Covariance matrix of the bias-corrected maximum likelihood estimator in generalized linear models

Barroso, Lúcia
University of São Paulo, Department of Statistics
Rua do Matão, 1010
São Paulo (05508-090), Brazil
E-mail: lbarroso@ime.usp.br

Cordeiro, Gauss
Federal University of Pernambuco, Department of Statistics
Recife (50740-540), Brazil
E-mail: gausscordeiro@uol.com.br

Botter, Denise
University of São Paulo, Department of Statistics
Rua do Matão, 1010
São Paulo (05508-090), Brazil
E-mail: botter@ime.usp.br

Cavalcanti, Alexsandro
Federal University of Campina Grande, Department of Mathematics and Statistics
Campina Grande (58109-970), Brazil
E-mail: alexbc@dme.ufcg.edu.br

Introduction

Generalized linear models (GLMs) (Nelder and Wedderburn, 1972) are regression models typically fitted by maximum likelihood. The methods for analysis of a GLM are based on the asymptotic properties of the maximum likelihood estimators (MLEs). Standard references such as McCullagh and Nelder (1989) and Fahrmeir and Tutz (1994) discussed both theory and applications of these models.

The random variables $Y_1, \ldots, Y_n$ in GLMs are assumed independent and each Y_l, $l = 1, \ldots, n$, has a density (or probability) function in the family of distributions

$$\pi(y; \theta_l, \phi) = \exp\{\phi[y\theta_l - b(\theta_l)] + a(y, \phi)\}, \quad (1)$$

where $a(\cdot,\cdot)$ and $b(\cdot)$ are known functions. The mean and the variance of Y_l are

$$\mathrm{E}(Y_l) = \mu_l = \mathrm{d}b(\theta_l)/\mathrm{d}\theta_l \text{ and } \mathrm{Var}(Y_l) = \phi^{-1}V_l,$$

where $V_l = \mathrm{d}\mu_l/\mathrm{d}\theta_l$ is called variance function and $\theta_l = \int V_l^{-1}\mathrm{d}\mu_l = q(\mu_l)$ is a known one-to-one function of the mean μ_l that varies in a subset of $\mathbb{R}$. The parameters θ_l and ϕ are called the canonical and precision parameters, respectively. The precision parameter ϕ is assumed to be a known constant. If ϕ is unknown, we assume that it can be replaced by a consistent estimate in order that the function (1) represents the exponential family of distributions with natural parameter θ_l. The inverse of ϕ is the dispersion parameter of the distribution. The choice of the variance function V_l determines the interpretation of ϕ. We consider GLMs having a systematic component defined by $c(\mu) = \eta = X\beta$, where X is a specified $n \times p$ model matrix of full rank $p < n$, $\beta = (\beta_1, \ldots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $\mu = (\mu_1, \ldots, \mu_n)^\top$, $\eta = (\eta_1, \ldots, \eta_n)^\top$ is the linear predictor and $x_l^\top = (x_{l1}, \ldots, x_{lp})$ is the lth row of X. We consider that $d(.)$ is a known one-to-one continuously twice differentiable monotonic function called the mean link function. The MLE $\hat{\beta}$ of β can be calculated using the Newton-Raphson's method.

Some attempts have been made to develop second-order asymptotic theory for GLMs in order to have better likelihood inference procedures about the vector of linear parameters β. Second-order bias correction in GLMs was investigated by Cordeiro and McCullagh (1991). They showed how the asymptotic bias vector of the MLE $\hat{\beta}$ can be obtained without iterative computation by means of a supplementary weighted linear regression calculation. Further, Cordeiro (2004) obtained an expression for the n^{-2} asymptotic covariance matrix of $\hat{\beta}$.

Ferrari et al. (1996) and Pace and Salvan (1997) derived second-order bias-corrected MLEs in general one-parametric models. They also obtained closed-form expressions for the variance of the corrected estimates. Pace and Salvan (1997) noted that "the general formula for the n^{-2} asymptotic covariance matrix of the MLE (Peers and Iqbal, 1985) allows one to obtain a multiparameter generalization of the n^{-2} asymptotic covariance matrix of the bias-corrected estimator". Following this point and based on the works by Cordeiro and McCullagh (1991) and Cordeiro (2004), we obtain the n^{-2} asymptotic covariance matrix of the (second-order) bias-corrected MLE of β in GLMs.

The plan of the paper is as follows. In Section 2, we obtain a general formula for the n^{-2} asymptotic covariance matrix of the bias-corrected MLE of β. We demonstrate that this matrix is given by a sum of terms that depend on the n^{-2} asymptotic covariance matrix of the MLE. In Section 3, we present a modified Wald statistic to test the vector β that is built up from the n^{-2} asymptotic covariance matrix of the bias-corrected MLE. In Section 4, we develop some simulations to investigate the covariance of the bias-corrected MLEs in GLMs based on second-order asymptotics and to motivate the use of the proposed formula. Section 5 provides some conclusion remarks.

n^{-2} Asymptotic Covariance Matrix

In this section, we obtain the asymptotic covariance matrix up to order n^{-2} of the bias-corrected MLEs (which are of order n^{-1}) in GLMs. Denote the total log likelihood function for β by $\ell = \ell(\beta)$ and the joint cumulants of log- likelihood derivatives by $\kappa_{rs} = \mathrm{E}(\partial^2\ell/\partial\beta_r\partial\beta_s), \kappa_{r,s} = \mathrm{E}(\partial\ell/\partial\beta_r\partial\ell/\partial\beta_s)$, $\kappa_{rst} = \mathrm{E}(\partial^3\ell/\partial\beta_r\partial\beta_s\partial\beta_t)$, $\kappa_{r,st} = \mathrm{E}(\partial\ell/\partial\beta_r\partial^2\ell/\partial\beta_s\partial\beta_t)$, etc. All κ's refer to a total over the sample and are, in general, of order n. The total expected information matrix $K_\beta(\beta)$ has elements $\kappa_{r,s} = -\kappa_{rs}$, and let $\kappa^{r,s} = -\kappa^{rs}$ be the corresponding elements of its inverse. When n increases, we assume that the MLE $\hat{\beta}$ converges to β and that its asymptotic distribution is multivariate normal with mean β and covariance matrix $\mathrm{Cov}(\hat{\beta}) = K_\beta^{-1}(\beta){=}\phi^{-1}(X^\top WX)^{-1}$, where $W = \mathrm{diag}\{(\mathrm{d}\mu/\mathrm{d}\eta)^2\mathrm{V}^{-1}\}$.

Let $\tilde{\beta} = \hat{\beta} - d(\hat{\beta})$ be the bias-corrected MLE of order n^{-1}, where $d(\beta)$ denotes the bias of order n^{-1} of $\hat{\beta}$. Let $\tilde{\beta}_r$ be the rth component of the vector $\tilde{\beta}$. Then, $\tilde{\beta}_r = \hat{\beta}_r - d^r(\hat{\beta})$, where $d^r(\hat{\beta})$ is the rth component of $d(\hat{\beta})$.

From the book of Pace and Salvan (1997) (p. 360), we can write

$$d^r(\hat{\beta}) = d^r(\beta) + \sum_v d^r_v\,(\hat{\beta}_v - \beta_v) + O_p(n^{-2}), \tag{2}$$

where

$$\begin{aligned} d^r_v &= \frac{\partial d^r}{\partial\beta_v} = \sum_{w,s,y,t,u}\{\kappa^{rw}\kappa^{sy}\kappa^{tu}(\kappa_{stu} + 2\kappa_{st,u})(\kappa_{vwy} + \kappa_{v,wy}) \\ &+ \frac{1}{2}\kappa^{rs}\kappa^{tu}(\kappa_{stuv} + \kappa_{stu,v} + 2\kappa_{stv,u} + 2(\kappa_{st,uv} + \kappa_{st,u,v}))\} \end{aligned}$$

is a term of order n^{-1} and

$$d^r(\beta) = \frac{1}{2}\sum_{s,t,u}\kappa^{rs}\kappa^{tu}(\kappa_{stu} + 2\kappa_{st,u}),\ r = 1, ..., p.$$

Here, we obtain a closed-form expression up to order n^{-2} for the (r,s)th component of the matrix $\mathrm{Cov}(\tilde{\beta})$ in GLMs, i.e.,

$$[\mathrm{Cov}(\tilde{\beta})]_{rs} = \mathrm{E}[(\tilde{\beta}_r - \beta_r)(\tilde{\beta}_s - \beta_s)].$$

We have

$$\begin{aligned}[\mathrm{Cov}(\tilde{\beta})]_{rs} &= \mathrm{E}[(\hat{\beta}_r - d^r(\hat{\beta}) - \beta_r)(\hat{\beta}_s - d^s(\hat{\beta}) - \beta_s)] \\ &= \mathrm{E}\{[(\hat{\beta}_r - \beta_r) - d^r(\hat{\beta})][(\hat{\beta}_s - \beta_s) - d^s(\hat{\beta})]\} \\ &= \mathrm{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] - \mathrm{E}[d^s(\hat{\beta})(\hat{\beta}_r - \beta_r)] \\ &- \mathrm{E}[d^r(\hat{\beta})(\hat{\beta}_s - \beta_s)] + \mathrm{E}[d^r(\hat{\beta})d^s(\hat{\beta})]. \end{aligned} \tag{3}$$

From (2) we have

$$\begin{aligned}\mathrm{E}[d^s(\hat{\beta})(\hat{\beta}_r - \beta_r)] &= \mathrm{E}\left\{(\hat{\beta}_r - \beta_r)\left[d^s(\beta) + \sum_v d^s_v\,(\hat{\beta}_v - \beta_v) + O_p(n^{-2})\right]\right\} \\ &= d^s(\beta)d^r(\beta) + \sum_v d^s_v(-\kappa^{rv}) + o(n^{-2})\end{aligned}$$

and $\mathrm{E}[d^r(\hat{\beta})d^s(\hat{\beta})] = d^s(\beta)d^r(\beta) + o(n^{-2})$.

Thus, expression (3) can be written up to order n^{-2} as

$$[\mathrm{Cov}(\tilde{\beta})]_{rs} = \mathrm{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] - d^r(\beta)d^s(\beta) + \sum_v d^r_v(\kappa^{sv}) + \sum_v d^s_v(\kappa^{rv}). \tag{4}$$

The first and second terms of equation (4) is the covariance up to order n^{-2} of $\hat{\beta}$ if the parameter ϕ is known, which was obtained by Cordeiro (2004). It is given in matrix notation by

$$\phi^{-1}(X^\top WX)^{-1} + \phi^{-2}P\Lambda P^\top + \phi^{-2}(X^\top WX)^{-1}\Delta(X^\top WX)^{-1},$$

where $\Lambda = HZ_d + \frac{3}{2}FZ^{(2)}F + GZ^{(2)}F - GZ^{(2)}G$, $P = (X^\top WX)^{-1}X^\top$, $Z = XP$, $H = \mathrm{diag}\{h_1,..,h_n\}$, $h_\ell = -\mu'_\ell\mu'''_\ell/V_\ell - \mu'^2_\ell\mu''_\ell V^{(1)}_\ell/V^2_\ell + \mu'^4_\ell V^{(1)2}_\ell/V^3_\ell$, $\mu'_\ell = \mathrm{d}\mu_\ell/\mathrm{d}\eta_\ell$, $\mu''_\ell = \mathrm{d}^2\mu_\ell/\mathrm{d}\eta^2_\ell$, $\mu'''_\ell = \mathrm{d}^3\mu_\ell/\mathrm{d}\eta^3_\ell$, $V^{(1)}_\ell = \mathrm{d}V_\ell/\mathrm{d}\mu_\ell$, $F = \mathrm{diag}\{f_1,...,f_n\}$, $f_\ell = V^{-1}_\ell\mu'_\ell\mu''_\ell$, $G = \mathrm{diag}\{g_1,...,g_n\}$, $g_\ell = V^{-1}_\ell\mu'_\ell\mu''_\ell - V^{-2}_\ell V^{(1)}_\ell\mu'^3_\ell$ and $Z^{(2)} = Z \odot Z$, where $\odot$ denotes the Hadamard product. Further, the index d indicates that the diagonal matrix is obtained from the original matrix. The matrix Δ is defined by

$$\Delta = \sum_{\ell=1}^{n} \Delta_\ell c_\ell,$$

where $\Delta_\ell = (f_\ell + g_\ell)x_\ell x^\top_\ell$, $c_\ell = \delta^\top_\ell Z_\beta Z_{\beta_d} F\mathbf{1}$, $x^\top_\ell = (x_{\ell 1},\ldots,x_{\ell p})$ is the ℓth row of the covariate matrix X, $Z_\beta = X(X^\top WX)^{-1}X^\top$, δ_ℓ is a vector of dimension $(n\times 1)$ with one in the position ℓ and zero in the other positions and $\mathbf{1}$ is an $(n\times 1)$ vector of ones.

The term $\sum_v d^s_v(\kappa^{rv})$ can be expressed in matrix notation as

$$\frac{1}{\phi^2}(X^\top WX)^{-1}\Delta(X^\top WX)^{-1} - \frac{1}{2\phi^2}PDZ_dP^\top,$$

where $D = \mathrm{diag}\left\{V^{-2}V^{(1)}\mu'^2\mu'' - V^{-1}\mu'\mu''' - V^{-1}\mu''^2\right\}$.

Then, the asymptotic covariance matrix up to order n^{-2} of the bias-corrected MLE reduces to

$$\begin{aligned}\mathrm{Cov}(\tilde{\beta}) &= \phi^{-1}(X^\top WX)^{-1} + \phi^{-2}P\Lambda P^\top + 3\phi^{-2}(X^\top WX)^{-1}\Delta(X^\top WX)^{-1} \\ &- \phi^{-2}PDZ_dP^\top.\end{aligned} \tag{5}$$

A Modified Wald Test

Here, we present two alternative versions of the classical Wald statistic: the corrected statistic by replacing only the estimate $\hat{\beta}$ by the bias-corrected estimate $\tilde{\beta}$ and the corrected statistic by replacing $\hat{\beta}$ by $\tilde{\beta}$ and the inverse of the information matrix by the covariance matrix of $\tilde{\beta}$ given by (5).

Suppose that we are interested in testing all elements of the parameter β. In this situation, the null hypothesis is $H_0 : \beta = \beta^{(0)}$ and the alternative hypothesis $H_1 : \beta \neq \beta^{(0)}$, where $\beta^{(0)}$ is a specified vector of dimension p. The classical Wald test statistic is

$$W_0 = (\hat{\beta} - \beta^{(0)})^\top K_\beta(\hat{\beta})(\hat{\beta} - \beta^{(0)}),$$

where $K_\beta(\beta) = \phi(X^\top WX)$ is determined at the MLE $\hat{\beta}$ of β.

We can modify this statistic by substituting the estimate $\hat{\beta}$ by its bias-corrected estimate $\tilde{\beta}$ which implies that

$$W_m = (\tilde{\beta} - \beta^{(0)})^\top K_\beta(\tilde{\beta})(\tilde{\beta} - \beta^{(0)})$$

Another modified statistic can be constructed by substituting $\hat{\beta}$ by the corresponding bias-corrected estimate $\tilde{\beta}$ and using its covariance matrix (obtained in the previous section) evaluated at $\tilde{\beta}$. In this case, the statistic becomes

$$W_c = (\tilde{\beta} - \beta^{(0)})^\top \left\{\mathrm{Cov}(\tilde{\beta})\right\}^{-1} (\tilde{\beta} - \beta^{(0)}).$$

Simulation Results

In order to check our theoretical results, we perform two simulation experiments. First, we consider a gamma regression model with systematic component $\mu_l = \exp(\beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l})$, where the values of the covariates x_{1l} and x_{2l}, for $l = 1, \ldots, n$, were chosen as random draws from a uniform $U(0,1)$ distribution, their values being held constant throughout the simulations with equal sample sizes. The number of observations was set at $n = 10, 20, 30$ and 40 and the precision parameter was fixed at $\phi = 2$. The true values of the linear parameters for the simulations were taken as $\beta_0 = 1$, $\beta_1 = 1$ and $\beta_2 = -1$.

The simulation was performed using the R programming environment. We carried out size simulations based on 10,000 replications. In each of these replications, we fitted the gamma model and computed the MLE $\hat{\beta}$, the bias-corrected estimate $\tilde{\beta}$, the asymptotic covariance matrices $\mathrm{Cov}(\hat{\beta})$ and $\mathrm{Cov}(\tilde{\beta})$ evaluated at $\hat{\beta}$ and $\tilde{\beta}$, respectively.

The first and third entries in Table 1 are the sample means of the asymptotic expansions based on the 10,000 replications, i.e.

$$\frac{1}{10,000}\sum_{i=1}^{10,000} \mathrm{Cov}(\hat{\beta}) \text{ and } \frac{1}{10,000}\sum_{i=1}^{10,000} \mathrm{Cov}(\tilde{\beta}).$$

The second entry in Table 1 refers to the sample means of the empirical mean square errors (EMSEs) evaluated at the true value of β and based on the 10,000 replications, i.e.

$$\frac{1}{10,000}\sum_{i=1}^{10,000} (\tilde{\beta}^{(i)} - \beta).$$

The figures in Table 1 show that the covariances obtained from equation (5) are closer to the sample means of the EMSEs of $\tilde{\beta}^{(1)}, \ldots, \tilde{\beta}^{(10,000)}$ than those quantities obtained from the expansion of $\mathrm{Cov}(\hat{\beta})$. Further, if we consider these quantities in absolute values, we note that the variances and covariances of the bias-corrected estimators are larger than those quantities of the uncorrected estimators.

Next, we consider a simulation experiment to show the performance of the Wald statistics W_0, W_m and W_c discussed in Section 3. The null hypothesis $H_0 : \beta_0^{(0)} = 1$, $\beta_1^{(0)} = 1$ and $\beta_2^{(0)} = -1$ is tested against the alternative hypothesis of violation of at least one equality. The number of observations was set at $n = 10, 20, 30, \ldots, 100$ and we report the results for three different nominal significance levels, namely $\alpha = 0.01, 0.05$ and 0.10. The estimated sizes of the three Wald tests are given in Table 2, where the entries are percentages.

The figures in Table 2 reveal the better performance of the statistic W_c compared with the statistics W_0 and W_m and show the importance in correcting not only the MLE $\hat{\beta}$ but also take into account the covariance of the bias-corrected estimator $\tilde{\beta}$. The three statistics are liberal, over-rejecting the null hypothesis more frequently than expected based on the selected nominal levels, specially for small sample sizes. When n increases, the empirical sizes of the three statistics converge to the true nominal levels and the values of the statistics W_0 and W_m converge to the value of W_c.

Table 1: $\mathrm{Cov}(\hat{\beta})$ ***evaluated at*** $\hat{\beta}$***,*** $\mathrm{EMSE}(\tilde{\beta})$ ***and*** $\mathrm{Cov}(\tilde{\beta})$ ***evaluated at*** $\tilde{\beta}$

	$n = 10$			$n = 20$		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
	0.63166	-0.48998	-0.69830	0.27125	-0.25231	-0.22918
$\hat{\beta}_0$	0.70838	-0.54588	-0.78027	0.29176	-0.26860	-0.24963
	0.70199	-0.55431	-0.78011	0.29294	-0.27414	-0.25052
		0.68931	0.29104		0.35014	0.14293
$\hat{\beta}_1$		0.76220	0.32982		0.36884	0.15764
		0.77835	0.33054		0.37606	0.15943
			1.15772			0.30545
$\hat{\beta}_2$			1.29040			0.33013
			1.28545			0.33062
	$n = 30$			$n = 40$		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
	0.13183	-0.10180	-0.12912	0.09770	-0.07414	-0.09273
$\hat{\beta}_0$	0.14195	-0.10864	-0.13982	0.10308	-0.07684	-0.09812
	0.13763	-0.10758	-0.13541	0.10055	-0.07659	-0.09610
		0.23933	-0.01453		0.13910	0.00690
$\hat{\beta}_1$		0.24785	-0.00902		0.14365	0.00764
		0.25167	-0.01437		0.14361	0.00729
			0.25562			0.17390
$\hat{\beta}_2$			0.27003			0.18252
			0.26736			0.18006

Concluding Remarks

In recent years there has been considerable interest in finding closed-form expressions for second-order biases and covariances of maximum likelihood estimators (MLEs) in some classes of regression models which do not involve cumulants of log-likelihood derivatives. We consider the important class of generalized linear models (GLMs) and derive a general formula for the second-order asymptotic covariance matrix of the bias-corrected maximum likelihood estimators of the linear parameters. The usefulness of the formula is illustrated in order to estimate the covariance matrix of these estimators and to construct improved Wald test statistics.

We define two corrected statistics as alternative to the classical Wald statistic. The performances of the three statistics in GLMs are evaluated through simulation studies. The simulations indicate that the elements of the inverse of the expected information matrix underestimate the variances and covariances of the bias-corrected estimator $\hat{\beta}$. We can also obtain more precise estimates of these variances and covariances if we use the elements of the bias-corrected estimate. We examine the performance of the Wald test statistic for testing the β parameter of a GLM with known dispersion. Instead of using the uncorrected estimate $\hat{\beta}$ in the estimated inverse of the expected information matrix, we can construct improved Wald statistics by considering the bias-corrected estimate $\tilde{\beta}$ and its estimated second-order covariance matrix.

Table 2: Estimated sizes of the three Wald tests W_0***,*** W_m ***and*** W_c***.***

n	$\alpha(\%)$	W_0	W_m	W_c	n	$\alpha(\%)$	W_0	W_m	W_c
	1.0	3.09	2.39	1.99		1.0	1.19	1.11	1.05
10	5.0	8.82	7.63	6.42	60	5.0	5.70	5.41	5.20
	10.0	14.71	13.05	11.40		10.0	10.96	10.73	10.35
	1.0	1.87	1.54	1.21		1.0	1.22	1.12	1.08
20	5.0	7.03	6.19	5.55	70	5.0	5.66	5.37	5.19
	10.0	12.45	11.96	11.03		10.0	11.15	10.98	10.69
	1.0	1.72	1.51	1.25		1.0	1.17	1.09	1.01
30	5.0	6.50	5.99	5.43	80	5.0	5.54	5.41	5.22
	10.0	11.93	11.33	10.50		10.0	11.02	10.73	10.46
	1.0	1.45	1.29	1.21		1.0	1.29	1.19	1.15
40	5.0	6.38	5.98	5.65	90	5.0	5.58	5.46	5.35
	10.0	11.51	11.13	10.63		10.0	10.78	10.54	10.30
	1.0	1.42	1.19	1.07		1.0	1.24	1.11	1.09
50	5.0	5.77	5.56	5.31	100	5.0	5.42	5.39	5.22
	10.0	11.04	10.99	10.38		10.0	10.16	10.15	9.97

Acknowledgements We gratefully acknowledge the partial financial support from the following Brazilian agencies: CNPq and FAPESP.

REFERENCES

Cordeiro, G.M. (2004). Second-order covariance matrix of maximum likelihood estimates in generalized linear models, *Statistics and Probability Letters*, **66**, 153–160.

Cordeiro, G.M. and McCullagh, P. (1991). Bias correction in generalized linear models, *Journal of the Royal Statistics Society, B*, **53**, 629–643.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer, New York.

Ferrari, S.L.P., Botter, D.A. and Cribari-Neto, F. (1996). Second- and third-order bias reduction for one-parameter family models, *Statistics and Probability Letters*, **30**, 339–345.

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

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society, A*, **135**, 370–384.

Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference*, World Scientific, Singapore.

Peers, H.W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications, *Journal of the Royal Statistics Society, B*, **47**, 547–554.