Stat 543 4-22-05

Recall

$\theta \in \mathbb{R}^k$ $\quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ $\quad \hat{\theta}_n$

$\theta^*_{n2}(\theta_1)$ maximizer of $\ell_n(\theta_1, \cdot)$

$$\ell^*_n(\theta_1) = \ell_n(\theta_1, \theta^*_{n2}(\theta_1)) = \underbrace{\max_{\theta_2} \ell_n(\theta_1, \theta_2)}_{\text{"profile loglikelihood"}}$$

Under appropriate conditions

Wilks' Theorem

$$2\left(\ell_n(\hat{\theta}_n) - \ell^*_n(\theta_{01})\right) \xrightarrow{\mathcal{L}_{\theta_0}} \chi^2_\ell$$

I can use this to set critical values for likelihood ratio tests (for large samples) of $H_0: \theta_1 = \theta_{01}$ and if c^* is the upper α pt of χ^2_ℓ

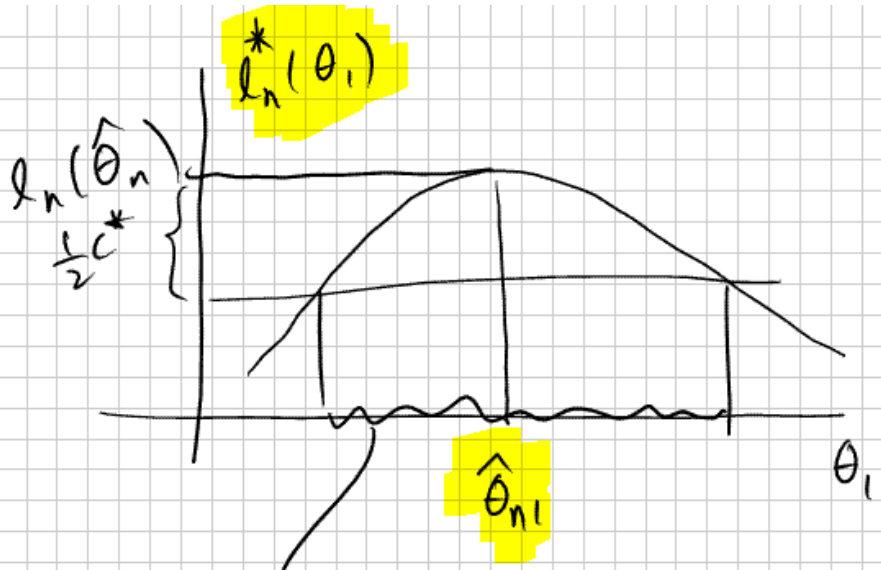
$$\underbrace{P_{\theta_0}\left[2\left(\ell_n(\hat{\theta}_n) - \ell_n^*(\theta_0)\right) < c^*\right]}_{P_{\theta_0}\left[\ell_n(\hat{\theta}_n) - \frac{1}{2}c^* < \ell_n^*(\theta_{01})\right]} \approx 1-\alpha$$

so with data $X = x$ and corresponding "MLE" $\hat{\theta}_n(x)$

$$\left\{\theta_1 \,\middle|\, \ell_n^*(\theta_1) > \ell_n^*(\hat{\theta}_n(x)) - \tfrac{1}{2}c^*\right\}$$

can be used as an approximate $(1-\alpha)\times 100\%$ confidence set for θ_1

$\ell = 1$ picture



approximate confidence set for θ_1

do a better job of holding a nominal coverage probability

Folklore says that usually this use of likelihood + profile loglikelihood to set confidence regions is more reliable than the use of large sample dsn of "MLE" (i.e. the confidence ellipsoids of HW 10)

Some large sample behavior of posterior dsns
(Section 5.5, 6.2.3) (to be followed by a brief intro to Score tests, etc. in B+D 6.3.2)

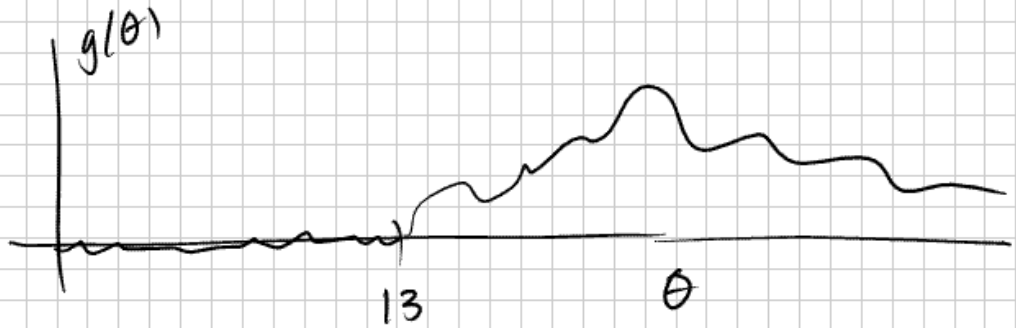
So $L(\theta)$ likelihood $g(\theta)$ prior

posterior $g(\theta \mid \text{data}) \propto L(\theta)\, g(\theta)$

what if the amount of data is large? (say, e.g., in an iid model $L_n(\theta)$ and $n \to \infty$?)

General Opening Comments

1) The posterior can not put any probability where the prior fails to do so



if $g(\theta) = 0$ Then $L(\theta)\, g(\theta) = 0$

(so Bayesians often try to "spread out" prior probability)

2) At least for larges samples, provided that $g(\theta)$ does spread its probability around

$$g(\theta \mid x)$$

posterior consistency

a) tend to "pile up" near θ_0 (under the θ_0 model)

b) tend to look approximately normal is cases where Θ is cont$^{\underline{s}}$ and this is possible

asymptotic normality of the posterior

$$L_n(\theta)\, g(\theta)$$

<u>Def</u> If X has pdf or pmf $f(x|\theta)$ the Kullback - Leibler Information in X is

$$K(\theta, \theta') = E_\theta \log \frac{f(X|\theta)}{f(X|\theta')}$$

(this is not usually symmetric in θ, θ')

<u>Example</u> $X \sim Bi(n, p)$

$$K(p, p') = E_p \log \left(\frac{p}{p'}\right)^X \left(\frac{1-p}{1-p'}\right)^{n-X}$$

$$= E_p \left[X \log \frac{p}{p'} + (n-X) \log \frac{(1-p)}{(1-p')} \right]$$

$$= np \log \frac{p}{p'} + (n - np) \log \left(\frac{1-p}{1-p'}\right)$$

$$= n \left[p \log\left(\frac{p}{p'}\right) + (1-p) \log\left(\frac{1-p}{1-p'}\right) \right]$$

<u>Example</u> $X \sim N(\theta, 1)$

$$K(\theta, \theta') = E_\theta \log \exp\left[-\tfrac{1}{2}(X-\theta)^2 + \tfrac{1}{2}(X-\theta')^2\right]$$

$$\vdots = \tfrac{1}{2}(\theta - \theta')^2$$

<u>Lemma</u> If X has pdf or pmf $f(x|\theta)$, then

$$K(\theta, \theta') \geq 0$$

and if θ and θ' produce different dsns for X, the inequality is strict