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,

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

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

Warmup

Work on the warmup activity. Solutions will be posted on the course website.

Some sufficient regularity conditions

Theorem: Under certain regularity conditions,

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

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

Conditions:

A counterexample

Suppose $Y_1, Y_2, \ldots \overset{iid}{\sim} \mathit{Uniform}(0, \theta)$.

A counterexample

Suppose that $Y_1, Y_2, \ldots \overset{iid}{\sim} Bernoulli(p)$.

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

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

Random vectors

Let $\mathbf{y} = (Y_1, ..., Y_d)^T \in \mathbb{R}^d$ be a **random vector**.

CDF: $F(y_1, ..., y_n) =$

Expected value:

Covariance matrix:

Properties of expectation and covariance matrix

Let $\mathbf{y} = (Y_1, ..., Y_d)^T$ be a random vector. Let $\mathbf{A}$ be a constant matrix, and $\mathbf{b}$ a constant vector.

Fisher information for the linear regression model

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

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

Fisher information for the linear regression model

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

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