Matriz de covariância do estimador de máxima verossimilhança corrigido pelo viés em modelos não-lineares da família exponencial

Tiago M. Magalhães, Denise A. Botter e Mônica C. Sandoval
Departamento de Estatística, Universidade de São Paulo

1. Introdução

Os modelos não-lineares da família exponencial (MNLFE) (Cordeiro e Paula, 1989) são uma extensão dos modelos linerares generalizados (MLG) (Nelder e Wedderburn, 1972), em que a componente sistemática não se restringe a uma função linear dos parâmetros da regressão.

Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes com função densidade de probabilidade (ou função de probabilidade) dada por

$$\pi\left(y_i;\theta_i,\phi\right) \quad = \quad \exp\left\{\phi\left[y_i\theta_i - b\left(\theta_i\right)\right] + a\left(y_i,\phi\right)\right\}, \tag{1}$$

em que $b\left(\cdot\right)$ e $a\left(\cdot,\cdot\right)$ são funções conhecidas. A média e a variância de Y_i são, respectivamente,

$$\mathbb{E}(Y_i) = \mu_i = db\left(\theta_i\right)/d\theta_i \text{ e } \mathrm{Var}\left(Y_i\right) = \phi^{-1}V_i,$$

em que $V = d\mu/d\theta$ é chamada de função de variância e $\theta = \int V^{-1}d\mu = q\left(\mu\right)$ é uma função um a um de μ, conhecida, que varia em um subconjunto do $\mathbb{R}$. Os parâmetros θ e $\phi > 0$ em (1) são chamados de parâmetro natural e de precisão, respectivamente. O inverso de ϕ é o parâmetro de dispersão $\sigma^2 = \phi^{-1}$. Para as distribuições da família exponencial bi-paramétrica com parâmetros naturais ϕ e $\phi\theta$, a função $a\left(y,\phi\right)$ em (1) pode ser escrita como $a\left(y,\phi\right) = \phi c\left(y\right) + d_1\left(\phi\right) + d_2\left(y\right)$.

Os modelos não-lineares de família exponencial têm um componente sistemático que é denotado por

$$g\left(\mu_i\right) = \eta_i = f\left(x_i;\beta\right), \tag{2}$$

em que $g(\cdot)$ é uma função de ligação biunívoca, diferenciável e conhecida, $f(x_i;\beta)$ é uma função conhecida, contínua e diferenciável, $\beta = (\beta_1, \ldots, \beta_p)^\top$, $p < n$, é um vetor de parâmetros desconhecidos a ser estimado e x_i é um vetor $q \times 1$ de variáveis explanatórias conhecidas associadas com a i-ésima resposta observada. Assumimos que diferentes valores de β implicam em diferentes valores de η. Isso garante que a matriz de derivadas $\widetilde{X} = \widetilde{X}(\beta) = \partial\eta/\partial\beta$ tenha posto p para todo β.

Cavalcanti (2009) obteve para o MLG, supondo ϕ conhecido, a matriz de covariância do EMV de β, corrigido pelo viés de ordem n^{-1}, utilizando os resultados de Pace e Salvan (1997), Cordeiro e McCullagh (1991) e Cordeiro (2004). Neste trabalho vamos fazer o mesmo para o MNLFE, utilizando os resultados de Paula (1992) e Rocha et al. (2010).

2. Matriz de covariância assintótica de ordem n^{-2}

Vamos denotar o logaritmo da função de verossimilhança para β por $\ell = \ell(\beta)$ e os cumulantes conjuntos das derivadas do logaritmo da função de verossimilhança por $\kappa_{rs} = \mathbb{E}(\partial^2\ell/\partial\beta_r\partial\beta_s)$, $\kappa_{r,s} = \mathbb{E}(\partial\ell/\partial\beta_r\partial\ell/\partial\beta_s)$, $\kappa_{rst} = \mathbb{E}(\partial^3\ell/\partial\beta_r\partial\beta_s\partial\beta_t)$, $\kappa_{r,st} = \mathbb{E}(\partial\ell/\partial\beta_r\partial^2\ell/\partial\beta_s\partial\beta_t)$, etc. Todos os κ's referem-se a um total sobre a amostra e são, em geral, de ordem $O(n)$. A matriz de informação total de Fisher tem elementos $\kappa_{r,s} = -\kappa_{rs}$; seja $\kappa^{r,s} = -\kappa^{rs}$ os correspondentes elementos de sua inversa. Assumimos que quando n cresce, o EMV $\hat{\beta}$ converge para seu verdadeiro valor β e que sua distribuição assintótica é normal multivariada com média β com matriz de covariância $\phi^{-1}\left(\widetilde{X}^\top W \widetilde{X}\right)^{-1}$, em que $W = \mathrm{diag}\left\{(d\mu/d\eta)^2 V^{-1}\right\}$.

Sejam $\widetilde{\beta} = \hat{\beta} - d(\hat{\beta})$, em que $d(\hat{\beta})$ é o viés de ordem n^{-1} de $\hat{\beta}$, EMV de β e $\widetilde{\beta}_r$, o r-ésimo elemento de $\widetilde{\beta}$. Temos então $\widetilde{\beta}_r = \hat{\beta}_r - d^r(\hat{\beta})$.

De Pace e Salvan (1997) vem que

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

em que

$$
\begin{aligned}
d_v^r &= \frac{\partial d^r}{\partial \beta_v} \\
&= \sum \kappa^{rw}\kappa^{sy}\kappa^{tu}(\kappa_{stu} + 2\kappa_{st,u})(\kappa_{vwy} + \kappa_{v,wy}) \\
&+ \frac{1}{2}\kappa^{rs}\kappa^{tu}(\kappa_{stuv} + \kappa_{stu,v} + 2\kappa_{stv,u} + 2(\kappa_{st,uv} + \nu_{st,u,v})) \\
&= O(n^{-1})
\end{aligned}
$$

e

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

Queremos encontrar uma expressão de ordem n^{-2} para

$$
\mathbb{E}\left[\left(\widetilde{\beta}_r - \beta_r)(\widetilde{\beta}_s - \beta_s\right)\right].
$$

Temos que

$$
\begin{aligned}
\mathbb{E}[(\widetilde{\beta}_r - \beta_r)(\widetilde{\beta}_s - \beta_s)] &= \mathbb{E}[(\hat{\beta}_r - d^r(\hat{\beta}) - \beta_r)(\hat{\beta}_s - d^s(\hat{\beta}) - \beta_s)] \\
&= \mathbb{E}\{[(\hat{\beta}_r - \beta_r) - d^r(\hat{\beta})][(\hat{\beta}_s - \beta_s) - d^s(\hat{\beta})]\} \\
&= \mathbb{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] - 2\mathbb{E}[d^s(\hat{\beta})(\hat{\beta}_r - \beta_r)] \\
&+ \mathbb{E}[d^r(\hat{\beta})d^s(\hat{\beta})] \\
&= \mathbb{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] - 2[\sum d^s(\beta)d^r(\beta) + \sum d_v^s(-\kappa^{rv})] \\
&+ \sum d^s(\beta)d^r(\beta) \\
&= \mathbb{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] - \sum d^s(\beta)d^r(\beta) \\
&- 2\sum d_v^s(-\kappa^{rv}). \qquad (3)
\end{aligned}
$$

A primeira parcela da soma em (3) é a covariância de $\hat{\beta}$ com termos até ordem n^{-2} obtida por Rocha et al. (2010) que, em notação matricial, é dada por

$$
\mathbb{E}[(\hat{\beta}_r - \beta_r)(\hat{\beta}_s - \beta_s)] = \phi^{-1}(\widetilde{X}^\top W \widetilde{X})^{-1} + \Sigma_L + \Sigma_{NL},
$$

em que

$$\begin{aligned}\Sigma_L &= \frac{1}{\phi^2} P\left\{H + \left(E\left[2E+6G-F\right] + 2G\left[2G-F\right]\right) Z\right\} Z_d P^\top \\ &+ \frac{1}{\phi^2} P\left\{F\left[-\frac{1}{2}F - 4G - E\right] + 2G\left[3G+5E\right] + 3EE\right\} Z^{(2)} P^\top\end{aligned}$$

e

$$\Sigma_{NL} = \frac{1}{\phi^2} P\left\{M + 7\left[2G+2E-F\right] W Z_d + \frac{7}{2} DWW\right\} DP^\top - \frac{1}{\phi^2} PW\Delta K^{-1},$$

com

$P = (X^\top W X)^{-1} X^\top$, $Z = XP$, $H = \mathrm{diag}\{-\frac{\mu'\mu'''}{V} - \frac{\mu'^2\mu'' V^{(1)}}{V^2} + \frac{\mu'^4 V^{(1)2}}{V^3}\}$, $Z_d = \mathrm{diag}\{Z\}$,

$E\left[2E+6G-F\right] + 2G\left[2G-F\right] = \mathrm{diag}\left\{\frac{2\mu'^2\mu''^2}{V^2} - \frac{V^{(1)}\mu'^4\mu''}{V^3}\right\}$,

$F\left[-\frac{1}{2}F - 4G - E\right] + 2G\left[3G+5E\right] + 3EE = \mathrm{diag}\left\{\frac{3\mu'^2\mu''^2}{2V^2} + \frac{V^{(1)}\mu'^4\mu''}{V^3} - \frac{V^{(1)2}\mu'^6}{V^4}\right\}$,

$M = \mathrm{diag}\left\{-\frac{V^{(1)}\mu'^3}{V^2} - \frac{3\mu'\mu''}{V}\right\}$, $7\left[2G+2E-F\right] = \mathrm{diag}\left\{\frac{7\mu'^3\mu''}{V^2}\right\}$, $\mu' = d\mu/d\eta$, $\mu'' = d^2\mu/d\eta^2$ e $\mu''' = d^3\mu/d\eta^3$.

A segunda e a terceira parcela de (3) são calculadas a partir da expressão do viés de ordem n^{-1} de $\hat{\beta}$ obtida por Paula (1992) que, em notação matricial, é dada por

$$d(\hat{\beta}) = \left(\widetilde{X}^\top W \widetilde{X}\right)^{-1} \widetilde{X}^\top W \left(\xi_1 + \xi_2\right),$$

em que $\xi_1 = -\left(2\phi\right)^{-1} Z_d W^{-1} G\mathbf{1}$, $\xi_2 = -\left(2\phi\right)^{-1} D\mathbf{1}$.

Neste trabalho obtemos a expressão (3) em notação matricial e realizamos estudos de simulação para verificar a viabilidade prática dos resultados obtidos.

3. Referências

Cavalcanti, A.B. (2009). *Aperfeiçoamento de métodos estatísticos em modelos de regressão da família exponencial*. Tese de doutorado. São Paulo: IME–USP.

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

Cordeiro, G.M., Paula, G.A. (1989).

Nelder, J.A., Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistics Society, A*, **135**, 370–384.

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

Paula, G.A. (1992). Bias correction for exponential family nonlinear models. *Journal of Statistic Computation and Simulation*, **40**, 43-54.

Rocha, A.V., Simas, A.B., Cordeiro, G.M. (2010). Second-order asymptotic expressions for the covariance matrix of maximum likelihood estimators in dispersion models. *Statistics and Probability Letters*, **80**, 718-725.