2-19-09 Math 152

but the distribution of θ^* need not be symmetric about $\hat{\theta}$.

A Sketch of Large Sample Theory for MLE's

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i | \theta).$$

$$l'(\hat{\theta}) = 0.$$

<u>Claim</u>: $\hat{\theta} \to \theta_0$ as $n \to \infty$

if f is sufficiently smooth.

(θ_0 is the "true" value of θ).

The argument goes as follows:

As $n \to \infty$

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

$$\to E(\log f(X|\theta))$$ by the law of large numbers.

$$= \int \log f(x|\theta)\, f(x|\theta_0)\, dx.$$

We argue that
the maximizer for $l(\theta)$
is well approximated by the
maximizer of $E(\log f(X|\theta))$.

$$\frac{\partial}{\partial\theta}\int \log f(x|\theta)\, f(x|\theta_0)\, dx$$

$$= \int \frac{\frac{\partial f}{\partial\theta}(x|\theta)}{f(x|\theta)}\, f(x|\theta_0)\, dx.$$

vanishes if $\theta = \theta_0$

since $\int \frac{\partial f}{\partial\theta}(x|\theta)\, dx$

$$= \frac{\partial}{\partial\theta}\left[\int f(x|\theta)\, dx\right] = \frac{\partial}{\partial\theta}(1) = 0.$$

f needs to be nice and smooth
to interchange differentiation and
integration as we have. $\square$

Further argument is required to show that $\theta = \theta_0$ is the global maximizer.

Now let

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right)^2\right].$$

<u>Claim:</u> If f is suff. smooth then

$$\sqrt{n I(\theta_0)}\,(\hat{\theta} - \theta_0) \longrightarrow N(0,1).$$

First we observe that.

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right].$$

"<u>pf:</u>" Since $\int f(X|\theta)\,dx = 1$.

$$\frac{\partial}{\partial\theta}\int f(X|\theta)\,dx = 0$$

(smoothness $\Rightarrow \int \frac{\partial}{\partial\theta} f(x|\theta)\,dx = 0$

$$\Rightarrow \int \frac{\partial}{\partial\theta}(\log f(x|\theta))\; f(x|\theta) = 0.$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int \frac{\partial}{\partial\theta}(\log f(x|\theta))\, f(x|\theta)\,dx = 0.$$

(smoothness)

$$\Rightarrow \int \frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\; f(x|\theta)\,dx$$

$$+ \int \frac{\partial}{\partial\theta}\log f(x|\theta)\; \frac{\partial}{\partial\theta}f(x|\theta)\,dx = 0.$$

$$= \int \frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\; f(x|\theta)\,dx$$

$$+ \int \left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2 f(x|\theta)\,dx = 0$$

$$\Rightarrow E\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] = - E\left(\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right)$$

Now the "proof" of the Claim.

$$0 = l'(\hat\theta) \simeq l'(\theta_0) + (\hat\theta-\theta_0)\, l''(\theta_0)$$

$$\Rightarrow (\hat\theta-\theta_0) \simeq -\frac{l'(\theta_0)}{l''(\theta_0)}.$$

$$\Rightarrow \sqrt{n}\,(\hat\theta-\theta_0) \simeq -\frac{\frac{1}{\sqrt{n}}\, l'(\theta_0)}{\frac{1}{n}\, l''(\theta_0)}.$$

Now (numerator).

$$E\left(\frac{1}{\sqrt{n}} l'(\theta_0)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta} \log f(X_i|\theta)\right]_{\theta=\theta_0}$$

$$= 0.$$

(again since $E\left[\frac{\partial}{\partial\theta} \log f(X|\theta)\right]$

$$= \int \frac{\partial}{\partial\theta} f(x|\theta)\, dx = 0.)$$

So $\mathrm{Var}\left(\frac{1}{\sqrt{n}} l'(\theta_0)\right)$

$$= \cancel{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{n} E\left(\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta)\right)\right)}$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\left(\left(\frac{\partial}{\partial\theta} \log f(X_i|\theta_0)\right)^2\right).$$

$$= I(\theta_0).$$

And, (denominator).

$$\frac{1}{n} l''(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2}\left(\log f(x_i|\theta=\theta_0)\right)$$

$$\to E\left[\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right]_{\theta=\theta_0} = -I(\theta_0)$$

by the law of large numbers.

We get

$$(*) \quad \sqrt{n}\,(\hat{\theta}-\theta_0) \approx \frac{\frac{1}{\sqrt{n}}\, l'(\theta_0)}{I(\theta_0)} \qquad n \text{ large.}$$

so that

$$E(\sqrt{n}\,(\hat{\theta}-\theta_0)) \approx 0.$$

and $\quad Var(\sqrt{n}\,(\hat{\theta}-\theta_0)) \approx \frac{1}{I(\theta_0)^2}\, I(\theta_0) = \frac{1}{I(\theta_0)}$

and $\therefore \quad Var(\hat{\theta}-\theta_0) \approx \frac{1}{n\,I(\theta_0)}.$

From $(*)$

$$\sqrt{n\,I(\theta_0)}\,(\hat{\theta}-\theta_0) \approx \frac{1}{\sqrt{n\,I(\theta_0)}}\, l'(\theta_0) = \frac{1}{\sqrt{n\,I(\theta_0)}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i \mid \theta_0).$$

and by C.L.T. the sum is approx normally distributed

So $\sqrt{n\,I(\theta_0)}\,(\hat{\theta}-\theta_0)$ is approx $N(0,1)$ for large n.

Note that the "asymptotic variance"

$$\mathrm{Var}(\hat{\theta}-\theta_0) \sim \frac{1}{nI(\theta_0)} = -\frac{1}{E(\ell''(\theta_0))}.$$