Asymptotic properties of maximum likelihood estimators

Ciaran Evans

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,

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

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

Application to regression models

Suppose that $(\mathbf{x}_1, Y_1), ..., (\mathbf{x}_n, Y_n)$ are iid from the linear regression model

$$Y_i | \mathbf{x}_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

The MLE is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Asymptotic normality of the MLE means that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}_1^{-1}(\boldsymbol{\beta}))$$

Fisher information for the linear regression model

$$Y_i|\mathbf{x}_i \sim N(\mathbf{x}_i^T\beta, \sigma^2)$$

Score: $U(\beta) = \frac{1}{\sigma^2}\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta)$ $= \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - x_i^T\beta) x_i$

Let $u_i = (Y_i - x_i^T\beta) x_i$

$\mathcal{I}_1(\beta) = Var\left(\frac{1}{\sigma^2} u_i\right) = \frac{1}{\sigma^4} Var(u_i)$

$Var(u_i) = Var(\mathbb{E}[u_i|x_i]) + \mathbb{E}[Var(u_i|x_i)]$

$\mathbb{E}[u_i|x_i] = 0$

$$\begin{aligned}
Var(u_i|x_i) &= \mathbb{E}\left[(u_i - \mathbb{E}[u_i|x_i])(u_i - \mathbb{E}[u_i|x_i])^T \,|\, x_i\right] \\
&= \mathbb{E}[u_i u_i^T | x_i] \\
&= \mathbb{E}\left[(Y_i - x_i^T\beta)^2 x_i x_i^T \,|\, x_i\right] \\
&= \mathbb{E}\left[(Y_i - x_i^T\beta)^2 | x_i\right] x_i x_i^T \\
&= Var(Y_i|x_i)\, x_i x_i^T \\
&= \sigma^2 x_i x_i^T
\end{aligned}$$

Fisher information for the linear regression model

$Y_i | \mathbf{x}_i \sim N(\mathbf{x}_i^T \beta, \sigma^2)$

Score: $U(\beta) = \frac{1}{\sigma^2} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta)$

Let $u_i = (Y_i - x_i^T \beta) x_i$

$\mathcal{I}_1(\beta) = Var\left(\frac{1}{\sigma^2} u_i\right) = \frac{1}{\sigma^4} Var(u_i)$

$Var(u_i) = Var(\mathbb{E}[u_i | x_i]) + \mathbb{E}[Var(u_i | x_i)]$

$Var(u_i) = Var(0) + \mathbb{E}[\sigma^2 x_i x_i^T]$

$= \sigma^2 \mathbb{E}[x_i x_i^T]$

$\Rightarrow \mathcal{I}_1(\beta) = \frac{1}{\sigma^2} \mathbb{E}[x_i x_i^T]$

Fisher information for the linear regression model

$Y_i | \mathbf{x}_i \sim N(\mathbf{x}_i^T \beta, \sigma^2)$

Fisher information: $\mathcal{I}_1(\beta) = \frac{1}{\sigma^2} \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^T]$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \sigma^2 \left(\mathbb{E}[x_i x_i^T]\right)^{-1}\right)$$

$$\hat{\beta} \approx N\left(\beta, \frac{\sigma^2}{n} \left(\mathbb{E}[x_i x_i^T]\right)^{-1}\right)$$

Need to estimate $\mathbb{E}[x_i x_i^T]$

Estimate $\mathbb{E}[x_i x_i^T]$ by $\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} X^T X$

(X = design matrix)

$$\Rightarrow \quad \hat{\beta} \approx N\left(\beta, \hat{\sigma}^2 (X^T X)^{-1}\right)$$