Binary choice – 3.3 Maximum likelihood estimation

Michel Bierlaire

Solution of the practice quiz.

The first order necessary optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial \beta_k}(\widehat{\beta}) = 0, \; k = 1, \ldots, K, \tag{1}$$

and the log likelihood is

$$\mathcal{L}(\beta_1, \ldots, \beta_K) = \sum_{n=1}^{N} (y_{in} \ln P_n(i) + y_{jn} \ln P_n(j)). \tag{2}$$

In the case of the logit model

$$P_n(i) = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}}, \tag{3}$$

where

$$V_{in} = \sum_k \beta_k x_{ink}, \tag{4}$$

and x_{in} is the vector of explanatory variables associated with alternative i. Note that

$$\ln P_n(i) = V_{in} - \ln(e^{V_{in}} + e^{V_{jn}}). \tag{5}$$

Therefore,

$$\frac{\partial \ln P_n(i)}{\partial V_{in}} = 1 - P_n(i), \tag{6}$$

$$\frac{\partial \ln P_n(i)}{\partial V_{jn}} = -P_n(j). \tag{7}$$

Consequently, the partial derivative of (2) with respect to V_{in} is

$$\frac{\partial \mathcal{L}}{\partial V_{in}} = \sum_{n=1}^{N} (y_{in}(1 - P_n(i)) - y_{jn} P_n(i)) = \sum_{n=1}^{N} y_{in} - P_n(i), \tag{8}$$

as $y_{in} + y_{jn} = 1$. The partial derivative of (2) with respect to β_k is

$$\frac{\partial \mathcal{L}}{\beta_k} = \sum_{n=1}^{N} \sum_{i} \frac{\partial \mathcal{L}}{\partial V_{in}} \frac{\partial V_{in}}{\partial \beta_k} = \sum_{n=1}^{N} (y_{in} - P_n(i)) x_{ink} + (y_{jn} - P_n(j)) x_{jnk}, \tag{9}$$

as the utility function (4) is linear. As $y_{in} + y_{jn} = 1$ and $P_n(i) + P_n(j) = 1$, this simplifies to

$$\frac{\partial \mathcal{L}}{\beta_k} = \sum_{n=1}^{N} (y_{in} - P_n(i))(x_{ink} - x_{jnk}). \tag{10}$$

Therefore, the first order necessary optimality conditions are

$$\sum_{n=1}^{N} (y_{in} - P_n(i))(x_{ink} - x_{jnk}) = 0, \ \forall k = 1, \ldots, K. \tag{11}$$