A Review of GLMs

Assume that n $(k+1)$-tuples $(y_i, x_{i1}, x_{i2}, \ldots, x_{ik})$, $i = 1, \ldots, n$, are observed. The values y_i are responses and components of $(x_{i1}, x_{i2}, \ldots, x_{ik})$ are predictors or covariates. Let $X^* = (x_{ij})_{n \times k}$ be a matrix of predictors.

Standard theory of linear regression proposes the linear model

$$y = X\beta + \epsilon, \tag{1}$$

where $y = (y_1, \ldots, y_n)'$ is the response vector, $X = [\mathbf{1} \ X^*]$ is a $n \times p$ design matrix, ($\mathbf{1}$ is a column vector of 1's), and ϵ is a $n \times 1$ vector od errors consisting of n iid normal $N(0, \sigma^2)$ random variables. Here $p = k + 1$. The variance σ^2 is common for all y_i's and independent of predictors. The β is a $p \times 1$ vector of parameters in the linear relationship,

$$\mathbb{E}y_i = x_i\beta = \beta_0 + \beta_1 x_{i1} + \ldots \beta_k x_{ik}, \ i = 1, \ldots, n$$

The term generalized linear model (GLM) refers to a large class of models, introduced by Nelder and Wedderburn (1972) and popularized by McCullagh and Nelder (1982, second edition 1989).
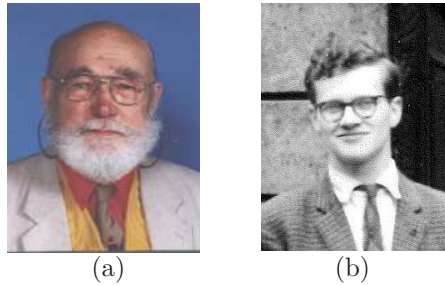

(a)   (b)

Figure 1: (a) John Nelder (1924–2010); (b) Robert Wedderburn (1947–1975)

In the canonical GLM model, the response variable y_i is assumed to follow an exponential family distribution with mean μ_i, which is assumed to be functionally dependent of $x_i'\beta$. This function can be nonlinear, but the distribution of y_i depends on covariates only via their linear combination, $\eta_i = x_i'\beta$, called a *linear predictor*. As in the linear regression, the epithet *linear* connotate linearity in parameters, not in the explanatory variables (predictors). Thus, the linear combination $\beta_0 + \beta_1 \ x_1 + \beta_2 \ x_2 + \beta_3 \ \log(x_1 + x_2) + \beta_4 \ x_1 \cdot x_2$, is a perfect linear predictor.

Question: What is generalized in model (1) by a GLM?

(1) Although y_i's remain independent, their (common) distribution is generalized. Instead of being the normal, the distribution of y_i is selected from the exponential family of distributions. This family is quite broad and versatile and includes normal, binomial, Poisson, negative binomial, gamma, inverse normal, etc, as special cases.

(2) In the linear model (1) the mean of y_i, $\mu_i = \mathbb{E}\, y_i$ was equal to $x_i'\beta$. The mean μ_i in GLM depends on the predictor $\eta_i = x_i'\beta$,

$$g(\mu_i) = \eta_i \quad (= x_i'\beta). \tag{2}$$

The function g is called the *link* function. For the model (1), the link is the identity function.

(3) The variance of y_i is not constant, but may depend on the mean μ_i.

(4) Much more (but beyond the scope of our course).

Some models and inference for categorical data are well unified in a large class of models which are GLM in nature.

For example in contingency tables, the cell counts N_{ij} have multinomial distribution $\mathcal{M}n(n, p_{ij})$. The hypothesis of interest in contingency tables is testing $H_0 : p_{ij} = \alpha_i\beta_j$ for some unknown α_i and β_j such that $\sum_i \alpha_i = \sum_j \beta_j = 1$.

Since the expected cell count, $\mathbb{E}N_{ij}$ is np_{ij}, and under H_0, $n\alpha_i\beta_j$, by taking the logarithm of both sides, we obtain,

$$\begin{aligned}
\log \mathbb{E}N_{ij} &= \log n + \log \alpha_i + \log \beta_j \\
&= \text{const} + a_i + b_j,
\end{aligned}$$

for some parameters a_i and b_j. Thus, the test of goodness of fit for this model linear in parameters a and b, is equivalent to the test of the original independence hypothesis H_0 from the contingency tables.

Before discussing more details of GLM we overview the Exponential Family of distributions.

Exponential Family of Distributions

Observations y_i follow the distribution form the exponential family if their density can be written as

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right\}, \tag{3}$$

for some given functions b and c. Parameter θ is called *canonical parameter*, and ϕ dispersion parameter.

Example. Normal $N(\mu, \sigma^2)$ distribution belongs to the exponential family, with $\theta = \mu$, $\phi = \sigma^2$, $b(\theta) = \theta^2/2$ and $c(y, \phi) = -1/2\, [y^2/\phi + \log(2\pi\phi)]$, since

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} = \exp\left\{\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left[y^2/\sigma^2 + \log(2\pi\sigma^2)\right]\right\}.$$

Links

In the GLM the predictors for y_i are summarized as linear predictor $\eta_i = x_i'\beta$. The link function is a monotone differentiable function g such that $\eta_i = g(\mu_i)$, where $\mu_i = EY_i$.

We note that in the normal case $\mu = \eta$ and the link is identity, $g(\mu) = \mu$.

For example, in analyzing counting data (e.g., contingency tables) Poisson model for ys is standardly assumed exponential family model. As $\mu > 0$, the identity link is inappropriate since η could be negative. However, if $\mu = e^\eta$ is assumed, the mean is always positive, and $\eta = \log(\mu)$ is an appropriate link.

A link is called *natural* if it is connecting θ (canonical parameter) and μ. In the Poisson case,

$$f(y|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} = \exp\left\{y\log\lambda - \lambda + \log y!\right\},$$

$\mu = \lambda$ and $\theta = \log\mu$. Thus, the log is the natural link for Poisson distribution.

The pmf for binomial distribution

$$f(y|\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y},$$

can be represented as

$$f(y|\pi) = \exp\left\{y\log\frac{\pi}{1-\pi} + n\log(1-\pi) + \log(\binom{n}{y}))\right\}.$$

From this, the natural link is $\eta = \log\frac{\pi}{1-\pi}$, called *logit* link.

With the binomial distribution, several more links are commonly used. For example, the *probit* link is $\eta = \Phi^{-1}(\pi)$ where Φ is a normal cdf, and the *complementary log-log* link is $\eta = \log\{-\log(1-\pi)\}$.

For the three popular links in binomial model, the probability parameter π is expressed as $\pi = e^\eta/(1+e^\eta)$, $\pi = \Phi(\eta)$, and $\pi = 1 - \exp\{-e^\eta\}$, respectively.

Distribution	$\theta(\mu)$	$b(\theta)$	ϕ
Normal $N(\mu,\sigma^2)$	μ	$\theta^2/2$	σ^2
Bernoulli $Bin(1,\pi)$	$\log(\pi/(1-\pi))$	$\log(1+\exp(\theta))$	1
Poisson $Poi(\lambda)$	$\log\lambda$	$\exp(\theta)$	1
Gamma $\Gamma(\mu,\nu)$	$-1/\mu$	$-\log(-\theta)$	$1/\nu$
Inv. normal $IN(\mu,\sigma^2)$	$1/\mu^2$	$-\sqrt{-2\theta}$	σ^2

Mean and Variance in Exponential Family

Let the $f(y|\theta)$ be a member of exponential family with natural parameter θ. Assume that θ is univariate. Then the log likelihood $\ell(\theta) = \sum_{i=1}^n \log f(y_i|\theta) = \sum_{i=1}^n \ell_i(\theta)$, where $\ell_i =$

$\log f(y_i|\theta)$. The MLE for θ is solution of the equation

$$\frac{\partial \ell}{\partial \theta} = 0.$$

Property (i): $\mathbb{E}(\frac{\partial \ell_i}{\partial \theta}) = 0.$

The following sequence proves property (i).

$$\int_{\mathbb{R}} f(y_i|\theta) dy_i = 1$$
$$\int_{\mathbb{R}} \frac{\partial f(y_i|\theta)}{\partial \theta} dy_i = 0$$
$$\int_{\mathbb{R}} \frac{1}{f(y_i, \theta)} \frac{\partial f(y_i|\theta)}{\partial \theta} f(y_i|\theta) dy_i = 0$$

Since $\frac{\partial \ell_i}{\partial \theta} = \frac{\partial \log f(y_i|\theta)}{\partial \theta} = \frac{1}{f(y_i,\theta)} \frac{\partial f(y_i|\theta)}{\partial \theta}$,

$$\int_{\mathbb{R}} \frac{\partial \ell_i}{\partial \theta} f(y_i|\theta) dy_i = \mathbb{E}\left(\frac{\partial \ell_i}{\partial \theta}\right) = 0$$

Property (ii): $\mathbb{E}(\frac{\partial^2 \ell_i}{\partial \theta^2}) + \mathbb{V}\mathbf{ar}\left(\frac{\partial \ell_i}{\partial \theta}\right) = 0.$

$$\begin{aligned}
0 &= \frac{\partial \mathbb{E}\left(\frac{\partial \ell_i}{\partial \theta}\right)}{\partial \theta} \\
&= \frac{\partial}{\partial \theta}\left[\int_{\mathbb{R}} \frac{\partial \ell_i}{\partial \theta} f(y_i|\theta) dy_i\right] \\
&= \int_{\mathbb{R}} \frac{\partial^2 \ell_i}{\partial \theta^2} f(y_i|\theta) dy_i + \int_{\mathbb{R}} \frac{\partial \ell_i}{\partial \theta} \frac{\partial f(y_i|\theta)}{\partial \theta} dy_i.
\end{aligned}$$

Using $\frac{\partial \ell_i}{\partial \theta} = \frac{\partial \log f(y_i|\theta)}{\partial \theta} = \frac{1}{f(y_i|\theta)} \frac{\partial f(y_i|\theta)}{\partial \theta}$ again, that is, $\frac{\partial f(y_i|\theta)}{\partial \theta} = \frac{\partial \ell_i}{\partial \theta} \times f(y_i|\theta)$,

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{\partial^2 \ell_i}{\partial \theta^2} f(y_i|\theta) dy_i + \int_{\mathbb{R}} \frac{\partial \ell_i}{\partial \theta} \frac{\partial \ell_i}{\partial \theta} f(y_i|\theta) dy_i \\
&= \mathbb{E}(\frac{\partial^2 \ell_i}{\partial \theta^2}) + \mathbb{E}\left(\frac{\partial \ell_i}{\partial \theta}\right)^2 = 0.
\end{aligned}$$

For the exponential family of distributions,

$$\ell_i = \ell(y_i, \theta, \phi) = \frac{y_i \theta - b(\theta)}{\phi} + c(y_i, \phi). \tag{4}$$

and $\frac{\partial \ell}{\partial \theta} = \frac{y - b'(\theta)}{\phi}$ and $\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{b''(\theta)}{\phi}$. By properties (i) and (ii),
$\mathbb{E}y = b'(\theta)$ and $\mathrm{Var}\, y = b''(\theta)\phi$.

The function $b''(\theta)$ is called *variance function* and denoted by $V(\mu)$ (since θ depends on μ).

When data y_i form the exponential family are given in a grouped form (from which an average is considered as the group response), then the distribution for y_i takes the form

$$f(y_i|\theta_i, \phi, \omega_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi}\omega_i + c(y_i, \phi, \omega_i)\right\}, \tag{5}$$

Weights ω_i are equal to 1 if individual responses are considered, $\omega_i = n_i$ if response y_i is an average of n_i responses, and $\omega_i = 1/n_i$ if the sum of n_i individual responses is considered.

The variance of y_i then takes the form

$$Var\ y_i = \frac{b''(\theta_i)\phi}{\omega_i} = \frac{\phi V(\mu_i)}{\omega_i}.$$

The *unit deviance* is defined as

$$d_i(y_i, \mu_i) = 2\int_{\mu_i}^{y_i} \frac{y_i - u}{V(u)} du,$$

and the total deviance, a measure of the distance between y and μ, is defined as

$$D(y, \mu) = \sum_{i=1}^{n} w_i d(y_i, \mu_i),$$

where the summation is over the data and w_i are the prior weights. The quantity $D(y, \mu)/\phi$ is called the scaled deviance. For the normal distribution, the deviance is equivalent to the residual sum-of-squares, $\sum_{i=1}^{n}(y_i - \mu_i)^2$.

Algorithm for Fitting the GLM

The algorithms for fitting generalized linear models are robust and well established (see Nelder and Wedderburn (1972) and McCullagh and Nelder (1994)). The maximum likelihood estimates of β can be obtained using iterative weighted least-squares, IWLS.

(i) Given vector $\hat{\mu}^{(k)}$, initial value of the linear predictor $\hat{\eta}^{(k)}$ is formed using the link function, and components of adjusted dependent variate (working response), $z_i^{(k)}$, can be formed as

$$z_i^{(k)} = \hat{\eta}_i^{(k)} + \left(y_i - \hat{\mu}_i^{(k)}\right)\left(\frac{d\eta}{d\mu}\right)_i^{(k)},$$

5

where the derivative is evaluated at the the available kth value.

(ii) The quadratic (working) weights, $W_i^{(k)}$, are defined so that

$$\frac{1}{W_i^{(k)}} = \left(\frac{d\eta}{d\mu}\right)_i^2 V_i^{(k)}.$$

where V is the variance function evaluated at the initial values.

(iii) The working response $z^{(k)}$ is then regressed onto the covariates x_i, with weights $W_i^{(k)}$ to produce new parameter estimates, $\hat{\beta}^{(k+1)}$. This vector is then used to form new estimates $\eta^{(k+1)} = X'\hat{\beta}^{(k+1)}$ and $\hat{\mu}^{(k+1)} = g^{-1}(\hat{\eta}^{(k+1)})$, and iterations are repeated until changes are sufficiently small. Starting values are obtained directly from the data, using $\hat{\mu}^{(0)} = y$, with occasional refinements in some cases (for example, to avoid evaluating $\log 0$ when fitting a log-linear model with zero counts).

By default, the scale parameter is estimated by the mean deviance, $\frac{1}{n}\sum_{i=1}^{n} D(y, \mu)$, in the case of the normal, gamma and inverse Gaussian distributions. The default value of the scale parameter for binomial and Poisson distributions is one. In the case of over- or under-dispersion, it can be estimated by the mean deviance.

Deviance Analysis in GLM

In GLM modeling goodness of fit of a proposed model can be assessed in several ways. The customary measure is *deviance* statistics. For a data set with n observations, assume the dispersion ϕ is known and equal to 1, and consider the two extreme models, the single parameter model stating $\mathbb{E}y_i = \hat{\mu}$ and the n parameter *saturated* model setting $\mathbb{E}y_i = \hat{\mu}_i = y_i$. Most likely, the interesting model is between the two extremes. Suppose $\mathcal{M}$ is the interesting model with $1 < p < n$ parameters.

If $\hat{\theta}_i^{\mathcal{M}} = \hat{\theta}_i^{\mathcal{M}}(\hat{\mu}_i)$ are predictions of the model $\mathcal{M}$ and $\hat{\theta}_i^{\mathcal{S}} = \hat{\theta}_i^{\mathcal{S}}(y_i) = y_i$ are the predictions of saturated model, then the deviance of the model $\mathcal{M}$ is

$$D_{\mathcal{M}} = 2\sum_{i=1}^{n} \left[(y_i\hat{\theta}_i^{\mathcal{S}}(y_i) - b(\hat{\theta}_i^{\mathcal{S}})) - (y_i\hat{\theta}_i^{\mathcal{M}} - b(\hat{\theta}_i^{\mathcal{M}}))\right]$$

When the dispersion ϕ is estimated and different than 1, the *scaled deviance* of the model $\mathcal{M}$ is defined as $D_{\mathcal{M}}^* = D_{\mathcal{M}}/\phi$.

Example. For $y_i \in \{0, 1\}$ and Binomial family,

$$D = 2\sum_{1=1}^{n} \left\{y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i) \log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right\}$$

- Deviance is minimized at saturated model $\mathcal{S}$. Equivalently, the log-likelihood $\ell^{\mathcal{S}} = \ell(y|y)$ is the maximal log-likelihood with the data y.

- The scaled deviance $D_{\mathcal{M}}^*$ is asymptotically distributed as χ_{n-p}^2 degrees of freedom. Significant deviance represents the deviation from good model fit.

- If a model $\mathcal{N}$ with q parameters, is a subset of model $\mathcal{M}$ with p parameters $(q < p)$, then

$$\frac{D_{\mathcal{N}}^* - D_{\mathcal{M}}^*}{\phi} \sim \chi_{p-q}^2.$$

Residuals of the model are critical for assessing the model. Residuals in the standard normal regression models are simply $y_i - \hat{\mu}_i$, but in the context of GLM's, both predicted values and residuals are more ambiguous. Fore example, for predictions is important to distinguish the scale: (i) predictions on scale of $\eta = x_i'\beta$ and on scale of observed responses y_i for which $\mathbb{E}y_i = g^{-1}(\eta_i)$.

As regards residuals, there are several approaches. *Response residuals* are defined as $r_i = y_i - g^{-1}(\eta_i) = y_i - \theta_i$. Also, the deviance residuals are defined as

$$r_i^D = \operatorname{sign}(y_i - \mu_i)\sqrt{d_i},$$

where d_i are observation specific contributions to the deviance D.

Deviance residuals are ANOVA-like decompositions,

$$\sum_i (r_i^D)^2 = D,$$

thus assessing the contribution of each observation to the model deviance. In addition, the deviance residuals increase with $y_i - \hat{\mu}_i$ and are distributed approximately as standard normals, irrespectively of the type of GLM.

Example. For $y_i \in \{0, 1\}$ and Binomial family,

$$r_i^D = \operatorname{sign}(y_i - \hat{y}_i)\sqrt{2\left\{y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right\}}$$

Second popular measure of goodness of fit of GLM is Pearson statistic

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

This statistics also has $n - p$ degrees of freedom.

Cellular Differentiation Data

In a biomedical study of the immuno-activating ability of two agents, TNF (tumor necrosis factor) and IFN (interferon), to induce cell differentiation, the number of cells that exhibited

Number of cells differentiating	Dose of TNF (U/ml)	Dose of IFN (U/ml)
11	0	0
18	0	4
20	0	20
39	0	100
22	1	0
38	1	4
52	1	20
69	1	100
31	10	0
68	10	4
69	10	20
128	10	100
102	100	0
171	100	4
180	100	20
193	100	100

Table 1: Cellular differentiation data.

markers of differentiation after exposure to TNF and/or IFN was recorded. At each of the 16 dose combinations of TNF/IFN, 200 cells were examined. The number y of cells differentiating corresponding to TNF/IFN combination are given in the Table 1.

The model suggested here is Poisson, of the form

$$\mu = E(y|\text{TNF, IFN}) = \exp\left\{\beta_0 + \beta_1 \, TNF + \beta_2 \, IFN + \beta_3 \, TNF * IFN\right\}.$$

```
model
{
for (i in 1:n)
   {
    numbercells[i] ~ dpois(lambda[i])
   lambda[i] <- exp(b0 + b1 *tnf[i]  + b2 *ifn[i])
   }
b0 ~ dnorm(0, 0.00001)
b1 ~ dnorm(0, 0.00001)
b2 ~ dnorm(0, 0.00001)
}


#data
list(n=16, numbercells = c(11,18,20,39,22,38,52,69,31,68,69,128,102,171,180,193),
tnf = c(0,0,0,0,  1,1,1,1,  10,10,10,10,  100,100,100,100),
ifn = c(0,4,20,100,  0,2, 20, 100,  0,4,20, 100,  0,4,20,100 ) )
```

8

```
#inits
list(b0=0, b1=0, b2=0)
```

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b0	3.573	0.05139	0.001267	3.473	3.574	3.672	1001	10000
b1	0.01313	5.921E-4	1.18E-5	0.01197	0.01314	0.01431	1001	10000
b2	0.00585	6.399E-4	1.142E-5	0.004585	0.005855	0.007086	1001	10000

```
model
{
for (i in 1:n)
   {
    numbercells[i] ~ dpois(lambda[i])
    lambda[i] <- exp(b0 + b1 *tnf[i]  + b2 *ifn[i] + eps[i])
    eps[i] ~ dnorm(0,tau)
    }
b0 ~ dnorm(0, 0.00001)
b1 ~ dnorm(0, 0.00001)
b2 ~ dnorm(0, 0.00001)
tau ~ dgamma(0.1, 0.1)
}
```

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b0	3.354	0.19	0.01375	2.979	3.35	3.738	1001	10000
b1	0.01501	0.002855	2.086E-4	0.009344	0.01496	0.02059	1001	10000
b2	0.00804	0.003409	2.615E-4	0.001794	0.007871	0.01531	1001	10000

Exercises

1. Gamma Distribution. Consider $Gamma(\alpha, \alpha/\mu)$ distribution. This parametrization was selected so that $Ey = \mu$.

Identify θ amd ϕ as functions of α and μ. Identify functions a, b and c.

Hint: The density can be represented as $\exp\left\{-\alpha \log \mu - \frac{\alpha y}{\mu} + \alpha \log(\alpha) + (\alpha - 1) \log y - \log(\Gamma(\alpha))\right\}$

2. Truncated Poisson. The zero-truncated Poisson Distribution is given by

$$f(y|\lambda) = \frac{\lambda^j}{j!(e^\lambda - 1)}, \ j = 1, 2, \dots$$

Show that f is a member of exponential family with canonical parameter $\log \lambda$.

3. Leukemia Example.

Feigl and Zelen (1965) published data concerning 33 patients suffering from leukemia. The white blood cell count (in thousands) and AG-factor (positive or negative) for each patient was recorded at the time of diagnosis, together with the time in weeks until the death from initial diagnosis. Feigl and Zelen considered the time to death (the response variable) as coming from an exponential distribution, using the logarithm of the white blood cell counts and the AG test results as covariates.

Hint: Noting that the exponential distribution is the gamma distribution with the scale parameter set to one, the model is easily fitted using glmlab. The data file comes with glmlab as `leuk.mat` and so can be loaded directly from the file, creating the variables Time, Ag and Wbc in the workspace. The error distribution can be set to gamma, the scale parameter to Fixed Value: 1, and the link function to the logarithm, log (the default link function for the gamma distribution is the reciprocal link). The variables are entered into the initial window as shown in Figure 3. Note the use of fac to indicate that Ag is a qualitative variable.

5. Shocks! An experiment was conducted to assess the effect of small electrical currents on farm animals, with the eventual goal of understanding the effects of high-voltage powerlines on livestock. The experiment was carried out with seven cows, and six shock intensities, 0, 1, 2, 3, 4, and 5 milliamps (shocks on the order of 15 milliamps are painful for many humans; see C. F. Dalziel, J.B. Lagen and J. L. Thurston, *Electric shocks*, *Trans IEEE* **60** (1941), 1073-1079). Each cow was given 30 shocks, five at each intensity, in random order. The entire experiment was then repeated, so each cow received a total of 60 shocks. For each shock the response, mouth movement, was either present or absent. The data as quoted give the total number of responses, out of 70 trials, at each shock level. We ignore cow differences and differences between blocks (experiments).

Current (milliamps) x	Number of Responses y	Number of Trials n	Proportion of Responses p
0	0	70	0.000
1	9	70	0.129
2	21	70	0.300
3	47	70	0.671
4	60	70	0.857
5	63	70	0.900

References

Dobson, A. J. (1983). An Introduction to Statistical Modelling. Chapman and Hall.

Dunn, P. K. and Smyth, G. K. (1996). Randomized Quantile Residuals. Journal of Computational and Graphical Statistics, 5:110.

Feigl, P and Zelen, M. (1965). Estimation of Exponential Survival Probabilities with Concomitant Information. Biometrics, 21:826838.

Francis, B, Green, M and Payne, C. (1993). The GLIM System: generalized linear interactive modelling. Release 4. Clarendon Press.

Hubbard, R. (1995). 53 Ways to Ask Questions in Mathematics and Statistics. Technical and Education Services.

McCullagh, P. and Nelder, J. A. (1994). Generalized Linear Models, second edition. Chapman and Hall.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. Journal of the Royal Statistical Society, A, 135:370384.

Nelder, J. A. and Pregibon, D. (1987). An Extended Quasi-Likelihood Function. Biometrika, 74:221232.

Nelder, J. A. (1990). Nearly Parallel Lines in Residual Plots. American Statistician, 44:221222.

Statistical Sciences (publisher) (1995). S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.3. StatSci, a division of MathSoft, Inc.

Thisted, R. A. (1988). Elements of Statistical Computing, Chapman and Hall.

Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. Biometrika, 61:439447

Dobson, A. J. (2001). An Introduction to Generalized Linear Models, Second Edition. Chapman and Hall/CRC (November 2001), London.

Hardin, J., and Hible, J. (2001). Generalized Linear Models and Extensions. Stata Press (May 2001), College Station, Texas.

McCulloch, C., and Searle, S. (2001). Generalized, linear and mixed models. Wiley, New York.

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (eds.) (2000). Generalized Linear Models: A Bayesian Perspective. Marcel Dekker, New York.

Wei, Bo-Cheng (1998). Exponential Family Nonlinear Models. Springer-Verlag, Sinpagore.

Jrgensen, B. (1997). Theory of Dispersion Models. Chapman and Hall, London

Applying Generalized Linear Models, James Lindsey (Limburghs Universitair Centrum, Belgium), Springer-Verlag, 1997.

Markov Chain Monte Carlo in Practice, W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Chapman & Hall, 1996.

Modelling Frequency and Count Data, J. K. Lindsey (University of Liege, Belgium), Oxford University Press, 1995.

Bayesian Data Analysis, A. Gelman, J. Carlin, H. Stern, D. Rubin, Chapman & Hall, 1995.

Multivariate Statistical Modelling Based on Generalized Linear Models, Tutz, L. Fahrmeir (University of Munich, Germany), and G. Tutz, (Technische Universitt Berlin, Germany), Springer Series in Statistics, New York, 1994.

Nonparametric Regression and Generalized Linear Models: A roughness penalty approach, Peter J. Green, Bernard W Silverman, Chapman & Hall, 1994.

Crawley, M. J. (1993). GLIM for Ecologists. Blackwell Scientific Publications, Oxford.

Statistical Models in S, John M Chambers, Trevor J Hastie, Chapman & Hall, 1992.

Generalized Additive Models, T.J. Hastie, R.J. Tibshirani, Chapman & Hall, 1990.

An Introduction to Generalized Linear Models, Second Edition, A.J. Dobson (University of Newcastle), Chapman & Hall, 1990.