9.3 MLE: The Case of Multiple Regression Model II

1. Regression model: $y = X\beta + u, \quad u \sim N(0, \sigma^2\Omega)$

   Transformation of Variables from u to y:

$$f_u(u) = (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2} \exp\Big(-\frac{1}{2\sigma^2}u'\Omega^{-1}u\Big)$$

$$\begin{aligned} f_y(y) &= f_u(y - X\beta)\left|\frac{\partial u}{\partial y'}\right| \\ &= (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2} \exp\Big(-\frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)\Big) \\ &= L(\theta; y, X), \end{aligned}$$

   where $\theta = (\beta, \sigma^2)$, because of $\dfrac{\partial u}{\partial y'} = I_n$.

   The log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta),$$

where $\theta = (\beta, \sigma^2)$.

2. $\max_{\theta} \quad \log L(\theta; y, X)$

$$\text{(FOC)} \quad \frac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$$

$$\text{(SOC)} \quad \frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'} \text{ is a negative definite matrix.}$$

Then, we obtain MLE of β and σ^2:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \qquad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta})}{n}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}\Big)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ, which is given by:

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'\Omega^{-1}X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

9.4 MLE: AR(1) Model

The pth-order Autoregressive Model, i.e., AR(p) Model (p 次の自己回帰モデル):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

AR(1) Model: $t = 2, 3, \cdots, n,$

$$y_t = \phi_1 y_{t-1} + u_t, \quad u_t \sim N(0, \sigma^2)$$

where $|\phi_1| < 1$ is assumed for now.

To obtain the joint density function of $y_1, y_2, \cdots, y_n$, $f(y_n, y_{n-1}, \cdots, y_1)$ is decomposed as follows:

$$f(y_n, y_{n-1}, \cdots, y_1) = f(y_1) \prod_{t=2}^{n} f(y_t | y_{t-1}, \cdots, y_1).$$

From $y_t = \phi_1 y_{t-1} + u_t$, we can obtain:

$$\mathrm{E}(y_t | y_{t-1}, \cdots, y_1) = \phi_1 y_{t-1}, \quad \text{and} \quad \mathrm{V}(y_t | y_{t-1}, \cdots, y_1) = \sigma^2.$$

Therefore, the conditional distribution $f(y_t | y_{t-1}, \cdots, y_1)$ is:

$$f(y_t | y_{t-1}, \cdots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_t - \phi_1 y_{t-1})^2 \right).$$

To obtain the unconditional distribution $f(y_t)$, y_t is rewritten as follows:

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + u_t \\
&= \phi_1^2 y_{t-2} + u_t + \phi_1 u_{t-1} \\
&\ \ \vdots \\
&= \phi_1^j y_{t-j} + u_t + \phi_1 u_{t-1} + \cdots + \phi_1^j u_{t-j} \\
&\ \ \vdots \\
&= u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots, \qquad \text{when } j \text{ goes to infinity.}
\end{aligned}$$

The unconditional expectation and variance of y_t is:

$$\mathrm{E}(y_t) = 0, \quad \text{and} \quad \mathrm{V}(y_t) = \sigma^2(1 + \phi_1^2 + \phi_1^4 + \cdots) = \frac{\sigma^2}{1 - \phi_1^2}.$$

Therefore, the unconditional distribution of y_t is given by:

$$f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}} \exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)} y_t^2\right).$$

Finally, the joint distribution of $y_1, y_2, \cdots, y_n$ is given by:

$$\begin{aligned} f(y_n, y_{n-1}, \cdots, y_1) &= f(y_1) \prod_{t=2}^{n} f(y_t | y_{t-1}, \cdots, y_1) \\ &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}} \exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)} y_1^2 \right) \\ &\qquad \times \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_t - \phi_1 y_{t-1})^2 \right) \end{aligned}$$

The log-likelihood function is:

$$\log L(\phi_1, \sigma^2; y_n, y_{n-1}, \cdots, y_1) = -\frac{1}{2}\log(2\pi\sigma^2/(1-\phi_1^2)) - \frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2$$
$$-\frac{n-1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(y_t - \phi_1 y_{t-1})^2.$$

Maximize $\log L$ with respect to ϕ_1 and σ^2.

Maximization Procedure:

- Newton-Raphson Method, or Method of Scoring
- Simple Grid Search (search maximization within the range $-1 < \phi_1 < 1$, changing the value of ϕ_1 by 0.01)