Asymptotic properties of maximum likelihood estimators

Ciaran Evans

Last time: Key results for the MLE

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

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

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

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

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

(b) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, v(\boldsymbol{\theta}))$

(where we still need to determine what the variance $v(\boldsymbol{\theta})$ should be!)

Last time: class activity

$$Y_1, ..., Y_n \overset{iid}{\sim} Bernoulli(p)$$

$$\ell_n(p) = \log(p)\left(\sum_{i=1}^n Y_i\right) + \log(1-p)\left(n - \sum_{i=1}^n Y_i\right)$$

$$\ell_n'(p) = \frac{\sum_{i=1}^n Y_i}{p} - \frac{\left(n - \sum_{i=1}^n Y_i\right)}{1-p} \qquad \approx \text{Normal}$$

$$\mathbb{E}[\ell_n'(p)] = 0 \qquad Var(\ell_n'(p)) = \frac{n}{p(1-p)}$$

$$\ell_n''(p) = -\frac{\sum_{i=1}^n Y_i}{p^2} - \frac{\left(n - \sum_{i=1}^n Y_i\right)}{(1-p)^2}$$

$$-\mathbb{E}[\ell_n''(p)] = \frac{n}{p(1-p)} = Var(\ell_n'(p))$$

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^n Y_i \qquad Var(\hat{p}) = \frac{p(1-p)}{n} = \frac{1}{Var(\ell_n'(p))}$$

$\approx$ Normal

The expected score

Let Y be a random variable with probability function $f(y|\boldsymbol{\theta})$.

Claim: Under regularity conditions,

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

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

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

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

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

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

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

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

Fisher information

Let Y be a random variable with probability function $f(y|\boldsymbol{\theta})$.

Definition: The **Fisher information** for a single sample Y, denoted $\mathcal{I}_1(\boldsymbol{\theta})$, is defined as

$$\mathcal{I}_1(\boldsymbol{\theta}) = Var\left(\frac{\partial}{\partial\boldsymbol{\theta}} \log f(Y|\boldsymbol{\theta})\right)$$

In general, for sample $Y_1, \dots, Y_n$,

$$\mathcal{I}_n(\theta) = Var\left(\frac{\partial}{\partial\theta} \log f(Y_1, \dots, Y_n|\theta)\right)$$

if iid: $\mathcal{I}_n(\theta) = \sum_{i=1}^n Var\left(\frac{\partial}{\partial\theta} \log f(Y_i|\theta)\right) = n\,\mathcal{I}_1(\theta)$

Claim: Under regularity conditions,

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

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,

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

Consider $\theta \in \mathbb{R}$

$$\frac{\partial^2}{\partial \theta^2} \log f(Y|\theta) = \frac{\partial}{\partial \theta}\left(\frac{1}{f(Y|\theta)}\left(\frac{\partial}{\partial \theta} f(Y|\theta)\right)\right)$$

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

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

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

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

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

Fisher information

Claim: Under regularity conditions,

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

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

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

(regularity)

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

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

Asymptotic normality: proof approach

We will sketch the proof in the case that $\theta \in \mathbb{R}$. Let $\ell_n'(\theta) = \frac{d}{d\theta}\ell_n(\theta)$, $\ell_n''(\theta) = \frac{d^2}{d\theta^2}\ell_n(\theta)$

Using a Taylor expansion of ℓ_n' around θ:

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

Numerator:

$$\frac{1}{\sqrt{n}}\ell_n'(\theta) = \sqrt{n}\left(\frac{1}{n}\ell_n'(\theta) \quad - 0\right)$$

$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\theta}\log f(Y_i|\theta) \quad - 0\right)$$

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

CLT: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \Rightarrow \sqrt{n}(\bar{X}-\mu) \xrightarrow{d} N(0,\sigma^2)$

$$\Rightarrow \frac{1}{\sqrt{n}}\ell_n'(\theta) \xrightarrow{d} N\left(0, Var\left(\frac{d}{d\theta}\log f(Y|\theta)\right)\right)$$

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

Asymptotic normality: proof approach

We will sketch the proof in the case that $\theta \in \mathbb{R}$. Let $\ell_n'(\theta) = \frac{d}{d\theta}\ell_n(\theta)$, $\ell_n''(\theta) = \frac{d^2}{d\theta^2}\ell_n(\theta)$

Using a Taylor expansion of ℓ_n' around θ:

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

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

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

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

Denominator:

$$-\frac{1}{n}\ell_n''(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta^2}\log f(Y_i|\theta)\right)$$

$$\underset{\text{(WLLN)}}{\xrightarrow{p}} -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(Y|\theta)\right] = \mathcal{I}_1(\theta)$$

$$\Rightarrow \frac{\frac{1}{\sqrt{n}}\ell_n'(\theta)}{-\frac{1}{n}\ell_n''(\theta)} \underset{\text{(Slutsky's)}}{\xrightarrow{d}} \frac{1}{\mathcal{I}_1(\theta)} N(0, \mathcal{I}_1(\theta)) = N(0, \mathcal{I}_1^{-1}(\theta))$$

So: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$

$$\hat{\theta}_n \approx N(\theta, (n\mathcal{I}_1(\theta))^{-1}) = N(\theta, \mathcal{I}_n^{-1}(\theta))$$