BS1a
Dr Eleni Matechou

Outline for Week 5

- Components of GLM
- Density function, mean and variance of response
- Canonical link
- Observed and expected information matrices
- Iterative (Re)Weighted Least Squares algorithm

Generalised Linear Models (GLM)

Until now we dealt with models involving a continuous response that depends linearly on the parameters.

However, a normal linear model is not always valid eg. the response is discrete (Binomial, Poisson etc.).

In the GLM model framework the response variables do not need to be normally distributed.

There are three components in a GLM:

- The **random component** which is the distribution of the independent random variables Y_i for $i = 1, \ldots, n$.
- The **systematic component** which is the linear predictor $\eta = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$.
- The **link function** $g(\mu) = \eta$ which specifies how $\mu =$E(Y) is linked to the explanatory variables.

Density, mean and variance of response

In a GLM, the response has density:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - \kappa(\theta)}{\phi} + c(y;\phi)\right\}$$

where,

- y: **natural observation**
- θ: **natural parameter**
- ϕ: **scale or dispersion parameter**

If $\phi = 1$, $f(y;\theta,\phi)$ is a member of the **natural exponential (NE) family**
If $\phi > 1$, $f(y;\theta,\phi)$ is a member of the **(NE) dispersion family**

If $Y \sim f(y;\theta,\phi)$ then:

Proof

- the **expected value** of Y is $\mathsf{E}(Y) = \kappa'(\theta)$ and
- the **variance** of Y is $\mathsf{Var}(Y) = \phi\kappa''(\theta)$

Poisson example

Canonical link

In GLM, instead of modelling the mean, we introduce a one-to-one continuous differentiable transformation $g(\mu_i)$. The function $g(.)$ is called the **link function**.

An important special case is the **canonical link** which is obtained when

$$\eta_i = \theta_i = \kappa'^{-1}(\mu_i)$$

i.e. it is the link function which makes the linear predictor η the same as the natural parameter θ. When the canonical link is used then the

log-likelihood can be written:

$$\ell(\beta; y) = \log\left(\prod_{i=1}^{n} f(y;\beta)\right) = \sum_{i=1}^{n} \frac{y_i\theta_i - \kappa(\theta_i)}{\phi} + c(y_i;\phi) =$$

$$\sum_{i=1}^{n} \frac{y_i x_i\beta - \kappa(x_i\beta)}{\phi} + c(y_i;\phi)$$

Poisson example cont.
and
Binomial example

Observed and expected information

The maximum likelihood estimates of β, $\widehat{\beta}$, are the values that maximise the likelihood and satisfy the **score equations**:

$$\frac{\partial \ell(\beta; y)}{\partial \beta_j} = 0, j = 1, \ldots, p$$

We check that $\widehat{\beta}$ gives a local maximum by verifying that $\mathcal{J}(\beta) = -\frac{\partial^2 \ell(\beta;y)}{\partial \beta \partial \beta^T}$ is positive definite at $\widehat{\beta}$.

- $\mathcal{J}(\beta)$ is the **observed information matrix**
- $-\mathcal{J}(\beta) = \frac{\partial^2 \ell(\beta;y)}{\partial \beta \partial \beta^T}$ is the **hessian operator**
- $\mathcal{I} = E(-\frac{\partial^2 \ell(\beta;y)}{\partial \beta \partial \beta^T})$ is the **expected information matrix** or Fisher information matrix.

$\mathcal{I}$ is called *expected information matrix* because it indicates the mean information that a sample will contain about the value of an unknown parameter.

If the likelihood curve is sharp it will have a high negative expected second derivative and therefore high information.

The amount of information is typically related to the sample size.

The estimate of β, $\widehat{\beta}$ becomes normal asymptotically in n:

$$\widehat{\beta} \sim N(\beta, \mathcal{I}^{-1})$$

Computing $\mathcal{I}$

$\mathcal{I}$ is the variance-covariance matrix of the score statistics $U_j = \frac{\partial \ell(\beta; y)}{\partial \beta_j}$, $j = 1, \ldots, p$ and is calculated as:

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

where $\mathbf{W}$ is a diagonal matrix with entries

Proof

$$W_{ii} = \frac{1}{g'(\mu_i)^2 \phi V(\mu_i)}$$

The function $V(\mu_i)$, called the **variance function**, is equal to $\kappa''(\theta_i)$ and it expresses the relationship between the variance and the mean.

Poisson example cont.

Iterative (Re)Weighted Least Squares

For GLM, the maximum likelihood estimates for the parameters are obtained by the **iterative (re)weighted least squares** algorithm.

This is a variant of the **Newton-Raphson** algorithm:

Given a starting value for β, $\beta^{(i)}$, we expand $\frac{\partial \ell(\beta^{(i+1)}; y)}{\partial \beta^{(i+1)}}$ by Taylor series about $\beta^{(i)}$ to obtain:

$$\frac{\partial \ell(\beta^{(i+1)}; y)}{\partial \beta^{(i+1)}} \simeq \frac{\partial \ell(\beta^{(i)}; y)}{\partial \beta^{(i)}} + \frac{\partial^2 \ell(\beta^{(i)}; y)}{\partial \beta^{(i)} \beta^{(i)T}} (\beta^{(i+1)} - \beta^{(i)}) = 0$$

$$\Rightarrow \beta^{(i+1)} = \beta^{(i)} + \mathcal{J}(\beta^{(i)})^{-1} U(\beta^{(i)})$$

where, $\mathcal{J}(\beta^{(i)})$ is the observed information matrix and $U(\beta^{(i)})$ is the score statistic, as before.

Start by setting:
$\mu^{(0)} = y$ so $X\beta^{(0)} = \eta^{(0)} = g(\mu^{(0)}) = g(y)$
$z^{(0)} = g(y)$ and $W^{(0)} =$diag$(g'(y)^2\phi V(y))^{-1}$

Then,

- In **step 1** set $\beta^{(i+1)} = (X^T W^{(i)} X)^{-1} X^T W^{(i)} z^{(i)}$ i.e. β is obtained by weighted linear regression of the vector z on the columns of X using weight matrix W.
- In **step 2** update
  - $\eta^{(i+1)} = X\beta^{(i+1)}$
  - $\mu^{(i+1)} = g^{-1}(\eta^{(i+1)})$
  - $z^{(i+1)} = \eta^{(i+1)} + g'(\mu^{(i+1)})(y - \mu^{(i+1)})$ and
  - $W^{(i+1)} =$diag$(g'(\mu^{(i+1)})^2\phi V(\mu^{(i+1)}))^{-1}$

Repeat steps 1 and 2 until the estimates are essentially unchanged between iterations (convergence has been reached).

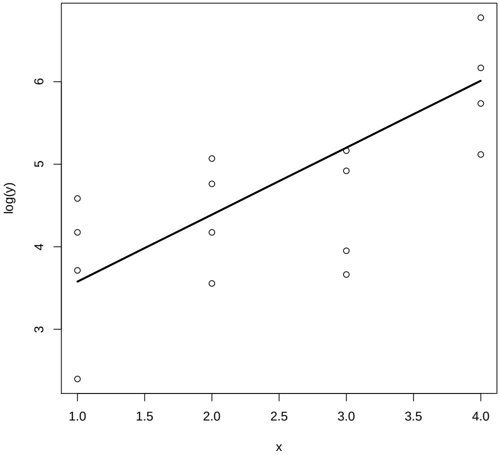
Remember that Var$(\widehat{\beta}_i) = (X^T W X)^{-1}_{ii}$ and hence they are computed in the IWLS algorithm.

See PS 5 Q1 for more details

Iterative (Re)Weighted Least Squares for Poisson data

The data are part of the data set used in exercise 9.2 in Dobson and Barnett.

y is the number of claims made for cars in various insurance policies and x is the age of the car.



Observed values of $\log(y)$ together with the resulting fitted line $\log(y) = \widehat{\beta}_1 + \widehat{\beta}_2 x$

```
##Call:
##glm(formula = y ~ x, family = poisson)

##Deviance Residuals:
##    Min       1Q   Median       3Q      Max
##-13.571   -5.248   -1.139    3.941   20.081

##Coefficients:
##            Estimate Std. Error z value Pr(>|z|)
##(Intercept)  2.76641    0.07481   36.98   <2e-16 ***
##x            0.81148    0.02153   37.69   <2e-16 ***
##---
##...

##    Null deviance: 2973.5  on 15  degrees of freedom
##Residual deviance: 1156.4  on 14  degrees of freedom
##AIC: 1263.7

##Number of Fisher Scoring iterations: 5
```