Fisher scoring and iterative reweighted least squares (IRWLS)

Fisher scoring

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + F^{-1}(\hat{\beta}^{(t)})\, s(\hat{\beta}^{(t)})$$

$$s(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\, x_i}{Var(Y_i)} \cdot h'(\eta_i) \qquad = X^T \underbrace{D}_{diag(h'(\eta_i))} \underbrace{\Sigma^{-1}}_{diag\ Var(Y_i)} (y - \mu)$$

$$F(\beta)_{[h,l]} = \sum_{i=1}^{n} \frac{x_{ih}\, x_{il}\, [h'(\eta_i)]^2}{Var(Y_i)} = X^T \underbrace{W}_{diag\left[\frac{[h'(\eta_i)]^2}{Var(Y_i)}\right]} X$$

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + \left(X^T W(\hat{\beta}^{(t)}) X\right)^{-1} X^T \underbrace{D(\hat{\beta}^{(t)})\, \Sigma(\hat{\beta}^{(t)})^{-1}}_{diag(h'(\eta_i))\, diag\left(\frac{1}{Var(Y_i)}\right)} \left(y - \mu(\hat{\beta}^{(t)})\right)$$

$$I = \underbrace{\left(X^T W(\hat{\beta}^{(t)}) X\right)^{-1}}\ \underbrace{X W(\hat{\beta}^{(t)}) X}$$

$$\underbrace{diag(h'(\eta_i))\ diag\left(\frac{1}{Var(Y_i)}\right)}$$

$$diag\left[\frac{h_i'(\eta_i)}{Var(Y_i)}\right]$$

$$\underbrace{diag\left[\frac{h_i'(\eta_i)^2}{Var(Y_i)}\right]}_{W(\hat{\beta}^{(t)})} \cdot \underbrace{diag\left[\frac{1}{h_i'(\eta_i)}\right]}_{D(\hat{\beta}^{(t)})^{-1}}$$

$$\hat{\beta}^{(t+1)} = \left(X^T W(\hat{\beta}^{(t)}) X\right)^{-1} X^T W(\hat{\beta}^{(t)}) \left[\underbrace{X\hat{\beta}^{(t)}}_{\eta(\hat{\beta}^{(t)})} + D(\hat{\beta}^{(t)})^{-1} \left(y - \mu(\hat{\beta}^{(t)})\right) \right]$$

$$= \left(X^T W(\hat{\beta}^{(t)}) X\right)^{-1} X^T W(\hat{\beta}^{(t)}) \tilde{y}^{(t)}$$

$\tilde{y}^{(t)} = (\ldots, \tilde{y}_i(\hat{\beta}^{(t)}), \ldots)^T$ working response vector

$$\tilde{y}_i(\hat{\beta}^{(t)}) = x_i^T \hat{\beta}^{(t)} + d_i^{-1}(\hat{\beta}^{(t)}) \left(y_i - \hat{\mu}_i(\hat{\beta}^{(t)})\right)$$

$$\frac{1}{h_i'(\eta_i)} = \frac{1}{h_i'(x_i^T \hat{\beta}^{(t)})}$$

$$= x_i^T \hat{\beta}^{(t)} + \frac{y_i - h(x_i^T \hat{\beta}^{(t)})}{h'(x_i^T \hat{\beta}^{(t)})}$$

Working weights: $w_i(x_i^T \hat{\beta}^{(t)}) = \dfrac{h'(x_i^T \hat{\beta}^{(t)})}{Var(Y_i)}$

$W = diag(w_i)$

evaluate at $\hat{\beta}^{(t)}$

might involve $\hat{\phi}$

Remark: we need to invert $X^T W(\hat{\beta}^{(t)}) X$

$\Rightarrow$ ok if all weights are positive
wi in W

Usually: convergence established in few steps.

Canonical link: unique maximum

In general: should try several starts to achieve global max