2.3 Models for binary responses:

1. Modeling

In practice, the formal model usually embodies assumptions such as zero correlation or independence, lack of interaction or additivity, linearity and so on. These assumptions can not be taken for granted and should, if possible, be checked.

For binary data, to express π as the linear combination

$$\pi = \sum_{j=1}^{p} \beta_j x_j$$

would be inconsistent with the law of probability. A simple and effective way of avoiding this difficulty is to use a transformation $g(\pi)$ that maps the unit interval $[0, 1]$ onto the whole real line $(-\infty, \infty)$. That is,

$$g(\pi) = \sum_{j=1}^{p} \beta_j x_j = \eta.$$

Several functions (link functions) commonly used in practice are:

1. **The logit or logistic function**

$$g_1(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = logit(\pi).$$

2. **The probit or inverse normal function**

$$g_2(\pi) = \Phi^{-1}(\pi).$$

3. **The complementary log-log function**

$$g_3(\pi) = log[-log(1-\pi)].$$

4. **The log-log function**

$$g_4(\pi) = -log[-log(\pi)].$$

Note:

$$g_1(\pi) = -g_1(1-\pi)$$

and

$$g_3(\pi) = -g_3(1-\pi).$$

Note:

The required inverse functions are

1. **The logit or logistic function:**

$$\pi_1(\eta) = \frac{exp(\eta)}{1+exp(\eta)}.$$

2. **The probit or inverse normal function:**

$$\pi_2(\eta) = \Phi(\eta).$$

3. **The complementary log-log function:**

$$\pi_3(\eta) = 1 - exp[-exp(\eta)].$$

4. The log-log function:

$$\pi_4(\eta) = exp[-exp(-\eta)].$$

Note:

The logistic function is most commonly used link function.

Note:

For the data in the motivating example, suppose the logistic link function is used. Then,

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\Leftrightarrow \pi = \frac{exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}$$

$$\Leftrightarrow \frac{\partial \pi}{\partial x_j} = \pi(1-\pi)\beta_j, j = 1, 2.$$

The last equation implies that a larger change in π due to the change of x_j as π is near 0.5 than π is near 0 or π.

2. Estimation

Suppose $Y_i \sim B(m_i, \pi_i), i = 1, \cdots, n,$

with link function

$$\eta_i = g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^{p} \beta_j x_{ij}.$$

Note that $E(Y_i) = \mu_i = m_i \pi_i$. The likelihood function is

$$f(\pi|y) = \prod_{i=1}^{n} \binom{m_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{m_i - y_i}$$

and the log-likelihood function is

$$l(\beta) = log[f(\pi|y)] = \sum_{i=1}^{n} l_i(\beta)$$

$$= \sum_{i=1}^{n} \left\{ log\left[\binom{m_i}{y_i}\right] + y_i \cdot log(\pi_i) + (m_i - y_i) \cdot log(1-\pi_i) \right\}$$

$$= \sum_{i=1}^{n} \left[y_i \cdot log\left(\frac{\pi_i}{1-\pi_i}\right) + m_i \cdot log(1-\pi_i) \right] + \sum_{i=1}^{n} log\left[\binom{m_i}{y_i}\right]$$

Thus,

$$U_r(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \pi_i} \frac{\partial \pi_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - m_i \pi_i)}{\pi_i(1-\pi_i)} \pi_i (1-\pi_i)\, x_{ir}$$

$$= \sum_{i=1}^{n} (y_i - m_i \pi_i)\, x_{ir}$$

since

$$\begin{aligned}
\frac{\partial l_i}{\partial \pi_i} &= y_i \left(\frac{1-\pi_i}{\pi_i}\right) \left[\frac{1}{1-\pi_i} + \frac{\pi_i}{(1-\pi_i)^2}\right] - \frac{m_i}{1-\pi_i} \\
&= y_i \left(\frac{1-\pi_i}{\pi_i}\right) \frac{1}{(1-\pi_i)^2} - \frac{m_i}{1-\pi_i} \\
&= \frac{y_i}{\pi_i(1-\pi_i)} - \frac{m_i}{1-\pi_i} \\
&= \frac{y_i - m_i \pi_i}{\pi_i(1-\pi_i)}
\end{aligned}$$

and

$$\frac{\partial \pi_i}{\partial \eta_i} = \frac{1}{\left(\frac{\partial \eta_i}{\partial \pi_i}\right)} = \frac{1}{\left\{\frac{\partial \left[\log\left(\frac{\pi_i}{1-\pi_i}\right)\right]}{\partial \pi_i}\right\}}$$

$$= \frac{1}{\left(\frac{1-\pi_i}{\pi_i}\right)\left[\frac{1}{1-\pi_i} + \frac{\pi_i}{(1-\pi_i)^2}\right]} = \frac{1}{\left(\frac{1-\pi_i}{\pi_i}\right)\frac{1}{(1-\pi_i)^2}}$$

$$= \pi_i(1-\pi_i).$$

On the other hand,

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_s \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial[(y_i - m_i\pi_i)x_{ir}]}{\partial \beta_s} = -\sum_{i=1}^{n} m_i \frac{\partial \pi_i}{\partial \beta_s} x_{ir} = -\sum_{i=1}^{n} m_i \frac{\partial \pi_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_s} x_{ir}$$

$$= -\sum_{i=1}^{n} m_i \pi_i (1-\pi_i) x_{is} x_{ir}.$$

Therefore,

$$I_{sr}(\boldsymbol{\beta}) = -E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_s \partial \beta_r}\right] = -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_s \partial \beta_r} = \sum_{i=1}^{n} m_i \pi_i (1-\pi_i) x_{is} x_{ir}.$$

Denote

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \mu(\beta) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} m_1\pi_1 \\ m_2\pi_2 \\ \vdots \\ m_n\pi_n \end{bmatrix}$$

$$W(\beta) = \begin{bmatrix} m_1\pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & m_2\pi_2(1-\pi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n\pi_n(1-\pi_n) \end{bmatrix},$$

Then, in matrix form,

$$U(\beta) = X^t[y - \mu(\beta)]$$

and

$$I(\beta) = X^t W(\beta) X.$$

The Fisher's scoring method is

$$\begin{aligned} & I(\widehat{\beta}_t)\widehat{\beta}_{t+1} = I(\widehat{\beta}_t)\widehat{\beta}_t + U(\widehat{\beta}_t) \\ \Leftrightarrow\ & X^t W(\widehat{\beta}_t) X \widehat{\beta}_{t+1} = X^t W(\widehat{\beta}_t) X \widehat{\beta}_t + X^t[y - \mu(\widehat{\beta}_t)] \\ \Leftrightarrow\ & X^t W(\widehat{\beta}_t) X \widehat{\beta}_{t+1} = X^t W(\widehat{\beta}_t)\{X\widehat{\beta}_t + W^{-1}(\widehat{\beta}_t)[y - \mu(\widehat{\beta}_t)]\} \\ \Leftrightarrow\ & X^t W_t X \widehat{\beta}_{t+1} = X^t W_t z_t \\ \Leftrightarrow\ & \widehat{\beta}_{t+1} = (X^t W_t X)^{-1} X^t W_t z_t \end{aligned}$$

$t = 0, 1, 2, \cdots,$ **where**

$$W_t = W(\widehat{\beta}_t), z_t = \begin{bmatrix} z_{t1} \\ z_{t2} \\ \vdots \\ z_{tn} \end{bmatrix} = X\widehat{\beta}_t + W^{-1}(\widehat{\beta}_t)[y - \mu(\widehat{\beta}_t)]$$

and

$$z_{ti} = \sum_{j=1}^{p} x_{ij}\widehat{\beta}_{tj} + \left[\frac{y_i - \mu_i}{m_i\pi_i(1-\pi_i)}\right]_{\beta=\widehat{\beta}_t}.$$

Note:

A good choice of starting value usually reduced the number of cycles by about one or perhaps two.

Note:

Let

$$W_t = W(\widehat{\beta}) = \begin{bmatrix} m_1\pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & m_2\pi_2(1-\pi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n\pi_n(1-\pi_n) \end{bmatrix}_{\beta=\widehat{\beta}}.$$

1. $E(\widehat{\beta} - \beta) = O(n^{-1}).$

2. $\boldsymbol{E(\hat{\beta} - \beta) = (X^t W X)^{-1}[1 + O(n^{-1})]}$.

Note:

The above results are also true for the alternative limit in which $\boldsymbol{n}$ is fixed and $\boldsymbol{m_i} \to \infty$.