M3S3/S4 STATISTICAL THEORY II

ASYMPTOTIC BEHAVIOUR OF THE MLE

ASSUMPTIONS: Consider a probability model defined on probability space $(\mathcal{X}, \mathcal{B}, P)$. Suppose that P is indexed by parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, and that the corresponding distribution function is $F_{X|\boldsymbol{\theta}}$, with density (with respect to measure ν) denoted $f_{X|\boldsymbol{\theta}}$. Suppose that the true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta_0}$.

A0. **Identifiability**

$$f_{X|\boldsymbol{\theta_1}}(x|\boldsymbol{\theta_1}) = f_{X|\boldsymbol{\theta_2}}(x|\boldsymbol{\theta_2}) \; \forall \; x \in \mathbb{X} \equiv \{x : f_{X|\boldsymbol{\theta}}(x|\boldsymbol{\theta}) > 0\} \qquad \Longleftrightarrow \qquad \boldsymbol{\theta_1} = \boldsymbol{\theta_2}$$

A1. The support of $f_{X|\boldsymbol{\theta}}$, $\mathbb{X}$, **does not depend on $\boldsymbol{\theta}$**.

A2. Random variables $X_1, \ldots, X_n$ are **i.i.d.** from $P_{\boldsymbol{\theta_0}}$ with distribution function $F_{X|\boldsymbol{\theta_0}}$.

A3. Θ contains an **open neighbourhood**, $\Theta_0 \subset \mathbb{R}^d$, of $\boldsymbol{\theta_0}$ on which

(i) $l(\boldsymbol{\theta}; x) = \log f_{X|\boldsymbol{\theta}}(x|\boldsymbol{\theta})$ is **twice continuously differentiable with respect to $\boldsymbol{\theta}$**, a.e. with respect to ν on $\mathbb{X}$.

(ii) Third derivatives of $l(\boldsymbol{\theta}; x)$ **exist** and are **absolutely bounded**, that is

$$\left| \dddot{l}_{jkl}(\boldsymbol{\theta}; x) \right| \leq M_{jkl}(x) \qquad \boldsymbol{\theta} \in \Theta_0$$

for all j, k, l, for some function $M_{jkl}(x)$ where

$$\dddot{l}_{jkl}(\boldsymbol{\theta}; x) = \frac{\partial^3 l(\boldsymbol{\theta}; x)}{\partial\theta_j \partial\theta_k \partial\theta_l}$$

and

$$E_{f_{X|\boldsymbol{\theta_0}}}[M_{jkl}(x)] < \infty$$

A4. Let

$$\dot{l}_j(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta}; x)}{\partial\theta_j} \qquad \ddot{l}_{jk}(\boldsymbol{\theta}; x) = \frac{\partial^2 l(\boldsymbol{\theta}; x)}{\partial\theta_j \partial\theta_k}$$

be components of the first partial derivative vector and second partial derivative matrix respectively. Then

(i) $E_{f_{X|\boldsymbol{\theta_0}}}\left[\dot{l}_j(\boldsymbol{\theta_0}; X)\right] = 0$ for $j = 1, \ldots, d$.

(ii) $E_{f_{X|\boldsymbol{\theta_0}}}\left[(\dot{l}_j(\boldsymbol{\theta_0}; X))^2\right] < \infty$ for $j = 1, \ldots, d$.

(iii) The $d \times d$ matrix $I(\boldsymbol{\theta_0})$ with $(j,k)^{\text{th}}$ entry

$$E_{f_{X|\boldsymbol{\theta_0}}}\left[-\ddot{l}_{jk}(\boldsymbol{\theta_0}; X)\right]$$

is **positive definitive**.

One approach to finding the MLE based on data $\boldsymbol{x} = (x_1, \ldots, x_n)$ is to solve the system of **likelihood equations**

$$\dot{\boldsymbol{l}}_n(\boldsymbol{\theta}) = 0 \qquad \text{(LE)}$$

that is, a system of d equations based on the first partial derivative vector $\dot{\boldsymbol{l}}_n$.

Theorem 2.1 ***Asymptotic Behaviour of Solutions to the Likelihood Equations***

Suppose that conditions A0 to A4 hold. Define $(d \times 1)$ *vector* $\boldsymbol{Z}_n$ *by*

$$\boldsymbol{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\boldsymbol{l}}(\boldsymbol{\theta_0}; X_i)$$

and the $(d \times 1)$ *vector* $\tilde{\boldsymbol{l}}(\boldsymbol{\theta_0}; X)$ *by*

$$\tilde{\boldsymbol{l}}(\boldsymbol{\theta_0}; X) = I(\boldsymbol{\theta_0})^{-1} \dot{\boldsymbol{l}}(\boldsymbol{\theta_0}; X)$$

so that

$$I(\boldsymbol{\theta_0})^{-1} \boldsymbol{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\boldsymbol{l}}(\boldsymbol{\theta_0}; X_i).$$

Then

(i) ***EXISTENCE AND CONSISTENCY:*** *As* $n \longrightarrow \infty$*, with probability converging to 1, there exist solutions* $\tilde{\boldsymbol{\theta}}_n$ *of the likelihood equations (LE) such that*

$$\tilde{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta_0}.$$

(ii) ***ASYMPTOTIC NORMALITY:*** *As* $n \longrightarrow \infty$*,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = I(\boldsymbol{\theta_0})^{-1} \boldsymbol{Z}_n + o_p(1)\mathbf{1} \xrightarrow{\mathfrak{L}} I(\boldsymbol{\theta_0})^{-1} \boldsymbol{Z} \stackrel{def}{=} \boldsymbol{D} \sim N(\mathbf{0}, I(\boldsymbol{\theta_0})^{-1})$$

Proof.

(i) Existence And Consistency: Let $\delta > 0$, and Q_δ be such that

$$Q_\delta = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta_0}\| \leq \delta\}.$$

Then, by a third order Taylor expansion around $\boldsymbol{\theta_0}$,

$$\begin{aligned} \frac{1}{n}(l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta_0})) &= \frac{1}{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} \dot{\boldsymbol{l}}_n(\boldsymbol{\theta_0}) &(1)\\ &\quad - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} \left(-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta_0})\right)(\boldsymbol{\theta} - \boldsymbol{\theta_0}) &(2)\\ &\quad + \frac{1}{6}\frac{1}{n} \sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d} (\theta_j - \theta_{j0})(\theta_k - \theta_{k0})(\theta_l - \theta_{l0}) \left\{\sum_{i=1}^{n} \gamma_{jkl}(X_i) M_{jkl}(X_i)\right\} &(3)\\ &= S_1 + S_2 + S_3 \end{aligned}$$

say, where by assumption A3(ii), $0 \leq |\gamma_{jkl}(x)| < 1$. Now, by assumption A3(ii), it follows that the first derivatives are also bounded at $\boldsymbol{\theta_0}$, so

$$S_1 \xrightarrow{p} 0 \tag{4}$$

as the term in equation (1) is a constant over n. Secondly, by assumption A4 and the Weak Law of Large Numbers (WLLN)

$$-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta_0}) \xrightarrow{\mathfrak{L}} I(\boldsymbol{\theta_0})$$

and hence

$$S_2 = -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} \left(\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta_0})\right)(\boldsymbol{\theta} - \boldsymbol{\theta_0}) \xrightarrow{p} -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} I(\boldsymbol{\theta_0})(\boldsymbol{\theta} - \boldsymbol{\theta_0})$$

Now, by properties of quadratic forms based on positive definite symmetric matrices, it can be shown that

$$(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} I(\boldsymbol{\theta_0})(\boldsymbol{\theta} - \boldsymbol{\theta_0}) \geq \lambda_d \|\boldsymbol{\theta} - \boldsymbol{\theta_0}\|^2$$

where λ_d is the smallest eigenvalue of $I(\boldsymbol{\theta_0})$. Then for $\boldsymbol{\theta} \in Q_\delta$

$$(\boldsymbol{\theta} - \boldsymbol{\theta_0})^\mathsf{T} I(\boldsymbol{\theta_0})(\boldsymbol{\theta} - \boldsymbol{\theta_0}) \geq \lambda_d \delta^2. \tag{5}$$

Finally, using the WLLN on the term in equation (3),

$$S_3 \xrightarrow{p} \frac{1}{6} \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} (\theta_j - \theta_{j0})(\theta_k - \theta_{k0})(\theta_l - \theta_{l0}) \left\{ \sum_{i=1}^{n} E[\gamma_{jkl}(X_i) M_{jkl}(X_i)] \right\} \tag{6}$$

By equation (4), for any given $\epsilon, \delta > 0$, the convergence in probability result ensures that for n large enough, with probability greater than $1 - \epsilon$, for all $\boldsymbol{\theta} \in \Theta$,

$$\|S_1\| \;<\; d\delta^3 \tag{7}$$

$$S_2 \;<\; -\lambda_d \delta^2 / 4 \tag{8}$$

$$\|S_3\| \;\leq\; \frac{1}{6}(d\delta)^3 \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} m_{jkl} \tag{9}$$

where $m_{jkl} = E[M_{jkl}(X)]$. Hence, combining results (7), (8) and (9),

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in Q_\delta} (S_1 + S_2 + S_3) &\leq \sup_{\boldsymbol{\theta} \in Q_\delta} \|S_1 + S_3\| + \sup_{\boldsymbol{\theta} \in Q_\delta} S_2 \\ &< d\delta^3 + M\delta^3 - \frac{\lambda_d}{4}\delta^2 \\ &= (d + M)\delta^3 - \frac{\lambda_d}{4}\delta^2 \qquad (10) \end{aligned}$$

where

$$M = \frac{1}{6} d^3 \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} m_{jkl}$$

Thus, if $\delta < \lambda_d / 4(M + d)$, the right hand side of equation (10) is negative, so

$$\sup_{\boldsymbol{\theta} \in Q_\delta} (S_1 + S_2 + S_3) < 0.$$

Thus, for n large enough, with probability at least $1 - \epsilon$

$$\frac{1}{n}(l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta_0})) < 0$$

or, equivalently,

$$P[\, l_n(\boldsymbol{\theta}) < l_n(\boldsymbol{\theta_0}) \text{ for all } \boldsymbol{\theta} \in Q_\delta \,] \longrightarrow 1 \qquad \text{as} \qquad n \longrightarrow \infty,$$

that is, l has a local maximum inside Q_δ. Therefore, as the likelihood equations (LE) are satisfied at local maxima, it follows that (with probability converging to 1 as $n \longrightarrow \infty$) there **exists** a solution, $\tilde{\boldsymbol{\theta}}_n(\delta)$, within Q_δ, for any $0 < \delta < \lambda_d / 4(M + d)$. As this holds for arbitrarily small δ, it follows that

$$\lim_{n \longrightarrow \infty} P[\, \|\tilde{\boldsymbol{\theta}}_n(\delta) - \boldsymbol{\theta_0}\| < \delta \,] = 1 \qquad \therefore \qquad \tilde{\boldsymbol{\theta}}_n(\delta) \xrightarrow{p} \boldsymbol{\theta_0}.$$

(ii) Consider the set G_n

$$G_n = \left\{ \tilde{\boldsymbol{\theta}}_n : \dot{\boldsymbol{l}}_n(\tilde{\boldsymbol{\theta}}_n) = 0 \text{ and } \|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}\| < \epsilon \right\}$$

then $P_{\boldsymbol{\theta_0}}(G_n) \longrightarrow 1$ as $n \longrightarrow \infty$. On this set, using a first order Taylor expansion of $\dot{\boldsymbol{l}}_n$ about $\boldsymbol{\theta_0}$,

$$0 = \frac{1}{\sqrt{n}}\dot{\boldsymbol{l}}_n(\tilde{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}}\dot{\boldsymbol{l}}_n(\boldsymbol{\theta_0}) - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})^{\mathsf{T}} \left(-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta}^\star_n) \right) \tag{11}$$

for some $\boldsymbol{\theta}^\star_n$ such that

$$\|\boldsymbol{\theta}^\star_n - \boldsymbol{\theta_0}\| \leq \|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}\|. \tag{12}$$

From assumption A4(i) and (iii),

$$\boldsymbol{Z}_n = \frac{1}{\sqrt{n}}\dot{\boldsymbol{l}}_n(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\boldsymbol{l}}(\boldsymbol{\theta_0}; X_i) \xrightarrow{\mathfrak{L}} N(\boldsymbol{0}, I(\boldsymbol{\theta_0})).$$

Now, by equation (12),

$$-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta}^\star_n) = -\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta_0}) + o_p(1)\boldsymbol{1}$$

as $\tilde{\boldsymbol{\theta}}_n(\delta) \xrightarrow{p} \boldsymbol{\theta_0}$, after considering another Taylor expansion of $\ddot{\boldsymbol{l}}$ about $\boldsymbol{\theta_0}$, and the boundedness of the third derivatives in assumption A3(ii). Thus, with high probability, the inverse matrix

$$\left(-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta}^\star_n) \right)^{-1}$$

exists and, by the continuous mapping result

$$\left(-\frac{1}{n}\ddot{\boldsymbol{l}}_n(\boldsymbol{\theta}^\star_n) \right)^{-1} \xrightarrow{p} I(\boldsymbol{\theta_0})^{-1}.$$

Hence, rearranging equation (11), we have that

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = I(\boldsymbol{\theta_0})^{-1}\boldsymbol{Z}_n + o_p(1)\boldsymbol{1} \xrightarrow{\mathfrak{L}} I(\boldsymbol{\theta_0})^{-1}\boldsymbol{Z} \sim N\left(0, I(\boldsymbol{\theta_0})^{-1}\right)$$

■

Corollary : Delta Method
Suppose that $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ where $\boldsymbol{g}$ is differentiable at $\boldsymbol{\theta_0}$. Then $\tilde{\boldsymbol{\phi}}_n = \boldsymbol{g}(\tilde{\boldsymbol{\theta}}_n)$ satisfies

$$\sqrt{n}(\tilde{\boldsymbol{\phi}}_n - \boldsymbol{\phi_0}) \xrightarrow{\mathfrak{L}} N(\boldsymbol{0}, \dot{\boldsymbol{g}}(\boldsymbol{\theta_0})^{\mathsf{T}} I(\boldsymbol{\theta_0})^{-1}\dot{\boldsymbol{g}}(\boldsymbol{\theta_0}))$$