MATH 523: ITERATIVELY REWEIGHTED LEAST SQUARES: EXAMPLE

In the Exponential-Dispersion family, with $a_i(\phi) = \phi/w_i$ for a set of known constants $w_1, \ldots, w_n$, the p *score equations* for parameter $(\beta_1, \ldots, \beta_p)^\mathsf{T}$ are obtained by differentiation of the log-likelihood partially with respect to the elements of β and equating to zero. For example, for β_j

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

We have that

$$\mathbb{E}[Y_i] = \dot{b}(\theta_i) = \mu_i \qquad \text{Var}[Y_i] = a_i(\phi)\ddot{b}(\theta_i) = \frac{\phi}{w_i}\upsilon(\mu_i) = V(\mu_i)$$

say. After simplification, the score equations arising using link function $g(t)$ are, for $j = 1, ..., p$

$$\sum_{i=1}^{n} \frac{w_i}{\phi}\frac{\left(y_i - \mu_i\right)x_{ij}}{\dot{g}\left(\mu_i\right)\upsilon\left(\mu_i\right)} = 0. \qquad \text{or equivalently} \qquad \sum_{i=1}^{n} \frac{w_i\left(y_i - \mu_i\right)}{\upsilon\left(\mu_i\right)}\frac{x_{ij}}{\dot{g}\left(\mu_i\right)} = 0 \tag{2}$$

Consider the random quantity $g(Y_i)$; by a Taylor approximation $g(Y_i) \doteq g(\mu_i) + (Y_i - \mu_i)\dot{g}(\mu_i)$, so let

$$Z_i = \eta_i + \left(Y_i - \mu_i\right)\dot{g}\left(\mu_i\right) = x_i\beta + \left(Y_i - \mu_i\right)\dot{g}\left(\mu_i\right) \tag{3}$$

As $\mathbb{E}\left[Y_i\right] = \mu_i$, we have that $\mathbb{E}\left[Z_i\right] = \eta_i$, and

$$\text{Var}\left[Z_i\right] = \text{Var}\left[Y_i\right]\left\{\dot{g}\left(\mu_i\right)\right\}^2 = \frac{\left\{\dot{g}\left(\mu_i\right)\right\}^2 \phi\upsilon\left(\mu_i\right)}{w_i} = v_i \tag{4}$$

say. Hence, from (3) the $\{Z_i\}$ form a **linear model** in β given $\{\mu_i\}$, and if observed values $\{z_i\}$ were available, β could be estimated using a **least squares** procedure, that is, by $\widehat{\beta}$ where

$$\widehat{\beta} = \left(\mathbf{X}^\mathsf{T}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{W}\boldsymbol{z} \tag{5}$$

and where $\mathbf{X}$ is the $n \times n$ matrix whose n rows are $x_1, ..., x_n$, with $x_i = (x_{i1}, \ldots, x_{ip})$ and from (4)

$$\mathbf{W} = \text{diag}\left(v_1^{-1}, \ldots, v_n^{-1}\right)$$

ALGORITHM:

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

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

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

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

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

for **tolerances** ϵ_μ and ϵ_η.

Example:- Poisson log-linear model.

Suppose $Y_i \sim Poisson(\mu_i)$, and $g(\mu_i) = \log \mu_i = \beta_0 + \beta_1 x_i$. Here $v(\mu_i) = \mu_i$, $\phi = 1$ and $w_i = 1$ for all i. Also note that $\dot{g}(t) = 1/t$, so that

$$v(\mu_i)\dot{g}(\mu_i) = \mu_i/\mu_i = 1.$$

We write by the Taylor approximation in (3)

$$z_i = (\beta_0 + \beta_1 x_i) + \frac{(y_i - \mu_i)}{\mu_i}$$

Step 0: Set $\widehat{\beta}_0^{(0)} = \widehat{\beta}_1^{(0)} = 0$, so that $\widehat{\eta}_i^{(0)} = 0$, $\widehat{\mu}_i^{(0)} = 1$, for all i.

Step 1: For $i = 1, \ldots, n$, let

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

and

$$\widehat{w}_i^{(1)} = \frac{w_i}{\left\{\dot{g}(\widehat{\mu}_i^{(0)})\right\}^2 v\left(\widehat{\mu}_i^{(0)}\right)} = \frac{1}{\left\{1/\widehat{\mu}_i^{(0)}\right\}^2 \widehat{\mu}_i^{(0)}} = \widehat{\mu}_i^{(0)} = 1$$

thus defining $\mathbf{W}^{(1)} = \mathbf{diag}(\widehat{w}_1^{(1)}, ..., \widehat{w}_n^{(1)}) = \mathbf{I}_n$. Then set

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

and

$$\widehat{\eta}_i^{(1)} = \widehat{\beta}_0^{(1)} + \widehat{\beta}_0^{(1)} x_i \qquad\qquad \widehat{\mu}_i^{(1)} = \exp\{\widehat{\beta}_0^{(1)} + \widehat{\beta}_0^{(1)} x_i\}$$

Step t+1: For $i = 1, \ldots, n$, let

$$\widehat{z}_i^{(t+1)} = \widehat{\eta}_i^{(t)} + \frac{(y_i - \widehat{\mu}_i^{(t)})}{\widehat{\mu}_i^{(t)}} = (\widehat{\beta}_0^{(t)} + \widehat{\beta}_1^{(t)} x_i) + \frac{(y_i - \widehat{\mu}_i^{(t)})}{\widehat{\mu}_i^{(t)}}$$

and

$$\widehat{w}_i^{(t+1)} = \widehat{\mu}_i^{(t)}$$

thus defining $\mathbf{W}^{(t+1)} = \text{diag}(\widehat{\mu}_1^{(t)}, \ldots, \widehat{\mu}_n^{(t)})$. Then set

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

Iterate Repeat until convergence; check that the score equations from (2)

$$\sum_{i=1}^{n}(y_i - \widehat{\mu}_i^{(t)}) = 0 \qquad\qquad \sum_{i=1}^{n}(y_i - \widehat{\mu}_i^{(t)})x_i = 0$$

holds at the solution.