# MATH 523: ITERATIVELY REWEIGHTED LEAST SQUARES

In the Exponential-Dispersion family, with weights $w_1, \ldots, w_n$ and identity dispersion parameter $\phi$ so that $a_i(\phi) = \phi/w_i$, the log-likelihood for independent data $\boldsymbol{y} = (y_1, \ldots, y_n)^\mathsf{T}$ with canonical parameters $\theta_i$ is

$$l\left(\theta, \phi; \boldsymbol{y}\right) = \sum_{i=1}^{n} \frac{w_i \left(y_i \theta_i - b(\theta_i)\right)}{\phi} + \sum_{i=1}^{n} c\left(y_i, \frac{\phi}{w_i}\right) \tag{1}$$

then the expectation and variance of observable $Y_i$ are

$$\mathbb{E}\left[Y_i\right] = \frac{\partial b(\theta_i)}{\partial \theta_i} = \dot{b}\left(\theta_i\right) = \mu_i \qquad \text{Var}\left[Y_i\right] = \frac{\phi}{w_i} \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} = \frac{\phi}{w_i} \ddot{b}\left(\theta_i\right) = \frac{\phi}{w_i} \frac{\partial \mu_i}{\partial \theta_i} = \frac{\phi}{w_i} V\left(\mu_i\right) \tag{2}$$

say. The GLM with link function $g$ sets

$$g\left(\mu_i\right) = g(\dot{b}(\theta_i)) = x_i \beta = \eta_i \tag{3}$$

where $x_i = (x_{i1}, \ldots, x_{ip})$, with parameters $\beta = (\beta_1, ..., \beta_p)^\mathsf{T}$. With the canonical link function $g$ we have merely that $g(\dot{b}(\theta_i)) = \theta_i$, so that $\theta_i = x_i \beta$.

A set of $p$ **score equations** is obtained by differentiation of (1) partially with respect to the elements of $\beta$ and equating to zero. For example, for $\beta_j$

$$\frac{\partial}{\partial \beta_j} \left\{l\left(\beta, \phi; \boldsymbol{y}\right)\right\} = \sum_{i=1}^{n} \frac{w_i}{\phi} \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j}\right) = \sum_{i=1}^{n} \frac{w_i}{\phi} \left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}\right) \frac{\partial \theta_i}{\partial \beta_j} = 0 \tag{4}$$

However, differentiating the LHS of (3) we have

$$\frac{\partial g\left(\mu_i\right)}{\partial \beta_j} = \frac{\partial g\left(\mu_i\right)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} = \dot{g}\left(\mu_i\right) V\left(\mu_i\right) \frac{\partial \theta_i}{\partial \beta_j}$$

by (2), and hence, differentiating the RHS of (3) and substituting back, we have that

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{1}{\dot{g}\left(\mu_i\right) V\left(\mu_i\right)} \frac{\partial \eta_i}{\partial \beta_j}.$$

In addition, in (4)

$$\frac{\partial b(\theta_i)}{\partial \theta_i} = \dot{b}(\theta_i) = \mu_i \qquad \frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$$

so that, finally, (4) becomes

$$\sum_{i=1}^{n} \frac{w_i}{\phi} \frac{\left(y_i - \mu_i\right) x_{ij}}{\dot{g}\left(\mu_i\right) V\left(\mu_i\right)} = 0. \tag{5}$$

Removing the constant $\phi$ and re-arranging gives the set of score equations

$$\sum_{i=1}^{n} \frac{w_i \left(y_i - \mu_i\right)}{V\left(\mu_i\right)} \frac{x_{ij}}{\dot{g}\left(\mu_i\right)} = 0 \qquad j = 1, ..., p \tag{6}$$

Note that if the canonical link is used

$$\theta_i = x_i \beta \quad \Longrightarrow \quad \frac{\partial \theta_i}{\partial \beta_j} = x_{ij}$$

and the equations simplify to

$$\sum_{i=1}^{n} w_i \left(y_i - \mu_i\right) x_{ik} = 0 \qquad \therefore \qquad \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} w_i \mu_i x_{ik} \qquad k = 1, ..., p \tag{7}$$

which we must solve simultaneously for $\beta_1, ..., \beta_k$ which appear in (7) as $\mu_i = g^{-1}\left(x_i\beta\right)$.

The system (6) cannot be solved analytically, but can be solved recursively. Let

$$Z_i = \eta_i + \left(Y_i - \mu_i\right)\dot{g}\left(\mu_i\right) = x_i\beta + \left(Y_i - \mu_i\right)\dot{g}\left(\mu_i\right) \tag{8}$$

then as $\mathbb{E}\left[Y_i\right] = \mu_i$ the expected value of $Z_i$ is

$$\mathbb{E}\left[Z_i\right] = \eta_i + \left(\mathbb{E}\left[Y_i\right] - \mu_i\right)\dot{g}\left(\mu_i\right) = \eta_i + 0 \times \dot{g}\left(\mu_i\right) = \eta_i.$$

and the variance of $Z_i$

$$\text{Var}\left[Z_i\right] = \text{Var}\left[Y_i\right]\left\{\dot{g}\left(\mu_i\right)\right\}^2 = \frac{\left\{\dot{g}\left(\mu_i\right)\right\}^2 \phi V\left(\mu_i\right)}{w_i}.$$

Hence, from (8) the $\{Z_i\}$ form a **linear model** in $\beta$ conditional on $\{\eta_i\}$ and $\{\mu_i\}$, and if observed values $\{z_i\}$ were available, the parameters in $\mu_i = g^{-1}\left(x_i^{\mathsf{T}}\beta\right)$ could be estimated using a **least squares** procedure, that is, by $\widehat{\beta}$ where

$$\widehat{\beta} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\boldsymbol{z} \tag{9}$$

and where $\mathbf{X}$ is the $n \times n$ matrix whose $n$ rows are $x_1, ..., x_n$, with $x_i = (x_{i1}, \ldots, x_{ip})^{\mathsf{T}}$ and

$$\mathbf{W} = \text{diag}\left(\frac{w_1}{\left\{\dot{g}\left(\mu_1\right)\right\}^2 V\left(\mu_1\right)}, \ldots, \frac{w_n}{\left\{\dot{g}\left(\mu_n\right)\right\}^2 V\left(\mu_n\right)}\right)$$

**ALGORITHM: ITERATIVELY REWEIGHTED LEAST SQUARES**

1 **Initialize**: choose any values for $\widehat{\mu}_i^{(0)}$ and $\widehat{\eta}_i^{(0)} = g(\widehat{\mu}_i^{(0)})$ for $i = 1, ..., n$

2 For $i = 1, ..., n$ set

$$\widehat{z}_i^{(1)} = \widehat{\eta}_i^{(0)} + (y_i - \widehat{\mu}_i^{(0)})\dot{g}(\widehat{\mu}_i^{(0)})$$

and

$$\widehat{w}_i^{(1)} = \frac{w_i}{\left\{\dot{g}(\widehat{\mu}_i^{(0)})\right\}^2 V\left(\widehat{\mu}_i^{(0)}\right)} \qquad \Longrightarrow \qquad \mathbf{W}^{(1)} = \mathbf{diag}\left(\widehat{w}_1^{(1)}, ..., \widehat{w}_n^{(1)}\right)$$

3 Compute estimates

$$\hat{\beta}^{(1)} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{W}^{(1)}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}^{(1)}z^{(1)}$$

4 Define vectors

$$\widehat{\eta}^{(1)} = \mathbf{X}\hat{\beta}^{(1)} \qquad\qquad \widehat{\mu}^{(1)} = g^{-1}\left(\widehat{\eta}^{(1)}\right)$$

5 Return to 2 and iterate through 2-5 using $\widehat{\mu}^{(1)}$ and $\widehat{\eta}^{(1)}$ as the updated starting values, eventually to obtain $\widehat{\mu}^{(2)}$ and $\widehat{\eta}^{(2)}$

6 Repeat until $\widehat{\mu}^{(t)}$ and $\widehat{\eta}^{(t)}$ satisfy

$$\left|\widehat{\mu}^{(t)} - \widehat{\mu}^{(t-1)}\right| < \epsilon_\mu \qquad\qquad \left|\widehat{\eta}^{(t)} - \widehat{\eta}^{(t-1)}\right| < \epsilon_\eta$$

for **tolerances** $\epsilon_\mu$ and $\epsilon_\eta$.