9.2 MLE: The Case of Multiple Regression Model I

1. Multivariate Normal Distribution: $X:\ n\times 1$ and $X \sim N(\mu, \Sigma)$

The density function of X is:

$$f(x) = (2\pi)^{n/2}|\Sigma|^{-1/2}\exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

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

Transformation of Variables from u to y:

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

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

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

Therefore, the log-likelihood function is:

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

Note that $|\Sigma|^{-1/2} = |\sigma^2 I_n|^{-1/2} = \sigma^{-n/2}$.

3. $\max_\theta \ \log L(\theta; y, X)$

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

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

We obtain MLE of β and σ^2:

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

where $\tilde\sigma^2$ is divided by n, not $n-k$.

4. Fisher's information matrix is:

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

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the

variance - covariance matrix for unbiased estimators of θ .

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

For large n, we approximately obtain: $\begin{pmatrix} \tilde\beta \\ \tilde\sigma^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}\right).$

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

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

(SOC) $\dfrac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}$ 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:

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

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 obian:

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

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:

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

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 $-1 < \rho < 1$, changing the value of ϕ_1 by 0.01)

9.5 MLE: Regression Model with AR(1) Error

When the error term is autocorrelated, the regression model is written as:

$$
y_t = x_t\beta + u_t, \qquad u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2).
$$

The joint distribution of $u_n, u_{n-1}, \cdots, u_1$ is:

$$
f_u(u_n, u_{n-1}, \cdots, u_1; \rho, \sigma_\epsilon^2) = f_u(u_1; \rho, \sigma_\epsilon^2) \prod_{t=2}^{n} f_u(u_t | u_{t-1}, \cdots, u_1; \rho, \sigma_\epsilon^2)
$$

$$
\begin{aligned}
&= (2\pi\sigma_\epsilon^2/(1-\rho^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)} u_1^2\right) \\
&\qquad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}(u_t - \rho u_{t-1})^2\right).
\end{aligned}
$$

By transformation of variables from $u_n, u_{n-1}, \cdots, u_1$ to $y_n, y_{n-1}, \cdots, y_1$, the joint distribution of $y_n, y_{n-1}, \cdots, y_1$ is:

$$
\begin{aligned}
&f_y(y_n, y_{n-1}, \cdots, y_1; \rho, \sigma_\epsilon^2, \beta) \\
&\quad = f_u(y_n - x_n\beta, y_{n-1} - x_{n-1}\beta, \cdots, y_1 - x_1\beta; \rho, \sigma_\epsilon^2) \left|\frac{\partial u}{\partial y'}\right|
\end{aligned}
$$

$$
\begin{aligned}
&= (2\pi\sigma_\epsilon^2/(1-\rho^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)}(y_1 - x_1\beta)^2\right) \\
&\qquad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}\Big((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta\Big)^2\right) \\
&= (2\pi\sigma_\epsilon^2)^{-1/2}(1-\rho^2)^{1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2}(\sqrt{1-\rho^2}y_1 - \sqrt{1-\rho^2}x_1\beta)^2\right) \\
&\qquad \times (2\pi\sigma_\epsilon^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}\Big((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta\Big)^2\right) \\
&= (2\pi\sigma_\epsilon^2)^{-n/2}(1-\rho^2)^{1/2} \exp\left(-\frac{1}{2\sigma_\epsilon^2}(y_1^* - x_1^*\beta)^2\right) \times \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}(y_t^* - x_t^*\beta)^2\right)
\end{aligned}
$$