theta,\alpha)$ が Hoadley (AMS vol 42) による漸近正規性の為の条件を満足すれば, G による mixture $f_{G,k}(x|\theta)=\int f_k(x|\theta,\alpha)\,dG(\alpha)$ がその条件を満足する事を示す。

又,先験分布 G が unknown である場合には $\tilde{\theta}_n$ は実際の推定量ではない(即ち, $\tilde{\theta}_n$ は unknown G を含む)。そこで $\tilde{\theta}_n$ にかわる推定量 $\hat{\theta}_n$ を提案する。

§4. で $f_k(x|\theta,\alpha)$ が平均ベクトル $\binom{\theta}{\alpha}$, 分散共分散行列が $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ である二次元正規分布の場合に $\hat{\theta}_n$ が $\tilde{\theta}_n$ と漸近的に同じ分布に従う事を示す。

§2. 記号と仮定.

$X_1, X_2, \cdots$ を独立な確率変数の列とし, X_k の密度関数を $f_k(x|\theta,\alpha)$ とする。ここに $\theta \in \Theta \subset R^p$, $\alpha \in \Omega \subset R^q$ であり, α の先験分布を G とする。又 θ の真の値を θ_0 とし, 次の量を定義する。

$$L_k(x,\theta,\alpha) = \log f_k(x|\theta,\alpha), \quad M_k(x,\theta) = \log f_{G,k}(x|\theta)$$

$$\dot{L}_k^{(i)}(x,\theta,\alpha) = \frac{\partial}{\partial\theta_i} L_k(x,\theta,\alpha), \quad \dot{M}_k^{(i)}(x,\theta) = \frac{\partial}{\partial\theta_i} M_k(x,\theta)$$

$$\dot{L}_k(x,\theta,\alpha) = (\dot{L}_k^{(1)}(x,\theta,\alpha), \cdots, \dot{L}_k^{(p)}(x,\theta,\alpha))^t, \quad \dot{M}_k = (\dot{M}_k^{(1)}, \cdots, \dot{M}_k^{(p)})^t$$

$$\ddot{L}_k^{(i,j)}(x,\theta,\alpha) = \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_k(x,\theta,\alpha), \quad \ddot{M}_k^{(i,j)} = \frac{\partial^2}{\partial\theta_i\partial\theta_j} M_k(x,\theta)$$

$$\ddot{L}_k(x,\theta,\alpha) = \left(\ddot{L}_k^{(i,j)}(x,\theta,\alpha) \right), \quad \ddot{M}_k(x,\theta) = \left(\ddot{M}_k^{(i,j)}(x,\theta) \right)$$

$$\Gamma_k(\theta,\alpha) = E\left[\dot{L}_k(X_k,\theta,\alpha)\dot{L}_k(X_k,\theta,\alpha)^t \,\middle|\, \alpha \right]$$

上のような記号の下に, 次の仮定をおく。

(1) 正数 K と δ が存在し, $E[\,|X_k|^{1+\delta}\,|\alpha] \leq K$, $k=1,2,\cdots$, a.s. α

(2) Θ は R^p の開集合である。

(3) $\tilde{\theta}_n \longrightarrow \theta_0$ in prob.

(4) a.s. (X_k,α) に対し, $\dot{L}_k$, $\ddot{L}_k$ が存在し, 有界である。

(5) a.s. (X_k,α) に対し, $\ddot{L}_k$ は k について一様に θ の連続関数である。

(6) $E[\dot{L}_k(X_k,\theta,\alpha)\,|\,\alpha] = 0 \quad$ a.s. α .

(7) $\Gamma_k(\theta,\alpha) = -E[\ddot{L}_k(X_k,\theta,\alpha)\,|\,\alpha] \quad$ a.s. α

(8) $\frac{1}{n}\sum_{k=1}^{n}\Gamma_k(\theta,\alpha) \longrightarrow \bar{\Gamma}(\theta,\alpha) \quad$ a.s. α

ここに、$E[\bar{\Gamma}(\theta,\alpha)] = \bar{\Gamma}(\theta)$ は正定値.

(9) 任意の $\xi \in R^p$ に対し

$$n^{-\frac{2+\delta}{2}}\sum_{k=1}^{n}E[\,|\xi^t\dot{L}_k(X_k,\theta_0,\alpha)|^{2+\delta}\,|\,\alpha] \longrightarrow 0, \text{ a.s. } \alpha$$

(10) 正数 ε と確率変数 $Z_{k,i,j} = Z_{k,i,j}(X_k,\alpha)$ が存在し,

(i) $\sup\{|\ddot{L}_k^{(i,j)}(X_k,t,\alpha)|\,;\,\|t-\theta_0\|\le\varepsilon\} \le Z_{k,i,j}$

(ii) $E[\,|Z_{k,i,j}|\,|\,\alpha] \le K$

が a.s. α で成立.

これ等の仮定は Hoadley の漸近正規性の仮定を a.s.α で書きかえたものである。

§.3. $\tilde{\theta}_n$ の漸近的正規性.

上に述べた仮定が $f_{G,k}$ について満足される事を示せば $\tilde{\theta}_n$ の漸近正規性は云える。

$E[|X_k|^{1+\delta}] = E[E\{|X_k|^{1+\delta}|\alpha\}]$ であるから (1) より才1番目の条件 $E[|X_k|^{1+\delta}] \le K$ なる定数の存在性は云える。(2), (3) は同じ。(4) については次の Lemma を示す。

Lemma 1. $\dot{L}_k(x_k,\theta,\alpha)$, $\ddot{L}_k(x_k,\theta,\alpha)$ が存在し且つ有界 a.s. (x_k,α), であるならば $\dot{M}_k(x_k,\theta)$, $\ddot{M}_k(x_k,\theta)$ も存

在し，且つ有界 a.s.x_k である。

証明

(#) $$\dot{M}_k^{(i)}(x,\theta)=E[\dot{L}_k^{(i)}(x,\theta,\alpha)\mid x]$$

であるから $\dot{M}_k(x,\theta)$ は存在して有界．更に，

$$\ddot{M}_k^{(i,j)}(x,\theta)=\frac{\partial}{\partial\theta_j}[\dot{M}_k^{(i)}(x,\theta)]$$
$$=\int\frac{\ddot{L}_k^{(i,j)}(x,\theta,\alpha)f_k(x|\theta,\alpha)}{f_{G,k}(x|\theta)}dG(\alpha)+\int\frac{\dot{L}_k^{(i)}(x,\theta,\alpha)\dot{f}_k^{(j)}(x|\theta,\alpha)}{f_{G,k}(x|\theta)}dG(\alpha)$$
$$-\int\frac{\dot{L}_k^{(i)}(x,\theta,\alpha)f_k(x|\theta,\alpha)\dot{f}_{G,k}^{(j)}(x|\theta)}{f_{G,k}(x|\theta)^2}dG(\alpha)$$

(b) $$\therefore\ \ddot{M}_k^{(i,j)}(x,\theta)=E[\ddot{L}_k^{(i,j)}(x,\theta,\alpha)|x]+E[\dot{L}_k^{(i)}(x,\theta,\alpha)\dot{L}_k^{(j)}(x,\theta,\alpha)\mid x]$$
$$-E[\dot{L}_k^{(i)}(x,\theta,\alpha)|x]\cdot E[\dot{L}_k^{(j)}(x,\theta,\alpha)|x].$$

従って $\ddot{M}_k$ は存在し且つ有界となる。

更に，(b) を用いて (5) より $\ddot{M}_k$ は k について一様に θ の連続関数となる。(6) より明らかに $E[\dot{M}_k^{(i)}(x,\theta)]=0$ となる。

Lemma 2. (7) を仮定すれば

$$I_{G,k}(\theta)\equiv E[\dot{M}_k(x,\theta)\cdot\dot{M}_k(x,\theta)^t]=-E[\ddot{M}_k(x,\theta)]$$

証明. (7) より

$$E[\ddot{L}_k^{(i,j)}(x,\theta,\alpha)]=E[E\{\ddot{L}_k^{(i,j)}(x,\theta,\alpha)|\alpha\}]$$
$$=E[-E\{\dot{L}_k^{(i)}(x,\theta,\alpha)\dot{L}_k^{(j)}(x,\theta,\alpha)|\alpha\}]$$
$$=-E[\dot{L}_k^{(i)}(x,\theta,\alpha)\dot{L}_k^{(j)}(x,\theta,\alpha)]$$

更に，(#) より

$$E[\dot{M}_k^{(i)}(x,\theta)\cdot\dot{M}_k^{(j)}(x,\theta)]=E[E\{\dot{L}_k^{(i)}(x,\theta,\alpha)|x\}\cdot E\{\dot{L}_k^{(j)}(x,\theta,\alpha)|x\}]$$

故に(b)を用いると

$$E[\dot{M}_k^{(i)}(x,\theta)\cdot\dot{M}_k^{(j)}(x,\theta)] = -E[\ddot{M}_k^{(i,j)}(x,\theta)]$$

$$\therefore\ \Gamma_{G,k}(\theta) = -E[\ddot{M}_k(x,\theta)]\ .$$

(8),(9),(10)に関しては α についての期待値を取る事により得られる。例ば、(10)の $Z_{k,i,j}$ としては $E[Z_{k,i,j}(X_k,\alpha)\,|\,X_k]$.

以上、簡単に述べたように、$f_{G,k}$ に関する Hoadley の条件が(1)〜(10)の仮定の下に成り立つので $\hat{\theta}_n$ は漸近的正規分布をする事がいえる。

ここでは $\hat{\theta}_n$ の consistency については(3)で仮定したが正規性の場合と同様に、$f_k(x|\theta,\alpha)$ についての Hoadley の consistency に関する条件が a.s. α で満足されゝば $f_{G,k}(x|\theta)$ についても満足されるので、次節で述べる unknown G に関しての $\hat{\theta}_n \xrightarrow{p} \theta_0$ や $\hat{\theta}_n$ の漸近正規性については $f_k(x|\theta,\alpha)$ の各 α に関する、それ等の条件が満足されているかどうかを調べればよい。

§.4. Unknown G の場合.

ここでは2次元正規分布でその第2変数の期待値が unknown な先験分布 G をもつ場合について考える。即ち、

$$f_j(x,y|\theta,\alpha) = f(x,y|\theta,\alpha) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{-\frac{1}{2(1-\rho^2)}\big[(x-\theta)^2 - 2\rho(x-\theta)(y-\alpha) + (y-\alpha)^2\big]\Big\}$$

$$j = 1, 2, \cdots\ .$$

ここに，ρ は known，且つ $\mathcal{A} \subset \{\alpha : |\alpha| < M\}$ と仮定する。

そうすれば $\dot{L}, \dot{M}$ はそれぞれ．

$$\dot{L}(x, y, \theta, \alpha) = \frac{1}{1-\rho^2}\{x - \theta - \rho(y - \alpha)\}$$

$$\dot{M}(x, y, \theta) = \frac{1}{1-\rho^2}\{x - \theta - \rho(y - E[\alpha | x, y])\}$$

ここで $E[\alpha | x, y]$ は G, θ に depend している事に注意。

$\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}$ の同時分布は $\prod_{k=1}^{n} f_G(x_k, y_k | \theta)$ である，従って，その log likelihood の θ に関する微分は．

$$(*) \quad \dot{M}_n(\theta) \equiv \frac{\partial}{\partial\theta}\left\{\log\left[\prod_{k=1}^{n} f_G(x_k, y_k | \theta)\right]\right\} = \sum_{k=1}^{n} \dot{M}(x_k, y_k, \theta)$$

$$= \frac{n}{1-\rho^2}\left\{\bar{x}_n - \theta - \rho\left(\bar{y}_n - \frac{1}{n}\sum_{k=1}^{n} E[\alpha | x_k, y_k]\right)\right\}$$

ここに，$\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$，$\bar{y}_n = \frac{1}{n}\sum_{k=1}^{n} y_k$ ．

そこで，$\dot{M}_n(\theta) = 0$ なる θ の値が θ の最尤推定となるが，$\dot{M}_n(\theta)$ の中に未知な α の先験分布 G が含まれているので，もっと正確には $E[\alpha | x_k, y_k]$ が G に depend しているので，この posterior mean を推定する事を考える。まず

$$\frac{\partial}{\partial y}[\log f_G(x, y | \theta)] = \frac{f_G'(x, y | \theta)}{f_G(x, y | \theta)} = \frac{1}{1-\rho^2}\{\rho(x-\theta) - y + E[\alpha | x, y]\}$$

従って

$$E[\alpha | x, y] = (1-\rho^2)\frac{f_G'(x, y | \theta)}{f_G(x, y | \theta)} - \rho(x - \theta) + y$$

これを $(*)$ に代入すると

$$\dot{M}_n(\theta) = n\left\{\bar{x}_n - \theta + \rho \cdot \frac{1}{n}\sum_{k=1}^{n} \frac{f_G'(x_k, y_k | \theta)}{f_G(x_k, y_k | \theta)}\right\}.$$

そこで，f_G'，f_G の $\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}$ に基づく推定量 f_n'，f_n を次のように与える。

$$p_j(x,y) = \Delta^{-2} K_0\left(\frac{X_j - x}{\Delta}, \frac{Y_j - y}{\Delta}\right),$$

$$f_n(x,y) = \frac{1}{n-1}\sum_{j=1}^{n} p_j(x,y)$$

$$g_j(x,y) = \Delta^{-3}\left\{\frac{1}{2} K_1\left(\frac{X_j - x}{\Delta}, \frac{Y_j - y}{2\Delta}\right) - K_1\left(\frac{X_j - x}{\Delta}, \frac{Y_j - y}{\Delta}\right)\right\}$$

$$f_n'(x,y) = \frac{1}{n-1}\sum_{j=1}^{n} g_j(x,y)$$

ここに，K_0, K_1 は kernel 関数である。

そこで $\dot{M}_n(\theta)$ の f_G, f_G' に f_n, f_n' を入れた式を 0 とおいて得られた θ を $\hat{\theta}_n$ とする，即ち

$$\hat{\theta}_n = \bar{x}_n + \rho \frac{1}{n}\sum_{k=1}^{n} \frac{f_n'(x_k, y_k)}{f_n(x_k, y_k)}$$

これは $\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}$ に基づく θ の推定量であって unknown な要素を含む最尤推定量 $\tilde{\theta}_n$ と漸近的に同じ分布に従う事が示される。証明の本質的な所は

$$\frac{f_n'(X_k, Y_k)}{f_n(X_k, Y_k)} - \frac{f_G'(X_k, Y_k \mid \theta_0)}{f_G(X_k, Y_k \mid \theta_0)} \longrightarrow 0 \quad \text{in } L^2$$

を示す事である。