3.1 Generalised Linear Models definition and examples

1. Generalised Linear Model Definition

The unity of many statistical methods was demonstrated by Nelder and Wedderburn (1972) using the idea of a generalised linear model. This model is defined in terms of a set of independent random variables $Y_1, \ldots, Y_N$ each with a distribution from the **exponential family** and the following properties:

1. The distribution of each Y_i has the canonical form and depends on a single parameter θ_i (the θ_i s do not all have to be the same), thus

$$f(y_i; \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$$

2. The distributions of all the Y_is are of the same form (e.g. all Gaussian, or all Poisson), so that the subscripts on b, c and d are not needed. The joint density function is

$$\begin{aligned} f(y_1, \ldots, y_N; \theta_1, \ldots, \theta_N) &= \prod_{i=1}^{N} \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ &= \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right] \end{aligned} \tag{3.1.1}$$

(Note that this means that responses (y_i) are independent random variables.) The N parameters θ_i are typically not of direct interest. We are usually interested in a smaller set of parameters $\beta_1, \ldots, \beta_p$, where $p < N$. Suppose that $\boldsymbol{E}(Y_i) = \mu_i$ is some function of θ_i.

For a generalised linear model there is a transformation of μ_i such that

$$\eta_i = g(\mu_i) = x_i^T \beta \tag{3.1.2}$$

where

- g is a monotone, differentiable function called the **link function**,
- x_i is a p vector of explanatory variables (or covariates)

$$x_i^T = (x_{i1}, \ldots, x_{ip})$$

and β is the p vector of parameters. x_i is the ith column of the design matrix $\boldsymbol{X}$.

For responses $Y_1, \ldots, Y_N$ we can write a GLM in matrix notation as

$$g[\boldsymbol{E}(y)] = \boldsymbol{X}\beta,$$

where $\boldsymbol{X}$ is a matrix whose elements are constants for levels of categorical explanatory variables or measured values of continuous explanatory variables. (see examples in A. J. Dobson & A. G. Barnett (2018), pp. 58-61).

Example: Normal linear model

The best known case of a genearlised linear model is the normal linear model

$$\boldsymbol{E}(Y_i) = \mu_i = x_i^T \beta; \quad Y_i \sim N(\mu_i, \sigma^2)$$

here the link function is the identity function $g(\mu_i) = \mu_i$. This model is usually written in the form

$$y = \boldsymbol{X}\beta + \varepsilon$$

where ϵ is a vector of i.i.d. random variables with $\varepsilon_i \sim N(0, \sigma^2)$.

In this form, the linear component $\mu = \boldsymbol{X}\beta$ represents the 'signal' and ε represents the 'noise'. Multiple regression and ANOVA (analysis of variance) are of this form. We will consider them later on in detail.

2. Maximum likelihood estimation for GLMs

Let's recall the following results from the previous slide:

The **joint distribution** is

$$
\begin{aligned}
f(\mathrm{Y}_1, \ldots, \mathrm{Y}_N | \theta_1, \ldots, \theta_N) &= \prod_{i=1}^{N} \exp[\mathrm{Y}_i b(\theta_i) + c(\theta_i) + d(\mathrm{Y}_i)] \\
&= \exp\left[\sum_{i=1}^{N} \mathrm{Y}_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(\mathrm{Y}_i)\right]
\end{aligned}
$$

For each Y_i, the **log-likelihood** is $\ell_i = \mathrm{Y}_i b(\theta_i) + c(\theta_i) + d(\mathrm{Y}_i)$, which gives

$$
\mathbb{E}(\mathrm{Y}_i) = \mu_i = -\tfrac{c'(\theta_i)}{b'(\theta_i)}, \quad \mathbb{V}\mathrm{ar}(\mathrm{Y}_i) = \tfrac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3}, \quad g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_i.
$$

The **log-likelihood** for all the Y_i's is then

$$
\ell(\boldsymbol{\theta}; \mathrm{Y}_1, \ldots, \mathrm{Y}_N) = \sum_{i=1}^{N} \ell_i = \sum_{i=1}^{N} \mathrm{Y}_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(\mathrm{Y}_i).
$$

The score function is then given by

$$
U_j = \sum_{i=1}^{N} \left[\frac{(Y_i - \mu_i)}{\mathbb{V}ar(Y_i)} \mathrm{X}_{ij} \left(\frac{d\mu_i}{d\eta_i}\right)\right] \tag{3.1.4}
$$

The variance-covariance matrix of the score is

$$
\mathcal{I}_{jk} = \sum_{i=1}^{N} \frac{\mathrm{X}_{ij}\mathrm{X}_{ik}}{\mathbb{V}\mathrm{ar}(\mathrm{Y}_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 \tag{3.1.6}
$$

Press on the button below to read more about how to apply the method of scoring to approximate the MLE: