CHAPTER 15

Further topics

15.1 Introduction

This chapter describes briefly a number of topics related to generalized linear models, some of which are of current research interest.

15.2 Bias adjustment

In large samples the bias of maximum-likelihood estimators is $O(n^{-1})$, and hence negligible compared with standard errors. For samples of more modest size, or for problems in which the number of parameters is appreciable compared with n, the bias may not be entirely negligible. In such cases the usual approximations can often be improved by making a bias adjustment to the maximum-likelihood estimate. In what follows we describe how the leading term in the asymptotic bias can be computed by weighted linear regression.

15.2.1 *Models with canonical link*

In the case of full exponential-family models with canonical link function, such as linear logistic models for binomial data, log-linear models for Poisson data, inverse linear models for exponential data, the approximate bias of the maximum-likelihood estimate can be obtained by a very simple supplementary computation, which we now describe.

Using tensor notation with implicit summation over indices that appear twice, the components of the approximate bias vector $\mathbf{b} = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ are given as follows:

$$E(\hat{\beta}^r - \beta^r) = b^r \simeq -\tfrac{1}{2}\kappa^{r,s}\kappa^{t,u}\kappa_{s,t,u}. \qquad (15.1)$$

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

For a derivation of this and related asymptotic formulae for maximum likelihood estimators, see McCullagh (1987, Chapter 7). In this formula $\kappa^{r,s}$ are the components of the inverse Fisher information matrix, elsewhere written using matrix notation as $(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$, where, for canonical-link models, $\mathbf{W} = \mathrm{cov}(\mathbf{Y})$. The expression for the components of the three-way array $\kappa_{s,t,u}$ in terms of the model matrix $\mathbf{X} = \{x_r^i\}$ is

$$\kappa_{s,t,u} = \sum_{i=1}^{n} x_s^i x_t^i x_u^i \kappa_{3i},$$

where κ_{3i} is the third cumulant of the ith component of the response vector.

As an intermediate step in the derivation it is helpful to consider the contracted array with components

$$b_s = -\tfrac{1}{2}\kappa_{s,t,u}\kappa^{t,u} = -\tfrac{1}{2}\sum_i x_s^i \kappa_{3i}\, x_t^i x_u^i \kappa^{t,u}. \qquad (15.2)$$

Using matrix notation, the product $x_t^i x_u^j \kappa^{t,u}$ is written as the $n \times n$ symmetric matrix

$$\mathbf{Q} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T,$$

which is the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$. Consequently the final factor appearing in (15.2) is the ith diagonal element of $\mathbf{Q}$, which we write as Q_{ii}. Thus

$$b_s = -\tfrac{1}{2}\sum_i x_s^i \kappa_{2i}\, Q_{ii}\frac{\kappa_{3i}}{\kappa_{2i}}.$$

Using matrix notation the components on the right of the above equation are just $\mathbf{X}^T\mathbf{W}\boldsymbol{\xi}$, where $\xi_i = -\frac{1}{2}Q_{ii}\kappa_{3i}/\kappa_{2i}$.

Evidently from (15.2) the bias vector $\mathbf{b}$ is obtained by premultiplying the components b_s by the inverse Fisher information matrix. This operation gives

$$\mathbf{b} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\boldsymbol{\xi}, \qquad (15.3)$$

which is easily obtained as the vector of regression coefficients in the formal linear regression of $\boldsymbol{\xi}$ on $\mathbf{X}$ with $\mathbf{W}$ as weight vector. In other words we retain the weights and the model formula from the log-linear or linear logistic model, but the link function becomes the identity and response vector becomes $\boldsymbol{\xi}$. The binomial index vector and any prior weights are assumed to be incorporated into $\mathbf{W}$.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.2.2 *Non-canonical models*

For non-canonical models the tensor expression for the first-order asymptotic bias of $\hat{\boldsymbol{\beta}}$ is a little more complicated because it involves the covariance between the vector of first-order derivatives and the matrix of second-order derivatives of the log-likelihood function. However, calculations similar to those given in the previous section show that the bias vector can be obtained by a similar supplementary regression computation. It is only necessary to re-define the formal response vector in this regression and then to use (15.3). We find that the components of $\boldsymbol{\xi}$ are given by

$$\xi_i = -\tfrac{1}{2}\Big(\frac{\mu_i''}{\mu_i'}\Big)Q_{ii}, \tag{15.4}$$

where $\mu_i' = \partial\mu_i/\partial\eta_i$ and $\mu_i'' = \partial^2\mu_i/\partial\eta_i^2$ are the derivatives of the inverse link function. Note that the weights in (15.3) are the usual quadratic weights, namely $W_i = \mu_i'^2/\kappa_{2i}$.

The following Table gives expressions for ξ_i for some common link functions.

Link	ξ_i
identity	0
log	$-Q_{ii}/2$
logit	$Q_{ii}(\pi_i - \frac{1}{2})$
probit	$Q_{ii}\eta_i/2$
c-log-log	$Q_{ii}(\exp(\eta_i) - 1)/2$

For binary regression models ξ_i has the same sign as η_i, though the vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are not co-linear in R^n. However, under conditions of approximate quadratic balance (Q_{ii} = const), and provided that $|\boldsymbol{\beta}|$ is small, it may be shown that the bias vector $\mathbf{b}$ and the parameter vector $\boldsymbol{\beta}$ are approximately co-linear. A very rough approximation for small $|\boldsymbol{\beta}|$ is

$$\mathbf{b} \simeq p\boldsymbol{\beta}/m_\cdot, \tag{15.5}$$

where $m_\cdot = \sum m_i$ and $p = \dim(\boldsymbol{\beta})$. Thus bias adjustment for binary regression models has an effect on the parameter estimates approximately the same as shrinkage towards the origin by the factor $1 - p/m_\cdot$.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.2.3 *Example: Lizard data (continued)*

To illustrate these computations we use the linear logistic model (4.24) applied to the data in Table 4.5. The parameter estimates shown in Table 4.8 lead to the following fitted quantities:

$$\begin{aligned}
\hat{\pi} &= (0.8749, 0.8977, 0.7699, 0.9558, 0.9645, 0.9120, \dots), \\
\hat{Q}_{ii} &= (0.1161, 0.1333, 0.1246, 0.1506, 0.1749, 0.1530, \dots), \\
\hat{\xi}_i &= (0.0435, 0.0530, 0.0336, 0.0687, 0.0812, 0.0630, \dots), \\
\hat{w}_i &= (2.4085, 0.8266, 1.4171, 0.5488, 0.2740, 0.9634, \dots).
\end{aligned}$$

Only the first six components of the fitted vectors are shown here: these correspond to the first two rows of Table 4.5. Note that $\hat{w}_i$ for linear logistic models is just $m_i\hat{\pi}_i(1-\hat{\pi}_i)$.

Weighted linear regression of $\hat{\boldsymbol{\xi}}$, using the same model formula, gives the bias vector $\hat{\mathbf{b}}$ shown together with $\hat{\boldsymbol{\beta}}$ in the following Table:

Parameter	*Estimate*	*S.E.*	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\beta}} - \hat{\mathbf{b}}$
μ	1.9447	0.3408	0.0436	1.9011
H	1.1300	0.2568	0.0238	1.1062
D	−0.7626	0.2112	−0.0090	−0.7536
S	−0.8473	0.3217	−0.0302	−0.8171
$T(2)$	0.2271	0.2500	−0.0009	0.2280
$T(3)$	−0.7368	0.2988	−0.0095	−0.7273

The largest biases here are about 10% of a standard error. In cases of marginal statistical significance biases of this magnitude could have a small effect on the conclusions, but they are unlikely to be of any consequence in this example. However an examination of the approximate biases is helpful here as a check on the significance of the factor S, which, though significant in the maximum-likelihood analysis, does not show up as significant in the preliminary analysis in Table 4.6 and Fig. 4.2. The bias adjustment indicated above reduces the significance of S, but not by an amount sufficient to alter the conclusions.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.3 Computation of Bartlett adjustments

15.3.1 *General theory*

A simple, or fully specified, null hypothesis $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ may be tested using the likelihood-ratio statistic, which is twice the difference between the maximum log likelihood and the value attained at $\boldsymbol{\theta}_0$. For generalized linear models in which the dispersion parameter is known, this difference may be written in terms of the deviance as follows:

$$\Lambda = 2l(\hat{\boldsymbol{\theta}}; Y) - 2l(\boldsymbol{\theta}_0; Y) = D(Y; \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}}).$$

We have assumed here for simplicity of notation that the dispersion parameter is equal to unity.

Under the usual asymptotic regularity conditions for large samples the asymptotic mean of this statistic is

$$\begin{aligned} E\{D(Y; \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}})\} &= p + \epsilon_p + O(n^{-2}), \\ &= p\{1 + b_p(\boldsymbol{\theta}_0)\} + O(n^{-2}), \end{aligned}$$

where $p = \dim(\boldsymbol{\theta})$ and $b(\boldsymbol{\theta})$ is known as the Bartlett adjustment factor. In fact it is possible to show that all cumulants up to any fixed order, r, are given to the same order of approximation by

$$\kappa_r\{D(Y; \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}})\} = (r-1)!\, 2^{r-1} p \{1 + b_p(\boldsymbol{\theta}_0)\}^r + O(n^{-2}). \tag{15.6}$$

For an outline proof in the single-parameter case, see Appendix C. The leading term in this expression is just the rth cumulant of the χ^2_p distribution. From the multiplicative property of cumulants it can be seen immediately that the cumulants of the adjusted statistic

$$\Lambda' = \frac{\Lambda}{1 + b_p}$$

agree with those of the χ^2_p distribution when terms of order $O(n^{-2})$ are ignored. Note that b_p and ϵ_p are both $O(n^{-1})$ by assumption.

Although convergence of the cumulants implies convergence in distribution provided that the asymptotic cumulants uniquely determine a distribution, the order of magnitude of the discrepancy in the cumulants is not necessarily the same as the size of the

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

error in the cumulative distribution function. Nevertheless it seems plausible to conclude that the distribution of the adjusted statistic is given by

$$\Lambda' \sim \chi^2_p + O(n^{-2}),$$

and in fact this claim is correct at least in the non-lattice case. In the lattice case the error cannot be reduced below $O(n^{-1/2})$ without resorting to discontinuous approximations, which are extremely inconvenient. It is unclear in that case whether the adjustment improves the approximation or not.

For composite null hypotheses, which are more common in applications, a similar adjustment can be made. The statistic can be written in the form

$$\Lambda(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_0) = 2l(\hat{\boldsymbol{\theta}}; Y) - 2l(\hat{\boldsymbol{\theta}}_0; Y) = D(Y; \hat{\boldsymbol{\theta}}_0) - D(Y; \hat{\boldsymbol{\theta}}), \quad (15.7)$$

where $\hat{\boldsymbol{\theta}}_0$ is the estimate of the nuisance parameters under H_0, and $\hat{\boldsymbol{\theta}}$ is the unrestricted estimate. Assuming that the hypotheses are nested, and that $q < p$ is the dimension of the parameter space under H_0, the mean of this statistic is

$$\begin{aligned} E\{\Lambda(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_0)\} &= p + \epsilon_p + O(n^{-2}) - \left(q + \epsilon_q + O(n^{-2})\right) \\ &= p - q + (\epsilon_p - \epsilon_q) + O(n^{-2}) \\ &= (p - q)\{1 + b_{pq}(\boldsymbol{\theta})\} + O(n^{-2}). \end{aligned}$$

Thus the required adjustment factor is now

$$b_{pq} = (pb_p - qb_q)/(p - q) = (\epsilon_p - \epsilon_q)/(p - q). \quad (15.8)$$

The cumulants of the maximized likelihood-ratio statistic (15.7) obey (15.6) with p replaced by $p - q$. Hence the cumulants of the adjusted statistic agree with those of χ^2_{p-q} apart from terms of order $O(n^{-2})$.

15.3.2 *Computation of the adjustment*

We focus here on computing ϵ_p as if the null hypothesis were simple. Differencing is necessary if nuisance parameters are present. All quantities are computed at $\hat{\boldsymbol{\theta}}_0$ rather than the true $\boldsymbol{\theta}$, which is usually unknown.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

The calculations that follow use the general expression (29) of McCullagh and Cox (1986). The aim here is to present that expression in a more readily computable form, by exploiting special properties of generalized linear models. Final expressions are presented in matrix notation although intermediate calculations make some use of index notation.

For the present discussion it is convenient to introduce the following diagonal matrices.

$$\begin{aligned}\mathbf{D}^{(1)} &= \operatorname{diag}\{\mu_i'\} = \operatorname{diag}\{d\mu_i/d\eta_i\},\\ \mathbf{D}^{(2)} &= \operatorname{diag}\{\mu_i'' - \mu_i'^2\, d\log V_i/d\mu_i\},\\ \mathbf{W} &= \operatorname{diag}\{\mu_i'^2/V_i\} = \mathbf{D}^{(1)}\mathbf{V}^{-1}\mathbf{D}^{(1)}.\end{aligned}$$

In addition to these, the following non-diagonal matrices arise naturally as a by-product of the weighted least squares algorithm used to compute parameter estimates.

$$\begin{aligned}\mathbf{Q} &= \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T,\\ \mathbf{P} &= \mathbf{D}^{(1)}\mathbf{Q}\mathbf{D}^{(1)}\mathbf{V}^{-1}.\end{aligned}$$

Note that $\sigma^2\mathbf{Q}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$ and $\sigma^2\mathbf{D}^{(1)}\mathbf{Q}\mathbf{D}^{(1)}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\mu}}$.

The first and second derivatives of the log likelihood with respect to the regression parameters β_r are

$$U_r = \partial l/\partial\beta_r = \sum_i x_r^i D_i^{(1)} V_i^{-1}(Y_i - \mu_i),$$

$$U_{rs} = \partial^2 l/\partial\beta_r\partial\beta_s = \sum_i x_r^i x_s^i D_i^{(2)} V_i^{-1}(Y_i - \mu_i) - x_r^i x_s^i W_i,$$

where x_r^i are the components of the model matrix $\mathbf{X}$. The Fisher information matrix is $\mathbf{X}^T\mathbf{W}\mathbf{X}$, which we write as $\kappa_{r,s}$, with inverse $\kappa^{r,s}$.

Evidently, for all generalized linear models, the log likelihood derivatives are linear functions of Y, a property that simplifies much of the subsequent calculations. A key step in deriving a simple expression for the Bartlett factor involves working with the residual second derivative matrix rather than U_{rs}. The covariance of U_{rs} and U_t is a three-way array whose components are

$$\kappa_{rs,t} = \sum_i x_r^i x_s^i D_i^{(2)} V_i^{-1} D_i^{(1)} x_t^i.$$

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

Since only arithmetic multiplication is involved, the order of terms in the above sum is immaterial. It is a remarkable property of generalized linear models that all such 'mixed' cumulants are symmetric under index permutation. In other words, for generalized linear models, but not in general, $\kappa_{rs,t} = \kappa_{rt,s} = \kappa_{st,r}$ and so on.

The residual matrix of second derivatives after linear regression on U_r is

$$V_{rs} = \sum_i x_r^i x_s^i D_i^{(2)} V_i^{-1} \{\delta_{ij} - P_{ij}\}(Y_j - \mu_j).$$

For purposes of computation we may write

$$V_{rs} \simeq \sum_i x_r^i x_s^i D_i^{(2)} V_i^{-1} (Y_i - \hat{\mu}_i),$$

but the former expression is simpler for theoretical calculations.

The Bartlett factor can now be given as a linear combination of six invariant functions of the joint cumulants of U_r and V_{rs}. The invariant fourth cumulant of U_r is

$$\begin{aligned} \rho_4 &= \sum_i x_r^i x_s^i x_t^i x_u^i \mu_i'^4 V_i^{-4} \kappa_{4i} \kappa^{r,s} \kappa^{t,u}, \\ &= \sum_i (P_{ii})^2 \rho_{4i}, \end{aligned} \tag{a}$$

where P_{ii} are the diagonal elements of the asymmetric projection matrix $\mathbf{P}$, and $\rho_{4i} = \kappa_{4i}/\kappa_{2i}^2$ is the usual standardized fourth cumulant of Y_i. The two quadratic skewness scalars are

$$\begin{aligned} \rho_{13}^2 &= \sum_{ij} x_r^i Q_{ii} \mu_i'^3 V_i^{-3} \kappa_{3i} \kappa^{r,s} x_r^j Q_{jj} \mu_j'^3 V_j^{-3} \kappa_{3j} \\ &= \sum_{ij} (P_{ii}\kappa_{3i}/\kappa_{2i}) V_i^{-1} P_{ij} (P_{jj}\kappa_{3j}/\kappa_{2j}); \end{aligned} \tag{b}$$

$$\rho_{23}^2 = \sum_{ij} (V_i^{-1} P_{ij})^3 \kappa_{3i} \kappa_{3j}. \tag{c}$$

These scalars can be computed easily using simple matrix operations.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

Similar calculations show that the two scalar measures of the variability of V_{rs} can be simplified as follows.

$$\nu_{rs,tu}\kappa^{r,s}\kappa^{t,u} = \boldsymbol{q}^T\mathbf{D}^{(2)}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P})\mathbf{D}^{(2)}\boldsymbol{q}, \qquad (d)$$

$$\nu_{rs,tu}\kappa^{r,t}\kappa^{s,u} = \sum_{ij} Q_{ij}Q_{ij}[\mathbf{D}^{(2)}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P})\mathbf{D}^{(2)}]_{ij}, \qquad (e)$$

where $\boldsymbol{q}$ is a vector with components Q_{ii}. Finally we have one further scalar

$$\nu_{r,s,tu}\kappa^{r,s}\kappa^{t,u} = \boldsymbol{q}^T\mathbf{D}^{(2)}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P})\boldsymbol{q}^*, \qquad (f)$$

where $q_j^* = q_j W_j \kappa_{3j}/\kappa_{2j}$. The notation on the left of the preceding three equations is taken from McCullagh and Cox (1986).

Although the tensor expressions on the left of (d)–(f) are algebraically more appealing than the matrix formulae, the latter expressions have the advantage for numerical purposes that they use only simple operations on matrices and vectors. Numerical computation involving higher-order arrays is thereby avoided.

In terms of these scalars the correction may be written as the linear combination

$$\epsilon_p = -\tfrac{1}{4}(a) + \tfrac{1}{4}(b) + \tfrac{1}{6}(c) - \tfrac{1}{4}(d) + \tfrac{1}{2}(e) - \tfrac{1}{2}(f). \qquad (15.9)$$

If the canonical link is used only the first three of these terms contribute.

15.3.3 *Example: exponential regression model*

Suppose that $Y_1, \ldots, Y_n$ are independent exponential random variables with $\eta_i = \log \mu_i$ satisfying the linear model $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. We find that the required matrices are as follows:

$$\mathbf{D}^{(1)} = \operatorname{diag}\{\mu_i\}, \qquad \mathbf{D}^{(2)} = \operatorname{diag}\{-\mu_i\}, \qquad \mathbf{W} = \mathbf{I},$$
$$\mathbf{Q} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \qquad \mathbf{P} = \operatorname{diag}\{\mu_i\}\mathbf{Q}\operatorname{diag}\{\mu_i^{-1}\},$$

so that $P_{ii} = Q_{ii}$ even though $\mathbf{P}$ is not symmetrical. These properties permit simplification of (15.9) giving

$$\epsilon_p = \tfrac{1}{6}\sum Q_{ij}^3 - \tfrac{1}{4}\boldsymbol{q}^T(\mathbf{I}-\mathbf{Q})\boldsymbol{q}. \qquad (15.10)$$

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

In this particular case, since the model is of the translation type, ϵ_p does not depend on the value of the parameter.

To take a simple numerical example, consider the data on survival times for 17 leukaemia patients given by Feigl and Zelen (1965). The data are discussed by Cox and Snell (1981, pp.148–150), who consider the model

$$\log \mu_i = \beta_0 + \beta_1 x_i \tag{15.11}$$

in which μ_i is the expected survival time, and x_i is the logarithm of the initial white blood cell count. The likelihood-ratio statistic for testing the hypothesis $H_0: \beta_1 = 0$ comes to 6.826 on one degree of freedom yielding a p-value of 0.89%. Here we compute the Bartlett-adjusted statistic as a check on the adequacy of the χ^2_1 approximation.

Since in this case

$$Q_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum (x_i - \bar{x})^2},$$

it follows that

$$\sum_{ij} Q_{ij}^3 = \frac{4}{n} + \frac{\left(\sum (x_i - \bar{x})^3\right)^2}{\left(\sum (x_i - \bar{x})^2\right)^3}.$$

For the leukaemia data this gives

$$\epsilon_2 = (4/17 + 0.00003849)/6 - 0.0688/4 = 0.0220.$$

Similar calculations under H_0 give $\epsilon_1 = 1/(6n) = 0.00980$. Thus $b_{21} = 0.0122$ giving a corrected statistic of 6.744. The p-value is increased to 0.94%—a trivial increase in this case. The usual χ^2 approximation appears to be quite accurate here even though the sample size is not especially large.

On the other hand if we were interested in testing the adequacy of the model (15.11) with no specific alternative in mind, the likelihood-ratio statistic is equal to 19.46 on 15 degrees of freedom. For the Bartlett adjustment we find $\epsilon_{17} = 17/6$ giving

$$b_{17,2} = (17/6 - 0.0220)/15 = 0.187.$$

Thus the adjusted statistic is $19.46/1.187 = 16.39$. The p-value is thereby increased from 19.4% to 35.6%. The 19% reduction in the value of the statistic is substantial and could in principle weaken the force of the conclusions, although that has not happened in this case. For example if the unadjusted statistic corresponds to a p-value of 1%, the p-value for the adjusted statistic is 4%.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.4 Generalized additive models

In choosing the set of terms to be included in the linear predictor of a generalized linear model we make the prior assumption that some subset of the terms chosen will give an adequate picture of the response surface generated by the covariates. If the actual shape of the surface is not expressible as a function of the terms and link function chosen then the fitted values will deviate systematically from the actual response surface. Model-checking techniques described in Chapter 12 can be used to detect this state of affairs, and perhaps to remedy the misfit. An alternative method, due to Hastie and Tibshirani (1986, 1987ab), is to fit *generalized additive models*, which, for continuous covariates x, replace the linear predictor $\eta = \sum_j x_j \beta_j$ by the additive model

$$\eta = \alpha + \sum_j f_j(x_j)$$

in which $f_j(x_j)$ are smooth functions estimated from the data. The model remains additive with respect to the covariates, but it is no longer linear in them. The functions $f_j(\cdot)$ are identifiable up to an arbitrary constant, much like the levels of a factor.

15.4.1 *Algorithms for fitting*

We consider first the fitting procedure for a single covariate x, using a generalized additive model with link function $g(\cdot)$ and variance function $V(\cdot)$. One method associates with each point (y_i, x_i) a set of neighbours on the x-axis, and estimates a value of the response function at x_i for each i by applying the standard GLM algorithm with linear predictor $\alpha + \beta x_i$ in that neighbourhood. Then $\hat{f}(x_i)$ is the fitted value of $\eta - \alpha$ at x_i, i.e. we use linear interpolation within each neighbourhood. This algorithm, also called the local-scoring algorithm, can be thought of as using a weighted running-lines smoother on the adjusted dependent variable z. The algorithm is efficient because the running-lines smoother can be updated easily as we pass from one neighbourhood to the next. Neighbourhoods are usually defined to include a certain fraction, or span, of the points. Commonly, spans of 40–50% are used, with appropriate contraction at the ends. The fit obtained depends on the span used: the shorter the span the rougher the fit.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

When more than one covariate is involved the algorithm acquires an additional loop in which each $f_j(\cdot)$ is fitted using the current estimates of the remaining functions. This is known as the back-fitting algorithm, and takes the form

for $j = 1, \ldots, p$
  form partial residual $r_j = z - \hat{\eta} + \hat{f}_j(x_j)$
  smooth r_j *with GLM weights* W *to update* $\hat{f}_j(x_j)$
repeat.

The full algorithm may be set out as follows:

Initialize: $f_j^{(0)}(x_j) = 0$, $\hat{\alpha}_0 = g(\bar{y})$
for $i = 0, 1, \ldots$
  form current estimates of $\hat{\eta}^{(i)}(\mathbf{x}_j)$, $\hat{\mu}^{(i)}$, $x^{(i)}$ *and* $W^{(i)}$
  perform back-fitting algorithm to obtain $\hat{\alpha}^{(i+1)}$ *and* $\hat{f}_j^{(i+1)}(x_j)$
repeat until deviance stabilizes.

This algorithm depends on the choice of span or fraction of points used for the local linear fit. At one extreme, if the span is 1, the algorithm produces the generalized linear fit. At the other extreme, if the span is equal to $1/n$ and if all the x-values are distinct then $\hat{\mu}_i = y_i$, and no smoothing occurs.

15.4.2 *Smoothing methods*

In the algorithm just described the running-lines smoother may be replaced by any other smoother, which may in turn be linear or non-linear. Running-lines and cubic-spline smoothers are linear in y, while a running-median smoother is non-linear. Another non-linear method involves maximizing the local likelihood for each neighbourhood. In practice, however, this usually gives results very close to that produced by the linear local scoring method.

The choice of span can be made data-dependent by choosing it to minimize the cross-validation deviance.

The effective number of parameters associated with an estimated smooth function can be calculated from the formula

$$\mathrm{tr}(2\mathbf{S} - \mathbf{S}^T\mathbf{W}\mathbf{S}\mathbf{W}^{-1})$$

in which $\mathbf{S}$ is the smoothing matrix applied to the vector $\mathbf{z}$ to produce $f(\cdot)$, and $\mathbf{W}$ is the weight matrix. For the running-lines smoother this formula simplifies to $\mathrm{tr}(\mathbf{S})$.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.4.3 *Conclusions*

Generalized additive models can be used either as a descriptive tool for expressing the joint effect of several explanatory variables as the sum of functions of them individually, or as an exploratory device to suggest a suitable class of transformations of covariates to be included in a generalized linear model. The restriction to additive terms can be relaxed by including product terms of the form $f_{12}(x_1x_2)$ in addition to $f_1(x_1)$ and $f_2(x_2)$. However not all functions of two variables are expressible in the form

$$f_1(x_1) + f_2(x_2) + f_{12}(x_1x_2).$$

Generalized partially additive models, in which some covariates enter linearly and others additively, may eventually prove to be the most useful application of these techniques.

15.5 Bibliographic notes

Bartlett (1937, 1954) gave explicit correction factors for a number of likelihood-ratio test statistics, including a number of test statistics that arise in multivariate analysis. Similar expressions for log-linear models were given by Williams (1976), and, for generalized linear models, by Cordeiro (1983, 1987). See also Ross (1987).

The matrix formulae in section 15.3 appear to be new. An alternative scheme for computing Bartlett adjustments is described by Barndorff-Nielsen and Blaesild (1986).

15.6 Further results and exercises 15

15.1 Compute the angle in R^p between the parameter vector $\hat{\boldsymbol{\beta}}$ and the bias vector $\mathbf{b}$ for the main-effects model fitted to the lizard data in section 15.2.3. Use the Fisher information as the inner product matrix. Comment briefly on the adequacy of the approximation (15.5).

15.2 Show that the matrix $\mathbf{P}$ defined in section 15.3.2 is a projection matrix. Describe the range space of $\mathbf{P}$, i.e. the set of vectors $\mathbf{x}$ such that $\mathbf{Px} = \mathbf{x}$. Explain how this space depends on the value of the parameter vector.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.

15.3 Derive expression (15.10) for the Bartlett adjustment for exponential regression models from the general expression (15.9). Show that the second term in (15.10) vanishes if $\mathbf{X}$ is the design matrix for a one-way layout.

15.4 Suppose $Y_i \sim \sigma_i^2 \chi^2_{f_i}/f_i$, independently for $i = 1, \ldots, k$. Specify the null hypothesis $H_0: \sigma_i^2 = \sigma^2$ and the unrestricted alternative as generalized linear models. Use (15.9) or (15.10) to compute the Bartlett adjustment factor. [Bartlett, 1937].

15.5 Fit the log-linear model (15.11) to the leukaemia data and check the calculations given in section 15.3.3.

Copyright © 1989. CRC Press LLC. All rights reserved.

McCullagh, P., and J. A. Nelder. Generalized Linear Models, CRC Press LLC, 1989. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=5631551.
Created from gatech on 2021-06-05 22:52:40.