Ex.4a: MLE POISSON REGRESSION solutions

Poisson regression:

$$\mathbb{P}((Y = y \mid x) = \frac{\mu^y}{y!} \exp(-\mu)$$
$$\mu = \exp(\alpha + \beta x)$$

sample $(x_1, y_1), \ldots, (x_n, y_n)$:

ML-ESTIMATION

log likelihood:

$$l(\alpha, \beta \mid \mathbf{y}, \mathbf{x}) = \sum_i ((\alpha + \beta x_i) y_i - \exp(\alpha + \beta x_i))$$

score function:

$$\frac{\partial}{\partial \alpha} l(\alpha, \beta \mid \mathbf{y}, \mathbf{x}) = \sum_i (y_i - \exp(\alpha + \beta x_i))$$
$$\frac{\partial}{\partial \beta} l(\alpha, \beta \mid \mathbf{y}, \mathbf{x}) = \sum_i x_i \; (y_i - \exp(\alpha + \beta x_i))$$

likelihood equations:

$$\frac{\partial l}{\partial \alpha}(\alpha, \beta) = 0$$
$$\frac{\partial l}{\partial \beta}(\alpha, \beta) = 0$$

mlest:

$$\frac{\partial l}{\partial \alpha}(\widehat{\alpha}, \widehat{\beta}) = 0$$
$$\frac{\partial l}{\partial \beta}(\widehat{\alpha}, \widehat{\beta}) = 0$$

information:

$$\mathbb{I}_n\left(\alpha,\beta \mid \mathbf{y},\mathbf{x}\right) = -\begin{bmatrix} \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial^2\alpha} & \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\alpha\partial\beta} \\ \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\alpha\partial\beta} & \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\beta^2} \end{bmatrix}$$

$$= -\begin{bmatrix} \sum_i \exp(\alpha+\beta x_i) & \sum_i x_i \exp(\alpha+\beta x_i) \\ \sum_i x_i \exp(\alpha+\beta x_i) & \sum_i x_i^2 \exp(\alpha+\beta x_i) \end{bmatrix}$$

(observed) information:

$$\widehat{\mathbb{I}}(\widehat{\alpha},\widehat{\beta}) = -\begin{bmatrix} \sum_i^n \frac{\partial^2 l(\widehat{\alpha},\widehat{\beta} \mid y_i,x_i)}{\partial\alpha^2} & \sum_i^n \frac{\partial^2 l(\widehat{\alpha},\widehat{\beta} \mid y_i,x_i)}{\partial\alpha\partial\beta} \\ \sum_i^n \frac{\partial^2 l(\widehat{\alpha},\widehat{\beta} \mid y_i,x_i)}{\partial\alpha\partial\beta} & \sum_i^n \frac{\partial^2 l(\widehat{\alpha},\widehat{\beta} \mid y_i,x_i)}{\partial\beta^2} \end{bmatrix}$$

estimator of the the **covariance matrix of the MLE**

$$\begin{bmatrix} \mathrm{var}(\widehat{\alpha}) & \mathrm{cov}(\widehat{\alpha},\widehat{\beta}) \\ \mathrm{cov}(\widehat{\alpha},\widehat{\beta}) & \mathrm{var}\ (\widehat{\beta}) \end{bmatrix} \approx \begin{bmatrix} \sum_i \exp(\widehat{\alpha}+\widehat{\beta} x_i) & \sum_i x_i \exp(\widehat{\alpha}+\widehat{\beta} x_i) \\ \sum_i x_i \exp(\widehat{\alpha}+\widehat{\beta} x_i) & \sum_i x_i^2 \exp(\widehat{\alpha}+\widehat{\beta} x_i) \end{bmatrix}^{-1}$$

NEWTON-METHOD

Second derivative matrix (**Hessian**)

$$H\left(\alpha,\beta\right) = -\begin{bmatrix} \frac{\partial^2 l(\alpha,\beta \mid \mathbf{y},\mathbf{x})}{\partial\alpha^2} & \frac{\partial^2 l(\alpha,\beta \mid \mathbf{y},\mathbf{x})}{\partial\alpha\partial\beta} \\ \frac{\partial^2 l(\alpha,\beta \mid \mathbf{y},\mathbf{x})}{\partial\alpha\partial\beta} & \frac{\partial^2 l(\alpha,\beta \mid \mathbf{y},\mathbf{x})}{\partial\beta^2} \end{bmatrix} = \begin{bmatrix} -\sum_i \exp(\alpha+\beta x_i) & -\sum_i x_i \exp(\alpha+\beta x_i) \\ -\sum_i x_i \exp(\alpha+\beta x_i) & -\sum_i x_i^2 \exp(\alpha+\beta x_i) \end{bmatrix}$$

Newton method:

$$\begin{bmatrix} \alpha^{(t+1)} \\ \beta^{(t+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(t)} \\ \beta^{(t)} \end{bmatrix} - H(\alpha^{(t)},\beta^{(t)}) \begin{bmatrix} \frac{\partial}{\partial\alpha} l\left(\alpha^{(t)},\beta^{(t)}\right) \\ \frac{\partial}{\partial\beta} l\left(\alpha^{(t)},\beta^{(t)}\right) \end{bmatrix}$$

It turns out that *in this special case* the Hessian equals its expectation (the Hessian is non-random in this case), i.e., Information *(=−Hessian)*

$$\mathbb{I}_n\left(\alpha,\beta \mid \mathbf{y},\mathbf{x}\right) = -\mathbb{E}H(\alpha,\beta) = -\begin{bmatrix} \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial^2\alpha} & \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\alpha\partial\beta} \\ \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\alpha\partial\beta} & \mathbb{E}\sum_i^n \frac{\partial^2 l(\alpha,\beta \mid y_i,x_i)}{\partial\beta^2} \end{bmatrix}$$

Generally, the Hessian can be considered the negative observed information. Define

$$\widehat{\mathbb{I}}_n\left(\alpha,\beta \mid \mathbf{y},\mathbf{x}\right) = -H(\alpha,\beta)$$

Using $\mathbb{I}(\alpha,\beta \mid \mathbf{y},\mathbf{x})$ instead of $H(\alpha,\beta)$ the Newton method is called the method of

"**Fisher scoring**":

$$\begin{bmatrix} \alpha^{(t+1)} \\ \beta^{(t+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(t)} \\ \beta^{(t)} \end{bmatrix} + \mathbb{I}_n^{-1}(\alpha^{(t)},\beta^{(t)}) \begin{bmatrix} \frac{\partial}{\partial\alpha} l\left(\alpha^{(t)},\beta^{(t)}\right) \\ \frac{\partial}{\partial\beta} l\left(\alpha^{(t)},\beta^{(t)}\right) \end{bmatrix}$$

Newton method:

$$\begin{bmatrix} \alpha^{(t+1)} \\ \beta^{(t+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(t)} \\ \beta^{(t)} \end{bmatrix} + \widehat{\mathbb{I}}_n^{-1}(\alpha^{(t)},\beta^{(t)}) \begin{bmatrix} \frac{\partial}{\partial\alpha} l\left(\alpha^{(t)},\beta^{(t)}\right) \\ \frac{\partial}{\partial\beta} l\left(\alpha^{(t)},\beta^{(t)}\right) \end{bmatrix}$$

Newton method for Poisson regression: with $\alpha = \alpha^{(old)},\ \beta = \beta^{(old)}$

$$\begin{bmatrix} \alpha^{(new)} \\ \beta^{(new)} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \sum_i e^{\alpha+\beta x_i} & \sum_i x_i e^{\alpha+\beta x_i} \\ \sum_i x_i e^{\alpha+\beta x_i} & \sum_i x_i^2 e^{\alpha+\beta x_i} \end{bmatrix}^{-1} \begin{bmatrix} \sum_i \left(y_i - e^{\alpha+\beta x_i}\right) \\ \sum_i x_i \left(y_i - e^{\alpha+\beta x_i}\right) \end{bmatrix}$$

Note: in the special case of Poisson regression $\widehat{\mathbb{I}}_n(\alpha,\beta) = \mathbb{I}_n(\alpha,\beta)$ (the expectation is taken of $\mathbf{y}$ conditional on $\mathbf{x}$) hence *the Newton method is identical to Fisher scoring in this case.*