Bui Tu Ha

Date No.

Theorem 10.1.12:

Let $X_1, X_2, \ldots$ be iid $f(x|\theta)$

Let $\hat{\theta}$ denote the MLE of θ and let $\tau(\theta)$ be a continuous function of θ.

Under the regularity conditions in Miscellanea 10.6.2 on $f(x|\theta)$ and, hence, $L(\theta|\underline{x})$, we have:

$$\sqrt{n}\,[\tau(\hat{\theta}) - \tau(\theta)] \to n[0; v(\theta)]$$

where $v(\theta) = \dfrac{1}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(x|\theta)\right)^2\right]}$ (the Cramér - Rao Lower Bound).

That is, $\tau(\hat{\theta})$ is a consistent and asymptotically efficient estimator of $\tau(\theta)$.

Proof:

Likelihood function: $\theta \to L(\theta|\underline{x}) := f(\underline{x}|\theta)$

Log likelihood function: $\ell(\theta|\underline{x}) = \log L(\theta|\underline{x}) = \sum \log f(x_i|\theta)$

Denote the derivatives with respect to θ by $\ell', \ell'', \ldots$

Using Taylor expansion around the true value θ_0,

$$\ell'(\theta|\underline{x}) = \ell'(\theta_0|\underline{x}) + (\theta-\theta_0)\ell''(\theta|\underline{x}) + \ldots,$$

where we are going to ignore the higher-order terms (under the regularity condition (A5) in 10.6.2)

Substitute the MLE $\hat{\theta}$ for θ and realize that $\ell'(\hat{\theta}) = 0$:

$$0 = \ell'(\theta_0|\underline{x}) + (\hat{\theta}-\theta_0)\ell''(\theta_0|\underline{x})$$

$$\Rightarrow \sqrt{n}\,(\hat{\theta}-\theta_0) = \sqrt{n}\,\frac{\ell'(\theta_0|\underline{x})}{-\ell''(\theta_0|\underline{x})} = \frac{\frac{1}{\sqrt{n}}\ell'(\theta_0|\underline{x})}{-\frac{1}{n}\ell''(\theta_0|\underline{x})} \quad (1)$$

Let $I(\theta_0) = E\{[\ell'(\theta_0|X_i)]^2\} = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right]\Bigg|_{\theta=\theta_0} = \dfrac{1}{v(\theta_0)}$ be the information number for one observation.

⊛ Consider the numerator in (1)

$$\frac{1}{\sqrt{n}}\ell'(\theta_0|\underline{X}) = \frac{1}{\sqrt{n}}\sum_i W_i = \sqrt{n}\left(\frac{1}{n}\sum_i W_i\right) = \sqrt{n}\,\overline{W}_n \text{ with } W_i(X_i) = \frac{\frac{d}{d\theta}f(X_i|\theta)}{f(X_i|\theta)}\Bigg|_{\theta=\theta_0} = \ell'(\theta_0|X_i) \text{ and } \overline{W}_n = \frac{1}{n}\sum_i W_i$$

⚠ W_i is considered as a random variable

Because $X_1, X_2, \ldots$ are iid, we also have $W_1, W_2, \ldots$ iid

$$\Rightarrow E_\theta(W_i) = E_\theta(\overline{W}_n) = \frac{1}{n}E_\theta(\ell'(\theta_0|\underline{X})) = \frac{1}{n}E_\theta\left[\frac{d}{d\theta}\left(\log f(\underline{X}|\theta)\right)\right] = \frac{1}{n}E_\theta\left[\frac{\frac{d}{d\theta}f(\underline{X}|\theta)}{f(\underline{X}|\theta)}\right]$$

Let $\mathcal{X}$ be the common support of the densities $f(x|\theta)$

$$\Rightarrow n E_\theta(W_i) = \int_{\mathcal{X}}\left(\frac{\frac{d}{d\theta}f(\underline{x}|\theta)}{f(\underline{x}|\theta)}\right)f(\underline{x}|\theta)\,d\underline{x} = \int_{\mathcal{X}}\frac{d}{d\theta}(f(\underline{x}|\theta))\,d\underline{x} = \frac{d}{d\theta}\int_{\mathcal{X}}1.\,f(\underline{x}|\theta)d\underline{x} = \frac{d}{d\theta}\Big(\underbrace{E_\theta(1)}_{=1}\Big) = 0$$

(see remark (⁎) in the end)

$$Var(W_i) = E_\theta(W_i^2) - [E_\theta(W_i)]^2 = I(\theta_0) - 0 = I(\theta_0)$$

Thus, W_i has mean 0 and variance $I(\theta_0)$

We have: $\frac{1}{\sqrt{n}}\, \ell'(\theta_0|\underline{X}) = \sqrt{n}\,(\overline{W_n})$

$(W_i)_{i\in\mathbb{N}}$ iid with $E_\theta(W_i) = 0$ and $Var_\theta(W_i) = I(\theta_0)$

By Central limit theorem, $Z_n = \sqrt{n}\,\frac{\overline{W_n}-0}{\sqrt{I(\theta_0)}}$ converges in distribution to the random variable having the standard normal distribution.

$$\Rightarrow \sqrt{n}\,\frac{\overline{W_n}}{\sqrt{I(\theta_0)}} \longrightarrow n[0;1] \qquad \text{(in distribution)}$$

$$\Rightarrow \frac{1}{\sqrt{n}}\,\ell'(\theta_0|\underline{X}) \longrightarrow n[0; I(\theta_0)] \qquad \text{(in distribution)} \quad (2)$$

⊛ Consider the denominator in (1)

$$-\frac{1}{n}\ell''(\theta_0|\underline{X}) = -\frac{1}{n}\frac{d}{d\theta}\left(\ell'(\theta_0|\underline{X})\right) = -\frac{1}{n}\frac{d}{d\theta}\left(\sum_i W_i\right) = -\frac{1}{n}\sum_i\left(\frac{d}{d\theta}W_i\right) = -\frac{1}{n}\sum_i\frac{d}{d\theta}\left(\frac{\frac{d}{d\theta}f(X_i|\theta)}{f(X_i|\theta)}\right)$$

$$= -\frac{1}{n}\sum_i\left[-\left(\frac{\frac{d}{d\theta}f(X_i|\theta)}{f(X_i|\theta)}\right)^2 + \frac{\frac{d^2}{d\theta^2}f(X_i|\theta)}{f(X_i|\theta)}\right] = \frac{1}{n}\sum_i W_i^2 - \sum_i\frac{\frac{d^2}{d\theta^2}f(X_i|\theta)}{f(X_i|\theta)}$$

$$E\left(\frac{1}{n}\sum_i W_i^2\right) = \frac{1}{n}\sum_i E(W_i^2) = \frac{1}{n}\sum_i E\left[(W_i-0)^2\right] = \frac{1}{n}\sum_i Var(W_i) = \frac{1}{n}\left(n\, I(\theta_0)\right) = I(\theta_0)$$

$$E\left(\frac{1}{n}\sum_i\frac{\frac{d^2}{d\theta^2}f(X_i|\theta)}{f(X_i|\theta)}\right) = \frac{1}{n}\sum_i E\left(\frac{\frac{d^2}{d\theta^2}f(X_i|\theta)}{f(X_i|\theta)}\right) = \frac{1}{n}\sum_i\int_{\mathcal{X}}\frac{\frac{d^2}{d\theta^2}f(x_i|\theta)}{\cancel{f(x_i|\theta)}}\,\cancel{f(x_i|\theta)}\,dx_i$$

$$= \frac{1}{n}\sum_i\frac{d^2}{d\theta^2}\underbrace{\left(\int_{\mathcal{X}}1.f(x_i|\theta)\,dx_i\right)}_{=E_\theta(1)=1} = 0$$

(see remark (⊛) in the end)

By weak law of large number, one has $\frac{1}{n}\sum_i W_i^2 \rightarrow I(\theta_0)$ in probability

and $\frac{1}{n}\sum_i\frac{\frac{d^2}{d\theta^2}f(X_i|\theta)}{f(X_i|\theta)} \rightarrow 0$ in probability

Therefore, one has:

$$-\frac{1}{n}\ell''(\theta_0|\underline{X}) \rightarrow I(\theta_0) \text{ in probability} \quad (3)$$

By Slutsky's theorem (appendix 4), from (1), (2) and (3) we have

$$\sqrt{n}\,(\hat{\theta}-\theta_0) \rightarrow n\left(0; \frac{1}{I(\theta_0)}\right) \text{ in distribution} \qquad \left(\frac{1}{I(\theta_0)} = v(\theta_0)\right)$$

Assume that $\tau(\theta)$ is differentiable at $\theta = \theta_0$. Using the delta method gives:

$$\sqrt{n}\,(\tau(\hat{\theta}) - \tau(\theta_0)) \rightarrow n\left[0; v(\theta).(\tau'(\theta))^2\big|_{\theta=\theta_0}\right] \text{ in distribution.}$$

Because θ_0 is not fixed, we can write: $\sqrt{n}\,(\tau(\hat{\theta}) - \tau(\theta)) \rightarrow n[0; v(\theta).(\tau'(\theta))^2]$ in distribution.

$\Rightarrow$ $\tau(\hat{\theta})$ is an asymptotically efficient estimator for $\tau(\theta)$.

Moreover,

normal distribution

$\sqrt{n}\,(\tau(\hat{\theta}) - \tau(\theta)) \rightarrow n(0; v(\theta))$ (in distribution)

$\tau(\hat{\theta}) - \tau(\theta) = \frac{\sqrt{v(\theta)}}{\sqrt{n}}\left(\sqrt{n}\,\frac{\tau(\hat{\theta}) - \tau(\theta)}{\sqrt{v(\theta)}}\right) \rightarrow \lim_{n\to\infty}\left(\frac{\sqrt{v(\theta)}}{\sqrt{n}}\, Z\right)$ with $Z \sim n(0;1)$ (in distribution)

$\Rightarrow \tau(\hat{\theta}) - \tau(\theta) \rightarrow 0$ in distribution

Theorem 5.5.13 [CB] states that: " The sequence of random variables $X_1, X_2, \ldots$ converges in probability to a constant μ if and only if the sequence also converges in distribution to μ."

Therefore, $\tau(\hat{\theta}) - \tau(\theta) \rightarrow 0$ in probability

$\Rightarrow$ $\tau(\hat{\theta})$ is a consistent estimator of $\tau(\theta)$.

Remark (*):

According to section 10.6.2("Suitable Regularity Conditions"), " These conditions mainly relate to differentiability of the density and the ability to interchange differentiation and integration." Therefore, we were able to use interchangeability between differentiation and integration twice in the proof.