Parameter estimation [F3.2]

Estimator for β [F3.2.1]

1) Maximum likelihood $\qquad Y = \mathbb{X}\beta + \varepsilon$

If $\varepsilon \sim N_n(0, \sigma^2 I)$ then $Y \sim N_n(\mathbb{X}\beta, \sigma^2 I)$

Alt 1: $Y_1, Y_2, \ldots, Y_n$ independent

$E(Y_i) = x_i^T \beta \quad , \quad Var(Y_i) = \sigma^2$

$$\rightarrow f(y_i; \overset{x_i^T\beta}{\overset{\downarrow}{\mu}}, \sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \cdot e^{-\frac{1}{2\sigma^2}(y_i - \underset{\overset{\|}{x_i^T\beta}}{\mu})^2}$$

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

joint density of $Y_1, \dots, Y_n$

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2\right\}$$

homework

$$\circledast = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \underbrace{(y - X\beta)^T (y - X\beta)}_{LS(\beta)}\right\}$$

maximizing L wrt β is the same a minimizing $LS(\beta)$ wrt β

with respect to

Alt 2: $Y \sim N_n(\mu, \Sigma)$

$$f(y; \mu, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left[\det(\Sigma)\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu)\right\}$$

Homework: $\mu = X\beta$, $\Sigma = \sigma^2 I \Rightarrow$ get the same $L(\beta, \sigma^2)$ as $\circledast$

2) Least squares: minimize $LS(\beta)$ wrt β

$$LS(\beta) = (y - X\beta)^T (y - X\beta)$$

i) $LS(\beta) = y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta$

observe: $\underbrace{y^T X\beta}_{\text{scalar}} = \beta^T X^T y$

$$LS(\beta) = y^T y - 2y^T X\beta + \beta^T X^T X \beta$$

ii) To minimize $LS(\beta)$ wrt β we may solve

$$\frac{\partial LS(\beta)}{\partial \beta} = 0$$

$$\begin{pmatrix} \frac{\partial LS(\beta)}{\partial \beta[1]} \\ \frac{\partial LS(\beta)}{\partial \beta[2]} \\ \vdots \\ \frac{\partial LS(\beta)}{\partial \beta[p]} \end{pmatrix}$$

"Need" two rules for derivatives wrt vector:

Rule 1: $\frac{\partial}{\partial \beta}(\underset{1\times p}{d^T}\underset{p\times 1}{\beta}) = \frac{\partial}{\partial \beta}\left(\sum_{i=1}^{p} d_i \beta_i\right)$

$$= d$$

Rule 2: $\frac{\partial}{\partial \beta}(\beta^T \underset{\{d_{jk}\}}{D} \beta) = \frac{\partial}{\partial \beta}\left(\sum_{j=1}^{p}\sum_{k=1}^{p} \beta_j d_{jk} \cdot \beta_k\right)$

$= (D + D^T)\beta$ and $2D\beta$ when $D = D^T$.

Using the two rules:

$$\frac{\partial LS(\beta)}{\partial \beta} = 0$$

$$\frac{\partial}{\partial \beta}\left(y^T y - 2y^T X\beta + \beta^T X^T X \beta \right) = 0$$

Rules

$$d^T = -2y^T X$$
$$D = X^T X$$

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

$$\boxed{X^T y = X^T X \beta \quad \text{normal equations}}$$

iii) Solving the normal equations (substitute β with $\hat{\beta}$)

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

iv) min or max

$$\frac{\partial^2}{\partial \beta^2} LS(\beta) = \frac{\partial}{\partial \beta}\left(-2X^T y - 2X^T X \beta\right)\Big|_{\beta=\hat{\beta}}$$

$$= 2X^T X$$

If this matrix has only positive eigenvalues this will be the minimum.