Model 1 (GLM):

$$Y_i \sim \text{Poisson}\left(e^{\beta_1+\beta_2 x_i}\right)$$

Log likelihood

$$l(\beta_1, \beta_2) = \beta_1 \sum_{i=1}^{N} y_i + \beta_2 \sum_{i=1}^{N} y_i x_i - e^{\beta_1} \sum_{i=1}^{N} e^{\beta_2 x_i} - \sum_{i=1}^{N} \ln y_i!$$

Model 0 (i.i.d.):

$$Y_i \sim \text{Poisson}(\theta)$$

Log likelihood

$$l(\theta) = (\ln \theta) \sum_{i=1}^{N} y_i - N\theta - \sum_{i=1}^{N} \ln y_i!$$

Newton-Raphson Model 0:

$$l(\theta) = (\ln \theta) \sum_{i=1}^{N} y_i - N\theta - \sum_{i=1}^{N} \ln y_i!$$

$$U(\theta) = \frac{\partial l}{\partial \theta} = \frac{\sum_{i=1}^{N} y_i}{\theta} - N$$

$$U'(\theta) = -\frac{\sum_{i=1}^{N} y_i}{\theta^2}$$

so iteration is

$$\theta^{(m)} = \theta^{(m-1)} - \frac{U(\theta^{(m-1)})}{U'(\theta^{(m-1)})} = 2\theta^{(m-1)} - \frac{N\left(\theta^{(m-1)}\right)^2}{\sum_{i=1}^{N} y_i}$$

In Table 4.3 we have $N = 9$ and $\sum_{i=1}^{N} y_i = 72$, giving

$$\theta^{(m)} = 2\theta^{(m-1)} - \frac{\left(\theta^{(m-1)}\right)^2}{8}$$

Starting with $\theta^{(0)} = 4$ we get the sequence

$$4,\ 6,\ 7.5,\ 7.9688,\ 7.9999,\ \ldots,\ 8(=\bar{y})$$

Newton-Raphson Model 1:

$$l(\beta_1, \beta_2) = \beta_1 \sum_{i=1}^{N} y_i + \beta_2 \sum_{i=1}^{N} y_i x_i - e^{\beta_1} \sum_{i=1}^{N} e^{\beta_2 x_i} - \sum_{i=1}^{N} \ln y_i!$$

$$\boldsymbol{U}(\boldsymbol{\beta}) = \begin{bmatrix} \sum_{i=1}^{N} y_i - e^{\beta_1} \sum_{i=1}^{N} e^{\beta_2 x_i} \\ \sum_{i=1}^{N} y_i x_i - e^{\beta_1} \sum_{i=1}^{N} x_i e^{\beta_2 x_i} \end{bmatrix}$$

$$\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} -e^{\beta_1} \sum_{i=1}^{N} e^{\beta_2 x_i} & -e^{\beta_1} \sum_{i=1}^{N} x_i e^{\beta_2 x_i} \\ -e^{\beta_1} \sum_{i=1}^{N} x_i e^{\beta_2 x_i} & -e^{\beta_1} \sum_{i=1}^{N} x_i^2 e^{\beta_2 x_i} \end{bmatrix}$$

Newton-Raphson iteration is

$$\boldsymbol{b}^{(m)} = \boldsymbol{b}^{(m-1)} - \left[\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{\beta}}\right]^{-1} |_{\boldsymbol{\beta} = \boldsymbol{b}^{(m-1)}} U(\boldsymbol{b}^{(m-1)})$$

With $\boldsymbol{b}^{(0)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ we get

$$\boldsymbol{b}^{(1)} = \boldsymbol{b}^{(0)} + \begin{bmatrix} 9e^2 & e^2 \\ e^2 & 5e^2 \end{bmatrix}^{-1} \begin{bmatrix} 72 - 9e^2 \\ 32 - e^2 \end{bmatrix} = \begin{bmatrix} 2.0088630 \\ 0.6643732 \end{bmatrix}$$

Distribution of score statistic

Model 0:

$$Y_i \sim \text{Poisson}(\theta)$$

$$E_\theta[Y_i] = Var_\theta[Y_i] = \theta.$$

$$U(\theta) = \frac{\partial l}{\partial \theta} = \frac{\sum_{i=1}^{N} Y_i}{\theta} - N$$

$$U'(\theta) = -\frac{\sum_{i=1}^{N} Y_i}{\theta^2}$$

so

$$E_\theta[U(\theta)] = \frac{\sum_{i=1}^{N} E_\theta[Y_i]}{\theta} - N = 0$$

$$E[-U'(\theta)] = \frac{\sum_{i=1}^{N} E_\theta[Y_i]}{\theta^2} = \frac{N}{\theta} = \mathcal{I}(\theta)$$

$$Var[U(\theta)] = \frac{\sum_{i=1}^{N} Var_\theta[Y_i]}{\theta^2} = \frac{N}{\theta} = \mathcal{I}(\theta)$$

Scoring algorithm for single parameter

Newton Raphson algorithm:

$$\theta^{(m)} = \theta^{(m-1)} - \frac{U(\theta^{(m-1)})}{U'(\theta^{(m-1)})}$$

Scoring algorithm:

$$\theta^{(m)} = \theta^{(m-1)} - \frac{U(\theta^{(m-1)})}{E_{\theta^{(m-1)}}U'(\theta^{(m-1)})} = \theta^{(m-1)} + \frac{U(\theta^{(m-1)})}{\mathcal{I}(\theta^{(m-1)})}$$

For Model 0:

$$\theta^{(m)} = \theta^{(m-1)} + \frac{\frac{\sum y_i}{\theta^{(m-1)}} - N}{\frac{N}{\theta^{(m-1)}}} = \frac{\sum y_i}{N}$$

(so no iteration is needed).

Scoring algorithm for Model 1

$$\mathcal{I}(\boldsymbol{\beta}) = E\left[-\frac{\partial U}{\partial \boldsymbol{\beta}}\right] = -\frac{\partial U}{\partial \boldsymbol{\beta}}$$

since $\frac{\partial U}{\partial \boldsymbol{\beta}}$ has no stochastic elements.

Thus Newton-Raphson and Scoring algorithm are the same.

Further check,

$$Cov(U(\boldsymbol{\beta})) = \begin{bmatrix} Var(U_1(\boldsymbol{\beta})) & Cov(U_1(\boldsymbol{\beta}), U_2(\boldsymbol{\beta})) \\ Cov(U_1(\boldsymbol{\beta}), U_2(\boldsymbol{\beta})) & Var(U_2(\boldsymbol{\beta})) \end{bmatrix} = (?)\mathcal{I}(\boldsymbol{\beta})$$

Here we get

$$Var(U_1(\boldsymbol{\beta})) = e^{\beta_1}\sum_{i=1}^{N} e^{\beta_2 x_i}$$

$$Var(U_2(\boldsymbol{\beta})) = e^{\beta_1}\sum_{i=1}^{N} x_i^2 e^{\beta_2 x_i}$$

$$\begin{aligned} Cov(U_1(\boldsymbol{\beta}), U_2(\boldsymbol{\beta})) &= Cov(\sum_{i=1}^{N} Y_i, \sum_{i=1}^{N} x_i Y_i) = \sum_{i=1}^{N} Cov(Y_i, x_i Y_i) \\ &= \sum_{i=1}^{N} x_i Var(Y_i) = e^{\beta_1}\sum_{i=1}^{N} x_i e^{\beta_2 x_i} \text{ so (?) is OK.} \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

Hence the score, given in (4.17), is

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