Assumptions 6.2.2 (Additional Regularity Condition). *Regularity condition (R5) is*

(R5) *The pdf $f(x;\theta)$ is three times differentiable as a function of θ. Further, for all $\theta \in \Omega$, there exist a constant c and a function $M(x)$ such that*

$$\left|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)\right| \le M(x),$$

with $E_{\theta_0}[M(X)] < \infty$, for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X.

Theorem 6.2.2. *Assume $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta_0)$ for $\theta_0 \in \Omega$ such that the regularity conditions (R0)–(R5) are satisfied. Suppose further that the Fisher information satisfies $0 < I(\theta_0) < \infty$. Then any consistent sequence of solutions of the mle equations satisfies*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right). \tag{6.2.18}$$

Proof: Expanding the function $l'(\theta)$ into a Taylor series of order 2 about θ_0 and evaluating it at $\widehat{\theta}_n$, we get

$$l'(\widehat{\theta}_n) = l'(\theta_0) + (\widehat{\theta}_n - \theta_0)l''(\theta_0) + \frac{1}{2}(\widehat{\theta}_n - \theta_0)^2 l'''(\theta_n^*), \tag{6.2.19}$$

where θ_n^* is between θ_0 and $\widehat{\theta}_n$. But $l'(\widehat{\theta}_n) = 0$. Hence, rearranging terms, we obtain

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{n^{-1/2}l'(\theta_0)}{-n^{-1}l''(\theta_0) - (2n)^{-1}(\widehat{\theta}_n - \theta_0)l'''(\theta_n^*)}. \tag{6.2.20}$$

By the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}l'(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\partial \log f(X_i;\theta_0)}{\partial\theta} \xrightarrow{D} N(0, I(\theta_0)), \tag{6.2.21}$$

because the summands are iid with $\mathrm{Var}(\partial \log f(X_i;\theta_0)/\partial\theta) = I(\theta_0) < \infty$. Also, by the Law of Large Numbers,

$$-\frac{1}{n}l''(\theta_0) = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log f(X_i;\theta_0)}{\partial\theta^2} \xrightarrow{P} I(\theta_0). \tag{6.2.22}$$

To complete the proof then, we need only show that the second term in the denominator of expression (6.2.20) goes to zero in probability. Because $\widehat{\theta}_n - \theta_0 \xrightarrow{P} 0$ by Theorem 5.2.7, this follows provided that $n^{-1}l'''(\theta_n^*)$ is bounded in probability. Let c_0 be the constant defined in condition (R5). Note that $|\widehat{\theta}_n - \theta_0| < c_0$ implies that $|\theta_n^* - \theta_0| < c_0$, which in turn by condition (R5) implies the following string of inequalities:

$$\left|-\frac{1}{n}l'''(\theta_n^*)\right| \le \frac{1}{n}\sum_{i=1}^n \left|\frac{\partial^3 \log f(X_i;\theta)}{\partial\theta^3}\right| \le \frac{1}{n}\sum_{i=1}^n M(X_i). \tag{6.2.23}$$

By condition (R5), $E_{\theta_0}[M(X)] < \infty$; hence, $\frac{1}{n}\sum_{i=1}^n M(X_i) \xrightarrow{P} E_{\theta_0}[M(X)]$, by the Law of Large Numbers. For the bound, we select $1 + E_{\theta_0}[M(X)]$. Let $\epsilon > 0$ be given. Choose N_1 and N_2 so that

$$n \ge N_1 \quad\Rightarrow\quad P[|\widehat{\theta}_n - \theta_0| < c_0] \ge 1 - \frac{\epsilon}{2} \tag{6.2.24}$$

$$n \ge N_2 \quad\Rightarrow\quad P\left[\left|\frac{1}{n}\sum_{i=1}^n M(X_i) - E_{\theta_0}[M(X)]\right| < 1\right] \ge 1 - \frac{\epsilon}{2}. \tag{6.2.25}$$

It follows from (6.2.23)–(6.2.25) that

$$n \ge \max\{N_1, N_2\} \Rightarrow P\left[\left|-\frac{1}{n}l'''(\theta_n^*)\right| \le 1 + E_{\theta_0}[M(X)]\right] \ge 1 - \frac{\epsilon}{2};$$

hence, $n^{-1}l'''(\theta_n^*)$ is bounded in probability. ■

We next generalize Definitions 6.2.1 and 6.2.2 concerning efficiency to the asymptotic case.

Definition 6.2.3. *Let $X_1, \ldots, X_n$ be independent and identically distributed with probability density function $f(x;\theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \ldots, X_n)$ is an estimator of θ_0 such that $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{D} N\left(0, \sigma^2_{\hat{\theta}_{1n}}\right)$. Then*

(a) *The* **asymptotic efficiency** *of $\hat{\theta}_{1n}$ is defined to be*

$$e(\hat{\theta}_{1n}) = \frac{1/I(\theta_0)}{\sigma^2_{\hat{\theta}_{1n}}}. \tag{6.2.26}$$

(b) *The estimator $\hat{\theta}_{1n}$ is said to be* **asymptotically efficient** *if the ratio in part (a) is 1.*

(c) *Let $\hat{\theta}_{2n}$ be another estimator such that $\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{D} N\left(0, \sigma^2_{\hat{\theta}_{2n}}\right)$. Then the* **asymptotic relative efficiency** *(ARE) of $\hat{\theta}_{1n}$ to $\hat{\theta}_{2n}$ is the reciprocal of the ratio of their respective asymptotic variances; i.e.,*

$$e(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \frac{\sigma^2_{\hat{\theta}_{2n}}}{\sigma^2_{\hat{\theta}_{1n}}}. \tag{6.2.27}$$

Hence, by Theorem 6.2.2, under regularity conditions, maximum likelihood estimators are asymptotically efficient estimators. This is a nice optimality result. Also, if two estimators are asymptotically normal with the same asymptotic mean, then intuitively the estimator with the smaller asymptotic variance would be selected over the other as a better estimator. In this case, the ARE of the selected estimator to the nonselected one is greater than 1.