ME720 - Modelos Lineares Generalizados

Parte 9 - Inferência em MLGs

Profa. **Larissa Avila Matos**

Estimação dos parâmetros

Uma vez definido cada componente do meodelo, obteremos expressões gerais para a função de verossimilhança e para as distribuições assintóticas dos estimadores de máxima verossimilhança dos parâmetros para os MLGs.

Para n observações independentes, temos que a função de verossimilhança do modelo é dada por $\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\beta}, \phi)$, onde

$$\ell_i(\boldsymbol{\beta}, \phi) = \frac{[\mathrm{y}_i \theta_i - b(\theta_i)]}{a(\phi)} + c(\mathrm{y}_i, \phi).$$

Então,

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \frac{[\mathrm{y}_i \theta_i - b(\theta_i)]}{a(\phi)} + \sum_{i=1}^{n} c(\mathrm{y}_i, \phi).$$

Para um GLM $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} = g(\mu_i)$ com função de ligação g, o sistema de equações de verossimilhança para $\boldsymbol{\beta}$ são

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_j =} \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} = 0, \quad \forall j.$$

Para diferenciar a log-verossimilhança, usamos a regra da cadeia,

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Uma vez que,

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[\mathrm{y}_i - b'(\theta_i)]}{a(\phi)} \quad \text{e} \quad \mu_i = b'(\theta_i), \quad \mathrm{Var}(Y_i) = b''(\theta_i) a(\phi),$$

temos que

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{(\mathrm{y}_i - \mu_i)}{a(\phi)} \quad \text{e} \quad \frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) = \frac{\mathrm{Var}(Y_i)}{a(\phi)}.$$

Também uma vez que $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}$, então $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$.

Finalmente, $\eta_i = g(\mu_i) \Rightarrow \frac{\partial \mu_i}{\partial \eta_i}$, depende da função de ligação para o modelo escolhido.

Resumindo, temos que

$$\begin{aligned} \frac{\partial \ell_i}{\partial \beta_j} &= \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \frac{(\mathrm{y}_i - \mu_i)}{a(\phi)} \frac{a(\phi)}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \frac{(\mathrm{y}_i - \mu_i)\, x_{ij}}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \end{aligned}$$

A soma das n observações produz o sistema de equações de verossimilhança para um MLG.

Equações de verossimilhança para um MLG

Equações de verossimilhança para um MLG:

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{(\mathrm{y}_i - \mu_i)\, x_{ij}}{\mathrm{Var}(\mathrm{Y}_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, 2, \ldots, p, \quad \text{e} \\
\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} &= 0,
\end{aligned}
$$

onde $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} = g(\mu_i)$ para uma função de ligação g.

Seja $\boldsymbol{V}$ a matriz diagonal de variâncias das n observações, e seja $\boldsymbol{D}$ uma matriz diagonal com os elementos de $\frac{\partial \mu_i}{\partial \eta_i}$.

Para as expressões GLM $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ com matriz de planejamento $\boldsymbol{X}$, as equações de verossimilhança têm a forma

$$\boldsymbol{X}\boldsymbol{D}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = 0.$$

Apesar de $\boldsymbol{\beta}$ não aparecer nessas equações, ele aparece implicitamente através de $\mu_i = g^{-1}(\sum_{j=1}^{p} \beta_j x_{ij})$.

Diferentes funções de ligação gera diferentes conjuntos de equações.

Essas equações são funções não lineares de $\boldsymbol{\beta}$, e esse problema deve ser resolvido iterativamente. Há várias opções de algoritmos numéricos para resolução de sistemas não lineares: Newton-Raphson, Escore de Fisher (EF), Nelder-Mead, BFGS entre outros.

Exercício

- Encontrar as equações de máxima verossimilhança para o Modelo Poisson Log Linear.

Relação entre média e variância

Os sistema de equações de máxima verossimilhança depende da distribuição de Y_i somente pela média ($\mathbb{E}(Y_i)$) e pela variância ($\mathrm{Var}(Y_i)$).

Além disso, a variância depende da média pela forma

$$\mathrm{Var}(Y_i) = V(\mu_i),$$

para alguma função $V(\cdot)$.

Ou seja, a relação entre a média e a variância caracteriza a distribuição de Y_i.

Exemplo: Se Y_i tem distribuição pertencente a família exponencial e $\mathrm{Var}(Y_i) = \mu_i$, então necessariamente Y_i tem distribuição de Poisson.

Distribuição Assintótica de $\widehat{\boldsymbol{\beta}}$

Através das propriedades de máxima verossimilhança, e sob condições de regularidade, para n grande o estimador de máxima verossimilhança $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ para um MLG é eficiente e tem distribuição Normal.

Temos que a matrix de informação $\boldsymbol{I}$, tem elementos dados por

$$\begin{aligned}
\mathbb{E}\left(-\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_k}\right) &= \mathbb{E}\left(\frac{\partial \ell_i}{\partial \beta_j}\frac{\partial \ell_i}{\partial \beta_k}\right) \\
&= \mathbb{E}\left(\frac{(Y_i-\mu_i)\,x_{ij}}{\mathrm{Var}(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}\frac{(Y_i-\mu_i)\,x_{ik}}{\mathrm{Var}(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}\right) \\
&= \frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \underbrace{\mathbb{E}\left((\mathrm{y}_i-\mu_i)^2\right)}_{=\mathrm{Var}(Y_i)} = \frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.
\end{aligned}$$

A matriz $\boldsymbol{I}$ é chamada de matriz de informação esperada.

Então,

$$\mathbb{E}\left(-\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right) = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

Seja, $\boldsymbol{W}$ uma matriz diagonal com elementos dados por

$$w_i = \frac{\left(\partial \mu_i / \partial \eta_i\right)^2}{\mathrm{Var}(Y_i)}$$

$$\Rightarrow \boldsymbol{I} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}$$

A forma de $\boldsymbol{W}$ depende da função de ligação $g(\cdot)$, uma vez que $\partial \mu_i / \partial \eta_i = g'(\mu_i)$.

Portanto, a matriz de covariância de $\widehat{\boldsymbol{\beta}}$ é dada pela inversa da matriz de informação $\boldsymbol{I}$.

Distribuição Assintótica de $\widehat{\beta}$ para um MLG $\boldsymbol{\eta} = \boldsymbol{X}\beta$:

$\widehat{\boldsymbol{\beta}}$ tem distribuição aproximadamente normal $N(\boldsymbol{\beta}, (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1})$,

onde $\boldsymbol{W}$ é a matriz diagonal com elementos $w_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{\mathrm{Var}(Y_i)}$.

A matriz de covariância assintótica é estimada por $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}$, onde $\widehat{\boldsymbol{W}}$ é $\boldsymbol{W}$ avaliado em $\hat{\boldsymbol{\beta}}$.

Obs:

1. Para a FE com parâmetro de escala, θ e ϕ são parâmetros ortogonais.
2. $\widehat{\boldsymbol{\beta}}$ e $\widehat{\phi}$ são assintoticamente independentes.

Matriz de covariância para os valores ajustados

O preditor linear estimado é dado por

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

Para n grande, temos

$$\text{Var}(\widehat{\boldsymbol{\eta}}) = \boldsymbol{X}'\text{Var}(\widehat{\boldsymbol{\beta}})\boldsymbol{X} \approx \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}.$$

Podemos obter a variância assintótica de $\widehat{\boldsymbol{\mu}}$ ($\text{Var}(\widehat{\boldsymbol{\mu}})$) por $\text{Var}(\widehat{\boldsymbol{\eta}})$, através do método delta. Então,

$$\text{Var}(\widehat{\boldsymbol{\mu}}) \approx \boldsymbol{D}'\text{Var}(\widehat{\boldsymbol{\eta}})\boldsymbol{D} \approx \boldsymbol{D}\boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{D},$$

onde $\boldsymbol{D}$ é uma matriz diadonal com elementos $\partial\mu_i/\partial\eta_i$.

Exercício: Pesquisar sobre o método delta.

Exercício

- Voltando ao exemplo da Poisson, encontre os elementos da matriz $\boldsymbol{W}$.

Estimação de $\boldsymbol{\beta}$

Como encontramos os estimadores de máxima verossimilhança dos parâmetros de um MLG?

Problema: O sistema de equações são geralmente não-lineares em $\boldsymbol{\beta}$.

Solução: Métodos iterativos para resolver sistema de equações não lineares. Focaremos em dois métodos:

- *Newton-Raphson*
- *Escore de Fisher*

Método de Newton-Raphson

O algoritmo Newton-Raphson, método de Newton-Raphson, foi desenvolvido por Isaac Newton e Joseph Raphson e tem o objetivo estimar as raízes de uma função.

Suponha que queremos encontrar a solução da equação $g(x_0) = 0$, onde g é uma função diferenciável. Dado um numero x próximo de $x0$, segue da expansão em série de Taylor em torno de x que

$$0 = g(x_0) \approx g(x) + g'(x)(x_0 - x).$$

Resolvendo para x_0, temos

$$x_0 \approx x - \frac{g(x)}{g'(x)}.$$

Assim, dado um valor estimado x_t, então podemos ter um novo valor estimado x_{t+1} por

$$x_{t+1} \approx x_t - \frac{g(x_t)}{g'(x_t)}.$$

Este procedimento é repetido para $t = 1, 2, 3, \ldots$ até $|g(x_t)/g'(x_t)|$ ser suficientemente pequeno.

Método de Newton-Raphson

Voltando ao nosso problema, queremos encontrar a solução da equação $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$.

No passo t, o processo iterativo $(t = 0, 1, 2, \ldots)$ aproxima $\ell(\boldsymbol{\beta})$ próximo de $\boldsymbol{\beta}^{(t)}$ pela expansão em série de Taylor de segunda ordem,

$$\ell(\boldsymbol{\beta}) \approx \ell(\boldsymbol{\beta}^{(t)}) + \boldsymbol{u}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) + \left(\frac{1}{2}\right)(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})'\boldsymbol{H}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}),$$

onde $\boldsymbol{u}^{(t)}$ e $\boldsymbol{H}^{(t)}$ são $\boldsymbol{u}$ e $\boldsymbol{H}$ avaliados em $\boldsymbol{\beta}^{(t)}$ respectivamente, com

- $\boldsymbol{u} = \left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1}, \frac{\partial \ell(\beta_2)}{\partial \beta_p}, \ldots, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)'$, e
- $\boldsymbol{H}$ a matrix *Hessiana*, onde $H_{ij} = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}$.

Obs: $\boldsymbol{\beta}^{(o)}$ valor inicial (chute inicial).

Resolvendo, $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \boldsymbol{u}^{(t)} + \boldsymbol{H}^{(t)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) = \mathbf{0}$, temos a seguinte aproximação

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\boldsymbol{H}^{(t)})^{-1}\boldsymbol{u}^{(t)},$$

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

O procedimento descrito é repetido até que mudanças em $\ell(\boldsymbol{\beta}^{(t)})$ entre ciclos sucessivos são suficientemente pequenas.

Para muitos MLGs, a matriz Hessian é negativa definida, e a log verossimilhança é uma função estritamente côncava. Então, as estimativas de máxima verossimilhança dos parâmetros do modelo existem e são únicas sob condições bastante gerais. A convergência de $\boldsymbol{\beta}^{(t+1)}$ para $\widehat{\boldsymbol{\beta}}$ na vizinhança de $\widehat{\boldsymbol{\beta}}$ é rápida.

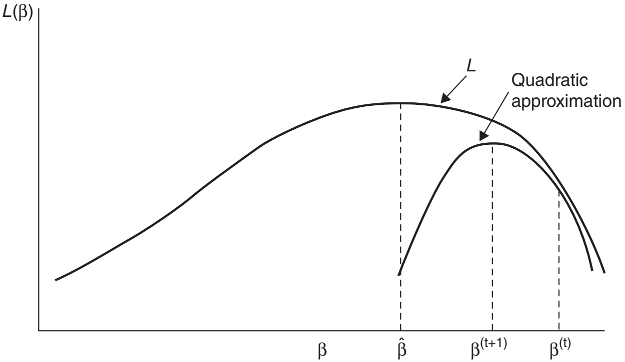

Figure 4.2 Illustration of a cycle of the Newton–Raphson method.

Figura ilustrativa do livro texto (Agresti, A. (2015)).

Método de Escore de Fisher

O método de Escore de Fisher é um método iterativo alternativo para resolver o sistema de equações de verossimilhança.

A diferença entre o método de Escore de Fisher e o método de Newton-Raphson está na maneira como escolhe a matriz *Hessiana*.

O método de Escore de Fisher usa o valor esperado da matriz *Hessiana*, chamada de matriz de informação esperada, enquanto método de Newton-Raphson usa a própria matriz hessiana, chamada de matriz de informação observada.

Portanto, temos a seguinte aproximação

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\boldsymbol{I}^{(t)})^{-1}\boldsymbol{u}^{(t)},$$

onde $\boldsymbol{I}^{(t)}$ é $\boldsymbol{I}$ avaliado em $\boldsymbol{\beta}^{(t)}$, ou seja $\boldsymbol{I}^{(t)}$ tem elementos $-\mathbb{E}(\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j})$ avaliado em $\boldsymbol{\beta}^{(t)}$.

Os algoritmos de Escore de Fisher e de Newton-Raphson são idênticos para os MLGs que usam a função de ligação canônica (Nelder e Wedderburn, 1972).

Estimação do parâmetro de escala

1. Maximizar $\ell(\boldsymbol{\beta}, \phi)$ com respeito a ϕ. (Muito sensível a suposição da distribuição)

2. Sabemos que $\text{Var}(Y_i) = a(\phi)V(\mu_i)$, então

$$\frac{\mathbb{E}((Y_i - \mu_i)^2)}{V(\mu_i)} = a(\phi)$$

$$\Rightarrow \widehat{a}(\phi) = \frac{1}{n-p}\sum_{i=1}^{n} \frac{\mathbb{E}((Y_i - \widehat{\mu}_i)^2)}{\text{Var}(\widehat{\mu}_i)}.$$

Referência

- Notas de aula do Prof. Gilberto de Paula.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley series in probability and statistics.