Consider independent random variables $Y_1, ..., Y_N$ satisfying the properties of a generalized linear model. We wish to estimate parameters $\boldsymbol{\beta}$ which are related to the Y_i's through $\mathrm{E}(Y_i) = \mu_i$ and $g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$.

For each Y_i, the log-likelihood function is

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i) \tag{4.13}$$

where the functions b, c and d are defined in (3.3). Also

$$\mathrm{E}(Y_i) = \mu_i = -c'(\theta_i)/b'(\theta_i) \tag{4.14}$$

$$\mathrm{var}(Y_i) = [b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)] \,/\, [b'(\theta_i)]^3 \tag{4.15}$$

$$\text{and } g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i \tag{4.16}$$

where $\mathbf{x}_i$ is a vector with elements $x_{ij}, j = 1, ...p$.

The log-likelihood function for all the Y_i's is

$$l = \sum_{i=1}^{N} l_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i).$$

To obtain the maximum likelihood estimator for the parameter β_j we need

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \theta_i} . \frac{\partial \theta_i}{\partial \mu_i} . \frac{\partial \mu_i}{\partial \beta_j} \right] \tag{4.17}$$

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \theta_i} . \frac{\partial \theta_i}{\partial \mu_i} . \frac{\partial \mu_i}{\partial \beta_j} \right] \tag{4.17}$$

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i)$$

$$\frac{\partial \theta_i}{\partial \mu_i} = 1 \bigg/ \left(\frac{\partial \mu_i}{\partial \theta_i} \right) .$$

$$\begin{aligned} \frac{\partial \mu_i}{\partial \theta_i} &= \frac{-c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i) b''(\theta_i)}{[b'(\theta_i)]^2} \\ &= b'(\theta_i) \mathrm{var}(Y_i) \end{aligned}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} . \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij.}$$

Hence the score, given in (4.17), is

$$U_j = \sum_{i=1}^{N} \left[\frac{(y_i - \mu_i)}{\mathrm{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] . \tag{4.18}$$

$$U_j = \sum_{i=1}^{N} \left[\frac{(y_i - \mu_i)}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right]. \tag{4.18}$$

The variance-covariance matrix of the U_j's has terms

$$\mathfrak{I}_{jk} = E\left[U_j U_k\right]$$

which form the **information matrix** $\mathfrak{I}$. From (4.18)

$$\mathfrak{I}_{jk} = \text{E}\left\{ \sum_{i=1}^{N} \left[\frac{(Y_i - \mu_i)}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{l=1}^{N} \left[\frac{(Y_l - \mu_l)}{\text{var}(Y_l)} x_{lk} \left(\frac{\partial \mu_l}{\partial \eta_l} \right) \right] \right\}$$

$$= \sum_{i=1}^{N} \frac{\text{E}\left[(Y_i - \mu_i)^2\right] x_{ij} x_{ik}}{\left[\text{var}(Y_i)\right]^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \tag{4.19}$$

because $\text{E}[(Y_i - \mu_i)(Y_l - \mu_l)] = 0$ for $i \neq l$ as the Y_i's are independent. Using $\text{E}\left[(Y_i - \mu_i)^2\right] = \text{var}(Y_i)$, (4.19) can be simplified to

$$\mathfrak{I}_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \tag{4.20}$$

The estimating equation (4.11) for the method of scoring generalizes to

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \left[\mathfrak{I}^{(m-1)}\right]^{-1} \mathbf{U}^{(m-1)} \tag{4.21}$$

$$\mathfrak{I}^{(m-1)} \mathbf{b}^{(m)} = \mathfrak{I}^{(m-1)} \mathbf{b}^{(m-1)} + \mathbf{U}^{(m-1)}. \tag{4.22}$$

From (4.20) $\mathfrak{I}$ can be written as

$$\mathfrak{I} = \mathbf{X}^T \mathbf{W} \mathbf{X}$$

where $\mathbf{W}$ is the $N \times N$ diagonal matrix with elements

$$w_{ii} = \frac{1}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \tag{4.23}$$

The expression on the right-hand side of (4.22) is the vector with elements

$$\sum_{k=1}^{p} \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)} + \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

evaluated at $\mathbf{b}^{(m-1)}$; this follows from equations (4.20) and (4.18). Thus the right-hand side of equation (4.22) can be written as

$$\mathbf{X}^T \mathbf{W} \mathbf{z}$$

where $\mathbf{z}$ has elements

$$z_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \tag{4.24}$$

with μ_i and $\partial \eta_i / \partial \mu_i$ evaluated at $\mathbf{b}^{(m-1)}$.

Hence the iterative equation (4.22), can be written as

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{b}^{(m)} = \mathbf{X}^T \mathbf{W} \mathbf{z}. \tag{4.25}$$

The information matrix $\mathcal{I}(\boldsymbol{\beta})$, Chapter 5.2 in book.

$$\boldsymbol{U}(\boldsymbol{\beta}) = \begin{bmatrix} U_1(\boldsymbol{\beta}) \\ \vdots \\ U_p(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix}$$

$$E_{\boldsymbol{\beta}}[\boldsymbol{U}(\boldsymbol{\beta})] = \begin{bmatrix} E_{\boldsymbol{\beta}}[U_1(\boldsymbol{\beta})] \\ \vdots \\ E_{\boldsymbol{\beta}}[U_p(\boldsymbol{\beta})] \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{0}$$

$$E_{\boldsymbol{\beta}}\left[-\frac{\partial \boldsymbol{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right] = E_{\boldsymbol{\beta}}\left[\boldsymbol{U}(\boldsymbol{\beta})\boldsymbol{U}(\boldsymbol{\beta})^T\right] = Cov_{\boldsymbol{\beta}}[\boldsymbol{U}(\boldsymbol{\beta})] \equiv \mathcal{I}(\boldsymbol{\beta})$$

The first equality above is the one that needs a proof, while the second follows directly from $E_{\boldsymbol{\beta}}[\boldsymbol{U}(\boldsymbol{\beta})] = \boldsymbol{0}$. We may also write this as

$$\mathcal{I}(\boldsymbol{\beta})_{ij} = E_{\boldsymbol{\beta}}\left[-\frac{\partial U_i(\boldsymbol{\beta})}{\partial \beta_j}\right] = Cov_{\boldsymbol{\beta}}[U_i(\boldsymbol{\beta}), U_j(\boldsymbol{\beta})] = E_{\boldsymbol{\beta}}[U_i(\boldsymbol{\beta})U_j(\boldsymbol{\beta})]$$

for $i, j = 1, \ldots, p$.

Distribution of MLE, Chapter 5.4 in book.

From the previous slide follows the asymptotic result (as $N \to \infty$):

$$\boldsymbol{U}(\boldsymbol{\beta}) \approx N_p(\boldsymbol{0}, \mathcal{I}(\boldsymbol{\beta}))$$

Now let $\boldsymbol{b}$ be MLE of $\boldsymbol{\beta}$, so that $\boldsymbol{U}(\boldsymbol{b}) = \boldsymbol{0}$. Let $\boldsymbol{\beta}$ be true value of parameter vector.

Taylor:

$$\begin{aligned} \boldsymbol{0} = \boldsymbol{U}(\boldsymbol{b}) &\approx \boldsymbol{U}(\boldsymbol{\beta}) + \frac{\partial \boldsymbol{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}(\boldsymbol{b} - \boldsymbol{\beta}) \\ &\approx \boldsymbol{U}(\boldsymbol{\beta}) - \mathcal{I}(\boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta}) \end{aligned}$$

so

$$\boldsymbol{b} - \boldsymbol{\beta} \approx \mathcal{I}(\boldsymbol{\beta})^{-1}\boldsymbol{U}(\boldsymbol{\beta})$$

Hence [using formula from Multivariate Analysis: $Cov(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\,Cov(\boldsymbol{X})\boldsymbol{A}^T$]:

$$Cov_{\boldsymbol{\beta}}(\boldsymbol{b} - \boldsymbol{\beta}) \approx \mathcal{I}(\boldsymbol{\beta})^{-1}\mathcal{I}(\boldsymbol{\beta})\mathcal{I}(\boldsymbol{\beta})^{-1} = \mathcal{I}(\boldsymbol{\beta})^{-1}$$

so

$$\boldsymbol{b} \approx N_p(\boldsymbol{\beta}, \mathcal{I}(\boldsymbol{\beta})^{-1})$$

Hence also [using result from Multivariate Analysis: If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi_p^2$].

$$(\boldsymbol{b} - \boldsymbol{\beta})^T \, \mathcal{I}(\boldsymbol{\beta}) \, (\boldsymbol{b} - \boldsymbol{\beta}) \approx \chi_p^2$$

Distribution of log-likelihood, Chapter 5.3 and 5.6 in book.

Taylor:

$$l(\boldsymbol{\beta}) \approx l(\boldsymbol{b}) + \boldsymbol{U}(\boldsymbol{b})^T (\boldsymbol{\beta} - \boldsymbol{b}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{b})^T \frac{\partial U}{\partial \boldsymbol{\beta}} |_{\boldsymbol{\beta} = \boldsymbol{b}} (\boldsymbol{\beta} - \boldsymbol{b})$$

so since $\boldsymbol{U}(\boldsymbol{b}) = \boldsymbol{0}$ and $\frac{\partial U}{\partial \boldsymbol{\beta}} |_{\boldsymbol{\beta} = \boldsymbol{b}} \approx -\mathcal{I}(\boldsymbol{b}) \approx -\mathcal{I}(\boldsymbol{\beta})$,

$$2[l(\boldsymbol{b}) - l(\boldsymbol{\beta})] \approx (\boldsymbol{\beta} - \boldsymbol{b})^T \mathcal{I}(\boldsymbol{\beta}) (\boldsymbol{\beta} - \boldsymbol{b}) \approx \chi_p^2$$

when $\boldsymbol{\beta}$ is the true parameter.