

MATH 523: ITERATIVELY REWEIGHTED LEAST SQUARES

In the exponential-dispersion family, with weights $w_1, \dots, w_n$ and identity dispersion function $a(\phi) = \phi$, the log-likelihood for independent data $\mathbf{y} = (y_1, \dots, y_n)^\top$ with canonical parameters θ_i is

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

For function $f(t)$, let $\dot{f}(t)$ and $\ddot{f}(t)$ denote the first and second partial derivatives of f with respect to t :

$$\dot{f}(t) = \frac{\partial f(t)}{\partial t} \quad \ddot{f}(t) = \frac{\partial^2 f(t)}{\partial t^2}$$

then the expectation and variance of observable Y_i are

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

say. The GLM with link function g sets

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

where $x_i = (x_{i1}, \dots, x_{ip})$, with parameters $\beta = (\beta_1, \dots, \beta_p)^\top$. 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 β and equating to zero. For example, for β_j

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

However, differentiating the LHS of (3) we have

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

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

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

In addition, in (4)

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

so that, finally, (4) becomes

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

Removing the constant ϕ and re-arranging gives the set of score equations

$$\sum_{i=1}^n \frac{w_i}{V(\mu_i)} \frac{(y_i - \mu_i) x_{ij}}{\dot{g}(\mu_i)} = 0 \quad j = 1, \dots, p \quad (6)$$

Note that if the canonical link is used

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

and the equations simplify to

$$\sum_{i=1}^n w_i (y_i - \mu_i) x_{ik} = 0 \quad \therefore \quad \sum_{i=1}^n w_i y_i = \sum_{i=1}^n w_i \mu_i x_{ik} \quad k = 1, \dots, p \quad (7)$$

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

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

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

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

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

and the variance of Z_i

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

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

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{z} \quad (9)$$

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

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

ALGORITHM: ITERATIVELY REWEIGHTED LEAST SQUARES

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

2 For $i = 1, \dots, n$ set

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

and

$$\hat{w}_i^{(1)} = \frac{w_i}{\{\dot{g}(\hat{\mu}_i^{(0)})\}^2 V(\hat{\mu}_i^{(0)})} \implies \mathbf{W}^{(1)} = \text{diag}(\hat{w}_1^{(1)}, \dots, \hat{w}_n^{(1)})$$

3 Compute estimates

$$\hat{\beta}^{(1)} = (\mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(1)} \mathbf{z}^{(1)}$$

4 Define vectors

$$\hat{\eta}^{(1)} = \mathbf{X} \hat{\beta}^{(1)} \quad \hat{\mu}^{(1)} = g^{-1}(\hat{\eta}^{(1)})$$

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

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

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

for **tolerances** ϵ_μ and ϵ_η .