Stat 543 4-13-05

large sample Thy for "maximum likelihood"

Theorem 6 (IID) Under some conditions a consistent root of the likelihood equation ($\ell_n'(\theta) = 0$) is asymptotically normal

Argument for Theorem

yet/still

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

abbreviate $\delta_n(X)$ to $\hat{\theta}_n$

Expand $\ell_n'(\cdot)$ around θ_0 in a Taylor Series

$$0 = \ell_n'(\hat{\theta}_n) = \ell_n'(\theta_0) + (\hat{\theta}_n - \theta_0)\,\ell_n''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\,\ell_n'''(\theta_1)$$

where θ_1 between θ_0, $\hat{\theta}_n$

So

$$-\ell_n'(\theta_0) = (\hat{\theta}_n - \theta_0)\left[\ell_n''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)\,\ell_n'''(\theta_1)\right]$$

and thus

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n}\,\ell_n'(\theta_0)}{\ell''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)\,\ell_n'''(\theta_1)}$$

So

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = -\frac{\underbrace{\frac{1}{\sqrt{n}} \ell_n'(\theta_0)}_{A_n}}{\underbrace{\frac{1}{n} \ell_n''(\theta_0)}_{B_n} + \underbrace{\frac{1}{n} \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right) \ell_n'''(\theta_1)}_{C_n}}$$

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i \mid \theta)\Big|_{\theta=\theta_0}$$

But the variables $\frac{d}{d\theta} \log f(X_i \mid \theta)\Big|_{\theta=\theta_0}$ under $\theta = \theta_0$ are iid with mean 0 (θ_0 mean score function at θ_0 for a single observation is 0) and variance

$$E_{\theta_0}\left(\frac{d}{d\theta} \log f(X_1|\theta)\Big|_{\theta=\theta_0} \right)^2 = I_1(\theta_0)$$

So $$A_n \xrightarrow{\mathcal{L}_{\theta_0}} N(0, I_1(\theta_0))$$

$$B_n = \frac{1}{n} l''_n(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log f(X_i|\theta)\Big|_{\theta=\theta_0}$$

But under θ_0 the variables $\frac{d^2}{d\theta^2} \log f(X_i|\theta)\Big|_{\theta=\theta_0}$

are iid with mean

$$E_\theta\left(\frac{d^2}{d\theta^2} \log f(X_1|\theta)\Big|_{\theta=\theta_0} \right) = -I_1(\theta_0)$$

☺ and the WLLN then says that

$$B_n \xrightarrow{P_{\theta_0}} -I_1(\theta_0)$$

To handle C_n, the hypotheses of the theorem are adequate to conclude

$$\hat{\theta}_n - \theta_0 \xrightarrow{P_{\theta_0}} 0$$

and to show that $\frac{1}{n} \ell_n'''(\theta_1)$ is "bounded in probability" so that

$$C_n = \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right) \frac{1}{n} \ell'''(\theta_1) \xrightarrow{P_{\theta_0}} 0$$

So then consider

$$g(u,v,w) = \frac{u}{-v-w}$$

and note that that it is cont$^{\underline{s}}$ except where $v+w=0$

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = g(A_n, B_n, C_n)$$

$$\Big\downarrow \mathcal{L}_{\theta_0}$$

$$N\left(0, \frac{1}{I_1(\theta)}\right) \qquad \frac{N(0, I_1(\theta_0))}{-(-I_1(\theta_0)) + 0}$$

This Theorem 6 from the handout doesn't quite yet give all we'd like for purposes of statistical inference

$$\sqrt{n}\,(\delta_n(X) - \theta) \;\dot{\sim}\; N\left(0, \frac{1}{I_1(\theta)}\right)$$

says

$$\frac{\delta_n(X) - \theta}{\frac{1}{\sqrt{n I_1(\theta)}}} \;\dot{\sim}\; N(0,1)$$

which e.g. leads to the unusable confidence limits for θ

$$\delta_n(X) \pm z \frac{1}{\sqrt{n I_1(\theta)}}$$

what is needed for a full tool inference is a way (or ways) to replace $I_1(\theta)$

2 possibilities

1) $I_1(\delta_n(X))$ > using the "Expected FI"

2) using observed curvature of log likelihood at $\delta_n(X)$ > using the "observed FI"