MZ
28.08.2017

Paremeter estimation (B.4)

MLR: β of interest σ^2: nuisence (not now)

$Y_1, Y_2, \dots, Y_n$ independent, each with distribution

$$f_i(y;\theta) \quad \leftarrow i=1,\dots,n$$

MLR: $Y_i \sim N(\mu_i, \sigma^2)$ where $\mu_i = x_i^T\beta \quad \leftarrow \theta: \beta$ vector

The likelihood $L(\theta; y)$ is the joint distribution
$y_1,\dots,y_n$

$$f(y;\theta) = \prod_{i=1}^{n} f_i(y_i;\theta) = L(\theta; y)$$
(unknown known) known
unknown

and $\ell(\theta;y) = \ln L(\theta;y) = \sum_{i=1}^{n} \underbrace{\ln f_i(y_i;\theta)}_{l_i(\theta)} = \sum_{i=1}^{n} l_i(\theta)$
natural log

We may estimate the peremeter θ by maximizing the likelihood (or equivalently the log-loglikelihood)

The maximum likelihood estimate $\hat\theta$ is the value of θ that
$$\ln L(\hat\theta;y) \geq \ln L(\theta;y) \quad \text{for all } \theta$$

MLR: $Y_i \sim N(x_i^T\beta, \sigma^2)$ $\quad i=1,\dots,n$ independent

$$L(\underset{\nearrow}{\theta}; y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2\right\}$$

β-vector

$$\ln L(\beta; y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 \quad \boxtimes$$

The vector of first partial derivatives of the log-likelihood is called the <u>score vector</u>

$$s(\theta) = \frac{\partial \ln L}{\partial \theta} = \left(\underbrace{\frac{\partial \ln L}{\partial \theta_1}}_{S_1(\theta)}, \frac{\partial \ln L}{\partial \theta_2}, \dots, \underbrace{\frac{\partial \ln L}{\partial \theta_p}}_{S_p(\theta)}\right)$$

and the observed Fisher information matrix is defined as the negative of the matrix of second derivatives of the log-likelihood

$$H(\theta) = -\begin{bmatrix} \frac{\partial^2 l}{\partial\theta_1\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_1\partial\theta_p} \\ & & \ddots & \\ & & & \frac{\partial^2 l}{\partial\theta_p\partial\theta_p} \end{bmatrix}_{p\times p}$$

$$= -\begin{bmatrix} \frac{\partial s_1(\theta)}{\partial\theta_1} & \frac{\partial s_1(\theta)}{\partial\theta_2} & \cdots & \frac{\partial s_1(\theta)}{\partial\theta_p} \\ \vdots & \vdots & & \\ & & & \frac{\partial s_p(\theta)}{\partial\theta_p} \end{bmatrix}$$

$H(\theta)$ may be considered as a local measure of information that the likelihood contains about the unknown param. θ. The higher the curvature of the log-likelihood near its maximum, the more information is provided by the likelihood about the unknown θ. Since $-H(\theta) \Rightarrow$ positive value near max.

To find the maximum likelihood (ML) estimator we solve

$$S(\theta) = 0$$

MLR:
$$\frac{\partial \ell}{\partial \beta} = 0$$

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial}{\partial \beta}\left\{ C - \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n} (y_i - x_i^T \beta)^2}_{(y - X\beta)^T (y - X\beta)} \right\} = 0$$

$$\Rightarrow \quad -2X^T y + 2X^T X \beta = 0$$

$$X^T y = X^T X \beta \qquad \text{normal eq.}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$