Maximum likelihood estimation

If we want to apply the method of scoring to approximate the MLE, the estimating equation is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{(m)} &= \hat{\boldsymbol{\beta}}^{(m-1)} + \left[\mathcal{I}^{(m-1)}\right]^{-1} \mathbf{u}^{(m-1)} \\
\left[\mathcal{I}^{(m-1)}\right] \hat{\boldsymbol{\beta}}^{(m)} &= \left[\mathcal{I}^{(m-1)}\right] \hat{\boldsymbol{\beta}}^{(m-1)} + \mathbf{u}^{(m-1)}
\end{aligned} \tag{3.1.7}
$$

From Equation $(3.1.6)$, the information matrix can be written as

$$
\mathcal{I} = \mathbf{X}^\top \mathbf{W} \mathbf{X} \tag{3.1.8}
$$

where $w_{ii} = \frac{1}{\mathbb{V}\mathrm{ar}(\mathrm{Y}_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2$ evaluated at $\boldsymbol{\beta}$.

Finally, the expression on the right hand side of $(3.1.7)$ can be written as

$$
\sum_{k=1}^{p} \sum_{i=1}^{N} \frac{\mathrm{X}_{ij}\mathrm{X}_{ik}}{\mathbb{V}\mathrm{ar}(\mathrm{Y}_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 \hat{\beta}_k^{(m-1)} + \sum_{i=1}^{N} \frac{(\mathrm{Y}_i - \mu_i)\mathrm{X}_{ij}}{\mathbb{V}\mathrm{ar}(\mathrm{Y}_i)} \left(\frac{d\mu_i}{d\eta_i}\right)
$$

which can be written in matrix terms as

$$
\mathcal{I}^{(m-1)} \hat{\boldsymbol{\beta}}^{(m-1)} + \mathbf{u}^{(m-1)} = \mathbf{X}^\top \mathbf{W} \mathbf{z} \tag{3.1.9}
$$

where

$$
\mathrm{Z}_i = \sum_{k=1}^{p} \mathrm{X}_{ik} \hat{\beta}_k^{(m-1)} + (\mathrm{Y}_i - \mu_i) \left(\frac{d\eta_i}{d\mu_i}\right)
$$

Therefore $(3.1.7)$ reads:

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

This has the same form as the weighted least squares equation. Note, however, this needs to be solved iteratively, since $\mathbf{W}$ and $\mathbf{z}$ have to be recalculated at each optimisation step.

Therefore, for generalised linear models, maximum likelihood estimators are obtained by an iterative weighted least squares procedure.