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

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

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:

- The dimension of θ does not change with n
- $f(y|\theta)$ is a sufficiently smooth function of θ
- we can swap integration & differentiation (C & B section 2.4)
- The support of $f(y|\theta)$ does not depend on θ
- θ is identifiable (if $\theta_1 \neq \theta_2$, then $F(y|\theta_1) \neq F(y|\theta_2)$ for at least one y)
- θ is not on the boundary of the parameter space (e.g. Bernoulli(θ), θ not 0 or 1)

A counterexample

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

$$\hat{\theta}_n = Y_{(n)}$$

$$P(n(\theta - \hat{\theta}_n) \le t) \to 1 - e^{-t/\theta} \quad \text{for } t > 0$$

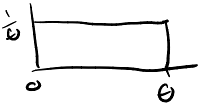
$$\Rightarrow \quad n(\theta - \hat{\theta}_n) \xrightarrow{d} \text{Exponential}(1/\theta)$$

$$\sqrt{n}(\hat{\theta}_n - \theta) = \underbrace{-\frac{1}{\sqrt{n}}}_{\to 0} \cdot \underbrace{n(\theta - \hat{\theta}_n)}_{\xrightarrow{d} \text{Exp}(1/\theta)} \xrightarrow{p} 0$$

(Slutsky's)

(not normal)

- $f(y|\theta)$ is not smooth at θ
  $\Rightarrow$ derivatives are not defined at θ
  $\mathcal{I}_1(\theta)$ is not defined
- Support of $f(y|\theta)$ depends on θ



A counterexample

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

$$\hat{p} = \bar{Y}$$

If $p = 0$: $\hat{p} = 0$ (all $Y_i = 0$)

$p = 1$: $\hat{p} = 1$ (all $Y_i = 1$)

$$\sqrt{n}(\hat{p} - p) \equiv 0 \overset{d}{\not\to} \text{Normal}$$

So: $p \in (0,1)$, then $\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p))$

but if $p = 0$ or $p = 1$, then $\sqrt{n}(\hat{p} - p) \overset{d}{\not\to} \text{Normal}$

Issue: $p = 0$ or 1 is on boundary of parameter space

Application to regression models

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

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

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

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

Need:

- random vectors
- multivariate normal distribution
- convergence for random vectors

Random vectors

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

$Y_1, Y_2, \ldots, Y_d$ are random variables

CDF: $F(y_1, \ldots, y_d) = P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_d \le y_d)$ (joint cdf)

Expected value:

$$\mathbb{E}[y] = \begin{bmatrix} \mathbb{E}[Y_1] \\ \mathbb{E}[Y_2] \\ \vdots \\ \mathbb{E}[Y_d] \end{bmatrix} \in \mathbb{R}^d$$

Covariance matrix:

$$\mathrm{Var}(y) = \begin{bmatrix} \mathrm{Var}(Y_1) & \mathrm{Cov}(Y_1, Y_2) & \cdots & \mathrm{Cov}(Y_1, Y_d) \\ \mathrm{Cov}(Y_2, Y_1) & \mathrm{Var}(Y_2) & & \vdots \\ \vdots & & \ddots & \mathrm{Var}(Y_d) \\ \mathrm{Cov}(Y_d, Y_1) & \cdots & & \end{bmatrix} \in \mathbb{R}^{d \times d}$$

<u>symmetric</u>

Let $\mu = \mathbb{E}[y]$

$$\mathrm{Var}(y) = \mathbb{E}[(y-\mu)(y-\mu)^T]$$

$$= \mathbb{E}[yy^T] - \mu\mu^T$$

Properties of expectation and covariance matrix

univariate case: $\mathbb{E}[aY + b] = a\mathbb{E}[Y] + b$

$\text{Var}(aY + b) = a^2 \text{Var}(Y)$

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.

Let $\mu = \mathbb{E}[y]$ and $\Sigma = \text{var}(y)$

① $\mathbb{E}[Ay + b] = A\mu + b$

② $\text{var}(Ay + b) = A\Sigma A^T$

Fisher information for the linear regression model

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

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

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

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

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

(law of total variance)

$\mathbb{E}[u_i | x_i] = \mathbb{E}[(Y_i - x_i^T\beta) x_i | x_i]$

$= \mathbb{E}[(Y_i - x_i^T\beta) | x_i] x_i$

$= (\mathbb{E}[Y_i | x_i] - x_i^T\beta) x_i$

$= 0$

$\mathbb{E}[Y_i | x_i] = x_i^T\beta$

Fisher information for the linear regression model

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

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