J. R. Statist. Soc. B (1991)
53, *No.* 3, *pp.* 629–643

Bias Correction in Generalized Linear Models

By GAUSS M. CORDEIRO† and PETER McCULLAGH

Universidade Federal de Pernambuco, Brazil *University of Chicago, USA*

[Received October 1988. Final revision March 1990]

SUMMARY

In this paper we derive general formulae for first-order biases of maximum likelihood estimates of the linear parameters, linear predictors, the dispersion parameter and fitted values in generalized linear models. These formulae may be implemented in the GLIM program to compute bias-corrected maximum likelihood estimates to order n^{-1}, where n is the sample size, with minimal effort by means of a supplementary weighted regression. For linear logistic models it is shown that the asymptotic bias vector of $\hat{\beta}$ is almost collinear with β. The approximate formula $\beta p/m_+$ for the bias of $\hat{\beta}$ in logistic models, where $p = \dim(\beta)$ and $m_+ = \Sigma m_i$ is the sum of the binomial indices, is derived and checked numerically.

Keywords: BIAS CORRECTION; CANONICAL LINK; DEVIANCE; DISPERSION PARAMETER; GENERALIZED LINEAR MODEL; LOGISTIC MODEL; MAXIMUM LIKELIHOOD ESTIMATE

1. INTRODUCTION

It is well known that maximum likelihood estimates (MLEs) may be biased when the sample size n or the total Fisher information is small. The bias is usually ignored in practice, the justification being that it is negligible compared with the standard errors. In small or moderate-sized samples, however, a bias correction can be appreciable, and it is helpful to have a rough estimate of its size.

For a likelihood involving only one parameter to be estimated, the n^{-1} bias of the MLE from a single random sample was first given by Bartlett (1953). Haldane and Smith (1956) gave similar order expressions for the first four cumulants of the MLE. Much work concerned with bias and accuracy of estimates has been carried out by Shenton and Bowman (1963, 1969, 1977).

Shenton and Bowman (1963) developed the one-parameter case further, giving the first four sampling moments of the MLE to orders n^{-2}, n^{-3}, n^{-3} and n^{-4} respectively. The n^{-1} multiparameter biases of the MLEs have been given by Shenton and Wallington (1962) and Cox and Snell (1968). Bowman and Shenton (1965) and Shenton and Bowman (1977), chapter 3, have given general formulae for the multiparameter n^{-2} biases and n^{-2} covariances of these estimates. There is a large number of terms in their formulae and any application may require a computer. If p is the number of parameters, the maximum number of terms in the formulae for the n^{-2} biases and the n^{-2} covariances are $B(p) = 2p^3 + 6p^5 + 23p^7 + 23p^9$ and $C(p) = 1 + 7p^4 + 16p^6$ respectively. Bowman and Shenton (1965) enumerated the n^{-2} covariances for the two-parameter case and there are 169 terms in each variance and 193 in the covariance.

†*Address for correspondence*: IMPA/CNPq, Estrada Dona Castorina 110, Jardin Botânico 22460, Rio de Janeiro, Brazil.

 0035-9246/91/53629 $2.00

A general method for calculating the biases in a class of non-linear least squares problems was presented by Box (1971). Also, Robertson and Fryer (1970) developed a general method for finding the biases and covariances of moment estimators to order n^{-2}. Formulae for the biases to order n^{-1} and covariances to order n^{-2} in a typical quantal response situation with application to the standard probit model were discussed by Sowden (1971), who showed that first-order biases can be appreciable even for moderate n, and that the usual asymptotic results for variances can be considerably in error. Again, Sowden (1972) made a comparison between the maximum likelihood, minimum χ^2 and modified minimum χ^2 methods to estimate the parameters in a quantal response model, in terms of first-order bias of parameter estimates, with particular reference to the standard probit model. His results showed that maximum likelihood is the most satisfactory procedure of the three alternatives in terms of first-order bias. Furthermore, Fryer and Robertson (1972) compared moment estimates and multinomial MLEs and minimum χ^2 estimates for bias to order n^{-1} and mean-squared error to n^{-2} for a variety of mixed distributions. Following the general formulae for the n^{-1} biases developed by Cox and Snell (1968), Anderson and Richardson (1979) and McLachlan (1980) found the biases of the MLEs in logistic discrimination problems and Young and Bakir (1987) derived those in the generalized log-gamma regression model. An important recent paper by Cook *et al.* (1986) presents the biases of the MLEs and of the residuals for normal non-linear regression models.

It is convenient here to give an example of bias correction estimation. Suppose that $Y \sim B(m, \pi)$, the binomial distribution with index m and parameter π, and that $\eta = \log\{\pi/(1-\pi)\}$ is the logistic transformation of π. The MLE of η is

$$\hat{\eta} = \log\{Y/(m-Y)\},$$

whereas the usual bias-corrected estimates have the form

$$\hat{\eta}_c = \log\{(Y+c)/(m-Y+c)\}$$

for some constant, usually taken to be $c = \frac{1}{2}$ (Cox and Snell, 1979).

Obviously if $c > 0$, $|\hat{\eta}_c| < |\hat{\eta}|$, so that bias correction has the effect of shrinking the MLE towards the origin. The general bias-corrected estimate of the linear predictor described in Section 4 is equivalent in this case to the choice $c = m/2(m-1)$. There is a finite probability that $Y = 0$ or $Y = m$, and consequently that $\hat{\eta}$ is infinite. It follows that the exact bias must be infinite for all m. However, $\hat{\eta}$ is consistent for η with asymptotic bias of order $O(m^{-1})$ for large m.

It is commonly believed that, for linear logistic models at least, this result extends also to vector-valued parameters in a sense that is usually left unspecified. The purpose of this paper is to clarify this claim for vector-valued parameters in generalized linear models (GLMs) (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), and to give a rough but simple expression for the bias of the estimates.

Suppose that the random variables $Y_1, \ldots, Y_n$ are independent and each Y_i has a density in the exponential family

$$\pi(y; \theta_i, \phi) = \exp[\phi\{y\theta_i - b(\theta_i) + c(y)\} + d(y, \phi)], \tag{1.1}$$

where $b(\cdot)$, $c(\cdot)$ and $d(\cdot, \cdot)$ are known functions and ϕ is a dispersion parameter, possibly unknown. If the distribution of Y_i involves only one unknown parameter, as in, for example, binomial and Poisson models, then ϕ can be taken to have the

value 1. For two-parameter full exponential family distributions with canonical parameters ϕ and $\phi\theta$, we have $d(y, \phi) = d_1(\phi) + d_2(y)$. Here $d_1(\phi) = \log(\phi/2\pi)/2$ and $d_2(y) = 0$ for the normal distribution with variance ϕ^{-1}, $d_1(\phi) = \phi \log \phi - \log \Gamma(\phi)$ and $d_2(y) = -\log y$ for the gamma distribution with index ϕ and $d_1(\phi) = \log \sqrt{\phi}$ and $d_2(y) = -\log(2\pi y^3)/2$ for the inverse Gaussian distribution with $\phi = E(Y)^3/\mathrm{var}(Y)$, where $\Gamma(\cdot)$ is the gamma function. For independent and identically distributed (IID) observations $n^{-1}\Sigma Y_i$ and $\Sigma d_2(Y_i)$ are the minimal sufficient statistics.

The mean and the variance of Y_i are respectively $E(Y_i) = \mu_i = \mathrm{d}b(\theta_i)/\mathrm{d}\theta_i$ and $\mathrm{var}(Y_i) = \phi^{-1}V_i$, where $V = \mathrm{d}\mu/\mathrm{d}\theta$ and $\theta = \int V^{-1}\mathrm{d}\mu = q(\mu)$ is a known one-to-one function of μ. The GLM is defined by a distribution in family (1.1) and the systematic component $f(\mu) = \eta$ and $\eta = X\beta$, where $f(\cdot)$ is a known one-to-one continuously twice-differentiable function, X is a specified $n \times p$ model matrix of full rank $p < n$ and $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a set of unknown parameters to be estimated.

Denote the sample values by $y_1, \ldots, y_n$ and the total log-likelihood function for a given GLM by $L(\phi, \beta)$. The parameters ϕ and β are orthogonal since

$$E\{\partial^2 L(\phi, \beta)/\partial\phi\, \partial\beta\} = 0.$$

Sweeting (1981) has given a definition of a dispersion parameter ϕ which includes that in the exponential family (1.1). He called ϕ a dispersion parameter if the log-likelihood is of the form $L(\phi, \beta) = \phi A(y, \beta) + B(y, \phi)$ and considered the estimation of ϕ and some approximations to the marginal posterior distribution of β.

Section 2 is devoted to estimating the dispersion ϕ in family (1.1) by maximum likelihood. In Section 3, the n^{-1} and n^{-2} biases of the MLE of a function of ϕ are given. In Section 4, we derive general expressions for the n^{-1} biases of the MLEs of the β, η and μ parameters. Section 5 provides the n^{-1} bias of the MLE of ϕ. In Section 6, we show how the asymptotic bias vector for GLMs can be computed without iterative computation by means of a supplementary weighted linear regression calculation. Section 7 gives the computations of the bias of the parameter estimates in two examples applied to real data analysed by McCullagh and Nelder (1989). Finally, in Section 8, we show that, in linear logistic regression models, the asymptotic bias vector of $\hat{\beta}$ and the parameter vector β are approximately collinear in $\mathbf{R}^p$.

2. ESTIMATION OF DISPERSION PARAMETER

Let $\hat{\phi}$, $\hat{\beta}$, $\hat{\eta} = X\hat{\beta}$ and $\hat{\mu} = f^{-1}(\hat{\eta})$ be the MLEs of ϕ, β, η and μ respectively. The dispersion parameter ϕ does not enter into the estimation equations $X^{\mathrm{T}}\hat{W}X\hat{\beta} = X^{\mathrm{T}}\hat{W}\hat{y}^*$ for $\hat{\beta}$, where $W = \mathrm{diag}\{w_1, \ldots, w_n\}$, $y^* = (y_1^*, \ldots, y_n^*)^{\mathrm{T}}$ with $w_i = V_i^{-1}(\mathrm{d}\mu_i/\mathrm{d}\eta_i)^2$ and $y_i^* = \eta_i + (y_i - \mu_i)\,\mathrm{d}\eta_i/\mathrm{d}\mu_i$.

The goodness-of-fit statistic for the GLM under investigation involves the dispersion parameter ϕ and is given by $S_p = \phi D_p(y, \hat{\mu})$, where

$$D_p(y, \hat{\mu}) = 2\sum_{i=1}^{n} \{v(y_i) - v(\hat{\mu}_i) + (\hat{\mu}_i - y_i)\, q(\hat{\mu}_i)\},$$

with $v(\mu) = \mu q(\mu) - b\{q(\mu)\}$, is called the deviance, and can be computed from the data and from the MLEs $\hat{\mu}_1, \ldots, \hat{\mu}_n$. Denote the efficient score $\partial L(\phi, \beta)/\partial\phi$ by $U(\phi, \mu)$. $D_p(y, \hat{\mu}) = 2\,U(\hat{\phi}, y)$, so the deviance is twice the efficient score at the

point $(\hat{\phi}, y)$. We can write the maximum likelihood equation for $\hat{\phi}$, $U(\hat{\phi}, \hat{\mu}) = 0$, as

$$d_1'(\hat{\phi}) = n^{-1} \sum_{i=1}^{n} [y_i q(\hat{\mu}_i) - b\{q(\hat{\mu}_i)\} + c(y_i)], \tag{2.1}$$

or

$$D_p(y, \hat{\mu}) = 2 \sum_{i=1}^{n} \{v(y_i) + c(y_i)\} + 2n\, d_1'(\hat{\phi}), \tag{2.2}$$

so that $\hat{\phi}$ is a function of the deviance of the model.

For the gamma model with index ϕ, equation (2.2) reduces to $\log \hat{\phi} - \psi(\hat{\phi}) = D_p/2n$, where

$$D_p = D_p(y, \hat{\mu}) = 2 \sum_i \{\log(\hat{\mu}_i/y_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}$$

and $\psi(\cdot)$ is the digamma function. An approximate solution for $\hat{\phi}$ is obtained from

$$\hat{\phi} = \frac{n\{1 + (1 + 2D_p/3n)^{1/2}\}}{2D_p}, \tag{2.3}$$

although for the ranges $0 \leqslant D_p/2n \leqslant 0.5772$ and $0.5772 \leqslant D_p/2n \leqslant 17.0$ we can use Greenwood and Durand's (1960) expressions as rational fraction approximations to obtain $\hat{\phi}$. Also, we can compute $\hat{\phi}$ from $D_p/2n$ by using Table 37 of Pearson and Hartley (1972).

From the inequality $1/2x < \log x - \psi(x) < 1/x$ it follows that

$$n/D_p < \hat{\phi} < 2n/D_p.$$

For the normal and inverse Gaussian models the MLE $\hat{\phi}$ is simply n/D_p, where $D_p = \Sigma_i (y_i - \hat{\mu}_i)^2$ is the residual sum of squares and $D_p = \Sigma_i (y_i - \hat{\mu}_i)^2/y_i\hat{\mu}_i^2$ respectively. Evidently, ϕ is a precision parameter for inverse Gaussian models and plays much the same role as $\phi = \sigma^{-2}$ in normal models.

Another possibility for estimating ϕ is to calculate $D_p(y, \hat{\mu})$ under a reasonable model with p parameters and to equate S_p to its expected value to order n^{-1}, given by Cordeiro (1983). It would seem preferable to reduce the bias of the estimate of ϕ by including the term of order n^{-1} in the expected deviance. This estimate, $\tilde{\phi}$ say, which is, in general, different from the MLE $\hat{\phi}$, is the solution of

$$\tilde{\phi} D_p = 2\tilde{\phi}\, h(\tilde{\phi}, \hat{\mu}) - (p + \hat{\alpha}_p/\tilde{\phi}), \tag{2.4}$$

where $h(\phi, \mu) = \Sigma_i [E\{v(Y_i)\} - v(\mu_i)]$ and $\alpha_p = \alpha_p(X, \mu)$ is a function of only μ and X obtained from

$$\alpha_p = \frac{1}{4} \operatorname{tr}(HZ_d^2) - \frac{1}{3} \mathbf{1}^{\mathrm{T}} G Z^{(3)} (F + G) \mathbf{1} + \frac{1}{12} \{\mathbf{1}^{\mathrm{T}} F (2Z^{(3)} + 3Z_d Z Z_d) F \mathbf{1}\} \tag{2.5}$$

and $\hat{\alpha}_p = \alpha_p(X, \hat{\mu})$. In this expression F, G and H are diagonal matrices of order n that depend on the variance and link functions and their first and second derivatives, $Z = \{z_{ij}\} = X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}$ is, apart from the dispersion parameter ϕ, the asymptotic covariance matrix for the estimates $\hat{\eta}_1, \ldots, \hat{\eta}_n$ of the linear predictors of the model, $Z_d = \operatorname{diag}\{z_{11}, \ldots, z_{nn}\}$, $Z^{(3)} = \{z_{ij}^3\}$ and $\mathbf{1}$ is an $n \times 1$ vector of ones. If

$\alpha_p = 0$, $\tilde{\phi}$ coincides with the suggestion of Baker and Nelder (1978) for estimating the dispersion parameter in GLMs.

Usually, $h(\phi, \mu)$ does not depend on μ and is easily calculated. For the normal and inverse Gaussian models it equals $n/2\phi$ and for the gamma model it reduces to $n\{\log\phi - \psi(\phi)\}$. For special model matrices the correction α_p may be very small and with normal and inverse Gaussian distributions $\tilde{\phi} \cong \hat{\phi}$. For the gamma distribution it is the solution of

$$\log\tilde{\phi} - \psi(\tilde{\phi}) = D_p/2n + p/2n\tilde{\phi},$$

which differs from the corresponding equation for $\hat{\phi}$ by the extra term $p/2n\phi$.

3. HIGHER ORDER BIAS FOR ESTIMATE OF FUNCTION OF DISPERSION PARAMETER

Here we work with a reparameterization $\delta = d_1'(\phi)$ (see expression (2.1)), whose MLE is $\hat{\delta}$. From equation (2.2) and by using the general formula for $E(S_p)$ to order n^{-1} given by Cordeiro (1983), we can obtain the expectation of $\hat{\delta}$ by noting that for a regular problem $E\{U(\phi, \mu)\} = 0$. Therefore

$$\sum_i [E\{c(Y_i)\} + v(\mu_i)] = -n\delta$$

and

$$E(\hat{\delta}) = \delta - \frac{(p + \alpha_p/\phi)}{2n\phi}. \tag{3.1}$$

The bias term

$$b(\delta, \mu) = -\frac{(p + \alpha_p/p)}{2n\phi}$$

may be estimated by $b(\hat{\delta}, \hat{\mu})$, so defining a corrected MLE $\hat{\delta}_c = \hat{\delta} - b(\hat{\delta}, \hat{\mu})$ to order n^{-2}. The bias-corrected MLE $\hat{\delta}_c$ would be expected to have better sampling properties than the uncorrected $\hat{\delta}$. The n^{-1} bias of $\hat{\delta}$, namely $-p/2n\phi$, will depend indirectly on the model matrix X since ϕ and δ should be estimated from $\hat{\mu}$, which is a function of X. We now illustrate applications of equation (3.1) to the normal, inverse Gaussian and gamma models.

For the normal distribution with variance $\phi^{-1} = \sigma^2$ we have $\delta = \sigma^2/2$ and this leads to the n^{-2} bias-corrected MLE of σ^2

$$\hat{\sigma}_c^2 = \frac{D_p}{n - p - \alpha_p D_p/n}, \tag{3.2}$$

where $\alpha_p = \frac{1}{4}\{\mathrm{tr}(HZ_d^2) + \mathbf{1}^{\mathrm{T}}F(Z_d Z Z_d - 2Z^{(3)})F\mathbf{1}\}$ and W, F and H are diagonal matrices defined by $(\mathrm{d}\mu/\mathrm{d}\eta)^2$, $\mathrm{d}\mu/\mathrm{d}\eta\, \mathrm{d}^2\mu/\mathrm{d}\eta^2$ and $(\mathrm{d}^2\mu/\mathrm{d}\eta^2)^2$ respectively. For the normal linear model $\alpha_p = 0$ and $\hat{\sigma}_c^2$ is the usual unbiased estimate of σ^2.

For the inverse Gaussian model with dispersion parameter ϕ defined earlier, $\delta = \phi^{-1}/2$, and it follows that the n^{-2} bias-corrected MLE of ϕ^{-1}, $\hat{\phi}_c^{-1}$ say, has the same expression (3.2), where α_p is computed from equation (2.5) with $V = \mu^3$. For a simple inverse Gaussian model $\eta = \mathbf{1}\beta$, $\alpha_1 = 0$ (Cordeiro, 1983) and then

$\hat{\phi}_c^{-1} = D_1/(n-1)$ is the uniform minimum variance unbiased estimate of ϕ (Folks and Chhikara, 1978), where $D_1 = \Sigma_i (1/y_i - 1/\bar{y})$ with $\bar{y}$ the sample mean.

For the gamma model with index ϕ we obtain an approximate n^{-2} bias-corrected MLE of $\delta = \log\phi - \psi(\phi)$ as

$$\hat{\delta}_c = D_p/2n + (p + \hat{\alpha}_p \hat{\phi}^{-1})/2n\hat{\phi},$$

where $\hat{\phi}$ comes from equation (2.3).

4. BIASES OF ESTIMATES OF PARAMETERS β, η AND μ

In this section our attention is directed to bias correction of the estimates of the parameters β and μ in GLMs. Consider a GLM defined as in Section 1. We shall use the following notation for the derivatives of the log-likelihood $L = L(\phi, \beta)$: $k_{rs} = E(\partial^2 L/\partial\beta_r \partial\beta_s)$, $k_{rs}^{(t)} = \partial k_{rs}/\partial\beta_t$, $k_{r,s} = E(\partial L/\partial\beta_r \partial L/\beta_s)$, $k_{rs,t} = E(\partial^2 L/\partial\beta_r \partial\beta_s \partial L/\partial\beta_t)$ and $k_{rst} = E(\partial^3 L/\partial\beta_r \partial\beta_s \partial\beta_t)$ with the indices being replaced by ϕ when the derivatives are with respect to this parameter. Note that $k_{r,s} = -k_{rs}$ and that $k_{rs,t}$ is the covariance of the first derivative of L with respect to β_t with the mixed second derivative with respect to β_r, β_s.

The joint information matrix for the parameters ϕ and β is

$$K = \begin{array}{c} \\ \beta \\ \phi \end{array} \begin{array}{c} \begin{array}{cc} \beta & \phi \end{array} \\ \begin{pmatrix} \{-k_{rs}\} & 0 \\ 0 & -n\, d_1''(\phi) \end{pmatrix} \end{array},$$

whose inverse is

$$K^{-1} = \begin{array}{c} \\ \beta \\ \phi \end{array} \begin{array}{c} \begin{array}{cc} \beta & \phi \end{array} \\ \begin{pmatrix} \{-k^{rs}\} & 0 \\ 0 & -k^{\phi\phi} \end{pmatrix} \end{array},$$

where here $\{-k^{rs}\}$ is the inverse Fisher information matrix for β, written in matrix notation as $\phi^{-1}(X^{\mathrm{T}}WX)^{-1}$. For canonical link models, W is the covariance structure of $Y_1, \ldots, Y_n$.

Let $B_1(\hat{\beta}_a)$ be the n^{-1} bias of $\hat{\beta}_a$ and $S = \{1, \ldots, p\}$. From the general expression for the multiparameter n^{-1} biases of the MLEs given by Cox and Snell (1968) and McCullagh (1987), we can write

$$B_1(\hat{\beta}_a) = \Sigma\, k^{ar} k^{tu} \{k_{rtu}/2 + k_{rt,u}\}, \tag{4.1}$$

where the summation is over all $p+1$ parameters $\beta_1, \ldots, \beta_p$ and ϕ.

For canonical models $k_{rt,u} = 0$ for r, t, u in S and the n^{-1} bias of $\hat{\beta}$ can be expressed only in terms of the three-way array $k_{r,t,u} = -k_{rtu}$. For non-canonical models the bias 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.

Since $k_{r\phi} = 0$ and $k_{r\phi\phi} = k_{r\phi}^{(\phi)} = k_{r\phi,\phi} = 0$ for $r \in S$, in equation (4.1) we have only to take into account one summation term involving the various combinations of the β parameters. It can be shown that the crucial quantity for the n^{-1} bias of $\hat{\beta}$ is equal to

$$\tfrac{1}{2} k_{rtu} + k_{rt,u} = -\tfrac{1}{2}\phi \sum_{i=1}^{n} f_i x_{ir} x_{it} x_{iu},$$

where here $f = V^{-1} \mathrm{d}\mu/\mathrm{d}\eta \, \mathrm{d}^2\mu/\mathrm{d}\eta^2$ is a typical element of the diagonal matrix F, which appears in expression (2.5). By rearranging the summation terms in equation (4.1) we have

$$B_1(\hat{\beta}_a) = -\tfrac{1}{2}\phi \sum_i f_i \left(\sum_r k^{ar} x_{ir} \right) \left(\sum_{t,u} k^{tu} x_{it} x_{iu} \right),$$

where r, t and u vary in S and i runs over the observations.

In matrix notation the n^{-1} bias of $\hat{\beta}$ reduces to the simple form

$$B_1(\hat{\beta}) = -(2\phi)^{-1}(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Z_d F\mathbf{1}, \tag{4.2}$$

with the definitions given in Section 2. It follows that the n^{-1} bias of $\hat{\eta}$ also has a simple expression:

$$B_1(\hat{\eta}) = -(2\phi)^{-1} Z Z_d F\mathbf{1}. \tag{4.3}$$

To evaluate the n^{-1} biases of $\hat{\beta}$ and $\hat{\eta}$ we need only the variance and link functions with their first and second derivatives. In the right-hand sides of equations (4.2) and (4.3), which are of order n^{-1}, consistent estimates of the parameters ϕ and μ can be inserted to define the corrected MLEs $\hat{\eta}_c = \hat{\eta} - \hat{B}_1(\hat{\eta})$ and $\hat{\beta}_c = \hat{\beta} - \hat{B}_1(\hat{\beta})$, which should have smaller biases than the corresponding $\hat{\eta}$ and $\hat{\beta}$. From now on $\hat{B}_1(\cdot)$ means the value of $B_1(\cdot)$ at the point $(\hat{\phi}, \hat{\mu})$. Expressions (4.2) and (4.3) are applicable even if the link is not the same for each observation. For the linear model with any distribution in the exponential family (1.1), $B_1(\hat{\beta})$ and $B_1(\hat{\eta})$ are zero. This is to be expected for the normal linear model or for the inverse Gaussian no-intercept linear regression model. However, it is not obvious that this happens for any distribution in family (1.1) with identity link since $\hat{\beta}$ is obtained, apart from these cases, from a non-linear equation $X^{\mathrm{T}}\hat{W}X\hat{\beta} = X^{\mathrm{T}}\hat{W}y$ with $W = \mathrm{diag}\{V^{-1}\}$ and because of the dependence of $\hat{\beta}$ on $\hat{W}$ and y. In general, the n^{-1} bias of $\hat{\eta}$ depends on μ, except for the gamma model with log-link when W and F reduce to the identity matrix, and for special forms of the model matrix X.

We now give the n^{-1} bias of $\hat{\mu}$. Let $G_1 = \mathrm{diag}\{\mathrm{d}\mu/\mathrm{d}\eta\}$ and $G_2 = \mathrm{diag}\{\mathrm{d}^2\mu/\mathrm{d}\eta^2\}$. Since μ_i is a one-to-one function of η_i, we find to order n^{-1} that

$$B_1(\hat{\mu}_i) = B_1(\hat{\eta}_i)\, \mathrm{d}\mu_i/\mathrm{d}\eta_i + \tfrac{1}{2}\, \mathrm{var}_1(\hat{\eta}_i)\, \mathrm{d}^2\mu_i/\mathrm{d}\eta_i^2,$$

where $\mathrm{var}_1(\hat{\eta}_i)$ is the n^{-1} term in the variance of $\hat{\eta}_i$. We can write in matrix notation

$$B_1(\hat{\mu}) = (2\phi)^{-1}(G_2 - G_1 Z F) Z_d \mathbf{1}, \tag{4.4}$$

and the corrected adjusted values are defined by $\hat{\mu}_c = \hat{\mu} - \hat{B}_1(\hat{\mu})$.

For the binomial and Poisson distributions, expressions (4.2), (4.3) and (4.4) do not involve ϕ.

5. n^{-1} BIAS FOR MAXIMUM LIKELIHOOD ESTIMATE OF DISPERSION PARAMETER IN GENERALIZED LINEAR MODELS

We now give a general formula for the n^{-1} bias of the MLE of the parameter ϕ derived from equation (2.1) or (2.2). As discussed in Section 4 we can find

$$B_1(\hat{\phi}) = k^{\phi\phi} \sum_{r,s \in S} k^{rs}(k_{\phi rs}/2 + k_{\phi r,s}) + k^{\phi\phi^2}(k_{\phi\phi\phi}/2 + k_{\phi\phi,\phi}).$$

The cumulants here are given in Cordeiro (1987). We obtain

$$k_{\phi rs} = -k_{\phi r,s} = \sum_{i=1}^{n} w_i x_{ir} x_{is}$$

for r, $s \in S$ and then the first sum in the last expression reduces to $-p/2\phi$, since

$$\sum_{i=1}^{n} w_i z_{ii} = \operatorname{tr}(WZ) = \operatorname{rank}(X) = p.$$

By noting that $k_{\phi\phi,\phi} = 0$ and $k_{\phi\phi\phi} = d_1'''(\phi)$ we obtain a simple formula for the n^{-1} bias of the MLE of the dispersion parameter ϕ in GLMs

$$B_1(\hat{\phi}) = \{\phi\, d_1'''(\phi) - p\, d_1''(\phi)\}/2n\phi\, d_1''(\phi)^2. \tag{5.1}$$

Expression (5.1) depends directly on the model matrix only through its rank. To define the corrected MLE to order $n^{-1}\hat{\phi}_c = \hat{\phi} - \hat{B}_1(\hat{\phi})$, it is necessary to obtain $\hat{\phi}$ from equation (2.1) or (2.2), which is a function of the deviance of the model. We now check equation (5.1) for some models. For the gamma model with index ϕ it reduces to

$$B_1(\hat{\phi}) = \frac{p\{\phi\,\psi'(\phi) - 1\} - \{1 + \phi\,\psi''(\phi)\}^2}{2n\{\phi\,\psi'(\phi) - 1\}^2} \tag{5.2}$$

which depends on the first and the second derivatives of the digamma function. Expression (5.2) may also be written in terms of the n^{-1} variance of $\hat{\phi}$, $\operatorname{var}_1(\hat{\phi}) = -k^{\phi\phi} = \phi/n\{\phi\,\psi'(\phi) - 1\}$. For $p = 1$ expression (5.2) coincides with the n^{-1} coefficient in the asymptotic expansion for the bias of $\hat{\phi}$ in the IID case obtained by Bowman and Shenton (1968) and Shenton and Bowman (1969).

For the normal model with variance inverse ϕ, equation (5.1) implies

$$B_1(\hat{\phi}) = (p+2)\phi/n, \tag{5.3}$$

and then the n^{-1} bias-corrected MLE of ϕ is smaller than $\hat{\phi}$ and is given by

$$\hat{\phi}_c = (n-p-2) \Big/ \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2.$$

For the inverse Gaussian model with parameter $\phi = E(Y)^3/\operatorname{var}(Y)$, formula (5.3) still holds and the corrected MLE is given by

$$\hat{\phi}_c = \frac{n-p-2}{\sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2/y_i\hat{\mu}_i^2}, \tag{5.4}$$

which reduces for the IID case to the uniform minimum variance unbiased estimate

$$\hat{\phi}_c = \frac{n-3}{\sum_{i=1}^{n} (1/y_i - 1/\bar{y})}$$

of ϕ (Folks and Chhikara, 1978).

6. COMPUTATION OF ASYMPTOTIC BIAS

In this section we show how the approximate bias $B_1(\hat{\beta})$ of the MLE $\hat{\beta}$ in GLMs can be obtained by a very simple supplementary weighted linear regression computation. We define the vector ξ with components given by

$$\xi_i = -\frac{1}{2\phi}\frac{\mu_i''}{\mu_i'} z_{ii}, \tag{6.1}$$

where $\mu_i' = \mathrm{d}\mu_i/\mathrm{d}\eta_i$ and $\mu_i'' = \mathrm{d}^2\mu_i/\mathrm{d}\eta_i^2$ are the derivatives of the inverse link function and z_{ii} is the ith diagonal element of Z, the asymptotic variance of $\hat{\eta}_i$. Table 1 gives expressions for ξ_i, apart from the dispersion parameter ϕ, for some common link functions.

For models with canonical link function, such as linear logistic models for binomial data, log-linear models for Poisson data and inverse linear models for exponential data, the components of ξ are just given by $-z_{ii}k_{3i}\phi/2V_i$, where k_{3i} is the third cumulant of the ith component of the response vector.

From equation (6.1) we have

$$\xi = -\frac{1}{2\phi} W^{-1} Z_d F \mathbf{1} \tag{6.2}$$

and from equation (4.2) the bias vector $B_1(\hat{\beta})$ reduces to

$$B_1(\hat{\beta}) = (X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W \xi. \tag{6.3}$$

Expression (6.3) is easily obtained as the vector of regression coefficients in the formal linear regression of ξ on X with W as weight vector. In other words we retain the weights and the model formula from the GLM, but the link function becomes the identity and the response vector becomes ξ. The binomial index vector and any prior weights are assumed to be incorporated into W.

The bias-corrected vector $\hat{\beta}_c = \hat{\beta} - \hat{B}_1(\hat{\beta})$ can be obtained directly via a one-step Newton approximation beginning at $\hat{\beta}$ and using the adjusted response $y' = y_i + \hat{z}_{ii}\hat{\mu}_i''/2\hat{\phi}$. In other words, we recycle from the maximum likelihood fitted values leaving the link, weights and model formula unchanged, but use y' in place of y. For the binomial models there is a risk, however, that y_i' may be negative or greater than m_i, which may cause apparent contradictions in the computations.

Result (6.2) is correct also for multinomial response models of the type considered by McCullagh (1980) in which the link function is applied to the cumulative

TABLE 1
Components of ξ for some links

Link	ξ_i
Identity	0
Log	$-\frac{1}{2} z_{ii}$
Logit	$z_{ii}(\pi_i - \frac{1}{2})$
Probit	$\frac{1}{2} z_{ii}\eta_i$
Complementary log-log	$\frac{1}{2} z_{ii}(\exp \eta_i - 1)$

probability $\gamma_j = \pi_1 + \ldots + \pi_j$. In that case z_{ii} are the diagonal elements of the covariance matrix of $\hat{\eta} = g(\hat{\gamma})$, and $\mu_j''/\mu_j' = 1 - 2\gamma_j$. The computations are then a little more complicated because the cumulative totals are not independent and W is a block diagonal matrix.

7. EXAMPLES

To illustrate these computations we fit the linear logistic model (4.24) discussed by McCullagh and Nelder (1989), section 4.6, to the data in Table 4.2 of their book, which concern a comparison of site preferences for two species of lizard. For these data $n = 24$ and the binomial indices range from zero to 69, so that there is a modest degree of imbalance in the data. The model matrix X corresponding to the model formula $H + D + S + T$ is the incidence matrix for a 3×2^3 factorial design with no interaction. The parameter estimates, shown in Table 4.5 of McCullagh and Nelder (1989), lead to the following fitted probabilities and other statistics:

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

Only the first six components of the fitted vectors are shown here: these correspond to the first two rows of Table 4.2 of McCullagh and Nelder (1989). Note that $\hat{w}_i$ for linear logistic models is just $m_i \hat{\pi}_i (1 - \hat{\pi}_i)$.

Weighted linear regression of $\hat{\xi}$ using the same model formula gives the estimated bias vector $B_1(\hat{\beta})$ shown together with $\hat{\beta}$ in Table 2. The largest biases here are about 9% of a standard error. Biases of this magnitude could have a small effect on the conclusions, but they are unlikely to be of any consequence in this example.

For a second example we consider the cumulative logistic model applied to the data in Table 5.1 of McCullagh and Nelder (1989) in which there are nine ordered response categories and four treatment groups. With the estimate for level 4 set to zero, the estimated biases of the three treatment contrasts $(\hat{\beta}_1\ \hat{\beta}_2\ \hat{\beta}_3)^{\mathrm{T}}$ ignoring the eight nuisance parameters are

$$\hat{B}_1(\hat{\beta}) = (-0.032 \quad -0.096 \quad -0.064)^{\mathrm{T}}.$$

These values range from 10% to 20% of the standard errors. To a very close

TABLE 2

Computation of the estimated biases and of the corrected MLEs for lizard data

Parameter	*Estimate*	*Standard error*	$B_1(\hat{\beta})$	$\hat{\beta}_c$
1	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

approximation the bias vector here is equal to $\hat{\beta}/50$. In other words the bias appears not to be greatly affected by the number of response categories.

8. BINARY REGRESSION MODELS

For binary regression models ξ_i has the same sign as η_i, though ξ and η are not collinear in $\mathbf{R}^n$. In this section we try to examine the nature and direction of the bias correction for linear logistic models and to justify the claim made in Section 1 that the bias vector and the parameter vector are approximately collinear in $\mathbf{R}^p$.

8.1. *Direction of Asymptotic Bias for Logistic Models*

The claim made in Section 1 suggests that the bias vector $b = B_1(\hat{\beta})$ and the parameter vector β ought to be collinear or approximately so. The aim in this section is to find an expression for the angle between these vectors. All such calculations are made with respect to the Fisher information metric on $\mathbf{R}^p$ so that the angle as computed here is unaffected by the parameterization.

The angle α between b and β is given by

$$\cos \alpha = b_r \beta_s k_{r,s} / \sqrt{(b_r b_s k_{r,s} \beta_r \beta_s k_{r,s})} \tag{8.1}$$

with implicit summation over indices that appear twice. With the usual kind of limiting assumptions $b_r = O(n^{-1})$, $k_{r,s} = O(n)$ and $\beta_r = O(1)$, so that $\alpha = O(1)$ in large samples.

In what follows it is helpful to consider the contracted array with components (McCullagh, 1987)

$$b^r = -\tfrac{1}{2} k_{r,t,u} k^{t,u} = -\tfrac{1}{2} \sum_i x_{ir} k_{3i} x_{it} x_{iu} k^{t,u} \tag{8.2}$$

and to work with the covariant version of the bias vector $b^r = k_{r,s} b_s$. The numerator in equation (8.1) becomes $b^s \beta_s$. From equation (8.2) we have

$$b^r = -\tfrac{1}{2} \sum_i x_{ir} z_{ii} \pi_i (1 - \pi_i)(1 - 2\pi_i). \tag{8.3}$$

Note that $P = ZW$ is a weighted projection matrix from $\mathbf{R}^n$ on to the column space of X. Further, $Z^- = WZW$ is a generalized inverse of Z—the Moore–Penrose inverse.

The numerator in equation (8.1) becomes

$$\begin{aligned} b^r \beta_r &= \sum_i \eta_i z_{ii} \pi_i (1 - \pi_i)(\pi_i - \tfrac{1}{2}), \\ &= \eta^{\mathrm{T}} Z^- \xi, \end{aligned} \tag{8.4}$$

where $\xi_i = z_{ii}(\pi_i - \frac{1}{2})$. Similar calculations for the denominator in equation (8.1) give

$$b^r b^s k^{r,s} = \xi^{\mathrm{T}} Z^- \xi \tag{8.5}$$

and

$$\beta_r \beta_s k_{r,s} = \eta^{\mathrm{T}} Z^- \eta. \tag{8.6}$$

Thus α is the angle in $\mathbf{R}^n$ between the vectors η and ξ, inner products being taken with respect to Z^-, which may be regarded as the Fisher information matrix for η.

For the first example in Section 7 we find

$$\cos \alpha = 2.3177/\sqrt{(0.04250 \times 143.60)} = 0.9381,$$

giving $\alpha \simeq 20°$, a fairly small angle in a space of dimension 6.

The vector ξ does not ordinarily lie in the column space of X. Consequently the angle between b and β with respect to the Fisher information metric $X^{\mathrm{T}} WX$ on $\mathbf{R}^p$ is not the same as the angle between ξ and η with respect to the metric W on $\mathbf{R}^n$. The latter angle turns out to be approximately 21.5° in this example. To calculate the correct angle it is necessary first to project ξ on to the column space of X using P: α is the angle in $\mathbf{R}^n$ between the projection of ξ and η using the metric W. The use of Z^- as metric in equations (8.4)–(8.6) effectively accomplishes this projection.

Clearly, from equation (8.4), $\cos \alpha > 0$ because η_i has the same sign as $\pi_i - \frac{1}{2}$ for each i. Further, since P is a projection matrix of rank p, we have

$$\operatorname{tr}(P) = \sum_{i=1}^{n} z_{ii} w_i = p.$$

Under conditions of approximate quadratic balance ($z_{ii} = z$, a constant), we have $z = p/\Sigma_i w_i$, so that

$$\xi_i \simeq p(\pi_i - \tfrac{1}{2}) \Big/ \sum_i w_i.$$

Furthermore, if η_i is small then

$$\eta_i \simeq (\pi_i - \tfrac{1}{2})/\pi_i(1 - \pi_i) \simeq n\xi_i/p.$$

Under these conditions $\cos \alpha \simeq 1$ so that $\alpha \simeq 0$. In other words the bias vector and the parameter vector are approximately collinear. Bias reduction then implies shrinkage towards the origin. To this order of approximation the bias is equal to

$$b \simeq p\beta/m_+. \tag{8.7}$$

The assumptions required to derive approximation (8.7) are rather strong. Consequently the approximation is at best a rough guide indicating whether bias corrections might be appreciable. For the first example discussed in Section 7, $m_+ = 564$ and $p = 6$, which gives

$$p\hat{\beta}/m_+ = (0.0207 \quad 0.0120 \quad -0.0081 \quad -0.0090 \quad 0.0024 \quad -0.0078)^{\mathrm{T}}.$$

The ratios of the $O(n^{-1})$ biases to these components are

$$\hat{b}_r m_{.}/\hat{\beta}_r p = (1.94, \quad 1.98, \quad 1.11, \quad 3.35, \quad -0.38, \quad 1.22).$$

These ratios, which would all be approximately equal to unity if approximation (8.7) were accurate, are not invariant under reparameterization.

8.2. *Binary Matched Pairs*

We consider the usual linear logistic model for binary matched pairs (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, in which

$$\begin{aligned} \operatorname{logit}(\pi_{i1}) &= \eta_{i1} = \lambda_i - \Delta/2 \\ \operatorname{logit}(\pi_{i2}) &= \eta_{i2} = \lambda_i + \Delta/2. \end{aligned} \tag{8.8}$$

This is a non-regular problem in which the dimension of the parameter space $(n+1)$ increases in proportion to the number of observations, $2n$. It is easily shown that the MLEs are

$$\hat{\lambda}_i = \begin{cases} -\infty & y_{i+} = 0, \\ 0 & y_{i+} = 1, \\ \infty & y_{i+} = 2 \end{cases} \tag{8.9}$$

and

$$\tfrac{1}{2}\hat{\Delta}_u = \log(\Sigma^* Y_{i2}/\Sigma^* Y_{i1}), \tag{8.10}$$

where Σ^* denotes summation over those n^* pairs for which

$$\hat{w}_{i+} = \hat{\pi}_{i1}(1 - \hat{\pi}_{i1}) + \hat{\pi}_{i2}(1 - \hat{\pi}_{i2}) > 0.$$

These are the so-called 'mixed pairs', for which $y_{i+} = 1$. Under suitable mild conditions on the sequence λ_i it can be shown that

$$\operatorname*{plim}_{n \to \infty}(\hat{\Delta}_u) = 2\Delta,$$

so that the asymptotic bias of $\hat{\Delta}_u$ is equal to Δ.

Despite the fact that this is a non-regular problem for which equation (4.1) is not valid as $n \to \infty$, it is of interest to compute the putative bias vector b, as a check on the reliability of bias corrections. To compute the component of b corresponding to $\hat{\Delta}$ we could use equation (4.1) directly, but it is simpler to use equation (6.3). Straightforward calculations reveal that

$$\hat{z}_{ii} = (n^* + 1)/n^* \hat{w}_{i+}, \tag{8.11}$$

where n^* is the number of mixed pairs. It can also be seen that $\hat{z}_{ii} = z$ because $\hat{\pi}(1 - \hat{\pi})$ is a constant for all the mixed pairs. Thus, for the mixed pairs,

$$\begin{aligned} \hat{\xi}_{i1} &= \frac{n^* + 1}{n^*} \frac{\hat{\pi}_{i1} - \frac{1}{2}}{\hat{w}_{i+}} \\ \hat{\xi}_{i2} &= -\hat{\xi}_{i1}. \end{aligned}$$

Therefore, from equation (6.3) we have

$$\begin{aligned} \hat{b}_\Delta &= \hat{\xi}_{i2} - \hat{\xi}_{i1} \\ &= \frac{n^* + 1}{n^*} \sinh\left(\frac{1}{2}\hat{\Delta}_u\right) \simeq \frac{n^* + 1}{n^*}\left(\frac{1}{2}\hat{\Delta}_u + \frac{1}{48}\hat{\Delta}_u^3 + \ldots\right). \end{aligned} \tag{8.12}$$

The bias-corrected estimate is then

$$\hat{\Delta}_u - \frac{n^* - 1}{n^*} \sinh\left(\frac{1}{2}\hat{\Delta}_u\right) \simeq \frac{n^* - 1}{2n^*} \hat{\Delta}_u - \frac{1}{48n^*} \hat{\Delta}_u^3 + \dots . \tag{8.13}$$

Thus the bias correction is effective for small $\hat{\Delta}_u$. The correction is too large for $\hat{\Delta}_u$ larger than about 2.0.

The conditional MLE for this problem is $\hat{\Delta}_c = \frac{1}{2}\hat{\Delta}_u$. Further, the bias-corrected conditional MLE $(n^* - 1)\hat{\Delta}_c/n^*$, which is the leading term in the Taylor expansion (8.13), differs from the unconditional bias-corrected estimate by $O(\hat{\Delta}^3)$.

8.3. *Non-canonical Models*

The results given in Section 8.1 for logistic models apply equally to linear probit and linear complementary log–log-models, essentially because ξ_i and η_i have the same sign. For example, if we fit the linear probit model to the first example discussed in Section 6, we find that the angle between $\hat{\beta}$ and b satisfies $\cos\alpha = 0.9335$, giving $\alpha \cong 20°$. The complementary log–log-link yields an angle of 17.3°.

The results given in Section 8.2 for matched pairs apply also to models for which the derivative of the link function is symmetric about $\pi = \frac{1}{2}$. The fitted values are independent of the link function provided that the link function is symmetric in this sense. Thus, replacing the logit function in equations (8.8) by the probit function, we find that the MLEs are given by equations (8.9) and

$$\tfrac{1}{2}\hat{\Delta}_u = \Phi^{-1}(\Sigma^* Y_{i2}/n^*) = \Phi^{-1}(\hat{\pi}_2),$$

say. The estimated bias of $\hat{\Delta}_u$ is given by equation (8.12), which reduces to

$$\hat{b}_\Delta = \frac{n^* + 1}{2n^* \hat{w}_{i+}} \hat{\Delta}_u .$$

The probit weight function

$$\hat{w}_{i+} = \frac{2\alpha^2(\hat{\Delta}_u/2)}{\hat{\pi}(1 - \hat{\pi})} = 1.273 + O(\hat{\Delta}_u^2)$$

tends to zero as $|\hat{\Delta}_u|$ becomes large. Thus, for small $\hat{\Delta}_u$, the estimated bias is approximately $\hat{\Delta}_u/2.55$. The actual bias of $\hat{\Delta}_u$ depends on the nuisance parameters λ_i, but if $\lambda_i \simeq 0$ for the mixed pairs, then

$$\operatorname{plim}(\hat{\Delta}_u) = 2\Delta .$$

Thus the approximate bias correction is insufficient for small $\hat{\Delta}$.

For asymmetric link functions $\hat{\lambda}_i$ is not a function of y_{i+} alone, as in equations (8.9). Consequently the results given in Section 8.2 do not apply to the complementary log–log-link.

REFERENCES

Anderson, J. A. and Richardson, S. C. (1979) Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, **21**, 71–78.

Baker, R. J. and Nelder, J. A. (1978) *The GLIM Manual: Release 3*, sect. 4.4. Oxford: Numerical Algorithms Group.

Bartlett, M. S. (1953) Approximate confidence intervals II. *Biometrika*, **40**, 306–317.

Bowman, K. O. and Shenton, L. R. (1965) Biases and covariances of maximum likelihood estimators. *Report K-1633*. Union Carbide Corporation, Oak Ridge.

——— (1968) Properties of estimators for the gamma distribution. *Report CTC-1*. Union Carbide Corporation, Oak Ridge.

Box, M. J. (1971) Bias in nonlinear estimation (with discussion). *J. R. Statist. Soc.* B, **33**, 171–201.

Cook, R. D., Tsai, C.-L. and Wei, B. C. (1986) Bias in nonlinear regression. *Biometrika*, **73**, 615–623.

Cordeiro, G. M. (1983) Improved likelihood ratio statistics for generalized linear models. *J. R. Statist. Soc.* B, **45**, 404–413.

——— (1987) On the corrections to the likelihood ratio statistics. *Biometrika*, **74**, 265–274.

Cox, D. R. and Snell, E. J. (1968) A general definition of residuals (with discussion). *J. R. Statist. Soc.* B, **30**, 248–275.

——— (1979) *Analysis of Binary Data*, p. 32. London: Chapman and Hall.

Folks, J. L. and Chhikara R. S. (1978) The inverse Gaussian distribution and its statistical applications—a review (with discussion). *J. R. Statist. Soc.* B, **40**, 263–289.

Fryer, J. G. and Robertson, C. A. (1972) A comparison of some methods for estimating mixed normal distributions. *Biometrika*, **59**, 639–648.

Greenwood, J. A. and Durand, D. (1960) Aids for fitting the gamma distribution by maximum likelihood. *Technometrics*, **2**, 55–65.

Haldane, J. B. S. and Smith, S. M. (1956) The sampling distribution of a maximum likelihood estimate. *Biometrika*, **63**, 96–103.

McCullagh, P. (1980) Regression models for ordinal data (with discussion). *J. R. Statist. Soc.* B, **42**, 109–142.

——— (1987) *Tensor Methods in Statistics*. London: Chapman and Hall.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. London: Chapman and Hall.

McLachlan, G. J. (1980) A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics*, **22**, 621–627.

Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.

Pearson, E. S. and Hartley, H. O. (1972) *Biometrika Tables for Statisticians*, vol. 2. Cambridge: Cambridge University Press.

Robertson, C. A. and Fryer, J. G. (1970) The bias and accuracy of moment estimators. *Biometrika*, **57**, 57–65.

Shenton, L. R. and Bowman, K. (1963) Higher moments of a maximum-likelihood estimate. *J. R. Statist. Soc.* B, **25**, 305–317.

——— (1969) Maximum likelihood estimate moments for the two-parameter gamma distribution. *Sankhya* B, **31**, 379–396.

——— (1977) *Maximum Likelihood Estimation in Small Samples*. London: Griffin.

Shenton, L. R. and Wallington, P. A. (1962) The bias of moment estimators with an application to the negative binomial distribution. *Biometrika*, **49**, 193–204.

Sowden, R. R. (1971) Bias and accuracy of parameter estimates in a quantal response model. *Biometrika*, **58**, 595–603.

——— (1972) On the first-order bias of parameter estimates in a quantal response model under alternative estimation procedures. *Biometrika*, **59**, 573–579.

Sweeting, T. (1981) Scale parameters: a Bayesian treatment. *J. R. Statist. Soc.* B, **43**, 333–338.

Young, D. H. and Bakir, S. T. (1987) Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183–191.