

CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

23 de agosto, 2017

Aula 6 - Estimação em modelos lineares generalizados

Estimação em modelos lineares generalizados

- Seja y_i uma única observação de uma distribuição na forma:

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi) \right\}. \quad (1)$$

- A log-verossimilhança correspondente a essa observação fica dada por:

$$l_i = l(\theta_i; \phi, y_i) = \log [f(y_i; \theta_i, \phi)] = \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi). \quad (2)$$

- Considerando n observações independentes y_1, y_2, \dots, y_n , a log-verossimilhança fica dada por:

$$l(\boldsymbol{\theta}; \phi, \mathbf{y}) = \sum_{i=1}^n l_i = \sum_{i=1}^n l(\theta_i; \phi, y_i) = \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi) \right]. \quad (3)$$

Estimação em modelos lineares generalizados

- Considere um modelo linear generalizado com função de ligação $g(\cdot)$ e preditor linear $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. \quad (4)$$

- A estimação de $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ por máxima verossimilhança baseia-se na determinação de $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)'$ tal que:

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p} &= 0 \\ \frac{\partial l(\beta)}{\partial \beta_1} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p} &= 0 \\ &\vdots \\ \frac{\partial l(\beta)}{\partial \beta_p} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p} &= 0 \end{aligned} \quad (5)$$

Estimação em modelos lineares generalizados

- Como y_1, y_2, \dots, y_n são independentes:

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta_j} = 0, \quad (6)$$

para todo j .

- Observe que estamos denotando a log-verossimilhança por $l(\beta)$ uma vez que $\mu_i = b'(\theta_i)$; $g(\mu_i) = \eta_i$ e $\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$.
- Assim, pela regra da cadeia:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j}, \quad (7)$$

para $j = 0, 1, 2, \dots, p$.

Estimação em modelos lineares generalizados

- Usando a definição e as propriedades de modelos lineares generalizados:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{(y_i - \mu_i)}{\text{var}(y_i)} \times \frac{\partial \mu_i}{\partial \eta_i} \times x_{ij}. \quad (8)$$

- Somando para as n observações, as equações de log-verossimilhança ficam dadas por:

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)}{\text{var}(y_i)} \times \frac{\partial \mu_i}{\partial \eta_i} \times x_{ij}, \quad j = 0, 1, 2, \dots, p, \quad (9)$$

onde $\eta_i = \sum_{j=0}^p \beta_j x_{ij} = g(\mu_i)$.

Estimação em modelos lineares generalizados

- Uma forma equivalente de escrever as equações de log-verossimilhança é a seguinte:

$$\mathbf{X}'\mathbf{D}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}, \quad (10)$$

em que \mathbf{V} é a matriz diagonal das variâncias das observações; \mathbf{X} é a matriz do modelo; \mathbf{D} é a matriz diagonal com entradas $\partial\mu_i/\partial\eta_i$ e \mathbf{y} e $\boldsymbol{\mu}$ são os vetores de observações e de médias, respectivamente.

- As equações de log-verossimilhança são funções não lineares dos β 's.
- Assim, a determinação das estimativas de máxima verossimilhança requer o uso de métodos iterativos. Vamos discutir mais adiante dois desses métodos: o **método de Newton-Raphson** e o **método Score de Fisher**.

Distribuição assintótica dos estimadores dos parâmetros de um MLG

- Os estimadores de máxima verossimilhança dos parâmetros de um MLG atendem às propriedades gerais de estimadores de máxima verossimilhança.
- Assim, assintoticamente:

$$\hat{\beta} \sim N(\beta, \mathcal{J}^{-1}), \quad (11)$$

em que \mathcal{J} é a matriz informação de Fisher (ou matriz informação esperada), com entradas $-E(\partial^2 l(\beta)/\partial\beta_r\partial\beta_s)$.

Distribuição assintótica dos estimadores dos parâmetros de um MLG

- Usando o fato que, sob condições de regularidade atendidas pela família exponencial de distribuições:

$$E \left(-\frac{\partial^2 l(\beta)}{\partial \beta_r \partial \beta_s} \right) = E \left[\left(\frac{\partial l(\beta)}{\partial \beta_r} \right) \left(\frac{\partial l(\beta)}{\partial \beta_s} \right) \right] \quad (12)$$

chega-se a:

$$E \left(-\frac{\partial^2 l(\beta)}{\partial \beta_r \partial \beta_s} \right) = \sum_{i=1}^n \frac{x_{ir} x_{is}}{\text{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \quad (13)$$

Distribuição assintótica dos estimadores dos parâmetros de um MLG

- Seja \mathbf{W} a matriz diagonal com elementos:

$$\omega_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{\text{var}(y_i)}. \quad (14)$$

- Então, generalizando para toda a matriz de informação, temos:

$$\mathbf{j} = \mathbf{X}'\mathbf{W}\mathbf{X}, \quad (15)$$

em que \mathbf{X} é a matriz do modelo. A matriz \mathbf{j} depende da função de ligação, uma vez que $\partial\eta_i/\partial\mu_i = g'(\mu_i)$.

Distribuição assintótica dos estimadores dos parâmetros de um MLG

- Assim, a distribuição assintótica de $\hat{\beta}$ é dada por:

$$\hat{\beta} \sim N(\beta, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}). \quad (16)$$

- A matriz de covariância assintótica é estimada por

$$\widehat{\text{var}}(\hat{\beta}) = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}, \quad (17)$$

sendo que $\hat{\mathbf{W}}$ é \mathbf{W} avaliado em $\hat{\beta}$.

Método de Newton-Raphson

- O método de Newton-Raphson é aplicado na solução de equações não lineares (no caso, na determinação do ponto em que a função assume seu máximo);
- O método inicia com um valor inicial como primeira aproximação para a solução;
- Na sequência, uma segunda aproximação é obtida aproximando a função, na vizinhança do valor inicial, por um polinômio de segundo grau, e encontrando o ponto de máximo do polinômio.
- Após a repetição de uma sequência de aproximações, o processo converge para a locação do máximo se a função é bem comportada e a aproximação inicial é boa.

Método de Newton-Raphson

- Formalizando: desejamos determinar $\hat{\beta}$ que maximiza $L(\beta)$. Seja:

$$\mathbf{s} = \left(\frac{\partial l(\beta)}{\partial \beta_0}, \frac{\partial l(\beta)}{\partial \beta_1}, \dots, \frac{\partial l(\beta)}{\partial \beta_p} \right)'. \quad (18)$$

- Seja \mathbf{H} a matriz hessiana, definida pelas derivadas parciais de segunda ordem:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_0} & \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 l(\beta)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 l(\beta)}{\partial \beta_p \partial \beta_p} \end{bmatrix}$$

Método de Newton-Raphson

- Sejam $\mathbf{S}^{(t)}$ e $\mathbf{H}^{(t)}$, respectivamente, \mathbf{S} e \mathbf{H} avaliados em $\beta^{(t)}$, a aproximação no passo t para $\hat{\beta}$.
- O método de Newton-Raphson aproxima $l(\beta)$ em torno de $\beta^{(t)}$ por meio de uma expansão em série de Taylor de segunda ordem:

$$l(\beta) \approx l(\beta^{(t)}) + \mathbf{S}^{(t)'}(\beta - \beta^{(t)}) + \left(\frac{1}{2}\right) (\beta - \beta^{(t)})' \mathbf{H}^{(t)} (\beta - \beta^{(t)}) \quad (19)$$

Método de Newton-Raphson

- Resolvendo $\partial L(\beta)/\partial \beta \approx \mathbf{S}^{(t)} + \mathbf{H}^{(t)}(\beta - \beta^{(t)}) = \mathbf{0}$ para β , temos como aproximação para $\hat{\beta}$ no passo $t + 1$:

$$\beta^{(t+1)} = \beta^{(t)} - (\mathbf{H}^{(t)})^{-1} \mathbf{S}^{(t)}, \quad (20)$$

assumindo que $\mathbf{H}^{(t)}$ é não singular.

- As iterações continuam até que a mudança em $\beta^{(t)}$ em passos sucessivos seja suficientemente pequena (convergência).

Método Score de Fisher

- A diferença do método Score de Fisher para o método de Newton Raphson é que o primeiro usa o valor esperado da matriz Hessiana, que é a matriz de *informação esperada*, enquanto o segundo usa a própria hessiana, que é a matriz de *informação observada*.
- Seja $\mathbf{j}^{(t)}$ a aproximação no passo t para a matriz de informação esperada, com entradas $-E(\partial^2 l(\boldsymbol{\beta})/\partial\beta_r\partial\beta_s)$ avaliado em $\boldsymbol{\beta}^{(t)}$. O algoritmo score de Fisher fica dado por:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{j}^{(t)})^{-1} \mathbf{S}^{(t)}. \quad (21)$$

Máxima verossimilhança e método de mínimos quadrados ponderados iterativamente

- Existe uma relação entre o algoritmo Score de Fisher, aplicado na estimação de máxima verossimilhança dos parâmetros de um MLG, e o método de mínimos quadrados ponderados.
- Após algumas passagens, as equações do algoritmo Score de Fisher podem ser expressas na seguinte forma:

$$(\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X})\boldsymbol{\beta}^{(t+1)} = \mathbf{X}'\mathbf{W}^{(t)}\mathbf{z}^{(t)}, \quad (22)$$

em que $\mathbf{z}^{(t)}$ é um vetor com elementos:

$$z_i^{(t)} = \sum_j x_{ij}\beta_j^{(t)} + (y_i - \mu_i^{(t)}) \frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}} = \eta_i^{(t)} + (y_i - \mu_i^{(t)}) \frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}}. \quad (23)$$

Máxima verossimilhança e método de mínimos quadrados ponderados iterativamente

- Isolando $\beta^{(t+1)}$:

$$\beta^{(t+1)} = (\mathbf{X}' \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{(t)} \mathbf{z}^{(t)} \quad (24)$$

- O vetor $\mathbf{z}^{(t)}$ é uma forma linearizada da função de ligação g avaliada em \mathbf{y} :

$$g(y_i) \approx g(\mu_i^{(t)}) + (y_i - \mu_i^{(t)})g'(\mu_i^{(t)}) = \eta_i^{(t)} + (y_i - \mu_i^{(t)}) \frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}}. \quad (25)$$

Máxima verossimilhança e método de mínimos quadrados ponderados iterativamente

- A *variável resposta ajustada* \mathbf{z} tem o i – *ésimo* elemento aproximado por $z_i^{(t)}$ no passo t do algoritmo.
- Neste passo, fazemos a regressão de $\mathbf{z}^{(t)}$ com relação a \mathbf{X} com pesos $\mathbf{W}^{(t)}$ para obter uma nova aproximação $\beta^{(t+1)}$.
- Com $\beta^{(t+1)}$ calcula-se $\eta^{(t+1)} = \mathbf{X}\beta^{(t+1)}$ e uma nova aproximação $\mathbf{z}^{(t+1)}$.
- O processo iterativo continua até se observar convergência para β .

Exercício 1

*Retome o Exemplo 4 da primeira aula (referente ao acasalamento de elefantes). Os dados estão disponíveis no pacote **Sleuth2** (sob o nome **case2201**). Vamos considerar um modelo de regressão Poisson com função de ligação logarítmica (canônica). Ajuste o MLG programando o algoritmo de estimação, maximizando a log-verossimilhança usando um otimizador do R e via função `glm`, conforme visto em aula.*