数理解析研究所講究録
第 749 巻 1991 年 61-75

Consistency of the maximum likelihood estimator : An approach from uniform consistency

筑波大　数学系　　赤平　昌文 (Masafumi Akahira)
筑波大　数学系　　鹿島　浩之 (Hiroyuki Kashima)

1．はじめに

最尤推定量が一致性を持つための十分条件については、Wald(1949) によって与えられたものが本質的であり、その後多くの研究者達によってその条件を変形したものが得られている（例えば Zacks(1971)）。ここでは少し見方を変えて次の方針で最尤推定量の一致性を示す。まず母数空間がコンパクト集合である場合に、最尤推定量が一様一致性を持つための十分条件を Wald 条件から抽出して簡単化する。次に一般の母数空間の場合に、その条件にさらにある条件を付加することによって最尤推定量が一致性を持つための十分条件を求める。ここで得られた十分条件は、Wald 条件と比べると単純である点と、ある分布が十分条件を満足しているかどうかを確認するときに、それが容易であるという点で意義があると思われる。

2．定義と Wald 条件

$(\mathcal{X},\mathcal{A})$を標本空間とし、$\{P_\theta;\theta\in\Theta\}$ を$(\mathcal{X},\mathcal{A})$上の確率測度の族とする。ここで$\Theta$は母数空間である。$\Theta$を R^kの部分集合と仮定し、$\|\cdot\|$ を k次元のユークリッドノルムとする。さらに、$(\mathcal{X},\mathcal{A})$の n 個の直積を$(\mathcal{X}^n,\mathcal{A}^n)$とし、その上の P_θの n 個の直積測度を P_θ^nとする。このときθの推定量を $\mathcal{X}^n$からΘへの $\mathcal{A}^n$-可測関数列 $\{\hat{\theta}_n\}$ として定義し、これを単に$\hat{\theta}_n$で表す。このとき、θの推定量が任意の$\theta\in\Theta$と任意の$\epsilon>0$ に対して、

$$\lim_{n\to\infty} P_\theta^n\{\|\hat{\theta}_n-\theta\|>\epsilon\}=0$$

を満たすとき、$\hat{\theta}_n$はθの一致推定量であるという。また母数空間Θの部分集合 Kについて、θの推定量$\hat{\theta}_n$が任意の$\epsilon>0$ に対して、

$$\lim_{n\to\infty}\sup_{\theta\in K} P_\theta^n\{\|\hat{\theta}_n-\theta\|>\epsilon\}=0$$

を満たすとき、$\hat{\theta}_n$は K上でθの一様一致推定量であるという。さらに、任意の$\theta \in \Theta$について P_θはσ-有限測度μに対して絶対連続で、その密度関数を

$$\frac{dP_\theta}{d\mu} = f(x:\theta)$$

とする。

Wald(1949) は、θ_0を真の母数とし、最尤推定量が一致性を持つための十分条件として次の (W1) ～ (W7) を上げている。

(W1)　Θは閉集合である。

(W2)

$$f(x:\theta,\rho) = \sup_{\|\theta-\theta'\|\leq\rho} f(x:\theta')$$

とし、さらに、

$$f^*(x:\theta,\rho) = \begin{cases} f(x:\theta,\rho) & \bigl(f(x:\theta,\rho)>1\bigr), \\ 1 & (\text{その他}) \end{cases}$$

とするとき、十分小さな$\rho > 0$ について　$E_{\theta_0}\bigl[log f^*(X:\theta,\rho)\bigr]$　は有限である。

(W3)

$$\phi(x;r) = \sup_{\|\theta\|>r} f(x:\theta)$$

とし、さらに、

$$\phi^*(x:r) = \begin{cases} \phi(x:r) & \bigl(\phi(x:r)>1\bigr), \\ 1 & (\text{その他}) \end{cases}$$

とするとき、十分大きな $r > 0$ について　$E_{\theta_0}\bigl[log\phi^*(X:r)\bigr]$　は有限である。

(W4)　μ に対してほとんど至る所の xについて、$f(x:\theta)$ はθ の連続関数で

ある。

(W5) $\theta_1 \neq \theta_2$のとき、$P_{\theta_1} \neq P_{\theta_2}$である。

(W6) $lim_{i\to\infty}\|\theta_i\| = \infty$ のとき、$lim_{i\to\infty} f(x:\theta_i) = 0$ が P_{θ_0}に対してほとんど至る所の xについて成り立つ。

(W7) $$E_{\theta_0}\left[|logf(X:\theta_0)|\right] < \infty$$

Wald はこれらの条件の下で最尤推定量の一致性を示すために、いくつかの補題を証明しているが、その中でここで後に必要となる２つの補題のみを証明なしで次に述べる。

補題 2.1. 条件 (W2)、(W5)、(W7) を仮定すれば、$\theta \neq \theta_0$となる任意の$\theta \in \Theta$について、

$$E_{\theta_0}\left[logf(X:\theta)\right] < E_{\theta_0}\left[logf(X:\theta_0)\right]$$

が成り立つ。

補題 2.2. 条件 (W2)、(W4) を仮定すれば、

$$\lim_{\rho\to 0} E_{\theta_0}\left[logf(X:\theta,\rho)\right] = E_{\theta_0}\left[logf(X:\theta)\right]$$

が成り立つ。

３．最尤推定量の一様一致性と一致性

$X_1, X_2, \cdots, X_n$を互いに独立にいずれも密度関数 $f(x:\theta)$ を持つ分布に従う確率変数であるとする。θの最尤推定量についてまず、Θがコンパクト集合の場合にその上で一様一致性を考え、さらに一般の集合の場合に一致性について考察する。

定理 3.1. 母数空間Θがコンパクト集合であるとき、条件 (W4)、(W5) と、さらに次の条件

(A1) μについて可積分な関数$\psi(x)$が存在して、任意の$\theta,\theta' \in \Theta$について、

$$|log f(x:\theta)| f(x:\theta') \leq \psi(x)$$

が成り立つならば、最尤推定量は、Θ上でθの一様一致推定量である。

証明 . 任意の$\epsilon > 0$、$\delta > 0$ について、

$$\sup_{\theta_0 \in \Theta} P_{\theta_0}^n\{\frac{sup_{\|\theta-\theta_0\|\geq\epsilon} f(X_1:\theta)\cdots f(X_n:\theta)}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > \delta\} \to 0 \quad (n \to \infty) \quad (3.1)$$

であることを示せば十分である。何故なら、もし (3.1) が成り立つならば、$\hat{\theta}_{ML,n}$を最尤推定量とすると、

$$f(x_1:\hat{\theta}_{ML,n})\cdots f(x_n:\hat{\theta}_{ML,n}) = \sup_{\theta\in\Theta} f(x_1:\theta)\cdots f(x_n:\theta) \qquad a.e.[\mu^n]$$

であり、従って任意の$\epsilon > 0$ に対して、

$$\begin{aligned}
&P_{\theta_0}^n\{\|\hat{\theta}_{ML,n} - \theta_0\| > \epsilon\} \\
&\leq P_{\theta_0}^n\{ \sup_{\|\theta-\theta_0\|\geq\epsilon} f(X_1:\theta)\cdots f(X_n:\theta) \\
&\qquad\qquad = f(X_1:\hat{\theta}_{ML,n})\cdots f(X_n:\hat{\theta}_{ML,n})\} \\
&\leq P_{\theta_0}^n\{ \sup_{\|\theta-\theta_0\|\geq\epsilon} f(X_1:\theta)\cdots f(X_n:\theta) \geq f(X_1:\theta_0)\cdots f(X_n:\theta_0)\} \\
&\leq P_{\theta_0}^n\{\frac{sup_{\|\theta-\theta_0\|\geq\epsilon} f(X_1:\theta)\cdots f(X_n:\theta)}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > 1\} \to 0 \quad (n \to \infty)
\end{aligned}$$

となるからである。ただし、μ^nはμの n 個の直積測度とする。よって、以下 (3.1) が成り立つことを示す。

$$p(\theta,\theta_0) = \int_{\mathcal{X}} f(x;\theta_0) log f(x;\theta) d\mu$$

とすると条件 (W4)、(A1) より、Lebesgue 収束定理によって、これは、2 変数の連続関数であることが分かる。さらに、

$$q(\theta,\theta_0) = p(\theta_0,\theta_0) - p(\theta,\theta_0)$$

とすれば、これも当然 2 変数の連続関数となる。今、

$$\Theta_\epsilon^2 = \left\{(\theta,\theta_0) \mid \|\theta-\theta_0\| \geq \epsilon, \quad \theta,\theta_0 \in \Theta\right\}$$

とすると、Θ_ϵ^2は R^{2k}のコンパクト部分集合であるので、

$$\mathop{inf}_{(\theta,\theta_0)\in\Theta_\epsilon^2} q(\theta,\theta_0) = q(\theta',\theta_0')$$

となる $(\theta',\theta_0') \in \Theta_\epsilon^2$が存在する。このとき、補題 2.1 より

$$q(\theta',\theta_0') > 0$$

である。ここで、

$$\alpha = \mathop{inf}_{(\theta,\theta_0)\in\Theta_\epsilon^2} q(\theta,\theta_0) > 0$$

とする。$E_{\theta_0}\left[logf(X:\theta,\rho)\right]$は$\rho \to 0$ について単調減少であることと、補題 2.2 より、

$$\Theta_0 = \left\{\theta \in \Theta \mid \|\theta-\theta_0\| \geq \epsilon\right\}$$

とすると、任意の$\theta \in \Theta_0$に対して、

$$E_{\theta_0}\left[logf(X:\theta,\rho_\theta)\right] - E_{\theta_0}\left[logf(X:\theta)\right] < \frac{\alpha}{2}$$

を満たすような$\rho_\theta > 0$ が存在する。従って、

$$\mu(\theta,\theta_0) = E_{\theta_0}\left[logf(X:\theta_0)\right] - E_{\theta_0}\left[logf(X:\theta,\rho_\theta)\right]$$

とすると、

$$\mu(\theta,\theta_0) > E_{\theta_0}\left[logf(X:\theta_0)\right] - E_{\theta_0}\left[logf(X:\theta)\right] - \frac{\alpha}{2} = q(\theta,\theta_0) - \frac{\alpha}{2} > \frac{\alpha}{2} > 0$$

また、(A1) より、

$$\mu(\theta,\theta_0) \leq E_{\theta_0}\big[logf(X:\theta_0)\big] - E_{\theta_0}\big[logf(X:\theta)\big] < +\infty \tag{3.2}$$

すなわち、

$$0 < \inf_{\theta_0 \in \Theta} \mu(\theta,\theta_0) < +\infty$$

である。ここで、

$$\mu = \inf_{\theta_0 \in \Theta} \mu(\theta,\theta_0)$$

とする。Θ_0はコンパクト集合であるので、

$$S(\theta:\rho) = \big\{\theta' \in R^k \mid \|\theta - \theta'\| < \rho\big\}$$

としたとき、

$$\Theta_0 \subset \bigcup_{i=1}^{h} S(\theta_i : \rho_{\theta_i})$$

を満たすような $\{\theta_i\}_{i=1\cdots h} \subset \Theta$が存在する。この $\{\theta_i\}_{i=1\cdots h}$を用いて、

$$\begin{aligned}
&\sup_{\theta \in \Theta_0} f(x_1:\theta)\cdots f(x_n:\theta) \\
&\leq \sum_{i=1}^{h} \sup_{\theta \in S(\theta_i:\rho_{\theta_i})} f(x_1:\theta)\cdots f(x_n:\theta) \\
&\leq \sum_{i=1}^{h} f(x_1:\theta_i,\rho_{\theta_i})\cdots f(x_n:\theta_i,\rho_{\theta_i})
\end{aligned}$$

が成り立つ。従って、任意の$\delta > 0$ について、

$$\begin{aligned}
&P_{\theta_0}^n\Big\{\frac{sup_{\theta\in\Theta_0} f(X_1:\theta)\cdots f(X_n:\theta)}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > \delta\Big\} \\
&\leq P_{\theta_0}^n\Big\{\frac{\sum_{i=1}^{h} f(X_1:\theta_i,\rho_{\theta_i})\cdots f(X_n:\theta_i,\rho_{\theta_i})}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > \delta\Big\} \\
&\leq \sum_{i=1}^{h} P_{\theta_0}^n\Big\{\frac{f(X_1:\theta_i,\rho_{\theta_i})\cdots f(X_n:\theta_i,\rho_{\theta_i})}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > \delta_i\Big\}
\end{aligned}$$

を満足する $0 < \delta_i \leq \delta \quad (i = 1 \cdots h)$ が存在する。各 i について、$log\delta_i = -\gamma_i$ とすると、$\gamma_i > 0$ と考えてよく、

$$\begin{aligned}
&P^n_{\theta_0}\{\frac{f(X_1:\theta_i,\rho_{\theta_i})\cdots f(X_n:\theta_i,\rho_{\theta_i})}{f(X_1:\theta_0)\cdots f(X_n:\theta_0)} > \delta_i\} \\
&= P^n_{\theta_0}\{\sum_{k=1}^{n}\left(logf(X_k:\theta_i,\rho_{\theta_i}) - logf(X_k:\theta_0)\right) > -\gamma_i\} \\
&= P^n_{\theta_0}\{\sum_{k=1}^{n}\frac{\mu(\theta_i,\theta_0) - \left(logf(X_k:\theta_0) - logf(X_k:\theta_i,\rho_{\theta_i})\right)}{n} \\
&\qquad > \mu(\theta_i,\theta_0) - \frac{\gamma_i}{n}\} \qquad (3.3)
\end{aligned}$$

が成立する。ここで、

$$n_0 > \frac{max_{1\leq i\leq h}\,\gamma_i}{\mu}$$

なる n_0 をとり、さらに、

$$\beta = \mu - \frac{max_{1\leq i\leq h}\,\gamma_i}{n_0}$$

とすると、$n \geq n_0$ のとき、

$$\mu(\theta_i,\theta_0) - \frac{\gamma_i}{n} \geq \beta > 0$$

であるので、このとき (3.2) より大数の弱法則が使えて、

$$\begin{aligned}
&(3.3)\text{式} \\
&\leq P^n_{\theta_0}\{\sum_{k=1}^{n}\frac{\mu(\theta_i,\theta_0) - \left(logf(X_k:\theta_0) - logf(X_k:\theta_i,\rho_{\theta_i})\right)}{n} > \beta\} \\
&\leq P^n_{\theta_0}\{|\sum_{k=1}^{n}\frac{\left(logf(X_k:\theta_0) - logf(X_k:\theta_i,\rho_{\theta_i})\right) - \mu(\theta_i,\theta_0)}{n}| > \beta\} \\
&\to 0 \quad (n \to \infty)
\end{aligned}$$

が成り立つ。以上で、(3.1) が示された。□

さらに、一般に、Θが R^kの部分集合である場合には、次の定理が成り立つ。

定理 3.2. 母数空間をΘとしたときのθ_0の最尤推定量を$\hat{\theta}_{ML,n}$とし、これが確率 1 で一意的に定まるとする。また、$K=1,2,\cdots$について、母数空間を$\Theta_K=\{\theta\in\Theta \mid \|\theta\|\le K\}$に制限したときの$\theta_0$の最尤推定量を$\hat{\theta}^{(K)}_{ML,n}$とし、$\hat{\theta}^{(K)}_{ML,n}$は$\Theta_K$上で一様一致推定量であるとする。さらに次の条件

$$\text{(A2)} \qquad \lim_{n\to\infty} P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}\|\ge K(n)\}=0$$

が成り立つと仮定する。このとき、$\hat{\theta}_{ML,n}$は一致推定量である。ここで $K(n)$とは、$\{\epsilon_K\}$ を$\epsilon_K\downarrow 0$ とし、$n_0(K)$ を、$n\ge n_0(K)$ ならば任意の$\theta\in\Theta_K$について、$P^n_\theta\{\|\hat{\theta}^{(K)}_{ML,n}-\theta\|>\epsilon_K\}<\epsilon_K$を満たす自然数としたとき、$K(n)=sup\{K:n>n_0(K)\}$のこととする。

証明 . $\hat{\theta}^{(K)}_{ML,n}$は、$\Theta_K$上の一様一致推定量であるから、各 Kについて、$n\ge n_0(K)$ ならば、任意の$\theta\in\Theta_K$について、

$$P^n_\theta\{\|\hat{\theta}^{(K)}_{ML,n}-\theta\|>\epsilon_K\}<\epsilon_K$$

を満足するような $n_0(K)$ が存在する。ここで、

$$\hat{\theta}^*_n=\begin{cases}\hat{\theta}^{\left(K(n)\right)}_{ML,n} & (\|\hat{\theta}_{ML,n}\|<K(n)) \\ \dfrac{K(n)\hat{\theta}_{ML,n}}{\|\hat{\theta}_{ML,n}\|} & (\|\hat{\theta}_{ML,n}\|\ge K(n))\end{cases}$$

と定義する。θ_0を真の母数として、$\|\theta_0\|<K_0$なる自然数 K_0をとる。$\epsilon>0$を任意にとると、$\frac{1}{2}\epsilon>\epsilon_{K_1}$かつ $K_1\ge K_0$なる K_1がとれ、さらに$\frac{1}{2}\epsilon_{K_1}>\epsilon_{K_2}$なる K_2をとることができる。(A2) より、$n\ge n_1$ならば、

$$P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}\|\ge K(n)\}<\frac{\epsilon}{4} \qquad (3.4)$$

なる n_1がとれる。以後、$n \geq max(n_0(K_2), n_1)$ なる n について、議論をすすめる。まず

$$
\begin{aligned}
&P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \theta_0\| > \epsilon\} \leq P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \theta_0\| > \epsilon_{K_1}\} \\
&\leq P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| + \|\hat{\theta}^*_n - \theta_0\| > \epsilon_{K_1}\} \\
&\leq P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}\} + P^n_{\theta_0}\{\|\hat{\theta}^*_n - \theta_0\| > \frac{1}{2}\epsilon_{K_1}\}
\end{aligned}
$$

となり、

$$
\begin{aligned}
&P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}\} \\
&= P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\} \\
&\quad + P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| \geq K(n)\}
\end{aligned}
$$

となる。次に$\hat{\theta}^*_n$の決め方より、

$$
\begin{aligned}
&P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\} \\
&= P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^{\left(K(n)\right)}_{ML,n}\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\} \quad (3.5)
\end{aligned}
$$

であるが、

$$\|\hat{\theta}_{ML,n}\| < K(n)$$

のとき、

$$\hat{\theta}_{ML,n} = \hat{\theta}^{\left(K(n)\right)}_{ML,n} \qquad a.e.[P^n_{\theta_0}]$$

であるので、

$$P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\} = 0$$

である。さらに、(3.4) より、

$$
\begin{aligned}
&P^n_{\theta_0}\{\|\hat{\theta}_{ML,n} - \hat{\theta}^*_n\| > \frac{1}{2}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| \geq K(n)\} \\
&\leq P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}\| \geq K(n)\} < \frac{\epsilon}{4}
\end{aligned}
$$

となる。従って、

$$P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}-\hat{\theta}^*_n\|>\frac{1}{2}\epsilon_{K_1}\}<\frac{\epsilon}{2}$$

が成り立つ。同様にして、

$$\begin{aligned}
&P^n_{\theta_0}\{\|\hat{\theta}^*_n-\theta_0\|>\frac{1}{2}\epsilon_{K_1}\}\\
&=P^n_{\theta_0}\{\|\hat{\theta}^*_n-\theta_0\|>\frac{1}{2}\epsilon_{K_1},\|\hat{\theta}_{ML,n}\|<K(n)\}\\
&\qquad+P^n_{\theta_0}\{\|\hat{\theta}^*_n-\theta_0\|>\frac{1}{2}\epsilon_{K_1},\|\hat{\theta}_{ML,n}\|\geq K(n)\}\\
&\leq P^n_{\theta_0}\{\|\hat{\theta}^{(K(n))}_{ML,n}-\theta_0\|>\frac{1}{2}\epsilon_{K_1}\}+P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}\|\geq K(n)\}\\
&\leq P^n_{\theta_0}\{\|\hat{\theta}^{(K(n))}_{ML,n}-\theta_0\|>\epsilon_{K_2}\}+\frac{\epsilon}{4}\\
&\leq\epsilon_{K_2}+\frac{\epsilon}{4}<\frac{1}{2}\epsilon_{K_1}+\frac{\epsilon}{4}<\frac{\epsilon}{2}
\end{aligned}$$

従って、

$$P^n_{\theta_0}\{\|\hat{\theta}_{ML,n}-\theta_0\|>\epsilon\}<\epsilon$$

が成立する。故に、$\hat{\theta}_{ML,n}$は、一致推定量であることが示された。□

Θが閉集合であれば、Θ_Kは言うまでもなくコンパクト集合であり、定理 3.1 より条件 (W1) と、任意の $K=1,2,\cdots$についてΘ_K上で条件 (W4)、(W5)、(A1) を満足すれば、定理 3.2 における前半の条件を満たすことになる。そこで、その後半の条件 (A2) を満たすための十分条件を考える。まず、次の補題が成り立つ。

補題 3.1. 条件 (W6) が成立するならば、

$$\text{(A3)}\qquad \|\hat{\theta}_{ML,n}\|\not\to\infty\quad(n\to\infty)$$

が成り立つ。

注意 3.1.　条件 (W6)、(A3) は共に $f(x:\theta)$ のθが原点から十分離れたところでの条件を与えている。Ibragimov-Has'minskii(1981) も最尤推定量が一致性を持つための十分条件のうちの一つに、やはり、$f(x:\theta)$ のθが原点から十分離れたところでの条件として、次のものを上げている。

(IH)　母数空間Θが非有界であるとき、任意の$\theta \in \Theta$に対して、

$$\lim_{c\to\infty} \int_{\mathcal{X}} \sup_{\|u\|\geq c} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:\theta+u)\right) d\mu < 1$$

が成り立つ。

以上3つの条件は、(W6) が成立するならば (IH) が成り立ち、(IH) が成立するならば (A3) が成り立つ、という関係にある。

補題 3.1 の証明．まず、(W6) が成り立つならば、(IH) が成り立つことを示す。仮定より、Θが非有界であるとき、

$$\lim_{c\to\infty} \sup_{\|u\|\geq c} f^{\frac{1}{2}}(x:\theta+u) = 0 \qquad a.e.[P_\theta]$$

となる。$sup_{\|u\|\geq c} f^{\frac{1}{2}}(x:\theta+u)$ は、$c\to\infty$ のとき単調減少であるので、

$$\mathcal{X}_0 = \left\{x \in \mathcal{X} | f(x:\theta) > 0\right\}$$

とすると、

$$\begin{aligned}
&\lim_{c\to\infty} \int_{\mathcal{X}} \sup_{\|u\|\geq c} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:\theta+u)\right) d\mu \\
&= \lim_{c\to\infty} \int_{\mathcal{X}_0} \frac{sup_{\|u\|\geq c} f^{\frac{1}{2}}(x:\theta+u)}{f^{\frac{1}{2}}(x:\theta)} dP_\theta = 0
\end{aligned}$$

となる。次に、(IH) が成り立つならば、(A3) が成り立つことを背理法で示す。(IH) が成り立つとき、Θが非有界であるならば、

$$\int_{\mathcal{X}} \sup_{\|u\|\geq R} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:\theta+u)\right) d\mu < 1$$

を満足するような Rが存在する。今、

$$\|\hat{\theta}_{ML,n}\| \to \infty \quad (n \to \infty)$$

であると仮定すると、$n \geq n_0$ならば、

$$\|\hat{\theta}_{ML,n}\| \geq R + \|\theta\|$$

を満足するような n_0が存在する。このとき、$n \geq n_0$なる任意の n について、

$$\begin{aligned}
&\int_{\mathcal{X}} \sup_{\|u\| \geq R} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:\theta+u)\right) d\mu \\
&\geq \int_{\mathcal{X}} \sup_{\|u+\theta\| \geq R+\|\theta\|} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:\theta+u)\right) d\mu \\
&\geq \int_{\mathcal{X}} \sup_{\|v\| \geq R+\|\theta\|} \left(f^{\frac{1}{2}}(x:\theta) f^{\frac{1}{2}}(x:v)\right) d\mu \\
&= \int_{\mathcal{X}} f^{\frac{1}{2}}(x;\theta) f^{\frac{1}{2}}(x:\hat{\theta}_{ML,n}) d\mu \geq \int_{\mathcal{X}} f(x:\theta) d\mu = 1
\end{aligned}$$

となり、矛盾する。従って、

$$\|\hat{\theta}_{ML,n}\| \not\to \infty \quad (n \to \infty)$$

であることが、示された。Θが有界のときは、$\|\hat{\theta}_{ML,n}\| \not\to \infty \quad (n \to \infty)$ となることは明白である。□

Markov の不等式から、

$$P_{\theta_0}^n\{\|\hat{\theta}_{ML,n}\| \geq K(n)\} \leq \frac{E_{\theta_0}[\|\hat{\theta}_{ML,n}\|]}{K(n)}$$

であり、$K(n) \to \infty (n \to \infty)$ であるので、補題 3.1 より (W6) ならば (A2) が成り立つことが分かる。すなわち、(W6) は (A2) より強い条件である。ところで、定理 3.2 において (A2) の代わりに (W6) を仮定すると、最尤推定量の一意性に関する条件がなくても、最尤推定量の一致性が次の系のように示される。

系 3.1. 任意の $K = 1, 2, \cdots$について、Θ_K上θ_0の最尤推定量$\hat{\theta}^{(K)}_{ML,n}$が$\Theta_K$上で一様一致推定量であり、条件 (W6) が成り立つと仮定するとき、Θ上θ_0の最尤推定量$\hat{\theta}_{ML,n}$は一致推定量である。

証明. 条件 (W6) より、

$$\|\hat{\theta}_{ML,n}\| \leq K(n^*)$$

が任意の n について成立するような自然数 n^*が存在する。ここで $K(n)$ は定理 3.2 で定義したものである。任意のθ_0について

$$\|\theta_0\| \leq K(n_0)$$

なる自然数 n_0が存在する。また、任意の$\epsilon > 0$について$\epsilon > \epsilon_{K'}$なる自然数 K' が存在する。ここで、$\{\epsilon_K\}$ は、同じく定理 3.2 で定義したものである。今、

$$K^* = max\big(K(n^*), K(n_0), K'\big)$$

とすると、十分大きな n について、

$$\begin{aligned}
&P^n_{\theta_0}\big\{\|\hat{\theta}_{ML,n} - \theta_0\| > \epsilon, \|\hat{\theta}_{ML,n}\| < K(n)\big\} \\
&\leq P^n_{\theta_0}\big\{\|\hat{\theta}_{ML,n} - \theta_0\| > \epsilon, \|\hat{\theta}_{ML,n}\| < K^*\big\} \\
&= P^n_{\theta_0}\big\{\|\hat{\theta}^{(K^*)}_{ML,n} - \theta_0\| > \epsilon\big\} \\
&\leq P^n_{\theta_0}\big\{\|\hat{\theta}^{(K^*)}_{ML,n} - \theta_0\| > \epsilon_{K^*}\big\} < \epsilon_{K^*} < \epsilon
\end{aligned}$$

が成り立つことが分かる。従って、

$$\begin{aligned}
&(3.5)\text{式} \\
&\leq P^n_{\theta_0}\big\{\|\hat{\theta}_{ML,n} - \theta_0\| > \frac{1}{4}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\big\} \\
&\qquad + P^n_{\theta_0}\big\{\|\hat{\theta}^{\big(K(n)\big)}_{ML,n} - \theta_0\| > \frac{1}{4}\epsilon_{K_1}, \|\hat{\theta}_{ML,n}\| < K(n)\big\} \\
&< \frac{1}{4}\epsilon_{K_1} + \epsilon_{K_3} < \frac{1}{2}\epsilon_{K_1} < \frac{1}{4}\epsilon
\end{aligned}$$

が十分大きな n について成立する。ここで、K_3は$\frac{1}{4}\epsilon_{K_1} > \epsilon_{K_3}$なる自然数であり、$K_1$は、定理 3.2 で定義したものである。以上より、後は定理 3.2 の証明と全く同様にして、$\hat{\theta}_{ML,n}$の一致性が示される。□

以上の定理 3.1 及び系 3.1 より、次の系が成り立つことが分かる。

系 3.2. 条件 (W1)、(W4)、(W5)、(W6) と任意の $K=1,2,\cdots$についてΘ_K上で条件 (A1) を満たしているとき、最尤推定量は一致推定量である。

(A1) は (W2)、(W7) を含んだやや強い条件であるが、(W3) は除去することができた。実際に、条件を確認するときには、(W2)、(W3) より (A1) の方が単純で良いのではないかと思われる。θが原点から十分離れたところに関する条件については、その条件を満たす分布のクラスを広げるには、一番ゆるい条件 (A2) を採用するのがよいが、条件の単純さ、直観的な分かり易さからすると (W6) や (A3) を採用するのがよいと思われる。

4．例

この節で Lehmann(1983) において Bahadur(1958)、LeCam(1979) による最尤推定量の不一致性の例として論じられていることについて考察する。彼らは次のような例を考えた。

h を左半開区間 (0,1] 上で定義された狭義の単調減少関数で、その区間において $h(x)\geq 1$ でかつ

$$\int_0^1 h(x)dx=\infty$$

を満たすとする。また $0<c<1$ となる定数 c に対して、定数列 $\{a_k\}_{k=0,1,\cdots}$ を次のように逐次的に定義する。$a_0=1$ とし $a_0,\cdots,a_{k-1}$が与えられたとき、定数 a_kを

$$\int_{a_k}^{a_{k-1}}\left[h(x)-c\right]dx=1-c \tag{4.1}$$

となるように定める。このとき (4.1) を満たす $0<a_k<a_{k-1}$となる a_kは一意的に決められ、$a_k\to 0(k\to\infty)$ も成り立つ。そこで、各 $k=1,2,\cdots$について、密度関数 f_kを

$$f_k(x)=\begin{cases} c & (x\leq a_k,\quad x>a_{k-1}),\\ h(x) & (a_k<x\leq a_{k-1}) \end{cases}$$

で定義し、$X_1, \cdots, X_n$を互いに独立に、いずれも f_kをもつ分布に従う確率変数であると仮定して、母数 kの推定問題を考える。もし $h(x)$ が、x のすべての十分小さな値に対して $h(x) \geq e^{\frac{1}{x^2}}$ を満たすならば、kの最尤推定量$\hat{k}_n$は確率１で一意的に定まり、さらに$\hat{k}_n$は無限大に確率収束する、すなわち任意の $k = 1, 2, \cdots$について、

$$\lim_{n \to \infty} P\{\hat{k}_n > k\} = 1$$

が成り立つことが示される。従って最尤推定量$\hat{k}_n$は kの一致推定量ではないことが分かる。この場合には、前節の条件 (A2) が満たされていないことが示される。従って、一致性を示すためには条件 (A2) は必要である。

参考文献

Bahadur, R. R.(1958). Examples of inconsistency of maximum likelihood estimates. Sankhya 20,207-210.

Ibragimov, I. A. and Has'minskii, R. Z.(1981). Statistical Estimation. Asymptotic Theory. Springer-Verlag, New York.

LeCam, L.(1979). Maximum Likelihood: An Introduction. Lecture Notes in Statiatics No.18, University of Maryland, College Park, Md.

Lehmann, E. L.(1983). Theory of Point Estimation. Wiley, New York.

Wald, A.(1949). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 595-601.

Zacks, S.(1971). The Theory of Statistical Inference. Wiley, New York.