1 Introduction

1.1 Generalized Linear Models

1.1 Definition: In a generalized linear model we have $y_1, y_2, \ldots, y_n$ which are observed values of independent random variables $Y_1, Y_2, \ldots, Y_n$. The generalized linear model has three components:

- The distribution of Y_i is

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{a(\phi)} + c(y_i; \phi)\right\}.$$

- The systematic model is specified by a linear predictor of the form

$$\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}.$$

  The β_j are unknown parameters and the x_{ij} are values of covariates.

- The link between the random and the systematic component is given by

$$\eta_i = g(\mu_i),$$

  where $\mu_i = E(Y_i)$. The link function is required to be monotonic and differentiable.

1.2 Example: For the normal distribution we have

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right\} = \exp\left\{\frac{y_i\mu_i - \frac{\mu_i^2}{2}}{\sigma^2} - \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}.$$

1.3 Note: It is easy to show that under weak conditions on f_{Y_i}

$$E(Y_i) = \left.\frac{db(\theta)}{d\theta}\right|_{\theta=\theta_i} \qquad \text{and} \qquad \text{var}(Y_i) = a(\phi) \times \left.\frac{d^2b(\theta)}{d\theta^2}\right|_{\theta=\theta_i}$$

Thus the mean depends only on θ_i, the canonical parameter. The variance depends on a function of the canonical parameter and the dispersion or scale parameter ϕ.

1.4 Definition: The link is called a canonical link if $\theta_i = \eta_i$.

1.5 Example: For the normal distribution we have $\theta_i = \eta_i$, and hence

$$\theta_i = E(Y_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}.$$

This is the usual general linear model for multiple regression and analysis of (co-)variance.

1.2 The Linear Model Structure

1.6 Example: Simple linear regression.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

1.7 Example: Polynomial regression.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

1.8 Example: Multiple linear regression.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

1.9 Example: Data transformations.

$$\begin{pmatrix} log(Y_1) \\ log(Y_2) \\ \vdots \\ log(Y_n) \end{pmatrix} = \begin{pmatrix} 1 & log(X_1) \\ 1 & log(X_2) \\ \vdots & \vdots \\ 1 & log(X_n) \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

1.10 Example: One-way analysis of variance.

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1J} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2J} \end{pmatrix}$$

1.11 Example: Two-way analysis of variance.

$$\begin{pmatrix} Y_{111} \\ \vdots \\ Y_{11K} \\ \hline Y_{121} \\ \vdots \\ Y_{12K} \\ \hline Y_{211} \\ \vdots \\ Y_{21K} \\ \hline Y_{221} \\ \vdots \\ Y_{22K} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \vdots \\ \varepsilon_{11K} \\ \hline \varepsilon_{121} \\ \vdots \\ \varepsilon_{12K} \\ \hline \varepsilon_{211} \\ \vdots \\ \varepsilon_{21K} \\ \hline \varepsilon_{221} \\ \vdots \\ \varepsilon_{22K} \end{pmatrix}$$

1.12 Example: Analysis of covariance.

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & (x_{11} - \bar{x}_{..}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & (x_{1J} - \bar{x}_{..}) \\ 1 & 0 & 1 & (x_{21} - \bar{x}_{..}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & (x_{2J} - \bar{x}_{..}) \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1J} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2J} \end{pmatrix}$$

The general linear model in matrix form:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Equivalent shorthand form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

$\mathbf{Y}$ $(n \times 1)$ is the response vector, $\mathbf{X}$ $(n \times p)$ is the design (or model or regression) matrix, $\boldsymbol{\beta}$ $(p \times 1)$ is the vector of regression coefficients, $\boldsymbol{\varepsilon}$ $(n \times 1)$ is the error vector (mean $\mathbf{0}$).

1.13 Note: Usually $x_{i0} = 1$ for all i, i.e., there is an intercept β_0 in the model.

1.14 Note: $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$ are called the predictor variables or regressor variables. They are known constants.

1.15 Note: The model is linear in the unknown regression coefficients $\beta_0, \beta_1, \ldots, \beta_{p-1}$, i.e.,

$$\mathbf{Y} = \sum_{j=0}^{p-1} \beta_j \mathbf{x}_j + \boldsymbol{\varepsilon},$$

where $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$.

1.16 Note: We will give assumptions on the distribution of ε when we discuss estimation methods.