Asymptotic Normality of MLE: Multivariate Case

Seung Bong Jung

October 2021

In this note, we prove the asymptotic normality of MLE $\hat{\theta}$ in case $\Theta \subset \mathbb{R}^d$, where Θ is a parameter space. Recall the following regular conditions.

- **(R0)** The parameter θ is identifiable in Θ.
- **(R1)** The density $f(\cdot;\theta)$ have common support $\mathfrak{X}$.
- **(R2)** The parameter space is open in $\mathbb{R}^d$.
- **(R3)** The log-density pdf $\log f(x;\theta)$ is twice differentable as a function of θ for all $x \in \mathfrak{X}$.
- **(R4)** For any statistic $u(X_1,\ldots,X_n)$ with finite expectation, the integral

$$\mathrm{E}_\theta(u(X_1,\ldots,X_n)) = \int_{\mathfrak{X}^n} u(x_1,\ldots,x_n)\prod_{i=1}^n f(x_i;\theta)\prod_{i=1}^n dx_i$$

  is twice differentiable under the integral sign.
- **(R5)** The Fisher Information $I(\theta)$ exists and is invertible for all $\theta \in \Theta$.
- **(R6)** The likelihood equation $\dot{l}(\theta) = 0$ has the unique solution $\hat{\theta}$ and the solution is a consistent estimator of θ.
- **(R7)** For all $\theta \in \Theta$, there exists a function $M(\cdot)$ with $\mathrm{E}_\theta M(X_1) < \infty$ such that

$$\max_{\theta\in\Theta}\max_{1\le h,i,j\le d}\left|\frac{\partial^3}{\partial\theta_h\partial\theta_j\partial\theta_j}\log f(X_1;\theta)\right| \le M(X_1), \mathrm{E}_\theta M(X_1) < \infty.$$

Under these conditions, we attain asymptotic normality of MLE when $d \ge 2$. The proof is essentially the same with the case when $d = 1$.

Theorem. *Under **(R0)**-**(R7)** and assuming $\hat{\theta}$ is MLE of θ, show that*

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, I(\theta)^{-1}) \text{ for all } \theta\in\Theta \text{ in } P_\theta - \text{probability}.$$

Proof. Let $S(\theta) = \frac{1}{n}\ell(\theta)$, where $\ell(\theta) = \sum_{i=1}^n \log f(X_i;\theta)$. Write $\hat{\theta} = (\hat{\theta}_1,\ldots,\hat{\theta}_d)^t$ and $\theta = (\theta_1,\ldots,\theta_d)^t$. Then, by definition of MLE,

$$0 = \dot{S}(\hat{\theta}) = \dot{S}(\theta) + \ddot{S}(\theta)(\hat{\theta}-\theta) + \frac{1}{2}\tilde{S}(\theta^*)(\hat{\theta}-\theta), \tag{1}$$

for some $\theta^* = (\theta_1^*,\ldots,\theta_d^*)^t \in \Theta$ such that $|\theta_i^* - \theta_i| \le |\hat{\theta}_i - \theta_i|$ for $i = 1, 2\ldots, d$. Here $\tilde{S}(\theta) = (s_{ij}(\theta))_{1\le i,j\le d}$ is $d \times d$ matrix for $\theta \in \Theta$, where

$$s_{ij}(\theta) = \frac{1}{n}\left(\frac{\partial^3\ell}{\partial\theta_j^2\partial\theta_i}(\hat{\theta}_j-\theta_j) + \sum_{l=j,k\ne j}\frac{\partial^3\ell}{\partial\theta_k\partial\theta_j\partial\theta_i}(\hat{\theta}_k-\theta_k) + \sum_{k=j,l\ne j}\frac{\partial^3\ell}{\partial\theta_j\partial\theta_l\partial\theta_i}(\hat{\theta}_l-\theta_l)\right)$$

Suppose $\tilde{S}(\theta) = o_p(1)$. Note that $\ddot{S}(\theta) = -I(\theta) + o_p(1)$ under P_θ by weak law of large numbers (WLLN) and $\sqrt{n}\dot{S}(\theta) \xrightarrow{\mathrm{d}} N(0, I(\theta))$ by CLT. With these facts, (1) implies that

$$\begin{aligned}
0 &= \dot{S}(\theta) + [\ddot{S}(\theta) + \frac{1}{2}\tilde{S}(\theta^*)](\hat{\theta} - \theta) \\
&= \dot{S}(\theta) + [-I(\theta) + o_p(1)](\hat{\theta} - \theta)
\end{aligned}$$

and so

$$\begin{aligned}
[I(\theta) + o_p(1)](\hat{\theta} - \theta) &= \dot{S}(\theta), \\
\sqrt{n}[I_d + o_p(1)](\hat{\theta} - \theta) &= \sqrt{n}I(\theta)^{-1}\dot{S}(\theta) \xrightarrow{\mathrm{d}} N(0, I(\theta)^{-1}). \qquad (2)
\end{aligned}$$

Here I_d is $d \times d$ identity matrix and we used Slutsky's theorem for the last equation in (2). Thus if we show that $\tilde{S}(\theta) = o_p(1)$, we're done. Hence it suffices to show $\tilde{S}(\theta) = o_p(1)$. Observe that $\tilde{S}(\theta) = o_p(1)$ if and only if $s_{ij}(\theta) = o_p(1)$. Take a sequence $\{\epsilon_n\}_{n=1}^{\infty}$ such that $\epsilon_n \geq \epsilon_{n+1} \geq 0$ and $\epsilon_n \to 0$ as $n \to \infty$. By **(R6)**, there exists $\delta > 0$ and $N \in \mathbb{N}$ such that $P_\theta(||\hat{\theta} - \theta|| > \delta) < \epsilon_n$ for all $n \geq N$. Then for any $C > 0$,

$$\begin{aligned}
P_\theta(|s_{ij}(\theta)| > C) &\leq P_\theta(|s_{ij}(\theta)| > C, ||\hat{\theta} - \theta|| > \delta) + P_\theta(|s_{ij}(\theta)| > C, ||\hat{\theta} - \theta|| \leq \delta) \\
&\leq P_\theta(||\hat{\theta} - \theta|| > \delta) + P_\theta(|s_{ij}(\theta)| > C, ||\hat{\theta} - \theta|| \leq \delta) \\
&\leq \epsilon_n + P_\theta(|s_{ij}(\theta)| > C, ||\hat{\theta} - \theta|| \leq \delta). \qquad (3)
\end{aligned}$$

Let $l_m(\theta) = \log f(X_m; \theta)$. If $|s_{ij}(\theta)| > C, ||\hat{\theta} - \theta|| \leq \delta$,

$$\begin{aligned}
C < |s_{ij}(\theta)| &\leq \frac{1}{n}\left(\left|\frac{\partial^3 \ell}{\partial\theta_j^2\partial\theta_i}\right| \cdot |\hat{\theta}_j - \theta_j| + \sum_{l=j,k\neq j}\left|\frac{\partial^3 \ell}{\partial\theta_k\partial\theta_j\partial\theta_i}\right| \cdot |\hat{\theta}_k - \theta_k| + \sum_{k=j,l\neq j}\left|\frac{\partial^3 \ell}{\partial\theta_j\partial\theta_l\partial\theta_i}\right| \cdot |\hat{\theta}_l - \theta_l|\right) \\
&\leq \frac{\delta}{n}\left(\sum_{m=1}^{n}\left|\frac{\partial^3}{\partial\theta_j^2\partial\theta_i}l_m(\theta)\right| + \sum_{m=1}^{n}\sum_{l=j,k\neq j}\left|\frac{\partial^3}{\partial\theta_k\partial\theta_j\partial\theta_i}l_m(\theta)\right| + \sum_{m=1}^{n}\sum_{k=j,l\neq j}\left|\frac{\partial^3}{\partial\theta_j\partial\theta_l\partial\theta_i}l_m(\theta)\right|\right) \\
&\leq \frac{\delta}{n}(\sum_{m=1}^{n} M(X_m) + 2(d-1)\sum_{m=1}^{n} M(X_m)) \\
&= \frac{\delta(2d-1)}{n}\sum_{m=1}^{n} M(X_m)
\end{aligned}$$

and this gives

$$\frac{C}{\delta(2d-1)} \leq \frac{1}{n}\sum_{m=1}^{n} M(X_m). \qquad (4)$$

By (3) and (4),

$$\begin{aligned}
\limsup_{n\to\infty} P_\theta(|s_{ij}(\theta)| > C) &\leq \limsup_{n\to\infty} P_\theta\left(\frac{C}{\delta(2d-1)} \leq \frac{1}{n}\sum_{m=1}^{n} M(X_m)\right) + \limsup_{n\to\infty}\epsilon_n \\
&\leq \frac{\delta(2d-1)}{C} \cdot \mathrm{E}_\theta M(X_1) \qquad (5)
\end{aligned}$$

Here we used Markov's inequality for the second inequality in (5). Since (5) holds for arbitrary $\delta > 0$, one may conclude that $\limsup_{n\to\infty} P_\theta(|s_{ij}(\theta)| > C) = 0$, which implies $\lim_{n\to\infty} P_\theta(|s_{ij}(\theta)| > C) = 0$ as $0 \leq \liminf_{n\to\infty} P_\theta(|s_{ij}(\theta)| > C) \leq \limsup_{n\to\infty} P_\theta(|s_{ij}(\theta)| > C) = 0$. Therefore, $s_{ij}(\theta) = o_p(1)$, which concludes the proof.

□