3

Generalized Linear Models

3.1 Introduction

Most models that have been proposed in the statistical literature for the analysis of discrete repeated measurements can be considered extensions of generalized linear models (McCullagh and Nelder 1989) to the context of correlated observations. In this chapter, these models will be introduced, inference will be briefly discussed, and several frequently used specific cases will be given special attention.

3.2 The Exponential Family

A random variable Y follows a distribution that belongs to the exponential family if the density is of the form

$$f(y) \quad \equiv \quad f(y|\theta, \phi) \; = \; \exp\left\{\phi^{-1}[y\theta - \psi(\theta)] + c(y, \phi)\right\} \tag{3.1}$$

for a specific set of unknown parameters θ and ϕ, and for known functions $\psi(\cdot)$ and $c(\cdot, \cdot)$. Often, θ and ϕ are termed 'natural parameter' (or 'canonical parameter') and 'scale parameter,' respectively.

The first two moments can easily be derived as follows. Starting from the property $\int f(y|\theta, \phi) dy = 1$ and taking the first- and second-order derivatives from both sides of the equation, we get that

$$\left\{ \begin{array}{rcl} \int [y - \psi'(\theta)] \; f(y|\theta, \phi) \; dy & = & 0, \\ \int \{\phi^{-1}[y - \psi'(\theta)]^2 - \psi''(\theta)\} \; f(y|\theta, \phi) \; dy & = & 0, \end{array} \right.$$

from which it directly follows that the average $\mu = \mathrm{E}(Y)$ equals $\psi'(\theta)$ and the variance $\sigma^2 = \mathrm{Var}(Y)$ is given by $\phi\psi''(\theta)$. An important implication is that, in general, the mean and variance are related through $\sigma^2 = \phi\psi''[\psi^{'-1}(\mu)] = \phi v(\mu)$ for an appropriate function $v(\mu)$, called the variance function.

In some of the models that will be discussed in this book, a quasi-likelihood perspective is taken. Although the above relation between the mean and the variance immediately follows from the density (3.1), one sometimes starts from specifying a mean and a variance function,

$$\begin{array}{rcl} \mathrm{E}(Y) & = & \mu, \\ \mathrm{Var}(Y) & = & \phi v(\mu). \end{array}$$

The variance function $v(\mu)$ can be chosen in accordance with a particular member of the exponential family. If not, then parameters cannot be estimated using maximum likelihood principles. Instead, a set of estimating equations needs to be specified, the solution of which is referred to as the quasi-likelihood estimates. Examples of this approach will be given in Chapter 8.

3.3 The Generalized Linear Model (GLM)

In a regression context, where one wishes to explain variability between outcome values based on measured covariate values, the model needs to incorporate covariates. This leads to so-called generalized linear models. Let $Y_1, \ldots, Y_N$ be a set of independent outcomes, and let $\boldsymbol{x_1}, \ldots, \boldsymbol{x_N}$ represent the corresponding p-dimensional vectors of covariate values. It is assumed that all Y_i have densities $f(y_i|\theta_i, \phi)$ which belong to the exponential family, but a different natural parameter θ_i is allowed per observation. Specification of the generalized linear model is completed by modeling the means μ_i as functions of the covariate values. More specifically, it is assumed that

$$\mu_i = h(\eta_i) = h(\boldsymbol{x_i}'\boldsymbol{\beta}),$$

for a known function $h(\cdot)$, and with $\boldsymbol{\beta}$ a vector of p fixed unknown regression coefficients. Usually, $h^{-1}(\cdot)$ is called the link function. In most applications, the so-called natural link function is used, i.e., $h(\cdot) = \psi'(\cdot)$, which is equivalent to assuming $\theta_i = \boldsymbol{x_i}'\boldsymbol{\beta}$. Hence, it is assumed that the natural parameter satisfies a linear regression model.

3.4 Examples

3.4.1 The Linear Regression Model for Continuous Data

Let Y be normally distributed with mean μ and variance σ^2. The density can be written as

$$f(y) \quad = \quad \exp\left\{\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) + \left(\frac{\ln(2\pi\sigma^2)}{2} - \frac{y^2}{2\sigma^2}\right)\right\},$$

which implies that the normal distribution belongs to the exponential family, with natural parameter θ equal to μ, scale parameter ϕ equal to σ^2, and variance function $v(\mu) = 1$. Hence, the normal distribution is very particular in the sense that there is no mean-variance relation, as will be shown to be present for other exponential family distributions. The natural link function equals the identity function, leading to the classical linear regression model $Y_i \sim N(\mu_i, \sigma^2)$ with $\mu_i = \boldsymbol{x_i}'\boldsymbol{\beta}$.

3.4.2 Logistic and Probit Regression for Binary Data

Let Y be Bernoulli distributed with success probability $P(Y = 1) = \pi$. The density can be written as

$$f(y) \quad = \quad \exp\left\{y \ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right\},$$

which implies that the Bernoulli distribution belongs to the exonential family, with natural parameter θ equal to the logit, i.e., $\ln[\pi/(1-\pi)]$, of π, scale parameter $\phi = 1$, with mean $\mu = \pi$ and with variance function $v(\pi) = \pi(1-\pi)$. The natural link function is the logit link, leading to the classical logistic regression model $Y_i \sim \text{Bernoulli}(\pi_i)$ with $\ln[\pi_i/(1-\pi_i)] = \boldsymbol{x_i}'\boldsymbol{\beta}$ or equivalently

$$\pi_i = \frac{\exp(\boldsymbol{x_i}'\boldsymbol{\beta})}{[1 + \exp(\boldsymbol{x_i}'\boldsymbol{\beta})]}.$$

Sometimes, the logit link function is replaced by the probit link, which is the inverse of the standard normal distribution function, Φ^{-1}. It has been repeatedly shown (Agresti 1990) that the logit and probit link functions behave very similarly, in the sense that for probabilities other than extreme ones (say, outside of the interval $[0.2; 0.8]$) logistic and probit regression provide approximately the same parameter estimates, up to a scaling factor equal to $\pi/\sqrt{3}$, the ratio of the standard deviations of a logistic and a standard normal variable.

3.4.3 Poisson Regression for Counts

Let Y be Poisson distributed with mean λ. The density can be written as

$$f(y) \quad = \quad \exp\{y \ln \lambda - \lambda - \ln y!\},$$

from which it follows that the Poisson distribution belongs to the exponential family, with natural parameter θ equal to $\ln\lambda$, scale parameter $\phi = 1$, and variance function $v(\lambda) = \lambda$. The logarithm is the natural link function, leading to the classical Poisson regression model $Y_i \sim \text{Poisson}(\lambda_i)$, with $\ln\lambda_i = \boldsymbol{x_i}'\boldsymbol{\beta}$.

3.5 Maximum Likelihood Estimation and Inference

Estimation of the regression parameters in $\boldsymbol{\beta}$ is usually done using maximum likelihood (ML) estimation. Assuming independence of the observations, the log-likelihood is given by

$$\ell(\beta, \phi) \;=\; \frac{1}{\phi}\sum_{i=1}^{N}[y_i\theta_i - \psi(\theta_i)] + \sum_i c(y_i, \phi).$$

The score equations obtained from equating the first-order derivatives of the log-likelihood to zero take the form

$$S(\boldsymbol{\beta}) \;=\; \sum_i \frac{\partial\theta_i}{\partial\boldsymbol{\beta}}[y_i - \psi'(\theta_i)] \;=\; 0.$$

Because $\mu_i = \psi'(\theta_i)$ and $v_i = v(\mu_i) = \psi''(\theta_i)$, we have that

$$\frac{\partial\mu_i}{\partial\boldsymbol{\beta}} \;=\; \psi''(\theta_i)\,\frac{\partial\theta_i}{\partial\boldsymbol{\beta}} \;=\; v_i\,\frac{\partial\theta_i}{\partial\boldsymbol{\beta}}$$

which implies the following score equations:

$$S(\boldsymbol{\beta}) \;=\; \sum_i \frac{\partial\mu_i}{\partial\boldsymbol{\beta}}\, v_i^{-1}\,(y_i - \mu_i) \;=\; 0.$$

In general, these score equations need to be solved iteratively, using numerical algorithms such as iteratively (re-)weighted least squares, Newton-Raphson, or Fisher scoring.

Once the ML estimates have been obtained, classical inference based on asymptotic likelihood theory becomes available, including Wald-type tests, likelihood ratio tests, and score tests, all asymptotically equivalent.

In some cases, such as in the logistic regression model, ϕ is a known constant. In other examples, such as the linear normal model, estimation of ϕ may be required to estimate the standard errors of the elements in $\boldsymbol{\beta}$. Because $\text{Var}(Y_i) = \phi v_i$, an obvious estimate for ϕ is given by

$$\widehat{\phi} \;=\; \frac{1}{N-p}\sum_i (y_i - \widehat{\mu_i})^2/v_i(\widehat{\mu_i}).$$

TABLE 3.1. *Toenail Data. Logistic regression, ignoring the association structure. Parameter estimates, associated standard errors, and inferences for the parameters in model (3.2).*

Parameter	Estimate	s.e.	p-value
β_0	-0.5571	0.1090	<0.0001
β_1	0.0240	0.1565	0.8780
β_2	-0.1769	0.0246	<0.0001
β_3	-0.0783	0.0394	0.0470

For example, under the normal model, this would yield

$$\widehat{\sigma}^2 = \frac{1}{N-p} \sum_i (y_i - \boldsymbol{x_i}'\widehat{\boldsymbol{\beta}})^2,$$

which is the mean squared error used in linear regression models to estimate the residual variance.

We refer to McCullagh and Nelder (1989) and to Agresti (1990) for more details on estimation and inference in the GLM's.

3.6 Logistic Regression for the Toenail Data

As an example of logistic regression, we analyze the toenail data introduced in Section 2.3, ignoring the correlation structure due to the repeated measurements within subjects. This would be correct if measurements at different time points would also be taken on different subjects. In Section 10.3, the results obtained here will be used as starting values in the fitting of more complicated models that do account for the association structure. Let Y_i be the binary outcome indicating severity of the toenail infection, for the ith observation. A logistic model will be assumed, with linear time trends, for both treatment groups separately. More specifically, the model is given by

$$\begin{array}{rcl} Y_i & \sim & \text{Bernoulli}(\pi_i), \\ \text{logit}(\pi_i) & = & \beta_0 + \beta_1 T_i + \beta_2 t_i + \beta_3 T_i t_i, \end{array} \qquad (3.2)$$

in which T_i is the treatment indicator for this observation, and t_i is the time-point at which the observation was taken. The results are shown in Table 3.1. The maximized log-likelihood value equals -905.91 and could be used in likelihood ratio tests for the validity of simpler models. Note the significant interaction ($p = 0.0470$) suggesting different trends in the two treatment groups.

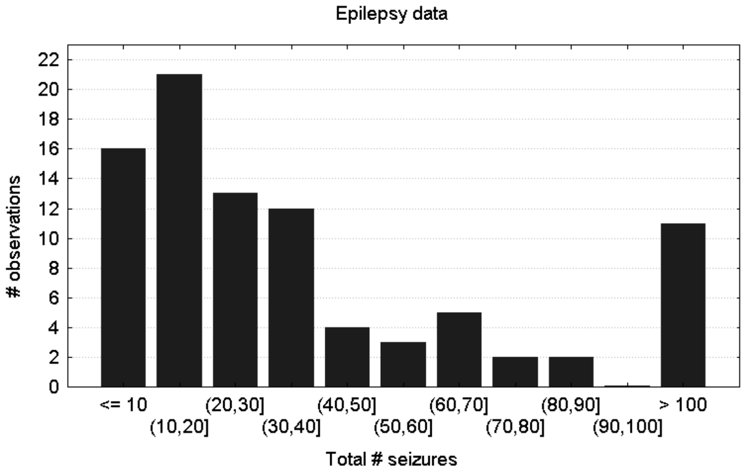

FIGURE 3.1. *Epilepsy Study. Frequency plot of the post-randomization total number of epileptic seizures, over both treatment groups.*

3.7 Poisson Regression for the Epilepsy Data

As an example of Poisson regression, we analyze the epilepsy data introduced in Section 2.5. Our response of interest will be the total number of seizures a patient has experienced during the study, after randomization took place. We want to test for a treatment effect on number of seizures, correcting for the average number of seizures during the 12-week baseline phase, prior to the treatment. Let Y_i be the total number of seizures for subject i. A histogram of the observed values is given in Figure 3.1. Note that this histogram does not correct for the fact that the subjects have not been followed for an equal number of weeks. Let n_i be the number of weeks subject i has been followed; we will correct for the differences in follow-up time by assuming that

$$\begin{aligned} Y_i &\sim \text{Poisson}(\lambda_i), \\ \ln(\lambda_i/n_i) &= \beta_0 + \beta_1 \text{Baseline}_i + \beta_2 T_i, \end{aligned} \tag{3.3}$$

in which T_i is the treatment indicator and where Baseline_i is the baseline seizure rate. Note that model (3.3) is equivalent to

$$\ln(\lambda_i) = \ln(n_i) + \beta_0 + \beta_1 \text{Baseline}_i + \beta_2 T_i \tag{3.4}$$

which is a traditional Poisson model with constant term $\ln(n_i)$ added to the linear predictor. This term is often called an ‘offset’.

The results are shown in Table 3.2. The maximized log-likelihood equals 14837.31. Note the highly significant positive effect of the baseline rate.

TABLE 3.2. *Epilepsy Study. Poisson regression for the total number of epileptic seizures. Parameter estimates, associated standard errors, and inferences for the parameters in model (3.4).*

Parameter	Estimate	S.e.	p-value
β_0	0.8710	0.0218	<0.0001
β_1	0.0172	0.0002	<0.0001
β_2	-0.4987	0.0341	<0.0001

Further, correcting for baseline rate, the treatment significantly reduces the average weekly number of epileptic seizures ($p < 0.0001$).