Modelos lineares generalizados

Disciplina ofertada pelo DECAT/UFS

Código: ESTAT0092
Nível: Graduação
Carga horária: 60h
Período: 2020.2

Professor responsável e ministrante: Luiz Henrique Dore

Aulas 5 e 6: estimação pelo método da máxima verossimilhança

Sumário

Metas

1. Introduzir o estimador de máxima verossimilhança para o modelo linear generalizado.

Objetivos

- Após estudar essa aula, o aluno ou aluna será capaz de:
  1. calcular a log-verossimilhança do modelo linear generalizado;
  2. calcular a função escore do modelo linear generalizado;
  3. calcular a informação de Fisher observada e esperada do modelo linear generalizado;
  4. obter as equações de estimação correspondentes ao estimador de máxima verossimilhança;
  5. encontrar a estimativa de máxima verossimilhança, resolvendo as equações de estimação pelo método de Newton-Raphson ou pelo método escore de Fisher;
  6. calcular a estimativa de máxima verossimilhança dos parâmetros de um modelo linear generalizado, usando a função `glm` da plataforma computational **R**.

Pré-requisitos

1. Aula 4.

Introdução

- Na aula 2, viu-se que o processo de modelagem de regressão é constituído por quatro etapas básicas: 1) especificação do modelo; 2) estimação dos parâmetros do modelo; 3) verificação da adequacidade do modelo; 4) aplicação do modelo.
- A aula 4 tratou da etapa 1, a formulação do MLG.
- O MLG é formulado especificando-se três componentes:
    - o componente aleatório, formado pela resposta e sua distribuição de probabilidade, a qual deve pertencer à família exponencial;
    - o componente sistemático, formado pelas variáveis preditoras e pelo preditor linear;
    - a função de ligação, a qual lineariza a relação funcional entre a média e as variáveis preditoras, além de, em geral, garantir que o modelo retorne apenas os valores que a média da resposta pode assumir, de acordo com a distribuição de probabilidade adotada.

Introdução

- As presentes aulas tratam da etapa 2, a estimação dos parâmetros do modelo.
- Os parâmetros do MLG são estimados pelo método da máxima verossimilhança.
- Na aula 5, são apresentadas as funções de log-verossimilhança e escore e as matriz es de informação de Fisher observada e esperada, elementos necessários para o uso de procedimentos de estimação e inferência por máxima verossimilhança.
- Na aula 6, o estimador de máxima verossimilhança é obtido.
- Em geral, as equações de estimação, que resultam do método da máxima verossimilhança, não possuem solução em forma fechada.

Introdução

- Desse modo, é necessário aplicar um procedimento numérico para obter as estimativas de máxima verossimilhança.
- Dois procedimentos são considerados.
- O primeiro é o método de Newton-Raphson, o qual utiliza a matriz de informação de Fisher observada;
- O segundo é o método escore de Fisher, o qual se baseia na matriz de informação de Fisher esperada.

Tópico 10: função de log-verossimilhança

- Seja Y a variável resposta e sejam $x_1, x_2, \cdots, x_p$ as variáveis preditoras.
- Supõe-se que Y e as variáveis preditoras possuem uma relação governada por um MLG.
- Isto é, Y é uma variável aleatória tal que
  1. a distribuição de Y pertence à família exponencial;
  2. $\eta(\boldsymbol{x}; \boldsymbol{\beta}) = \boldsymbol{x}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$;
  3. $g(\mu) = \eta(\boldsymbol{x}; \boldsymbol{\beta})$;

  onde μ é a média de Y, g é a função de ligação,
  $$\boldsymbol{x} = (1, x_1, \cdots, x_p)^T$$
  é o vetor de variáveis preditoras e
  $$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$$
  é o vetor dos coeficientes de regressão.

Tópico 10: função de log-verossimilhança

- Seja $\{Y_1, \cdots, Y_n\}$ uma amostra aleatória simples de Y.
- Para cada $i \in \{1, \cdots, n\}$, sejam $x_{i1}, \cdots, x_{ip}$ os valores das preditoras $x_1, \cdots, x_p$, respectivamente, correspondentes a Y_i.
- Nesse caso, supõe-se 1que
  1. a distribuição de Y_i pertence à família exponencial;
  2. $\eta(\boldsymbol{x}_i; \boldsymbol{\beta}) = \boldsymbol{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$;
  3. $g(\mu_i) = \eta(\boldsymbol{x}_i; \boldsymbol{\beta}) = \eta_i$;

  onde μ_i é a média de Y_i, g é a função de ligação,

  $$\boldsymbol{x}_i = (1, x_{i1}, \cdots, x_{ip})^T$$

  é o vetor da i-ésima observação das variáveis preditoras e

  $$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$$

  é o vetor dos coeficientes de regressão.

Tópico 10: função de log-verossimilhança

- Como a distribuição de Y_i pertence à família exponencial, sua função de densidade (ou função de probabilidade) pode ser escrita como

$$f(y_i;\theta_i,\phi) = \exp\left\{\phi^{-1}\left[y_i\theta_i - b(\theta_i)\right] + c(y_i,\phi)\right\}.$$

- O parâmetro canônico varia com as unidades amostrais, pois, como se sabe, ele se relaciona com a média por meio da equação

$$b'(\theta_i) = \mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i^T\boldsymbol{\beta}).$$

- Entretanto, o modelo assume que o parâmetro de dispersão é constante ao longo da amostra.
- Isto é, ϕ não varia de uma unidade amostral para a outra.

Tópico 10: função de log-verossimilhança

- A função de verossimilhança é dada por

$$L(\boldsymbol{\beta}, \phi) = \prod_{i=1}^{n} f(Y_i; \theta_i, \phi) =$$

$$= \exp\left\{ \phi^{-1} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} c(Y_i, \phi) \right\}. \quad (1)$$

- A função de log-verossimilhança é dada por

$$l(\boldsymbol{\beta}, \phi) = \ln\left[L(\boldsymbol{\beta}, \phi)\right] = \sum_{i=1}^{n} \ln\left[f(Y_i; \theta_i, \phi)\right] =$$

$$= \phi^{-1} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} c(Y_i, \phi). \quad (2)$$

Tópico 11: função escore

- A função escore é o gradiente da log-verossimilhança.
- Isto é, denotando por $\boldsymbol{U}(\boldsymbol{\beta}, \phi)$ a função escore, tem-se que

$$\boldsymbol{U}(\boldsymbol{\beta}, \phi) = \left(\frac{\partial l}{\partial \beta_0}, \frac{\partial l}{\partial \beta_1}, \cdots, \frac{\partial l}{\partial \beta_p}, \frac{\partial l}{\partial \phi} \right)^T . \tag{3}$$

- As derivadas em (3) são (ver as notas complementares)

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \phi^{-1} \sqrt{\frac{\omega_i}{V_i}} (Y_i - \mu_i) x_{ij}, \tag{4}$$

onde $\omega_i = (\partial \mu_i / \partial \eta_i)^2 / V_i$, $V_i = V(\mu_i)$ e

$$\frac{\partial l}{\partial \phi} = -\phi^{-2} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} \frac{\partial c(Y_i, \phi)}{\partial \phi}. \tag{5}$$

Tópico 11: função escore

- Portanto, a função escore é o vetor

$$\boldsymbol{U}(\boldsymbol{\beta}, \phi) = \left(\boldsymbol{U}_{\boldsymbol{\beta}}^T, U_\phi\right)^T, \tag{6}$$

onde

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \left(\frac{\partial l}{\partial \beta_0}, \frac{\partial l}{\partial \beta_1}, \cdots, \frac{\partial l}{\partial \beta_p}\right)^T, \tag{7}$$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \phi^{-1} \sqrt{\frac{\omega_i}{V_i}} (Y_i - \mu_i) x_{ij}, \tag{8}$$

$$\omega_i = (\partial \mu_i / \partial \eta_i)^2 / V_i, \quad V_i = V(\mu_i) \tag{9}$$

e

$$U_\phi = \frac{\partial l}{\partial \phi} = -\phi^{-2} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} \frac{\partial c(Y_i, \phi)}{\partial \phi}. \tag{10}$$

Tópico 11: função escore

- Sejam $\boldsymbol{Y}^T = (Y_1, \cdots, Y_n)$ o vetor dos valores de Y a serem observados e $\boldsymbol{X}$ a matriz $n \times (p+1)$ definida como

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{11}$$

- O vetor $\boldsymbol{U}_{\boldsymbol{\beta}}$ em (7) pode ser escrito como [2, p. 21]

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \phi^{-1} \boldsymbol{X}^T \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu}), \tag{12}$$

onde $\boldsymbol{W} = \mathrm{diag}\{\omega_1, \cdots, \omega_n\}$, $\boldsymbol{V} = \mathrm{diag}\{V_1, \cdots, V_n\}$ e

$$\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T \quad \text{e} \quad \boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T.$$

Tópico 11: função escore

- Se a ligação é canônica, então $g(\mu_i) = \eta_i = \theta_i$ (ver tópico 5).
- Nesse caso, $\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \theta_i} = V(\mu_i) = V_i$ (ver tópico 4).
- Portanto, $\omega_i = (\partial \mu_i / \partial \eta_i)^2 / V_i = V_i^2 / Vi = V_i$ e

$$\boldsymbol{W} = \text{diag}\{\omega_1, \cdots, \omega_n\} = \text{diag}\{V_1, \cdots, V_n\} = \boldsymbol{V}.$$

- Ou seja, se a função de ligação é canônica, então a derivada da log-verossimilhança com relação a β_l em (4) e (8) se torna

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \phi^{-1}(Y_i - \mu_i)x_{ij} \tag{13}$$

e $\boldsymbol{U}_{\boldsymbol{\beta}}$ em (12) fica dada por

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \phi^{-1}\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{\mu}). \tag{14}$$

Tópico 12: informação de Fisher

- A matriz de informação de Fisher observada, denotada por $\boldsymbol{J}(\boldsymbol{\beta}, \phi)$, é o negativo da matriz hessiana da log-verossimilhança.
- Ou seja,

$$\boldsymbol{J}(\boldsymbol{\beta}, \phi) = -\begin{pmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_p} & \frac{\partial^2 l}{\partial \beta_0 \partial \phi} \\ \frac{\partial^2 l}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_1^2} & \cdots & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_p} & \frac{\partial^2 l}{\partial \beta_1 \partial \phi} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 l}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_p \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_p^2} & \frac{\partial^2 l}{\partial \beta_p \partial \phi} \\ \frac{\partial^2 l}{\partial \phi \partial \beta_0} & \frac{\partial^2 l}{\partial \phi \beta_1} & \cdots & \frac{\partial^2 l}{\partial \phi \partial \beta_p} & \frac{\partial^2 l}{\partial \phi^2} \end{pmatrix} \tag{15}$$

Tópico 12: informação de Fisher

- As derivadas parciais em (15) são (ver as notas complementares)

$$\frac{\partial^2 l}{\partial \beta_j \beta_l} = \phi^{-1} \sum_{i=1}^{n} (Y_i - \mu_i) \frac{\partial^2 \theta_i}{\partial \mu_i^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ij} x_{il} + \\ + \phi^{-1} \sum_{i=1}^{n} (Y_i - \mu_i) \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial^2 \mu_i}{\partial \eta_i^2} x_{ij} x_{il} - \phi^{-1} \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{\partial \theta_i}{\partial \mu_i} x_{ij} x_{il}; \tag{16a}$$

$$\frac{\partial^2 l}{\partial \beta_j \phi} = -\phi^{-2} \sum_{i=1}^{n} \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) x_{ij}; \tag{16b}$$

$$\frac{\partial^2 l}{\partial \phi^2} = 2\phi^{-3} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} \frac{\partial^2 c(Y_i, \phi)}{\partial \phi^2}. \tag{16c}$$

Tópico 12: informação de Fisher

- Pode-se particionar $\boldsymbol{J}(\boldsymbol{\beta},\phi)$ como $\boldsymbol{J}(\boldsymbol{\beta},\phi) = \begin{pmatrix} \boldsymbol{J}_{\boldsymbol{\beta\beta}} & \boldsymbol{J}_{\boldsymbol{\beta}\phi} \\ \boldsymbol{J}_{\boldsymbol{\beta}\phi}^T & J_{\phi\phi} \end{pmatrix}$, onde

$$\boldsymbol{J}_{\boldsymbol{\beta\beta}} = -\begin{pmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 l}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_1^2} & \cdots & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_p \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_p^2} \end{pmatrix}; \tag{17a}$$

$$\begin{aligned} \boldsymbol{J}_{\boldsymbol{\beta}\phi} &= -\left(\frac{\partial^2 l}{\partial \phi \partial \beta_0}, \frac{\partial^2 l}{\partial \phi \beta_1}, \cdots, \frac{\partial^2 l}{\partial \phi \partial \beta_p}\right)^T = \\ &= -\phi^{-2}\boldsymbol{W}^{1/2}\boldsymbol{V}^{-1/2}(\boldsymbol{Y}-\boldsymbol{\mu})\,\text{(em notação matricial)}; \end{aligned} \tag{17b}$$

$$J_{\phi\phi} = -\frac{\partial^2 l}{\partial \phi^2}. \tag{17c}$$

Tópico 12: informação de Fisher

- A matriz de informação de Fisher esperada, a qual é denotada por $\boldsymbol{K}(\boldsymbol{\beta}, \phi)$, é o valor esperado da matriz de informação de Fisher observada.
- Isto é, $\boldsymbol{K}(\boldsymbol{\beta}, \phi) = E\left[\boldsymbol{J}(\boldsymbol{\beta}, \phi)\right] =$

$$= - \begin{pmatrix} E\left[\frac{\partial^2 l}{\partial \beta_0^2}\right] & E\left[\frac{\partial^2 l}{\partial \beta_0 \partial \beta_1}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_0 \partial \beta_p}\right] & E\left[\frac{\partial^2 l}{\partial \beta_0 \partial \phi}\right] \\ E\left[\frac{\partial^2 l}{\partial \beta_1 \partial \beta_0}\right] & E\left[\frac{\partial^2 l}{\partial \beta_1^2}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_1 \partial \beta_p}\right] & E\left[\frac{\partial^2 l}{\partial \beta_1 \partial \phi}\right] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ E\left[\frac{\partial^2 l}{\partial \beta_p \partial \beta_0}\right] & E\left[\frac{\partial^2 l}{\partial \beta_p \beta_1}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_p^2}\right] & E\left[\frac{\partial^2 l}{\partial \beta_p \partial \phi}\right] \\ E\left[\frac{\partial^2 l}{\partial \phi \partial \beta_0}\right] & E\left[\frac{\partial^2 l}{\partial \phi \beta_1}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \phi \partial \beta_p}\right] & E\left[\frac{\partial^2 l}{\partial \phi^2}\right] \end{pmatrix}. \quad (18)$$

Tópico 12: informação de Fisher

- As esperanças em (18) são (ver as notas complementares)

$$E\left[\frac{\partial^2 l}{\partial \beta_j \beta_l}\right] = -\phi^{-1}\sum_{i=1}^{n}\omega_i x_{ij} x_{il}; \tag{19a}$$

$$E\left[\frac{\partial^2 l}{\partial \beta_j \phi}\right] = 0; \tag{19b}$$

$$E\left[\frac{\partial^2 l}{\partial \phi^2}\right] = 2\phi^{-3}\left(\sum_{i=1}^{n}\mu_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} E\left(\frac{\partial^2 c(Y_i,\phi)}{\partial \phi^2}\right). \tag{19c}$$

Tópico 12: informação de Fisher

- Pode-se particionar $\boldsymbol{K}(\beta, \phi)$ como $\boldsymbol{K}(\beta, \phi) = \begin{pmatrix} \boldsymbol{K}_{\boldsymbol{\beta\beta}} & \boldsymbol{K}_{\boldsymbol{\beta}\phi} \\ \boldsymbol{K}_{\boldsymbol{\beta}\phi}^T & K_{\phi\phi} \end{pmatrix}$, onde

$$\boldsymbol{K}_{\boldsymbol{\beta\beta}} = -\begin{pmatrix} E\left[\frac{\partial^2 l}{\partial \beta_0^2}\right] & E\left[\frac{\partial^2 l}{\partial \beta_0 \partial \beta_1}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_0 \partial \beta_p}\right] \\ E\left[\frac{\partial^2 l}{\partial \beta_1 \partial \beta_0}\right] & E\left[\frac{\partial^2 l}{\partial \beta_1^2}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_1 \partial \beta_p}\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[\frac{\partial^2 l}{\partial \beta_p \partial \beta_0}\right] & E\left[\frac{\partial^2 l}{\partial \beta_p \beta_1}\right] & \cdots & E\left[\frac{\partial^2 l}{\partial \beta_p^2}\right] \end{pmatrix} = E\left[\boldsymbol{J}_{\boldsymbol{\beta\beta}}\right]; \quad (20a)$$

$$\boldsymbol{K}_{\boldsymbol{\beta}\phi} = -\left(E\left[\frac{\partial^2 l}{\partial \phi \partial \beta_0}\right], E\left[\frac{\partial^2 l}{\partial \phi \beta_1}\right], \cdots, E\left[\frac{\partial^2 l}{\partial \phi \partial \beta_p}\right]\right)^T = E\left[\boldsymbol{J}_{\boldsymbol{\beta}\phi}\right]; \qquad (20b)$$

$$K_{\phi\phi} = -E\left[\frac{\partial^2 l}{\partial \phi^2}\right] = E\left[J_{\phi\phi}\right]. \qquad (20c)$$

Tópico 12: informação de Fisher

- O termo $\boldsymbol{K}_{\boldsymbol{\beta\beta}}$ pode ser escrito como [2, p. 22]

$$\boldsymbol{K}_{\boldsymbol{\beta\beta}} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}. \tag{21}$$

- Pela equação (19b),

$$\boldsymbol{K}_{\boldsymbol{\beta}\phi} = \underset{(p+1)\times 1}{\boldsymbol{0}}, \tag{22}$$

onde $\underset{(p+1)\times 1}{\boldsymbol{0}}$ é o vetor-coluna de zeros de dimensão $p+1$.

- Logo, $\boldsymbol{K}(\boldsymbol{\beta}, \phi)$ é

$$\boldsymbol{K}(\boldsymbol{\beta}, \phi) = \begin{pmatrix} \boldsymbol{K}_{\boldsymbol{\beta\beta}} & \boldsymbol{K}_{\boldsymbol{\beta}\phi} \\ \boldsymbol{K}_{\boldsymbol{\beta}\phi}^T & K_{\phi\phi} \end{pmatrix} = \begin{pmatrix} \phi^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \underset{(p+1)\times 1}{\boldsymbol{0}} \\ \underset{1\times(p+1)}{\boldsymbol{0}} & K_{\phi\phi} \end{pmatrix}, \tag{23}$$

onde $K_{\phi\phi}$ é dado por (20c) e (19c).

Tópico 12: informação de Fisher

- Se a ligação é canônica, então

$$\boldsymbol{J}_{\boldsymbol{\beta}\phi} = -\phi^{-2}\boldsymbol{X}^T(\boldsymbol{Y}-\boldsymbol{\mu}), \tag{24a}$$

$$\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}; \tag{24b}$$

$$\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}. \tag{24c}$$

- De fato, conforme visto no tópico 11, supondo uma ligação canônica, tem-se que $\boldsymbol{W} = \boldsymbol{V}$.
- As expressões (24a) e (24b) são obtidas fazendo a substituição de $\boldsymbol{W}$ por $\boldsymbol{V}$ em (17b) e (21), respectivamente.
- Além disso, supondo que a ligação é canônica, a derivada da log-verossimilhança com relação a β_j fica dada por (13).

Tópico 12: informação de Fisher

- Calculando $\frac{\partial^2 l}{\partial \beta_j \partial \beta_l}$ a partir de (13), obtém-se

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} \phi^{-1} \frac{\partial \mu_i}{\partial \beta_l} x_{ij}.$$

- Dessa forma,

$$E\left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_l}\right] = \frac{\partial^2 l}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} \phi^{-1} \frac{\partial \mu_i}{\partial \beta_l} x_{ij},$$

pois $\frac{\partial^2 l}{\partial \beta_j \partial \beta_l}$ não depende de $Y_1, \cdots, Y_n$.

- Segue-se que (ver as expressões 17a e 20a)

$$\boldsymbol{K}_{\boldsymbol{\beta\beta}} = E\left[\boldsymbol{J}_{\boldsymbol{\beta\beta}}\right] = \boldsymbol{J}_{\boldsymbol{\beta\beta}}.$$

Tópico 12: informação de Fisher

Exemplo 1

Para o modelo linear normal (exemplo 1, tópico 4), tem-se que

$$\theta_i = \mu_i,\ \ b(\theta_i) = \frac{\theta_i^2}{2},\ \ \phi = \sigma^2 \text{ e } c(y_i, \phi) = -\frac{y_i^2}{2\phi} - \frac{1}{2}\ln(2\pi\phi).$$

Além disso, $V_i = V(\mu_i) = 1$ e $g(\mu_i) = \mu_i = \eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$. Logo,

$$\boldsymbol{V} = \operatorname{diag}\{1, \cdots, 1\} = I_n \quad \text{e} \quad \boldsymbol{\mu} = \left(\boldsymbol{x}_1^T \boldsymbol{\beta}, \cdots, \boldsymbol{x}_n^T \boldsymbol{\beta}\right).$$

Como a ligação é canônica (exemplo 4, tópico 5), tem-se que

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \phi^{-1} \boldsymbol{X}^T (\boldsymbol{Y} - \boldsymbol{\mu}) \quad \text{e} \quad \boldsymbol{J}_{\boldsymbol{\beta}\phi} = -\phi^{-2} \boldsymbol{X}^T (\boldsymbol{Y} - \boldsymbol{\mu}).$$

Tópico 12: informação de Fisher

Continuação do exemplo 1

Tem-se também que

$$\boldsymbol{J}_{\boldsymbol{\beta\beta}} = \boldsymbol{K}_{\boldsymbol{\beta\beta}} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{X}.$$

Resta calcular U_ϕ, $J_{\phi\phi}$ e $K_{\phi\phi}$. A derivada de $c(y_i, \phi)$ com relação a ϕ é

$$\frac{\partial c(y_i, \phi)}{\partial \phi} = \frac{y_i^2}{2\phi^2} - \frac{1}{2\phi}.$$

Com isso,

$$U_\phi = -\phi^{-2}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} \frac{\partial c(Y_i, \phi)}{\partial \phi} =$$

$$= -\frac{1}{\phi^2}\left(\sum_{i=1}^{n} Y_i\mu_i - \sum_{i=1}^{n} \frac{\mu_i^2}{2}\right) + \sum_{i=1}^{n} \frac{Y_i^2}{2\phi^2} - \frac{n}{2\phi}.$$

Tópico 12: informação de Fisher

Continuação do exemplo 1

A última expressão acima pode ser escrita como

$$U_\phi = \frac{1}{2\phi^2}\sum_{i=1}^{n}(-2Y_i\mu_i + \mu_i^2 + Y_i^2) - \frac{n}{2\phi} =$$

$$= \frac{1}{2\phi^2}\sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{n}{2\phi}.$$

A segunda derivada de $c(y_i, \phi)$ com relação a ϕ é

$$\frac{\partial^2 c(y_i,\phi)}{\partial\phi^2} = -\frac{y_i^2}{\phi^3} + \frac{1}{2\phi^2}.$$

Tópico 12: informação de Fisher

Continuação do exemplo 1

Logo,

$$
\begin{aligned}
J_{\phi\phi} &= -2\phi^{-3}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) - \sum_{i=1}^{n} \frac{\partial^2 c(Y_i, \phi)}{\partial \phi^2} = \\
&= -\frac{2}{\phi^3}\left(\sum_{i=1}^{n} Y_i\mu_i - \sum_{i=1}^{n} \frac{\mu_i^2}{2}\right) + \sum_{i=1}^{n} \frac{Y_i^2}{\phi^3} - \frac{n}{2\phi^2} = \\
&= \frac{1}{\phi^3}\sum_{i=1}^{n}(-2Y_i\mu_i + \mu_i^2 + Y_i^2) - \frac{n}{2\phi^2} = \\
&= \frac{1}{\phi^3}\sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{n}{2\phi^2}.
\end{aligned}
$$

Tópico 12: informação de Fisher

Continuação do exemplo 1

Segue-se que

$$K_{\phi\phi} = E\left[J_{\phi\phi}\right] = \frac{1}{\phi^3}\sum_{i=1}^{n} E[(Y_i - \mu_i)]^2 - \frac{n}{2\phi^2} =$$

$$= \frac{1}{\phi^3}\sum_{i=1}^{n} Var(Y_i) - \frac{n}{2\phi^2}.$$

Como $Var(Y_i) = \sigma^2 = \phi$ (exemplo 1, tópico 4), tem-se que

$$K_{\phi\phi} = \frac{n}{\phi^2} - \frac{n}{2\phi^2} = \frac{n}{2\phi^2}.$$

Tópico 12: informação de Fisher

Exemplo 2

Para o modelo de regressão Poisson (exemplo 2, aula 3)

$$\theta_i = \ln(\mu_i), \quad b(\theta_i) = e^{\theta_i}, \quad \phi = 1 \text{ e } \quad c(y_i, \phi) = -\ln(y_i!)\,(2\pi\phi)\,.$$

Além disso, $V_i = V(\mu_i) = \mu_i$ e $\mu_i = e^{\eta_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$. Logo,

$$\boldsymbol{V} = \text{diag}\left\{e^{\boldsymbol{x}_1^T\boldsymbol{\beta}}, \cdots, e^{\boldsymbol{x}_n^T\boldsymbol{\beta}}\right\} \quad \text{e} \quad \boldsymbol{\mu} = \left(e^{\boldsymbol{x}_1^T\boldsymbol{\beta}}, \cdots, e^{\boldsymbol{x}_n^T\boldsymbol{\beta}}\right).$$

Como ϕ é conhecido, os termos da função escore e das matrizes de informação que contêm ϕ devem ser ignorados. Ou seja,

$$\boldsymbol{U}(\beta, \phi) = \boldsymbol{U}_{\boldsymbol{\beta}}, \quad \boldsymbol{J}(\beta, \phi) = \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} \quad \text{e} \quad \boldsymbol{K}(\beta, \phi) = \boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}.$$

Tópico 12: informação de Fisher

Continuação do exemplo 2

A ligação é canônica (exemplo 6, tópico 5) e $\phi = 1$. Logo,

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{\mu}) \quad \text{e} \quad \boldsymbol{J}_{\boldsymbol{\beta\beta}} = \boldsymbol{K}_{\boldsymbol{\beta\beta}} = \boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}.$$

Exemplo 3

Para o modelo de regressão logística (exemplo 3, tópico 4)

$$\theta_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right), \; b(\theta_i) = \ln\left(1 + e^{\theta_i}\right), \; \phi = 1 \text{ e } c(y_i;\phi) = 0.$$

Além disso, $V_i = V(\mu_i) = \mu_i(1-\mu_i)$ e $\mu_i = \frac{e^{\eta_i}}{e^{\eta_i}+1}$.

Tópico 12: informação de Fisher

Continuação do exemplo 3

Portanto,

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n) \quad \text{e} \quad \boldsymbol{V} = \text{diag}\left\{V_1, \cdots, V_n\right\},$$

onde $\mu_i = \frac{e^{\eta_i}}{e^{\eta_i}+1}$ e $V_i = \mu_i(1-\mu_i) = \frac{e^{\eta_i}}{(e^{\eta_i}+1)^2}$. Como ϕ é conhecido, os termos da função escore e das matrizes de informação que contêm ϕ devem ser ignorados. Ou seja,

$$\boldsymbol{U}(\boldsymbol{\beta}, \phi) = \boldsymbol{U}_{\boldsymbol{\beta}}, \quad \boldsymbol{J}(\boldsymbol{\beta}, \phi) = \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} \quad \text{e} \quad \boldsymbol{K}(\boldsymbol{\beta}, \phi) = \boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}.$$

A ligação é canônica (exemplo 6, tópico 5) e $\phi = 1$. Logo,

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{\mu}) \quad \text{e} \quad \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}.$$

Tópico 12: informação de Fisher

Exemplo 4

Para o modelo de regressão gama, com ligação logarítmica, tem-se que (exemplos 7 e 8, aula 3)

$$\theta_i = -\frac{1}{\mu_i}, \quad b(\theta_i) = -\ln(-\theta_i), \quad \phi = a^{-1}$$

e

$$c(y_i, \phi) = \phi^{-1} \ln(\phi^{-1}) - \ln \Gamma(\phi^{-1}) + (\phi^{-1} - 1) \ln(y_i).$$

Além disso, $V_i = V(\mu_i) = \mu_i^2$ e $\mu_i = e^{\eta_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$. Portanto,

$$\boldsymbol{V} = \text{diag}\{e^{2\boldsymbol{x}_1^T \boldsymbol{\beta}}, \cdots, e^{2\boldsymbol{x}_n^T \boldsymbol{\beta}}\}.$$

A ligação não é canônica. Logo, é necessário calcular $\boldsymbol{W}$.

Tópico 12: informação de Fisher

Continuação do exemplo 4

Como $\frac{\partial \mu_i}{\partial \eta_i} = \mu_i$, tem-se que

$$\omega_i = (\partial \mu_i / \partial \eta_i)^2 / V_i = \mu_i^2 / \mu_i^2 = 1.$$

Segue-se que $\boldsymbol{W} = \text{diag}\{1, \cdots, 1\} = I_n$. Com isso,

$$\boldsymbol{U_\beta} = \phi^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu}),$$

$$\boldsymbol{J_{\beta\phi}} = \phi^{-2} \boldsymbol{X}^T \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu}) \quad \text{e} \quad \boldsymbol{K_{\beta\beta}} = \phi^{-1} \boldsymbol{X}^T \boldsymbol{X}.$$

A derivada de $c(y_i, \phi)$ com relação a ϕ é

$$\frac{\partial c(y_i, \phi)}{\partial \phi} = -\phi^{-2} \left(\ln(\phi^{-1}) + 1 - \psi(\phi^{-1}) + \ln(y_i) \right),$$

onde $\psi(x) = \frac{\partial \ln \Gamma(x)}{\partial x} = \frac{\Gamma'(x)}{\Gamma(x)}$ é a função digama.

Tópico 12: informação de Fisher

Continuação do exemplo 4

Lembrando que $\theta_i = -\frac{1}{\mu_i}$ e $b(\theta_i) = -\ln(-\theta_i) = \ln(\mu_i)$, tem-se

$$U_\phi = -\phi^{-2}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} \frac{\partial c(Y_i,\phi)}{\partial \phi} =$$

$$= \phi^{-2}\left(\sum_{i=1}^{n} \frac{Y_i}{\mu_i} + \sum_{i=1}^{n} \ln(\mu_i)\right) - \phi^{-2}\sum_{i=1}^{n} \ln(Y_i) - n\phi^{-2} -$$

$$- n\phi^{-2}\left(\ln(\phi^{-1}) - \psi(\phi^{-1})\right) =$$

$$= \phi^{-2}\sum_{i=1}^{n}\left[\frac{Y_i - \mu_i}{\mu_i} + \ln\left(\frac{\mu_i}{Y_i}\right)\right] - n\phi^{-2}\left(\ln(\phi^{-1}) - \psi(\phi^{-1})\right).$$

Tópico 12: informação de Fisher

Continuação do exemplo 4

As derivadas necessárias para obter $\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ são

$$\frac{\partial \mu_i}{\partial \eta_i} = \mu_i, \quad \frac{\partial^2 \mu_i}{\partial \eta_i^2} = \mu_i, \quad \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\mu_i^2}, \quad \text{e} \quad \frac{\partial^2 \theta_i}{\partial \mu_i^2} = -\frac{2}{\mu_i^3}.$$

Substituindo esses valores em (16a), obtém-se

$$\frac{\partial^2 l}{\partial \beta_j \beta_l} = -\phi^{-1} \sum_{i=1}^{n} (Y_i - \mu_i) x_{ij} x_{il} - \phi^{-1} \sum_{i=1}^{n} x_{ij} x_{il}.$$

Portanto, o elemento (j, l) da matriz $\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ em (17a) é

$$J_{\beta_j \beta_l} = -\frac{\partial^2 l}{\partial \beta_j \beta_l} = \phi^{-1} \sum_{i=1}^{n} (Y_i - \mu_i) x_{ij} x_{il} + \phi^{-1} \sum_{i=1}^{n} x_{ij} x_{il}.$$

Tópico 12: informação de Fisher

Continuação do exemplo 4

Em notação matricial, $\boldsymbol{J}_{\boldsymbol{\beta\beta}}$ pode ser escrita como

$$\boldsymbol{J}_{\boldsymbol{\beta\beta}} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{\mathcal{E}}\boldsymbol{X} + \phi^{-1}\boldsymbol{X}^T\boldsymbol{X},$$

onde

$$\boldsymbol{\mathcal{E}} = \mathrm{diag}\{Y_1 - \mu_1, \cdots, Y_n - \mu_n\}.$$

A segunda derivada de $c(Y_i, \phi)$ com relação a ϕ é

$$\frac{\partial^2 c(Y_i, \phi)}{\partial \phi^2} = -2\phi^{-1}\frac{\partial c(Y_i, \phi)}{\partial \phi} + \phi^{-4}(\phi - \psi'(\phi^{-1})),$$

onde $\psi'(x) = \frac{\partial \psi(x)}{\partial x}$ é a função trigama.

Tópico 12: informação de Fisher

Continuação do exemplo 4

Substituindo em (16c), obtém-se

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \phi^2} &= 2\phi^{-3}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} \frac{\partial^2 c(Y_i,\phi)}{\partial \phi^2} = \\
&= 2\phi^{-3}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) - 2\phi^{-1}\sum_{i=1}^{n} \frac{\partial c(Y_i,\phi)}{\partial \phi} + \\
&+ n\phi^{-4}(\phi - \psi'(\phi^{-1})) = \\
&= -2\phi^{-1}\left[-\phi^{-2}\left(\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right) + \sum_{i=1}^{n} \frac{\partial c(Y_i,\phi)}{\partial \phi}\right] + \\
&+ n\phi^{-4}(\phi - \psi'(\phi^{-1})).
\end{aligned}
$$

Tópico 12: informação de Fisher

Continuação do exemplo 4

Comparando essa última expressão com (10), tem-se que

$$\frac{\partial^2 l}{\partial \phi^2} = -2\phi^{-1} U_\phi + n\phi^{-4}(\phi - \psi'(\phi^{-1})).$$

Portanto, $J_{\phi\phi}$ em (17c) é

$$J_{\phi\phi} = -\frac{\partial^2 l}{\partial \phi^2} = 2\phi^{-1} U_\phi + n\phi^{-4}(\psi'(\phi^{-1}) - \phi).$$

Segue-se que

$$K_{\phi\phi} = E\left(J_{\phi\phi}\right) = 2\phi^{-1} E(U_\phi) + n\phi^{-4}(\psi'(\phi) - \phi) = n\phi^{-4}(\psi'(\phi) - \phi),$$

pois $E(U_\phi) = 0$ (ver as notas complementares).

Tópico 14: estimador de máxima verossimilhança

- Os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ são os valores desses parâmetros que maximizam a log-verossimilhança.
- Sejam $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ, respectivamente.
- Os valores de $\widehat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são obtidos como solução da equação

$$\boldsymbol{U}(\boldsymbol{\beta}, \phi) = 0.$$

- Usando (6), (10) e (12), a equação de estimação acima pode ser escrita como

$$\begin{cases} \boldsymbol{U}_{\boldsymbol{\beta}} & = \phi^{-1} \boldsymbol{X}^T \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu}) = 0 \\ U_{\phi} & = -\phi^{-2} \left(\sum_{i=1}^{n} Y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} \frac{\partial c(Y_i, \phi)}{\partial \phi} = 0. \end{cases}$$

Tópico 14: estimador de máxima verossimilhança

- A equação $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ equivale a

$$\boldsymbol{X}^T \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu}) = 0.$$

- Como $\boldsymbol{W}$, $\boldsymbol{V}$ e $\boldsymbol{\mu}$ dependem apenas de $\boldsymbol{\beta}$, o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ pode ser encontrado como a solução da equação $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$, independentemente de ϕ.
- Ou seja, $\widehat{\boldsymbol{\beta}}$ é a solução da equação

$$\boldsymbol{X}^T \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} \boldsymbol{\mu} = \boldsymbol{X}^T \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} \boldsymbol{Y}. \tag{26}$$

- Quando a ligação é canônica, $\boldsymbol{W} = \boldsymbol{V}$ (ver tópico 11).
- Logo, se a ligação é canônica, $\widehat{\boldsymbol{\beta}}$ é a solução da equação

$$\boldsymbol{X}^T \boldsymbol{\mu} = \boldsymbol{X}^T \boldsymbol{Y}. \tag{27}$$

Tópico 14: estimador de máxima verossimilhança

- Já o estimador de máxima verossimilhança de ϕ, em geral, depende do valor de $\boldsymbol{\beta}$.
- A equação $U_\phi = 0$ equivale a
$$\sum_{i=1}^{n} Y_i\theta_i - \sum_{i=1}^{n} b(\theta_i) = \phi^2 \sum_{i=1}^{n} \frac{\partial c(Y_i, \phi)}{\partial \phi}. \tag{28}$$
- Como $b'(\theta_i) = \mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i^T \boldsymbol{\beta})$, pode-se concluir que θ_i depende de $\boldsymbol{\beta}$.
- $\hat{\phi}$ é a solução da equação (28) com $\widehat{\boldsymbol{\beta}}$ no lugar de $\boldsymbol{\beta}$.
- Em resumo, os estimadores de máxima verossimilhança são obtidos da seguinte forma:
  1. calcula-se $\widehat{\boldsymbol{\beta}}$ como solução de (26) ou de (27);
  2. calcula-se $\hat{\phi}$ como solução de (28) com $\widehat{\boldsymbol{\beta}}$ no lugar de $\boldsymbol{\beta}$.

Tópico 14: estimador de máxima verossimilhança

- Os estimadores de máxima verossimilhança de η_i e de μ_i são

$$\hat{\eta}_i = \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}} \tag{29}$$

e

$$\hat{\mu}_i = g^{-1}(\hat{\eta}_i) = g^{-1}(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}). \tag{30}$$

- O estimador de máxima verossimilhança de $\boldsymbol{\mu}$ é

$$\widehat{\boldsymbol{\mu}} = (\hat{\mu}_1, \cdots, \hat{\mu}_n)^T. \tag{31}$$

- Seja $\boldsymbol{\eta}$ o vetor de preditores lineares

$$\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)^T = \left(\boldsymbol{x}_1^T \boldsymbol{\beta}, \cdots, \boldsymbol{x}_n^T \boldsymbol{\beta}\right)^T. \tag{32}$$

Tópico 14: estimador de máxima verossimilhança

- Em notação matricial, $\boldsymbol{\eta}$ pode ser escrito como

$$\boldsymbol{\eta} = \boldsymbol{X}^T\boldsymbol{\beta}, \tag{33}$$

  onde $\boldsymbol{X}$ é a matriz (11).
- O estimador de máxima verossimilhança de $\boldsymbol{\eta}$ é

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{X}^T\widehat{\boldsymbol{\beta}}. \tag{34}$$

- Quando ϕ é conhecido, como no caso das distribuições de Poisson e de Bernoulli, apenas $\boldsymbol{\beta}$ precisa ser estimado.

Tópico 14: estimador de máxima verossimilhança

Exemplo 5

No caso do modelo linear normal, a função de ligação é canônica. Logo, $\widehat{\boldsymbol{\beta}}$ é a solução de (27). Para o modelo linear normal tem-se que $\mu_i = \eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$. Em notação matricial,

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)^T = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta},$$

Substituindo em (27), obtém-se

$$\boldsymbol{X}^T \boldsymbol{\mu} = \boldsymbol{X}^T \boldsymbol{Y} \Leftrightarrow \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^T \boldsymbol{Y} \Leftrightarrow \boldsymbol{\beta} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{Y}.$$

Portanto, no caso do modelo linear normal,

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{Y}.$$

Tópico 14: estimador de máxima verossimilhança

Continuação do exemplo 5

Viu-se no exemplo 1 que

$$U_\phi = \frac{1}{2\phi^2}\sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{n}{2\phi}.$$

$U_\phi = 0$ implica que

$$\frac{1}{2\phi^2}\sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{n}{2\phi} = 0 \Leftrightarrow \phi = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_i)^2.$$

Ou seja, no caso do modelo linear normal

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}_i)^2,$$

onde $\hat{\mu}_i = \hat{\eta}_i = \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$.

Tópico 14: estimador de máxima verossimilhança

Exemplo 6

No caso do modelo de regressão Poisson, a função de ligação é canônica. Logo, $\widehat{\boldsymbol{\beta}}$ é a solução de (27). No modelo de regressão Poisson, $\mu_i = e^{\eta_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$. Ou seja, $\widehat{\boldsymbol{\beta}}$ é a solução de

$$\boldsymbol{X}^T \boldsymbol{\mu} = \boldsymbol{X}^T \boldsymbol{Y}.$$

com

$$\boldsymbol{\mu} = \left(e^{\boldsymbol{x}_1^T \boldsymbol{\beta}}, \cdots, e^{\boldsymbol{x}_n^T \boldsymbol{\beta}} \right)^T.$$

Essa equação não tem solução analítica. Nesse caso, $\widehat{\boldsymbol{\beta}}$ deve ser calculado utilizando um procedimento numérico. Não é preciso estimar ϕ, pois, no caso da distribuição de Poisson, $\phi = 1$.

Tópico 14: estimador de máxima verossimilhança

Exemplo 7

No caso do modelo de regressão logística, a função de ligação é canônica. Logo, $\widehat{\boldsymbol{\beta}}$ é a solução de (27). No modelo de regressão logística, $\mu_i = \frac{e^{\eta_i}}{e^{\eta_i}+1} = \frac{e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}+1}$. Ou seja, $\widehat{\boldsymbol{\beta}}$ é a solução de

$$\boldsymbol{X}^T\boldsymbol{\mu} = \boldsymbol{X}^T\boldsymbol{Y}, \quad \text{com} \quad \boldsymbol{\mu} = \left(\frac{e^{\boldsymbol{x}_1^T\boldsymbol{\beta}}}{e^{\boldsymbol{x}_1^T\boldsymbol{\beta}}+1}, \cdots, \frac{e^{\boldsymbol{x}_n^T\boldsymbol{\beta}}}{e^{\boldsymbol{x}_n^T\boldsymbol{\beta}}+1} \right)^T.$$

Essa equação não tem solução analítica. Nesse caso, $\widehat{\boldsymbol{\beta}}$ deve ser calculado utilizando um procedimento numérico. Não é preciso estimar ϕ, pois, no caso da distribuição de Bernoulli, $\phi = 1$.

Tópico 14: estimador de máxima verossimilhança

Exemplo 8

Como a ligação exponencial não é a ligação canônica da distribuição gama, no caso do modelo de regressão Gama com ligação exponencial, $\widehat{\boldsymbol{\beta}}$ é a solução de (26). Sabe-se que, nesse modelo, $\mu_i = e^{\eta_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$. Viu-se no exemplo 4 que

$$\boldsymbol{V} = \text{diag}\{\mu_1^2, \cdots, \mu_n^2\} \quad \text{e} \quad \boldsymbol{W} = \text{diag}\{1, \cdots, 1\} = \boldsymbol{I}_n$$

Portanto, $\widehat{\boldsymbol{\beta}}$ é a solução de

$$\boldsymbol{X}^T \boldsymbol{V}^{-1/2} \boldsymbol{\mu} = \boldsymbol{X}^T \boldsymbol{V}^{-1/2} \boldsymbol{Y}.$$

com

$$\boldsymbol{\mu} = \left(e^{\boldsymbol{x}_1^T \boldsymbol{\beta}}, \cdots, e^{\boldsymbol{x}_n^T \boldsymbol{\beta}}\right)^T \quad \text{e} \quad \boldsymbol{V} = \left\{e^{2\boldsymbol{x}_1^T \boldsymbol{\beta}}, \cdots, e^{2\boldsymbol{x}_n^T \boldsymbol{\beta}}\right\}.$$

Tópico 14: estimador de máxima verossimilhança

Continuação do exemplo 8

Essa equação não tem solução analítica. Nesse caso, $\widehat{\boldsymbol{\beta}}$ deve ser calculado utilizando um procedimento numérico. Uma vez que $\widehat{\boldsymbol{\beta}}$ tenha sido calculado, passa-se ao cálculo de $\hat{\phi}$. Viu-se, no exemplo 4, que

$$U_\phi = \phi^{-2}\left[\sum_{i=1}^{n}\frac{Y_i-\mu_i}{\mu_i}+\sum_{i=1}^{n}\ln\left(\frac{\mu_i}{Y_i}\right)\right]-n\phi^{-2}(\ln(\phi^{-1})-\psi(\phi^{-1})).$$

$\hat{\phi}$ é a solução de $U_\phi = 0$ com $\widehat{\boldsymbol{\beta}}$ no lugar de $\boldsymbol{\beta}$. $U_\phi = 0$ implica

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{Y_i-\mu_i}{\mu_i}+\ln\left(\frac{\mu_i}{Y_i}\right)\right]=\ln(\phi^{-1})-\psi(\phi^{-1}).$$

Tópico 14: estimador de máxima verossimilhança

Continuação do exemplo 8

Portanto, $\hat{\phi}$ é a solução de

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i} + \ln\left(\frac{\hat{\mu}_i}{Y_i}\right)\right] = \ln(\phi^{-1}) - \psi(\phi^{-1}).$$

Essa equação não tem solução analítica. Nesse caso, $\hat{\phi}$ deve ser calculado utilizando um procedimento numérico.

Tópico 15: método de Newton-Raphson

- O estimador de máxima verossimilhança de $\boldsymbol{\beta}$, conforme visto no tópico 14, é a solução da equação (26).
- Viu-se também que $\hat{\phi}$, o estimador de máxima verossimilhança de ϕ, é a solução da equação (28), com $\widehat{\boldsymbol{\beta}}$ no lugar de $\boldsymbol{\beta}$.
- Acontece que nem sempre essas equações possuem uma solução analítica, isto é, uma solução em forma fechada.
- Isso foi mencionado nos exemplos (7) e (8).
- Em casos como esses, é necessário utilizar um procedimento numérico para encontrar as soluções dessas equações.
- Um método que pode ser utilizado com essa finalidade é o método de Newton-Raphson [1, p. 72].

Tópico 15: método de Newton-Raphson

- Seja $h : (a, b) \to \mathbb{R}$ uma função real diferenciável em (a, b).
- Deseja-se encontrar a solução da equação $h(x) = 0$.
- Seja $x_0 \in (a, b)$ um valor aproximado para essa solução.
- A aproximação de Taylor de ordem 1 de h, em torno de x_0, é

$$h(x) \approx h(x_0) + h'(x_0)(x - x_0).$$

- O método de Newton-Raphson propõe resolver $h(x) = 0$, de forma aproximada, substituindo $h(x)$ por sua aproximação de Taylor de ordem 1.
- Nesse caso,

$$h(x_0) + h'(x_0)(x - x_0) = 0 \Rightarrow x = x_0 - \frac{h(x_0)}{h'(x_0)}.$$

Tópico 15: método de Newton-Raphson

- Se x_0 é suficientemente próximo da solução de $h(x) = 0$ e se a função h satisfaz certas condições, então $x = x_0 - \frac{h(x_0)}{h'(x_0)}$ é uma aproximação para a solução de $h(x) = 0$ melhor do que x_0.
- Fazendo $x_1 = x_0 - \frac{h(x_0)}{h'(x_0)}$, com base nesse mesmo raciocínio, obtém-se que $x_2 = x_1 - \frac{h(x_1)}{h'(x_1)}$ é uma aproximação para a solução de $h(x) = 0$ melhor do que x_1.
- Isso leva ao seguinte processo iterativo:

$$x_m = x_{m-1} - \frac{h(x_{m-1})}{h'(x_{m-1})},$$

no qual x_{m-1} é uma solução aproximada para $h(x) = 0$ e x_m é uma aproximação melhor do que x_{m-1}.

Tópico 15: método de Newton-Raphson

- O método de Newton-Raphson consiste em repetir esse processo iterativo até que se obtenha um valor suficientemente próximo da solução de $h(x) = 0$.
- Isso ocorre quando algum critério de precisão pré-estabelecido é satisfeito como, por exemplo, quando a distância entre duas aproximações sucessivas se torna menor do que um dado ϵ.
- De acordo com esse critério, se $|x_m - x_{m-1}| < \epsilon$, então x_m é suficientemente próximo da solução de $h(x) = 0$.
- Nesse caso, toma-se x_m como sendo a solução dessa equação.
- Conforme já mencionado, $\widehat{\boldsymbol{\beta}}$ é a solução da equação $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$.
- A solução da equação $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ é obtida utilizando-se uma versão multidimensional do método de Newton-Raphson [2, p. 25].

Tópico 15: método de Newton-Raphson

- Seja $\beta^{(0)}$ uma solução aproximada para $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$.
- A aproximação de Taylor de primeira ordem de $\boldsymbol{U}_{\boldsymbol{\beta}}$, em torno de $\boldsymbol{\beta}^{(0)}$, é

$$\boldsymbol{U}_{\boldsymbol{\beta}} \approx \boldsymbol{U}_{\boldsymbol{\beta}^{(0)}} + \boldsymbol{U}'_{\boldsymbol{\beta}^{(0)}} \left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\right),$$

  onde $\boldsymbol{U}'_{\boldsymbol{\beta}}$ é a matriz $(p+1) \times (p+1)$, cujas linhas são as derivadas do vetor $\boldsymbol{U}_{\boldsymbol{\beta}}$ com relação a $\beta_0, \beta_1, \cdots, \beta_p$.
- Isto é, pelas fórmulas (7) e (17a), tem-se que $\boldsymbol{U}'_{\boldsymbol{\beta}} = -\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$.
- A aproximação de Taylor de primeira ordem de $\boldsymbol{U}_{\boldsymbol{\beta}}$, em torno de $\boldsymbol{\beta}^{(0)}$, pode, portanto, ser escrita como

$$\boldsymbol{U}_{\boldsymbol{\beta}} \approx \boldsymbol{U}_{\boldsymbol{\beta}^{(0)}} - \boldsymbol{J}_{\boldsymbol{\beta}^{(0)}\boldsymbol{\beta}^{(0)}} \left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\right).$$

Tópico 15: método de Newton-Raphson

- Substituindo $\boldsymbol{U}_{\boldsymbol{\beta}}$, na equação $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$, por sua aproximação de Taylor de primeira ordem, obtém-se

$$\boldsymbol{U}_{\boldsymbol{\beta}^{(0)}} - \boldsymbol{J}_{\boldsymbol{\beta}^{(0)}\boldsymbol{\beta}^{(0)}}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\right) = 0 \Leftrightarrow$$
$$\Leftrightarrow \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \left(\boldsymbol{J}_{\boldsymbol{\beta}^{(0)}\boldsymbol{\beta}^{(0)}}\right)^{-1}\boldsymbol{U}_{\boldsymbol{\beta}^{(0)}}.$$

- Se $\boldsymbol{\beta}^{(0)}$ é suficientemente próximo da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ e se a função $\boldsymbol{U}_{\boldsymbol{\beta}}$ satisfaz certas condições, então

$$\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} + \left(\boldsymbol{J}_{\boldsymbol{\beta}^{(0)}\boldsymbol{\beta}^{(0)}}\right)^{-1}\boldsymbol{U}_{\boldsymbol{\beta}^{(0)}}$$

  é uma aproximação da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ melhor do que $\boldsymbol{\beta}^{(0)}$.

Tópico 15: método de Newton-Raphson

- Fazendo $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(0)} + \left(\boldsymbol{J}_{\boldsymbol{\beta}^{(0)}\boldsymbol{\beta}^{(0)}}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\beta}^{(0)}}$, com base nesse mesmo raciocínio, obtém-se que $\boldsymbol{\beta}^{(2)}$, dado por

$$\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)} + \left(\boldsymbol{J}_{\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(1)}}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\beta}^{(1)}},$$

é uma aproximação da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ melhor do que $\boldsymbol{\beta}^{(1)}$.

- Isso leva ao seguinte procedimento iterativo:

$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} + \left(\boldsymbol{J}_{\boldsymbol{\beta}^{(m-1)}\boldsymbol{\beta}^{(m-1)}}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\beta}^{(m-1)}}, \qquad (35)$$

no qual $\boldsymbol{\beta}^{(m-1)}$ é uma solução aproximada para $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$ e $\boldsymbol{\beta}^{(m)}$ é uma aproximação melhor do que $\boldsymbol{\beta}^{(m-1)}$.

Tópico 15: método de Newton-Raphson

- O método de Newton-Raphson consiste em repetir esse processo iterativo até que se obtenha um valor suficientemente próximo da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$.
- Isso ocorre quando algum critério de precisão pré-estabelecido é satisfeito como, por exemplo, quando a distância entre duas aproximações sucessivas se torna menor do que um dado ϵ.
- Por esse critério, se $||\boldsymbol{\beta}^{(m)} - \boldsymbol{\beta}^{(m-1)}|| < \epsilon$, então considera-se que $\boldsymbol{\beta}^{(m)}$ é suficientemente próximo da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$.
- Nesse caso, toma-se $\boldsymbol{\beta}^{(m)}$ como sendo a solução dessa equação.
- Isto é, toma-se tal $\boldsymbol{\beta}^{(m)}$ como sendo o valor calculado de $\widehat{\boldsymbol{\beta}}$.
- De posse de $\widehat{\boldsymbol{\beta}}$, pode-se usar o método de Newton-Raphson para calcular $\hat{\phi}$.

Tópico 15: método de Newton-Raphson

- O método é aplicado para resolver $U_\phi = 0$.
- Nesse caso, o processo iterativo, deduzido de forma análoga ao caso unidimensional, fica dado por

$$\phi_m = \phi_{m-1} - \frac{U_{\phi_{m-1}}}{U'_{\phi_{m-1}}} = \phi_{m-1} + \frac{U_{\phi_{m-1}}}{J_{\phi_{m-1}\phi_{m-1}}}, \tag{36}$$

no qual, pelas fórmulas (10) e (17c), $J_{\phi\phi} = -U'_\phi = -\frac{\partial^2 l}{\partial \phi^2}$.

- Além disso, em (36), $\boldsymbol{\beta}$ é substituído por $\widehat{\boldsymbol{\beta}}$.
- O método de Newton-Raphson consiste em repetir esse processo iterativo até que se obtenha um valor suficientemente próximo da solução de $U_\phi = 0$.

Tópico 15: método de Newton-Raphson

- Isso ocorre quando algum critério de precisão pré-estabelecido é satisfeito como, por exemplo, quando a distância entre duas aproximações sucessivas se torna menor do que um dado ϵ.
- De acordo com esse critério, se $|\phi_m - \phi_{m-1}| < \epsilon$, então ϕ_m é suficientemente próximo da solução de $U_\phi = 0$.
- Nesse caso, toma-se ϕ_m como sendo a solução dessa equação.
- Isto é, toma-se tal ϕ_m como sendo o valor calculado de $\hat{\phi}$.

Tópico 16: método escore de Fisher

- O método escore de Fisher é idêntico ao método de Newton-Raphson, com exceção do fato de que, ao invés de se utilizar a matriz de informação de Fisher observada, usa-se a matriz de informação de Fisher esperada.
- Isto é, na expansão de Taylor de ordem 1 de $\boldsymbol{U}_{\boldsymbol{\beta}}$, faz-se a substituição de $\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ por seu valor esperado $\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ em (20a).
- Nesse caso, o procedimento iterativo (35) fica

$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} + \left(\boldsymbol{K}_{\boldsymbol{\beta}^{(m-1)}\boldsymbol{\beta}^{(m-1)}}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\beta}^{(m-1)}}, \qquad (37)$$

- Como o método de Newton-Raphson, o método escore de Fisher repete esse processo iterativo até que se obtenha um valor suficientemente próximo da solução de $\boldsymbol{U}_{\boldsymbol{\beta}} = 0$.

Tópico 16: método escore de Fisher

- Analogamente, na expansão de Taylor de ordem 1 de U_ϕ, faz-se a substituição de $J_{\phi\phi}$ por seu valor esperado $K_{\phi\phi}$ em (20c).
- Nesse caso, o procedimento iterativo (36) fica

$$\phi_m = \phi_{m-1} - \frac{U_{\phi_{m-1}}}{U'_{\phi_{m-1}}} = \phi_{m-1} + \frac{U_{\phi_{m-1}}}{K_{\phi_{m-1}\phi_{m-1}}}. \tag{38}$$

- Assim como o método de Newton-Raphson, o método escore de Fisher repete esse processo iterativo até que se obtenha um valor suficientemente próximo da solução de $U_\phi = 0$.
- Por apresentar diversas vantagens com relação ao método de Newton-Raphson [1, p. 88], recomenda-se, em geral, o uso do método escore de Fisher.

Tópico 16: método escore de Fisher

- Entretanto, conforme afirma a equação (24c), se a ligação é canônica, então as matrizes $\boldsymbol{K}_{\boldsymbol{\beta\beta}}$ e $\boldsymbol{J}_{\boldsymbol{\beta\beta}}$ se igualam, igualando os métodos de Newton-Raphson e escore de Fisher, quando aplicados ao parâmetro $\boldsymbol{\beta}$.
- O procedimento iterativo (37) pode ser escrito de uma maneira mais conveniente [1, p. 74].
- Note que, a cada passo, o vetor $\boldsymbol{\eta}$ em (32) e o vetor $\boldsymbol{\mu}$, a matriz $\boldsymbol{V}$ e a matriz $\boldsymbol{W}$ em (12) são atualizados, junto com a atualização de $\boldsymbol{\beta}$.
- No passo $m-1$, o vetor de preditores lineares é

$$\boldsymbol{\eta}^{(m-1)} = \left(\eta_1^{(m-1)}, \cdots, \eta_n^{(m-1)}\right) = \\ = \left(\boldsymbol{x}_1^T \boldsymbol{\beta}^{(m-1)}, \cdots, \boldsymbol{x}_n^T \boldsymbol{\beta}^{(m-1)}\right)^T = \boldsymbol{X}\boldsymbol{\beta}^{(m-1)}.$$

Tópico 16: método escore de Fisher

- O vetor de médias é

$$\boldsymbol{\mu}^{(m-1)} = \left(\mu_1^{(m-1)}, \cdots, \mu_n^{(m-1)}\right),$$

onde $\mu_i^{(m-1)} = g^{(-1)}\left(\eta_i^{(m-1)}\right) = g^{-1}\left(\boldsymbol{x}_1^T \boldsymbol{\beta}^{(m-1)}\right)$.

- A matriz de funções de variância é

$$\boldsymbol{V}^{(m-1)} = \text{diag}\left\{V_1^{(m-1)}, \cdots, V_n^{(m-1)}\right\},$$

onde $V_i^{(m-1)} = V\left(\mu_1^{(m-1)}\right)$.

Tópico 16: método escore de Fisher

- A matriz de pesos é

$$\boldsymbol{W}^{(m-1)} = \text{diag}\left\{\omega_1^{(m-1)}, \cdots, \omega_n^{(m-1)}\right\},$$

onde $\omega_i^{(m-1)} = \frac{1}{V_i^{(m-1)}} \frac{\partial \mu_i}{\partial \eta_i}\Big|_{\eta_i = \eta_i^{(m-1)}}$.

- Com isso, $\boldsymbol{U}_{\boldsymbol{\beta}^{(m-1)}}$ é dado por

$$\boldsymbol{U}_{\boldsymbol{\beta}^{(m-1)}} = \phi^{-1}\boldsymbol{X}^T\left[\boldsymbol{W}^{(m-1)}\right]^{1/2}\left[\boldsymbol{V}^{(m-1)}\right]^{-1/2}\left(\boldsymbol{Y} - \boldsymbol{\mu}^{(m-1)}\right).$$

- $\boldsymbol{K}_{\boldsymbol{\beta}^{(m-1)}\boldsymbol{\beta}^{(m-1)}}$ é dado por

$$\boldsymbol{K}_{\boldsymbol{\beta}^{(m-1)}\boldsymbol{\beta}^{(m-1)}} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(m-1)}\boldsymbol{X}.$$

Tópico 16: método escore de Fisher

- Utilizando essas expressões, pode-se escrever (37) como

$$\beta^{(m)} = \left[\boldsymbol{X}^T \boldsymbol{W}^{(m-1)} \boldsymbol{X}\right]^{-1} \boldsymbol{X}^T \boldsymbol{W}^{(m-1)} \boldsymbol{z}^{(m-1)}, \qquad (39)$$

onde

$$\boldsymbol{z}^{(m-1)} = \boldsymbol{\eta}^{(m-1)} + \left[\boldsymbol{W}^{(m-1)}\right]^{-1/2} \left[\boldsymbol{V}^{(m-1)}\right]^{-1/2} \left(\boldsymbol{Y} - \boldsymbol{\mu}^{(m-1)}\right). \qquad (40)$$

- Detalhes sobre a obtenção de (39) podem ser vistos nas notas complementares.
- Nota-se que $\beta^{(m)}$ é o ajuste de mínimos quadrados ponderados dos coeficientes de uma regressão, na qual $\boldsymbol{X}$ representa as preditoras e $\boldsymbol{z}^{(m-1)}$ a resposta e $\boldsymbol{W}^{(m-1)}$ a matriz de pesos.

Tópico 16: método escore de Fisher

- Por conta disso, o método escore de Fisher, quando aplicado no cálculo de $\widehat{\boldsymbol{\beta}}$, é conhecido como procedimento de mínimos quadrados iterativos reponderados [2, p. 25].
- Para iniciar o procedimento iterativo, é necessário fornecer uma estimativa inicial de $\widehat{\boldsymbol{\beta}}$.
- Ao invés de fornecer uma estimativa inicial para $\widehat{\boldsymbol{\beta}}$, pode-se fornecer uma estimativa inicial para é necessária uma estimativa inicial de $\boldsymbol{\mu}$.
- Uma escolha comum é considerar o valor observado de Y_i como estimativa inicial de μ_i.
- Nesse caso, $\boldsymbol{\mu}^{(0)} = \boldsymbol{y} = (y_1, \cdots, y_n)^T$.
- Com base em $\boldsymbol{\mu}^{(0)}$, calcula-se $\boldsymbol{\eta}^{(0)}$, $\boldsymbol{V}^{(0)}$, $\boldsymbol{W}^{(0)}$ e $\boldsymbol{z}^{(0)}$.

Tópico 16: método escore de Fisher

- Substituindo esses valores em (39) , obtém-se

$$\boldsymbol{\beta}^{(1)} = \left[\boldsymbol{X}^T\boldsymbol{W}^{(0)}\boldsymbol{X}\right]^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(0)}\boldsymbol{z}^{(0)}.$$

- A partir de $\boldsymbol{\beta}^{(1)}$, calcula-se $\boldsymbol{\eta}^{(1)}$, $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{V}^{(1)}$, $\boldsymbol{W}^{(1)}$ e $\boldsymbol{z}^{(1)}$.
- Novamente, substituindo esses valores em (39), obtém-se

$$\boldsymbol{\beta}^{(2)} = \left[\boldsymbol{X}^T\boldsymbol{W}^{(1)}\boldsymbol{X}\right]^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(1)}\boldsymbol{z}^{(1)}.$$

- Repete-se esse processo até que se chegue a um valor $\boldsymbol{\beta}^{(m)}$ suficientemente preciso.
- Em [1, p. 76], comenta-se sobre possíveis critérios de precisão.
- Em [1, p. 75] e [1, p. 76], encontram-se sugestões sobre como inicializar o método escore de Fisher.

Tópico 16: método escore de Fisher

- Como aproximação inicial para ϕ, pode-se utilizar o valor do seguinte estimador de Pearson [1, p. 113]

$$\hat{\phi}_P = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \frac{(Y_i - \mu_i)^2}{V_i}, \tag{41}$$

  onde $p+1$ é o número de coeficientes de regressão.
- Na prática, pode-se utilizar o estimador (41) para estimar ϕ, ao invés do estimador de máxima verossimilhança.
- Na plataforma computacional **R**, a função que realiza o ajuste do MLG é a função `glm`.
- A função `gamma.shape`, do pacote `MASS`, calcula a estimativa de máxima verossimilhança de ϕ^{-1}, no caso do modelo gama.

Tópico 14: estimador de máxima verossimilhança

Exemplo 9

Considere o conjunto de dados **store.dat**, cuja descrição foi feita no exemplo 2 da aula 2. O objetivo é ajustar um modelo de regressão Poisson, no qual a variável resposta é o número de clientes e as variáveis preditoras são a distância ao concorrente mais próximo e a distância à loja. Os detalhes desse exemplo podem ser vistos no jupyter notebook **exemplo_9.ipynb** ou no arquivo **exemplo_9.html**, localizados na pasta das aulas 5 e 6.

Tópico 14: estimador de máxima verossimilhança

Exemplo 10

Considere o conjunto de dados **icu.csv**, que se encontra descrito no exemplo 3 da aula 2. O objetivo é ajustar um modelo de regressão logística, tendo a variável **sta** (Sobrevivência) como variável resposta e a variável **age** (Idade) como variável preditora. Os detalhes desse exemplo podem ser vistos no jupyter notebook **exemplo_10.ipynb** ou no arquivo **exemplo_10.html**, localizados na pasta das aulas 5 e 6.

Tópico 14: estimador de máxima verossimilhança

Exemplo 11

Considere o conjunto de dados **trees.dat**, que se encontra descrito no exemplo 4 da aula 2. O objetivo é ajustar um modelo de regressão gama, com ligação logarítmica, tendo a variável volume como variável resposta e as variáveis altura e diâmetro como variáveis preditoras. Os detalhes desse exemplo podem ser vistos no jupyter notebook **exemplo_11.ipynb** ou no arquivo **exemplo_11.html**, ambos na pasta das aulas 5 e 6.

Referências I

[1] G. M. Cordeiro e C. G. B. Demétrio, *Modelos lineares generalizados e extensões*, 2013, disponível em https://docs.ufpr.br/~taconeli/CE22518/LivClarice.pdf.

[2] G. A. Paula, *Modelos lineares generalizados com apoio computacional*, 2013, disponível em https://www.ime.usp.br/~giapaula/texto_2013.pdf.