

## 1.4 Model estimation: IRLS:

### 1. Score function and information matrix:

Let the responses  $Y_i, i = 1, \dots, n$ , have the distribution in the exponential family, taking the form

$$f(y_i, \theta_i, \phi) = \exp \left\{ \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi) \right\}$$

with link function

$$g(\mu_i) = g[E(Y_i)] = \eta_i = x_i \beta = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip},$$

where

$$x_i = [x_{i1} \quad x_{i2} \quad \dots \quad x_{ip}].$$

Thus, the log-likelihood function for the *i*th observation is

$$l_i(\beta) = \log\{f(\beta, \phi | y_i)\} = \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi).$$

Since

$$\begin{aligned} b'(\theta_i) &= \mu_i \\ \Rightarrow \frac{\partial \theta_i}{\partial \mu_i} &= \frac{1}{\left(\frac{\partial \mu_i}{\partial \theta_i}\right)} = \frac{1}{b''(\theta_i)} \\ \Rightarrow \frac{\partial l_i}{\partial \beta_j} &= \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{[y_i - b'(\theta_i)]}{a(\phi)} \cdot \frac{1}{V_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ij} \end{aligned}$$

for  $j = 1, \dots, p$ , where we denote  $V_i = b''(\theta_i)$ . Define

$$w_i(\beta) = \frac{1}{V_i \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2}.$$

Then,

$$\begin{aligned} \frac{\partial l_i}{\partial \beta_j} &= \frac{[y_i - b'(\theta_i)]}{a(\phi)} \cdot \frac{1}{V_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ij} = \left( \frac{y_i - \mu_i}{a(\phi)} \right) \cdot \frac{1}{V_i \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2} \cdot \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ij} \\ &= \left( \frac{y_i - \mu_i}{a(\phi)} \right) \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \end{aligned}$$

since

$$\frac{\partial \eta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{\left( \frac{\partial \mu_i}{\partial \eta_i} \right)} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 1.$$

The likelihood function is

$$\prod_{i=1}^n f(\beta, \phi | y_i)$$

and

$$l = \log \left[ \prod_{i=1}^n f(\beta, \phi | y_i) \right] = \sum_{i=1}^n l_i(\beta).$$

Thus, the score function is

$$U_j(\beta) = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^n \left( \frac{y_i - \mu_i}{a(\phi)} \right) \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij}.$$

Further,

$$\begin{aligned} \frac{\partial^2 l_i}{\partial \beta_r \partial \beta_j} &= \frac{\partial \left( \frac{\partial l_i}{\partial \beta_j} \right)}{\partial \beta_r} = \frac{\partial \left[ \left( \frac{y_i - \mu_i}{a(\phi)} \right) \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \right]}{\partial \beta_r} \\ &= \frac{\partial \left[ \frac{y_i - \mu_i}{a(\phi)} \right]}{\partial \beta_r} \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} + \frac{y_i - \mu_i}{a(\phi)} \cdot \frac{\partial \left[ w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \right]}{\partial \beta_r}. \end{aligned}$$

and

$$\begin{aligned} &-E \left( \frac{\partial^2 l_i}{\partial \beta_r \partial \beta_j} \right) \\ &= \frac{1}{a(\phi)} \cdot \frac{\partial \mu_i}{\partial \beta_r} \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \\ &= \frac{1}{a(\phi)} \cdot \frac{\partial \eta_i}{\partial \beta_r} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ij} \\ &= \frac{1}{a(\phi)} \cdot w_i(\beta) \cdot x_{ij} \cdot \frac{\partial \eta_i}{\partial \beta_r} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} \right) \\ &= \frac{1}{a(\phi)} \cdot w_i(\beta) \cdot x_{ij} \cdot x_{ir} \end{aligned}$$

since

$$E \left[ \frac{Y_i - \mu_i}{a(\phi)} \right] = \frac{E(Y_i - \mu_i)}{a(\phi)} = 0$$

and

$$\frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} = 1.$$

Therefore,

$$I_{rj}(\beta) = -E \left( \frac{\partial^2 l}{\partial \beta_r \partial \beta_j} \right) = -\sum_{i=1}^n E \left( \frac{\partial^2 l_i}{\partial \beta_r \partial \beta_j} \right) = \sum_{i=1}^n \frac{1}{a(\phi)} \cdot w_i(\beta) \cdot x_{ij} \cdot x_{ir}$$

## 2. Fisher's scoring method:

The Fisher's scoring method is

$$[I(\hat{\beta}_t)\hat{\beta}_{t+1}]_r = [I(\hat{\beta}_t)\hat{\beta}_t]_r + U_r(\hat{\beta}_t), r = 1, 2, \dots, p,$$

where  $[I(\hat{\beta}_t)\hat{\beta}_{t+1}]_r$  and  $[I(\hat{\beta}_t)\hat{\beta}_t]_r$  are the  $r$ th component of  $I(\hat{\beta}_t)\hat{\beta}_{t+1}$  and  $I(\hat{\beta}_t)\hat{\beta}_t$ , respectively. Then,

$$\begin{aligned} [I(\hat{\beta}_t)\hat{\beta}_t]_r &= \sum_{j=1}^p I_{rj}(\hat{\beta}_t)\hat{\beta}_{tj} \\ &= \sum_{j=1}^p \sum_{i=1}^n \frac{w_{ti}x_{ir}x_{ij}\hat{\beta}_{tj}}{a(\phi)} \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti}x_{ir} \left( \sum_{j=1}^p x_{ij}\hat{\beta}_{tj} \right) \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti}x_{ir}\hat{\eta}_{ti} \end{aligned}$$

where

$$I_{rj}(\hat{\beta}_t) = \sum_{i=1}^n \frac{w_{ti}x_{ir}x_{ij}}{a(\phi)}, w_{ti} = w_i(\hat{\beta}_t),$$

$$\hat{\eta}_{ti} = \sum_{j=1}^p x_{ij}\hat{\beta}_{tj} = x_{i1}\hat{\beta}_{t1} + x_{i2}\hat{\beta}_{t2} + \dots + x_{ip}\hat{\beta}_{tp}.$$

Similarly,

$$[I(\hat{\beta}_t)\hat{\beta}_{t+1}]_r = \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti}x_{ir}\hat{\eta}_{(t+1)i} = \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti}x_{ir} \left( \sum_{j=1}^p x_{ij}\hat{\beta}_{(t+1)j} \right),$$

where

$$\hat{\eta}_{(t+1)i} = \sum_{j=1}^p x_{ij}\hat{\beta}_{(t+1)j} = x_{i1}\hat{\beta}_{(t+1)1} + x_{i2}\hat{\beta}_{(t+1)2} + \dots + x_{ip}\hat{\beta}_{(t+1)p}.$$

Further,

$$U_r(\hat{\beta}_t) = \sum_{i=1}^n \left[ \left( \frac{y_i - \mu_i}{a(\phi)} \right) \cdot w_i(\beta) \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ir} \right]_{\beta=\hat{\beta}_t}$$

$$= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \left[ (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \right]_{\beta=\hat{\beta}_t}$$

The **right** side of Fisher's scoring equation is

$$\begin{aligned} & [I(\hat{\beta}_t) \hat{\beta}_t]_r + U_r(\hat{\beta}_t) \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \hat{\eta}_{ti} + \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \left[ (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \right]_{\beta=\hat{\beta}_t} \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \left\{ \hat{\eta}_{ti} + \left[ (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \right]_{\beta=\hat{\beta}_t} \right\} \\ &= \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} z_{ti} \end{aligned}$$

while the **left** side of Fisher's scoring equation is

$$[I(\hat{\beta}_t) \hat{\beta}_{t+1}]_r = \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \hat{\eta}_{(t+1)i}$$

where

$$z_{ti} = \hat{\eta}_{ti} + \left[ (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \right]_{\beta=\hat{\beta}_t}.$$

Thus, Fisher's scoring equation can be reduced to

$$\begin{aligned} & [I(\hat{\beta}_t) \hat{\beta}_{t+1}]_r = [I(\hat{\beta}_t) \hat{\beta}_t]_r + U_r(\hat{\beta}_t) \\ & \Leftrightarrow \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} \left( \sum_{j=1}^p x_{ij} \hat{\beta}_{(t+1)j} \right) = \frac{1}{a(\phi)} \sum_{i=1}^n w_{ti} x_{ir} z_{ti} \\ & \Leftrightarrow \sum_{i=1}^n w_{ti} x_{ir} \left( z_{ti} - \sum_{j=1}^p x_{ij} \hat{\beta}_{(t+1)j} \right) = \mathbf{0} \end{aligned}$$

**Note:**

$z_{ti}$  depends only on  $\hat{\beta}_t$ .

### 3. Iterated reweighted least squares (IRLS):

The above equation can be expressed in matrix form,

$$X^t W_t (z_t - X \hat{\beta}_{t+1}) = 0,$$

where

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, W_t = \begin{bmatrix} w_{t1} & 0 & \cdots & 0 \\ 0 & w_{t2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{tn} \end{bmatrix}, z_t = \begin{bmatrix} z_{t1} \\ z_{t2} \\ \vdots \\ z_{tn} \end{bmatrix}.$$

Then, the maximum likelihood estimate at the  $(t + 1)$ th iteration is

$$\begin{aligned} X^t W_t z_t &= X^t W_t X \hat{\beta}_{t+1} \\ \Leftrightarrow \hat{\beta}_{t+1} &= (X^t W_t X)^{-1} X^t W_t z_t \end{aligned}$$

Note that  $\hat{\beta}_{t+1}$  can be thought as the weighted least squares estimate with weight matrix  $W_t$ , covariate matrix  $X$  and the response vector  $z_t$ . Thus,  $\hat{\beta}_t, t = 0, 1, 2, \dots$  can be generated by

$$\begin{aligned} \hat{\beta}_1 &= (X^t W_0 X)^{-1} X^t W_0 z_0 \\ \hat{\beta}_2 &= (X^t W_1 X)^{-1} X^t W_1 z_1 \\ &\vdots \\ \hat{\beta}_{t+1} &= (X^t W_t X)^{-1} X^t W_t z_t \\ &\vdots \end{aligned}$$

Note that the weight matrix  $W_t$  is **reweighted** (changed) at each iteration. Therefore, we also refer to  $\hat{\beta}_t, t = 0, 1, 2, \dots$ , as **iterated reweighted least squares estimate**.