Lecture 14: Asymptotic properties of the MLE

Ciaran Evans

Logistics

- HW 5 due Monday, February 24
- Exam 1 on Canvas; due Monday, March 3
- No other assignments due before spring break

Recap: Convergence in probability

Definition: A sequence of random variables $X_1, X_2, \ldots$ *converges in probability* to a random variable X if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

We write $X_n \xrightarrow{p} X$.

Convergence in distribution

Definition: A sequence of random variables $X_1, X_2, \ldots$ *converges in distribution* to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points where $F_X(x)$ is continuous. We write $X_n \xrightarrow{d} X$.

Convergence of the MLE

Suppose that we observe $Y_1, Y_2, Y_3, \ldots$ iid from a distribution with probability function $f(y|\theta)$, where $\theta \in \mathbb{R}^d$ is the parameter(s) we are trying to estimate. Let

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

log-likelihood

$$\widehat{\theta}_n = \text{argmax}_\theta \ \ell_n(\theta)$$

MLE for first n observations

$$\mathcal{I}_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(Y_i|\theta)\right]$$

under regularity conditions, this is Fisher information for one observation

Theorem: Under certain regularity conditions (to be discussed later),

(a) $\widehat{\theta}_n \xrightarrow{p} \theta$ (consistency)

(b) $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$ (asymptotic normality)

Asymptotic normality: proof approach

$(\theta \in \mathbb{R})$

Let $\ell_n'(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta)$, $\ell_n''(\theta) = \frac{\partial^2}{\partial\theta^2}\ell_n(\theta)$

Begin with a Taylor expansion of ℓ_n' around θ:

1st order Taylor expansion: $g(x) \approx g(a) + g'(a)(x-a)$ when x is close to a

$$\ell_n'(\widehat{\theta}_n) \approx \ell_n'(\theta) + (\hat{\theta}_n - \theta)\ell_n''(\theta)$$

$\hat{\theta}$ = MLE, so $\ell_n'(\hat{\theta}) = 0$

$$\Rightarrow \quad 0 \approx \ell_n'(\theta) + (\hat{\theta} - \theta)\ell_n''(\theta)$$

$$\Rightarrow \quad (\hat{\theta} - \theta) \approx \frac{\ell_n'(\theta)}{-\ell_n''(\theta)}$$

$$\Rightarrow \quad \sqrt{n}(\hat{\theta} - \theta) \approx \frac{\sqrt{n}\,\ell_n'(\theta)}{-\ell_n''(\theta)} = \frac{\frac{1}{\sqrt{n}}\ell_n'(\theta)}{-\frac{1}{n}\ell_n''(\theta)}$$

Asymptotic normality: proof approach

$X \sim N(0, \sigma^2)$

$aX \sim N(0, a^2\sigma^2)$

Using the Taylor expansion,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \approx \frac{\frac{1}{\sqrt{n}}\ell_n'(\theta)}{-\frac{1}{n}\ell_n''(\theta)}$$

Next, look at limits for the numerator and denominator:

- $\frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$
- $-\frac{1}{n}\ell_n''(\theta) \xrightarrow{p} \mathcal{I}_1(\theta)$

$$\frac{1}{\mathcal{I}_1(\theta)} N(0, \mathcal{I}_1(\theta)) = N(0, \mathcal{I}_1^{-1}(\theta))$$

Asymptotic normality: the numerator

Want to show: $\frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$

- CLT: for iid $X_1, X_2, \ldots$, under mild conditions
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i]\right) \xrightarrow{d} N(0, Var(X_i))$$
- $\ell_n'(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(Y_i|\theta)$

Applying CLT to $\ell_n'(\theta)$:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(Y_i|\theta) - \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right]\right)$$

$$\xrightarrow{d} N\left(0, Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right)\right)$$

Asymptotic normality: the numerator

Want to show: $\frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$

CLT gives

$$\sqrt{n}\left(\frac{1}{n}\ell_n'(\theta) - \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right]\right) \xrightarrow{d} N\left(0, Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right)\right)$$

Need to show:

- $\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right] =$ 0
- $Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right) =$ $\mathcal{I}_1(\theta)$ $= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)\right]$

The expected score

Claim: Under regularity conditions,

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right] = 0$$

$$\frac{\partial}{\partial\theta}\log f(Y_i|\theta) = \frac{1}{f(Y_i|\theta)}\left(\frac{\partial}{\partial\theta} f(Y_i|\theta)\right)$$

$$\Rightarrow \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y|\theta)\right] = \int_{-\infty}^{\infty} \frac{1}{\cancel{f(y|\theta)}}\left(\frac{\partial}{\partial\theta} f(y|\theta)\right) \cancel{f(y|\theta)}\, dy$$

$$= \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta} f(y|\theta)\right) dy$$

regularity conditions required to switch derivative & integral (C & B 2.4)

$$= \frac{\partial}{\partial\theta}\left(\int_{-\infty}^{\infty} f(y|\theta)\, dy\right)$$

$$= \frac{\partial}{\partial\theta}(1) = 0$$

//

Fisher information

From previous slide:

$$\frac{\partial}{\partial\theta}\log f(Y|\theta) = \frac{\frac{\partial}{\partial\theta} f(Y|\theta)}{f(Y|\theta)}$$

Claim: Under regularity conditions,

$$\overbrace{Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right)}^{\text{Fisher info (definition)}} = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)\right]$$

$$\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta) = \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta} f(Y|\theta)}{f(Y|\theta)}\right)$$

$$= \frac{\frac{\partial^2}{\partial\theta^2} f(Y|\theta)}{f(Y|\theta)} - \left(\frac{\frac{\partial}{\partial\theta} f(Y|\theta)}{f(Y|\theta)}\right)^2$$

$$= \frac{\frac{\partial^2}{\partial\theta^2} f(Y|\theta)}{f(Y|\theta)} - \left(\frac{\partial}{\partial\theta}\log f(Y|\theta)\right)^2$$

Fisher information

(continued)

$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Claim: Under regularity conditions,

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

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

$$Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(Y|\theta)\right)^2\right] - \left(\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y|\theta)\right]\right)^2 \leftarrow 0$$

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

$$\mathbb{E}\left[\frac{\frac{\partial^2}{\partial\theta^2} f(Y|\theta)}{f(Y|\theta)}\right] = \int_{-\infty}^{\infty}\left(\frac{\partial^2}{\partial\theta^2} f(y|\theta)\right)\frac{1}{\cancel{f(y|\theta)}}\cancel{f(y|\theta)}\,dy$$

(regularity)

$$= \frac{\partial^2}{\partial\theta^2}\left(\int_{-\infty}^{\infty} f(y|\theta)\,dy\right) = \frac{\partial^2}{\partial\theta^2}(1) = 0$$

Fisher information

Claim: Under regularity conditions,

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

Conclude:

$$-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right] = \underbrace{-\mathbb{E}\left[\frac{\frac{\partial^2}{\partial\theta^2} f(Y|\theta)}{f(Y|\theta)}\right]}_{=0 \text{ (regularity)}} + \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(Y|\theta)\right)^2\right]$$

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

$$= Var\left(\frac{\partial}{\partial\theta}\log f(Y|\theta)\right)$$

//

Numerator: putting everything together

Want to show: $\frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$

- CLT gives

$$\sqrt{n}\left(\frac{1}{n}\ell_n'(\theta) - \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right]\right) \xrightarrow{d} N\left(0, Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right)\right)$$

- Under regularity conditions,

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right] = 0$$

- Under regularity conditions,

$$Var\left(\frac{\partial}{\partial\theta}\log f(Y_i|\theta)\right) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)\right] = \mathcal{I}_1(\theta)$$

$$\Rightarrow \quad \sqrt{n}\left(\frac{1}{n}\ell_n'(\theta) - 0\right) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$$

$$\Rightarrow \quad \frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta)) \quad \checkmark$$

Now the denominator

Want to show: $-\frac{1}{n}\ell_n''(\theta) \xrightarrow{p} \mathcal{I}_1(\theta)$

Question: What big theorem do we have for convergence in probability?

The denominator: WLLN

Want to show: $-\frac{1}{n}\ell_n''(\theta) \xrightarrow{p} \mathcal{I}_1(\theta)$

- WLLN: For iid $X_1, X_2, \dots$, under mild conditions $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X_i]$
- $-\frac{1}{n}\ell_n''(\theta) = \frac{1}{n}\sum_{i=1}^{n} -\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)$

Applying WLLN to $-\frac{1}{n}\ell_n''(\theta)$:

$$-\frac{1}{n}\ell_n''(\theta) \xrightarrow{p} \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)\right]$$

$$= \mathcal{I}_1(\theta) \quad \checkmark$$