Lecture 18 Asymptotic Efficiency

18.1 Basic formulation and examples

Let $X_1, X_2, ...$ be iid samples from $f(x|\theta)$, $\theta \in \Theta \subset I\!R$, and T_n be a CANE of $q(\theta)$. It is reasonable to expect that

$$\liminf_{n\to\infty} \frac{Var_\theta T_n}{[q'(\theta)]^2/[nI_{X_1}(\theta)]} \geq 1 \quad \forall\ \theta,$$

where $I_{X_1}(\theta) = E_\theta \left[\frac{d}{d\theta} \log f(x|\theta)\right]^2$. The following definition provides a criterion for choosing a CANE.

Definition 1 *A CANE T_n is called the best asymptotically normal estimator (BANE), or an asymptotically efficient estimator (AEE), if*

$$\lim_{n\to\infty} \frac{Var_\theta T_n}{[q'(\theta)]^2/[nI_{X_1}(\theta)]} = 1 \quad \forall\ \theta. \tag{18.1}$$

Example 18.1 In the case of exponential families, $f(x|\theta) = \exp[\theta T(x) - \psi(\theta)]$. Recall that $T_n = \frac{1}{n}\sum_{i=1}^n T(X_i)$ is the UMVUE of $q(\theta) = \psi'(\theta)$, which implies for every positive integer n, we have $\frac{Var_\theta T_n}{[q'(\theta)]^2/[nI_{X_1}(\theta)]} = 1 \quad \forall\ \theta$ (note that $Var_\theta T_n = \psi''(\theta)/n$ and $I_{X_1}(\theta) = \psi''(\theta)$).

Example 18.2 Let $X_1 \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$. Recall that $T_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is the UMVUE for $q(\theta) = \sigma^2$, and $Var_\theta T_n = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$ (the C-R lower bound). Nevertheless, T_n is an AEE of σ^2, i.e. (18.1) holds.

Example 18.3 Let X_1 follow a Poisson distribution with parameter θ. Then the sample mean $T_n^{(1)} = \bar{X}$ is the UMVUE of θ, thus an AEE. The sample variance $T_n^{(2)} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased CANE of θ, but not an AEE, because $Var_\theta T_n^{(2)} = \frac{\theta(1+2\theta)}{n}$, $I_{X_1}(\theta) = 1/\theta$, and $\frac{Var_\theta T_n^{(2)}}{[nI_{X_1}(\theta)]^{-1}} = 1 + 2\theta > 1\ \forall\ \theta > 0$.

18.2 Consistency and asymptotic efficiency of MLEs

Theorem 1 *Let $X_1, X_2, ...$ be iid samples from $f(x|\theta)$. Assume that the MLE $\hat{\theta}$ of θ based on X^n is the unique solution of the likelihood equation, i.e. $\left[\frac{d}{d\theta} \log \prod_{i=1}^n f(x_i|\theta)\right]_{\theta=\hat{\theta}} = 0$. Under certain regularity conditions, $\hat{\theta}$ is an AEE.*

Proof:

Write $L(\theta|X_i) = \log f(X_i|\theta)$, $i = 1, ..., n$, then $L(\theta|X^n) = \sum_{i=1}^n L(\theta|X_i)$. We refer to the book *Mathematical Statistics* (2nd edition, Volume I, 2001) by Bickel and Doksum for the proof of consistency of $\hat{\theta}$, and focus on showing that $\hat{\theta}$ is a BANE (AEE).

A Taylor expansion yields

$$\left[\frac{d}{d\theta}L(\theta|X^n)\right]_{\theta=\hat{\theta}} - \frac{d}{d\theta}L(\theta|X^n) = \frac{d^2}{d\theta^2}L(\theta|X^n)(\hat{\theta}-\theta) + \frac{1}{2}\left[\frac{d^3}{d\theta^3}L(\theta|X^n)\right]_{\theta=\theta^*}(\hat{\theta}-\theta)^2,$$

where θ^* is between θ and $\hat{\theta}$. Hence

$$\sqrt{n}(\hat{\theta}-\theta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{d}{d\theta}L(\theta|X_i)}{-\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}L(\theta|X_i) - \frac{1}{2}(\hat{\theta}-\theta)\cdot\frac{1}{n}\sum_{i=1}^n\left[\frac{d^3}{d\theta^3}L(\theta|X_i)\right]_{\theta=\theta^*}}$$

Note that

(a) $\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{d}{d\theta}L(\theta|X_i) \xrightarrow{\mathcal{D}} Y \sim N(0, I_{X_1}(\theta))$;

(b) $-\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}L(\theta|X_i) \xrightarrow{P} I_{X_1}(\theta)$;

(c) Assuming $\left|\frac{d^3}{d\theta^3}L(\theta|x)\right| \leq M(x)\ \forall\ \theta \in \Theta$ and $\forall\ x \in \mathcal{X}$, with $E_\theta M(X_1) < \infty$, then

$$\left|\frac{1}{n}\sum_{i=1}^n \frac{d^3}{d\theta^3}L(\theta|X_i)\right| \leq \frac{1}{n}\sum_{i=1}^n M(X_i) \xrightarrow{P} E_\theta M(X_1),$$

which along with the consistency of $\hat{\theta}$ implies

$$\frac{-1}{2}(\hat{\theta}-\theta)\ \frac{1}{n}\sum_{i=1}^n\left[\frac{d^3}{d\theta^3}L(\theta|X_i)\right]_{\theta=\theta^*} \xrightarrow{P} 0.$$

Finally, it follows from (a), (b), (c) and the Slutsky's Theorem that

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{\mathcal{D}} Z \sim N(0, [I_{X_1}(\theta)]^{-1}).$$

QED.

Note: The results presented in this lecture are only for 1D parameter $\theta \in \mathbb{R}$. The case for multidimensional parameter $\theta = (\theta_1, ..., \theta_k) \in \mathbb{R}^k$ can be treated similarly with more technicalities (vector calculus, etc.) involved.

Example 18.4 (Hodges' superefficiency) Let $X_1 \sim N(\theta, 1)$. Then the MLE $\hat{\theta} = \bar{X}$ and $I_{X_1}(\theta) = 1$. Consider an alternative estimator $\tilde{\theta} = \bar{X} I_{\{|\bar{X}| > n^{-1/4}\}}$. Note that

$$P_\theta(|\bar{X}| \leq n^{-1/4}) = P(|Z + \sqrt{n}\theta| \leq n^{1/4}) = \Phi(n^{1/4} - \sqrt{n}\theta) - \Phi(-n^{1/4} - \sqrt{n}\theta)$$

where $Z \sim N(0,1)$. If $\theta \neq 0$, then $P_\theta(|\bar{X}| \leq n^{-1/4}) \to 0$, and $P_\theta(\tilde{\theta} = \bar{X}) \to 1$; if $\theta = 0$, then $P_\theta(\tilde{\theta} = 0) \to 1$. Therefore, $\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{D}} Z' \sim N(0, \sigma^2(\theta))$, where $\sigma^2(\theta) = 1 = [I_{X_1}(\theta)]^{-1}$ for $\theta \neq 0$, and $\sigma^2(0) = 0 < [I_{X_1}(\theta)]^{-1}$. Such a phenomenon, referred to as *superefficiency*, seems to have us question the asymptotic efficiency of MLEs even in the most standard setting (a normal population with an unknown mean). However, the following scrutiny will reveal a flaw of the superefficient estimator $\tilde{\theta}$, and suggest a minimax criterion under which MLEs are still preferred.

Let $\theta_n = \frac{c}{\sqrt{n}}$ with $c > 1$. Then

$$\begin{aligned} E_{\theta_n}[\sqrt{n}(\tilde{\theta} - \theta_n)]^2 &= E_{\theta_n}\left\{\sqrt{n}\left[\left(\frac{Z}{\sqrt{n}} + \frac{c}{\sqrt{n}}\right) I_{\left\{\left|\frac{Z}{\sqrt{n}} + \frac{c}{\sqrt{n}}\right| > n^{-1/4}\right\}} - \frac{c}{\sqrt{n}}\right]\right\}^2 \\ &= E\left[(Z+c)I_{\{|Z+c| > n^{1/4}\}} - c\right]^2 \\ &\xrightarrow{n\to\infty} c^2 > 1. \quad \text{(can you justify the limit?)} \end{aligned}$$

Note that $E_\theta[\sqrt{n}(\bar{X} - \theta)]^2 = 1 \ \forall\ \theta$, which means in the vicinity of $\theta = 0$ (here $\theta = 0$ is identified as a point for superefficiency) there always exist an infinite number of θ_n at which $\tilde{\theta}$ behaves worse than $\hat{\theta} = \bar{X}$ in terms of standardized mean squared errors. More precisely, in any neighborhood Θ_0 of the origin and for all sufficiently large n,

$$\sup_{\theta_n \in \Theta_0} E_{\theta_n}[\sqrt{n}(\tilde{\theta} - \theta_n)]^2 > \sup_{\theta \in \Theta_0} E_\theta[\sqrt{n}(\bar{X} - \theta)]^2,$$

hence

$$\sup_{\theta \in \Theta_0} E_\theta[\sqrt{n}(\tilde{\theta} - \theta)]^2 > \sup_{\theta \in \Theta_0} E_\theta[\sqrt{n}(\hat{\theta} - \theta)]^2.$$