Feynman Liang
STAT 210A
Lecture #21
11/3/2016

1 Review

Recall that for the *log likelihood* $l(\theta; x) = \log p_\theta(x)$, we defined the *score* $\nabla_\theta l(\theta; x)$.

$$1 = \int p_\theta(x) d\mu(x) = \int e^{l(\theta;x)} d\mu(x) \tag{1.1}$$

$$\frac{\partial}{\partial\theta_j} \implies 0 = \int \frac{\partial}{\partial\theta_j} l(\theta; x) e^{l(\theta;x)} d\mu(x) = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_j} l(\theta; x)\right] \tag{1.2}$$

$$\frac{\partial}{\partial\theta_i} \implies 0 = \int \frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta; x) e^{l(\theta;x)} + \frac{\partial}{\partial\theta_i} l(\theta; x) \frac{\partial}{\partial\theta_j} l(\theta; x) e^{l(\theta;x)} d\mu(x) \tag{1.3}$$

$$= -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta; x)\right] = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_i} l(\theta; x) \frac{\partial}{\partial\theta_j} l(\theta; x)\right] \tag{1.4}$$

$$= \mathrm{Cov}_\theta\left(\frac{\partial}{\partial\theta_i} l(\theta; x), \frac{\partial}{\partial\theta_j} l(\theta; x)\right) \tag{1.5}$$

In 1 variable

$$\implies -\mathbb{E}_\theta[l''(\theta; x)] = \mathrm{Var}_\theta(l'(\theta; x)) = \underbrace{J(\theta)}_{\text{Fisher information}} \tag{1.6}$$

The CRLB: If $\hat\theta$ unbiased, $\mathrm{Var}(\hat\theta) \geq 1/J(\theta)$.
We showed that for exponential families

$$\sqrt{n}(\hat\theta - \theta) \Rightarrow N(0, J_1(\theta)^{-1}) \tag{1.7}$$

$$\therefore \hat\theta \approx N(\theta, (nJ_1(\theta)^{-1}) = N(\theta, J(\theta)^{-1}) \tag{1.8}$$

Can we show this generally for MLEs?

2 Asymptotic distribution of MLE

Setup: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} p_\theta(x)$, "smooth" in θ. Then under some conditions

$$\hat\theta \approx N(\theta, (nJ(\theta))^{-1}) \tag{2.1}$$

as long as $\hat\theta$ is consistent.

Proof. "Part worth remembering". Let θ_0 denote the true value. Then

$$\frac{1}{\sqrt{n}} l'(\theta_0; x) = \frac{1}{\sqrt{n}} \sum_i l'(\theta_0; x_i) \Rightarrow N(0, J_1(\theta_0)) \tag{2.2}$$

$$\frac{1}{n} l'(\theta_0; x) \xrightarrow{p} -J_1(\theta_0) \text{ by WLLN} \tag{2.3}$$

Do a Taylor expansion on $l'(\hat{\theta}; x)$

$$0 = \underbrace{l'(\hat{\theta}_j; x)}_{\hat{\theta} \text{ MLE}} = l'(\theta_0; x) + (\hat{\theta} - \theta_0) l''(\theta_0; x) + H.O.T. \tag{2.4}$$

so $\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{\frac{1}{\sqrt{n}} l'(\theta_0;x)}{-\frac{1}{n} l''(\theta_0;x)} \Rightarrow N(0, J_1(\theta_0)^{-1})$ by Slutsky's theorem.

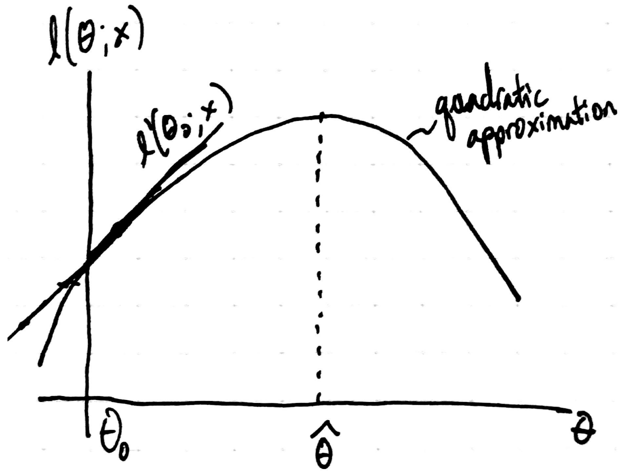

Figure 1: Taylor expansion to second term is a best quadratic approximation at $\theta = \theta_0$

TODO for completing the proof:

(a) Need to control H.O.T. (higher order terms)

(b) Need asymptotic consistency

□

Example 2.1 (Gaussian)**.**

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta_0, \sigma^2) \tag{2.5}$$

$$l(\theta; x) = n\bar{X}_n \frac{\theta}{\sigma^2} - n\frac{\theta^2}{2\sigma^2} - h(x) \tag{2.6}$$

$$l'(\theta; x) = \frac{n}{\sigma^2}(\bar{X} - \theta) \sim N(0, \frac{n}{\sigma^2}) \tag{2.7}$$

So

$$J_1(\theta) = \sigma^{-2}, \qquad l''(\theta; x) = -\frac{n}{\sigma^2} \tag{2.8}$$

and the polynomial approximation (in the Taylor series expansion) holds exactly

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, \sigma^2) \tag{2.9}$$

Now let's do it for real.

Theorem 2.2 (Keener 9.14). *Let $X_1, \dots, X_n \sim p_\theta(x)$. Suppose we have a dominated family $\mathcal{P} = \{p_\theta : \theta \in \Theta \subset \mathbb{R}\}$ and the following conditions hold:*

Twice-diff log-likelihood $\forall \theta \in \Theta, x \in \mathcal{X}$ *st* $p_\theta(x) > 0, l(\theta; x) = \log p_\theta(x)$ *has* 2 *continuous derivatives wrt* θ

Fisher information $\mathbb{E}_\theta l'(\theta; x) = 0$ *and* $Var_\theta(l'(\theta; x)) = -\mathbb{E} l''(\theta; x) = J(\theta)$

"Tame" 2nd derivative $\forall \theta \in \Theta^\circ$, $\exists \varepsilon > 0$ *st* $\mathbb{E}_\theta \sup_{\tilde{\theta} \in [\theta - \varepsilon, \theta + \varepsilon]} l''(\theta; x) < \infty$

MLE is consistent $\forall \theta \in \Theta$, $\hat{\theta}_n = \arg\max_\theta p_\theta(x) \xrightarrow{p} \theta$

Then $\forall \theta \in \Theta^\circ$

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, J_1(\theta)^{-1}) \tag{2.10}$$

First, a technical lemma:

Lemma 2.3. *Suppose $X_n \Rightarrow X$, $P(B_n) \to 1$ as $n \to \infty$. Let Z_n be arbitrary RVs. Then*

$$\tilde{X}_n = X_n 1_{B_n} + Z_n 1_{B_n^c} \Rightarrow X \tag{2.11}$$

Proof of lemma. Fix $\varepsilon > 0$.

$$P(|Z_n 1_{B_n^c} > \varepsilon) \leq P(B_n^c) \to 0 \tag{2.12}$$

$$P(|1_{B_n} - 1| > \varepsilon) \leq P(B_n^c) \to 0 \tag{2.13}$$

$$\therefore Z_n 1_{B_n^c} \xrightarrow{\text{P}} 0 \text{ and } 1_{B_n} \xrightarrow{\text{P}} 1 \tag{2.14}$$

so by Slutsky $\tilde{X} \Rightarrow X$. □

Proof of theorem. Fix $\theta \in \Theta^\circ$. Choose $\varepsilon > 0$ for which

(a) $[\theta - \varepsilon, \theta + \varepsilon] \subset \Theta^\circ$

(b) $\mathbb{E} \sup_{\tilde{\theta} \in [\theta - \varepsilon, \theta + \varepsilon]} l''(\theta; x) < \infty$

If $B_n = \{|\hat{\theta} - \theta| < \varepsilon\}$ then $P(B_n) \to 1$.

On B_n,

$$0 = l'(\hat{\theta}_n; x) = l'(\theta; x) + (\hat{\theta}_n - \theta) l''(\tilde{\theta}_n; x) \tag{2.15}$$

where (by Taylor's theorem) $\tilde{\theta}_n$ is between θ and θ_n, so $\tilde{\theta} \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ and therefore

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{\frac{1}{\sqrt{n}} l'(\theta; x)}{\frac{1}{n} l''(\hat{\theta}_n; x)} \tag{2.16}$$

By CLT, the numerator $\Rightarrow N(0, J_1(\theta))$.

Want denominator $\xrightarrow{\text{P}} J_1(\theta)$. If $\hat{\theta}_n \xrightarrow{\text{P}} \theta$ (i.e. MLE is consistent, assumed), then $\tilde{\theta} \xrightarrow{\text{P}} \theta$. □