

Introdução geral aos MLG: parte 1

Prof. Caio Azevedo

Modelo linear generalizado

- Consideraremos que $Y \sim FE(\theta, \phi)$ e que temos $Y_i \stackrel{ind.}{\sim} FE(\theta_i, \phi), i = 1, 2, \dots, n$, ou seja

$$f(y_i; \theta_i) = \exp \{ \phi [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \} \mathbb{1}_A(y_i)$$

Modelo linear generalizado

- $Y_i \sim \text{FE}(\theta_i, \phi), i = 1, 2, \dots, n, \theta_i = h(\mu_i).$
- $g(\mu_i) = \eta_i, \eta_i = \sum_{j=1}^p X_{ji}\beta_j, \mathcal{E}(Y_i) = \mu_i = b'(\theta_i), X_{ji} :$ covariável j associada ao indivíduo i (fixa e conhecida) e $\beta = (\beta_1, \dots, \beta_p)'$, $\phi :$ parâmetros desconhecidos.
- $\mathcal{V}(Y_i) = \phi^{-1}V(\mu_i),$ em que $V(\mu_i) = \frac{d\mu_i}{d\theta_i}.$
- $g(\cdot)$ é uma função de ligação (inversível e duplamente diferenciável). Quando $\theta = g(\cdot)$ temos a função de ligação canônica.
- Note que $\theta_i = h(g^{-1}(\eta_i)).$ Se considerarmos a função de ligação canônica, temos que $\theta_i = \eta_i.$

Modelo linear generalizado

$$Y_i \stackrel{ind.}{\sim} FE(\theta_i, \phi) \quad , \quad \theta_i = h(\mu_i), i = 1, \dots, n$$
$$g(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta} = \sum_{j=1}^p X_{ji} \beta_j; \mathbf{x}_i = (X_{1i}, \dots, X_{pi})'$$

em que $g(\cdot)$ é chamada de função de ligação e η_i é o preditor linear (relacionado ao indivíduo i).

Escolhas naturais para funções de ligação

Distribuição	$g(\mu)$
Normal	μ (canônica)
Poisson	$\ln \mu$ (canônica)
Binomial	$\ln(\mu/(1 - \mu))$ (canônica) (a fda de qualquer variável contínua definida na reta)
Gama	$\ln \mu$
N.Inversa	$\ln \mu$

Outras funções de ligação

- Seja μ a proporção de sucessos de uma binomial.
- A ligação probito é dada por

$$\Phi^{-1}(\mu) = \eta$$

ou, de modo equivalente, $\mu = \Phi(\eta)$, em que $\Phi(\cdot)$ é a fda de uma distribuição normal padrão.

Outras funções de ligação

- Novamente, seja μ a proporção de sucessos de uma binomial.
- A fda de uma distribuição do valor extremo padrão (ou Gumbell padrão, a qual corresponde ao logaritmo natural de uma distribuição exponencial com seu parâmetro igual a 1) é dada por:

$$F(x) = 1 - \exp\{-\exp(x)\}$$

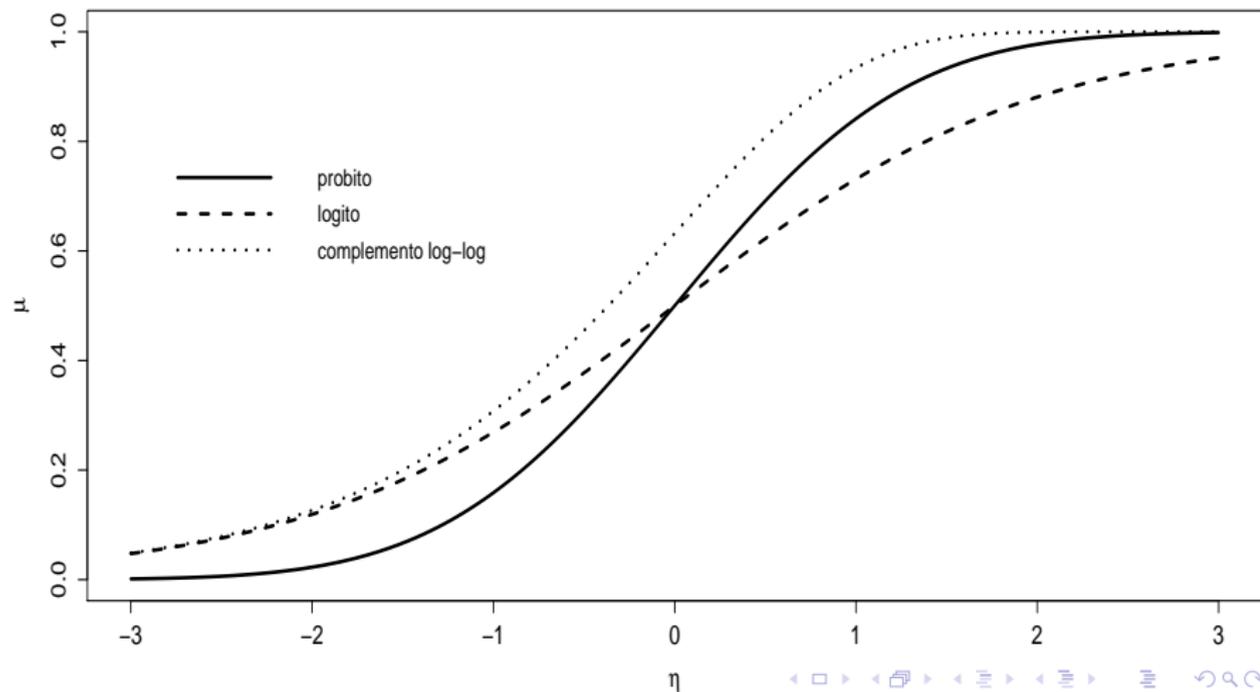
- Assim, o modelo binomial com ligação log-log é dado por

$$\mu = 1 - \exp\{-\exp(\eta)\}$$

ou de modo equivalente,

$$\ln(-\ln(1 - \mu)) = \eta.$$

Funções de ligação para médias no intervalo (0,1)



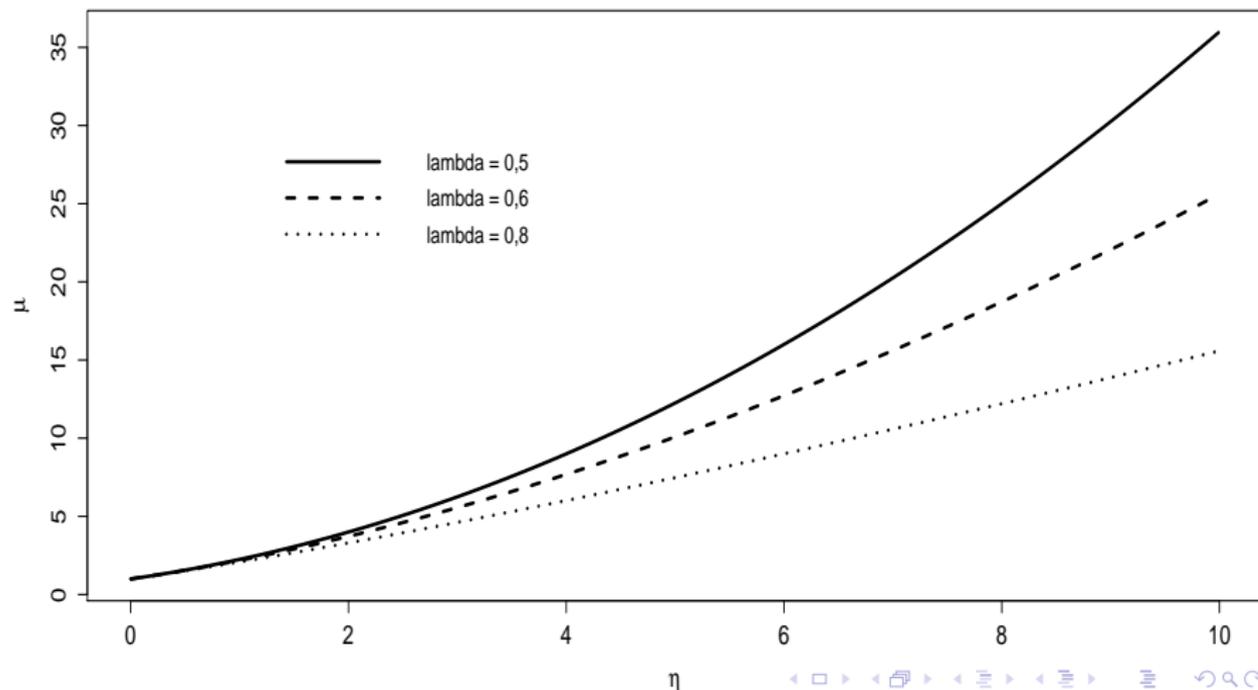
Funções de ligação Box-Cox

- Uma classe importante de ligações (para observações positivas) é a classe de ligações de Box-Cox definida por

$$\eta = \begin{cases} (\mu^\lambda - 1)/\lambda, & \text{se } \lambda \neq 0 \\ \ln(\mu), & \text{se } \lambda = 0 \end{cases}$$

- Podemos considerar vários valores para λ , ajustando o modelo para cada um deles, e utilizar algum critério de seleção de modelos para escolher o valor mais apropriado.
- Podemos também estimar λ em concomitância com os outros parâmetros.

Funções de ligação da família Box-Cox



Resumo: principais funções de ligação

Distribuição	$g(\mu)$
Normal	$\mu, (\text{canônica}), 1/\mu \text{ e } \ln \mu \text{ (se } \mu > 0)$
Poisson	$\ln \mu (\text{canônica}), \sqrt{\mu}$
Binomial	$\ln(\mu/(1 - \mu)) \text{ (canônica)}, \Phi^{-1}(\mu), \ln(-\ln(1 - \mu))$
Gama	$1/\mu (\text{canônica}), \ln \mu$
N.Inversa	$1/(2\mu^2) \text{ (canônica)}, \ln \mu, 1/\mu$

Estimação paramétrica

- Utilizaremos o método de máxima verossimilhança (MV).
- Outros métodos de estimação podem ser considerados, p.e.: métodos bayesianos, MQO, MQP, MV ponderada, dentre outros.
- Verossimilhança

$$L(\boldsymbol{\beta}, \phi) = \exp \left\{ \phi \left[\sum_{i=1}^n y_i \theta_i - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi) \right\}$$

- Log-verossimilhança

$$l(\boldsymbol{\beta}, \phi) = \phi \left[\sum_{i=1}^n y_i \theta_i - \sum_{i=1}^n b(\theta_i) \right] + \sum_{i=1}^n c(y_i, \phi)$$

- Notação: $\hat{\boldsymbol{\beta}}$ estimador, $\tilde{\boldsymbol{\beta}}$ estimativa.

Derivadas matriciais úteis

- Vamos começar considerando ligações canônicas. Note que temos um total de $p + 1$ parâmetros (família bi-paramétrica) ou p (família uni-paramétrica).
- Sejam $\mathbf{A}_{(m \times n)}$ e $\mathbf{x}_{(n \times 1)}$ tais que

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}; \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Derivadas matriciais úteis

- Sejam $\mathbf{A}_{(m \times n)}$ e $\mathbf{x}_{(n \times 1)}$, e defina $\mathbf{y} = \mathbf{Ax}$ (\mathbf{A} não depende de \mathbf{x}).

Então:

$$\mathbf{y} = \begin{bmatrix} \sum_{k=1}^n a_{1k} x_k \\ \sum_{k=1}^n a_{2k} x_k \\ \vdots \\ \sum_{k=1}^n a_{mk} x_k \end{bmatrix}$$

■ Logo

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \sum_{k=1}^n a_{1k} x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{1k} x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{1k} x_k}{\partial x_n} \\ \frac{\partial \sum_{k=1}^n a_{2k} x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{2k} x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{2k} x_k}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{k=1}^n a_{mk} x_k}{\partial x_1} & \frac{\partial \sum_{k=1}^n a_{mk} x_k}{\partial x_2} & \cdots & \frac{\partial \sum_{k=1}^n a_{mk} x_k}{\partial x_n} \end{bmatrix}$$
$$= \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Derivadas matriciais úteis

- Alguns resultados:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}, (\mathbf{A}', \text{ se } \mathbf{A} \text{ for um vetor linha})$$

$$\frac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}', (\mathbf{A}, \text{ se } \mathbf{A} \text{ for um vetor coluna})$$

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}, (2\mathbf{Ax}, \text{ se } \mathbf{A} \text{ for simétrica})$$

Estimação paramétrica (ligação canônica)

- Log-verossimilhança (lembre que $\theta_i = \eta_i = \mathbf{X}'_i\boldsymbol{\beta}$).

$$l(\boldsymbol{\beta}, \phi) = \phi \left[\sum_{i=1}^n y_i \eta_i - \sum_{i=1}^n b(\eta_i) \right] + \sum_{i=1}^n c(y_i, \phi)$$

- Temos que encontrar o vetor de derivadas de $l(., .)$ com relação à $\boldsymbol{\beta}$,

$$\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} \text{ e a derivada com relação à } \phi, S(\phi) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \phi},$$

em que

$$\mathbf{S}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_1} \\ \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_2} \\ \vdots \\ \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_p} \end{bmatrix} \quad (1)$$



- Depois, devemos resolver o sistema de equações

$$\begin{cases} \mathbf{S}(\tilde{\boldsymbol{\beta}}) = \mathbf{0}_{(p \times 1)} \\ S(\tilde{\phi}) = 0 \end{cases}, \quad (2)$$

em que $\mathbf{0}_{(p \times 1)}$ é um vetor de zeros de dimensão $(p \times 1)$.

- Naturalmente, espera-se que a solução encontrada seja ponto de máximo no espaço $(p+1)$ (matriz Hessiana tem de ser negativa definida).
- Calcular $S(\phi)$ é relativamente simples. Para obtermos $\mathbf{S}(\boldsymbol{\beta})$ temos duas opções. Derivar $\mathbf{S}(\boldsymbol{\beta})$ diretamente com relação ao vetor $\boldsymbol{\beta}$ ou obter a derivada com relação à cada parâmetro $\beta_j, j = 1, \dots, p$ e deduzir a forma matricial (equação (1)) dessas derivadas.
- Vamos considerar a segunda opção.

- Temos que:

$$\begin{aligned}\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_j} &= \phi \sum_{i=1}^n \left(y_i \frac{\partial \eta_i}{\partial \beta_j} - \frac{\partial b(\eta_i)}{\partial \beta_j} \right) - \sum_{i=1}^n \frac{\partial c(y_i, \phi)}{\partial \beta_j} \\ &= \phi \sum_{i=1}^n \left(y_i X_{ji} - \frac{\partial b(\eta_i)}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right)\end{aligned}$$

- Lembremos que $E(Y_i) = \mu_i = \frac{\partial b(\eta_i)}{\partial \eta_i}$ (pois nesse caso $\theta_i = \eta_i$).

Além disso, $\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \sum_{s=1}^p X_{si} \beta_s}{\partial \beta_j} = X_{ji}$. Logo:

$$\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_j} = \phi \sum_{i=1}^n X_{ji} (y_i - \mu_i) = \phi \left(\sum_{i=1}^n X_{ji} y_i - \sum_{i=1}^n X_{ji} \mu_i \right)$$

■ Agora note que, se $\mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}$

(matriz de planejamento considerando-se todos os indivíduos) e

$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, então

$$\begin{aligned}
 \mathbf{X}'\mathbf{y} &= [\mathbf{X}'_1 \ \mathbf{X}'_2 \ \dots \ \mathbf{X}'_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
 &= \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n X_{1i}y_i \\ \sum_{i=1}^n X_{2i}y_i \\ \vdots \\ \sum_{i=1}^n X_{ip}y_i \end{bmatrix}
 \end{aligned}$$

Analogamente,

$$\mathbf{X}'\boldsymbol{\mu} = \begin{bmatrix} \sum_{i=1}^n X_{1i}\mu_i \\ \sum_{i=1}^n X_{2i}\mu_i \\ \vdots \\ \sum_{i=1}^n X_{ip}\mu_i \end{bmatrix}$$

em que $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$, $\mu_i = g^{-1}(\mathbf{X}_i'\boldsymbol{\beta})$.

Logo

$$\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}'(\mathbf{y} - \boldsymbol{\mu})$$

- Com relação à ϕ , temos que

$$S(\phi) = \sum_{i=1}^n (y_i \eta_i - b(\eta_i)) + \sum_{i=1}^n c'(y_i, \phi),$$

em que $c'(y_i, \phi) = \frac{\partial c(y_i, \phi)}{\partial \phi}$.

- Não é necessário estimar ϕ para os modelos de Poisson e Binomial. Para os modelos normal e normal inverso a solução é explícita (em função de $\hat{\beta}$) (exercício). Para o modelo gama não há solução explícita e veremos os detalhes específicos pertinentes quando da apresentação desse modelo.

- Podemos notar que o sistema de equações (2) não tem, em geral, solução explícita pois, via de regra, μ_i é uma função não linear de β (quando se considera a função de ligação identidade, é possível obter solução explícita).
- Há várias opções de algoritmos numéricos para resolução de sistemas não lineares: Newton-Raphson, Escore de Fisher (EF), Nelder-Mead, BFGS entre outros.
- Por simplicidade e por apresentar, em geral, um excelente desempenho, consideraremos o algoritmo EF.

- Para isso precisaremos da informação de Fisher (ela também será útil na obtenção de resultados assintóticos):

$$\begin{aligned} \mathbf{I}(\boldsymbol{\beta}, \phi) &= -\mathcal{E}(\mathbf{H}(\boldsymbol{\beta}, \phi)) = \begin{bmatrix} -\mathcal{E}(\mathbf{H}_{11}(\boldsymbol{\beta}, \phi)) & -\mathcal{E}(\mathbf{H}_{12}(\boldsymbol{\beta}, \phi)) \\ -\mathcal{E}(\mathbf{H}_{21}(\boldsymbol{\beta}, \phi)) & -\mathcal{E}(\mathbf{H}_{22}(\boldsymbol{\beta}, \phi)) \end{bmatrix} \\ &= \begin{bmatrix} -\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right) & -\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \phi}\right) \\ -\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi \partial \boldsymbol{\beta}'}\right) & -\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi^2}\right) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{11}(\boldsymbol{\beta}, \phi) & \mathbf{I}_{12}(\boldsymbol{\beta}, \phi) \\ \mathbf{I}_{21}(\boldsymbol{\beta}, \phi) & I_{22}(\boldsymbol{\beta}, \phi) \end{bmatrix} \end{aligned}$$

- A obtenção das componentes matriciais $\mathbf{I}_{11}(\boldsymbol{\beta}, \phi)$ e $\mathbf{I}_{12}(\boldsymbol{\beta}, \phi)$ podem ser feitas de modo análogo ao que foi feito para obtermos $\mathbf{S}(\boldsymbol{\beta})$.
- Note que

$$\mathbf{H}_{11}(\boldsymbol{\beta}, \phi) = \begin{bmatrix} \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_1^2} & \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_1 \partial \beta_p} \\ \cdot & \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_2^2} & \cdots & \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_p^2} \end{bmatrix}$$

$$\mathbf{H}_{12}(\boldsymbol{\beta}, \phi) = \begin{bmatrix} \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_1 \partial \phi} \\ \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_2 \partial \phi} \\ \vdots \\ \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_p \partial \phi} \end{bmatrix}$$

- Por outro lado, temos que

$$\begin{aligned}\frac{l(\boldsymbol{\beta}, \phi)}{\partial \beta_j \partial \beta_r} &= \frac{\partial}{\partial \beta_r} \left(\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_j} \right) = \frac{\partial}{\partial \beta_r} \left[\phi \left(\sum_{i=1}^n X_{ji} y_i - \sum_{i=1}^n X_{ji} \mu_i \right) \right] \\ &= -\phi \sum_{i=1}^n X_{ji} \frac{\partial \mu_i}{\partial \beta_j} = -\phi \sum_{i=1}^n X_{ji} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r} \\ &= -\phi \sum_{i=1}^n X_{ji} \frac{\partial^2 b(\eta_i)}{\partial \eta_i^2} X_{ri} = -\phi \sum_{i=1}^n X_{ji} \frac{\partial \mu_i}{\partial \eta_i} X_{ri}\end{aligned}$$

lembrando que $\mu_i = \frac{\partial b(\eta_i)}{\partial \eta_i}$.

- Assim, vem que:

$$\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\phi \mathbf{X}' \mathbf{V} \mathbf{X}$$

em que $\mathbf{V} = \text{diag} \left(\frac{\partial \mu_1}{\partial \eta_1}, \frac{\partial \mu_2}{\partial \eta_2}, \dots, \frac{\partial \mu_n}{\partial \eta_n} \right)$.

- Além disso, temos que:

$$\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta} \partial \phi} = \frac{\partial}{\partial \phi} [\phi \mathbf{X}' (\mathbf{y} - \boldsymbol{\mu})] = \mathbf{X}' (\mathbf{y} - \boldsymbol{\mu})$$

- Finalmente,

$$\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi^2} = \frac{\partial}{\partial \phi} \left(\sum_{i=1}^n (y_i \eta_i - b(\eta_i)) + \sum_{i=1}^n c'(y_i, \phi) \right) = \sum_{i=1}^n c''(y_i, \phi)$$

em que $c''(y_i, \phi) = \frac{\partial^2 c(y_i, \phi)}{\partial \phi^2}$.

- Portanto, as componentes da informação de Fisher são dadas por:

$$I_{11}(\boldsymbol{\beta}, \phi) = -\mathcal{E}(-\phi \mathbf{X}' \mathbf{V} \mathbf{X}) = \phi \mathbf{X}' \mathbf{V} \mathbf{X}$$

$$I_{12}(\boldsymbol{\beta}, \phi) = -\mathcal{E}(\mathbf{X}'(\mathbf{Y} - \boldsymbol{\mu})) = -\mathbf{X}'(\mathcal{E}(\mathbf{Y}) - \boldsymbol{\mu}) = \mathbf{0}_{(p \times 1)}$$

$$I_{22}(\boldsymbol{\beta}, \phi) = -\mathcal{E} \left(\sum_{i=1}^n c''(Y_i, \phi) \right) = -\sum_{i=1}^n \mathcal{E} [c''(Y_i, \phi)]$$

- O algoritmo Escore de Fisher é definido como:

Sejam $\beta^{(0)}$ e $\phi^{(0)}$ estimativas iniciais de β e ϕ (chute inicial), respectivamente, então faça

$$\begin{aligned}
 \begin{bmatrix} \beta^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix} &= \begin{bmatrix} \beta^{(t)} \\ \phi^{(t)} \end{bmatrix} \\
 &+ \begin{bmatrix} \phi \mathbf{X}' \mathbf{V}^{(t)} \mathbf{X} & \mathbf{0}_{(p \times 1)} \\ \mathbf{0}_{(1 \times p)} & -\mathcal{E} \left(\sum_{i=1} c''(Y_i, \phi^{(t)}) \right) \end{bmatrix}^{-1} \\
 &\times \begin{bmatrix} \mathbf{S}(\beta^{(t)}) \\ S(\phi^{(t)}) \end{bmatrix}, t = 0, 1, 2, \dots
 \end{aligned} \tag{3}$$

até que algum critério de convergência seja satisfeito.

■ Em que

$$\mathbf{S}(\boldsymbol{\beta}^{(t+1)}) = \left(\phi^{(t+1)}\right) \mathbf{X}' \left(\mathbf{y} - \boldsymbol{\mu}^{(t)}\right)$$

$$S\left(\phi^{(t)}\right) = \sum_{i=1}^n \left(y_i \eta_i^{(t)} - b\left(\eta_i^{(t)}\right)\right) + \sum_{i=1}^n c'\left(y_i, \phi^{(t)}\right)$$

$$\boldsymbol{\mu}^{(t)} = \mathbf{g}^{-1}\left(\mathbf{X}\boldsymbol{\beta}^{(t)}\right), \eta_i^{(t)} = \mathbf{X}'_i \boldsymbol{\beta}^{(t)}$$

$$\mathbf{V}^{(t)} = \text{diag}\left(\left(\frac{\partial \mu_1}{\partial \eta_1}\right)^{(t)}, \left(\frac{\partial \mu_2}{\partial \eta_2}\right)^{(t)}, \dots, \left(\frac{\partial \mu_n}{\partial \eta_n}\right)^{(t)}\right)$$

$$\left(\frac{\partial \mu_j}{\partial \eta_j}\right)^{(t)} = \left.\frac{\partial \mu_j}{\partial \eta_j}\right|_{\eta_j = \eta_j^{(t)}}, j = 1, 2, \dots, n$$

Por exemplo, se $\mu_i = \exp(\eta_i)$ então $\left.\frac{\partial \mu_j}{\partial \eta_j}\right|_{\eta_j = \eta_j^{(t)}} = \exp(\eta_j^{(t)})$.

- Como a informação de Fisher é bloco diagonal, temos que
($t = 0, 1, 2, \dots$)

$$\begin{aligned}
 \begin{bmatrix} \boldsymbol{\beta}^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix} \\
 &+ \begin{bmatrix} (\phi^{(t)})^{-1} (\mathbf{X}'\mathbf{V}^{(t)}\mathbf{X})^{-1} (\phi^{(t)}) \mathbf{X}' (\mathbf{y} - \boldsymbol{\mu}^{(t)}) \\ -\frac{\sum_{i=1}^n (y_i \eta_i^{(t)} - b(\eta_i^{(t)})) + \sum_{i=1}^n c'(y_i, \phi^{(t)})}{\sum_{i=1}^n \mathcal{E}(c''(Y_i, \phi^{(t)}))} \end{bmatrix} \\
 &= \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix} \\
 &+ \begin{bmatrix} (\mathbf{X}'\mathbf{V}^{(t)}\mathbf{X})^{-1} \mathbf{X}' (\mathbf{y} - \boldsymbol{\mu}^{(t)}) \\ -\frac{\sum_{i=1}^n (y_i \eta_i^{(t)} - b(\eta_i^{(t)})) + \sum_{i=1}^n c'(y_i, \phi^{(t)})}{\sum_{i=1}^n \mathcal{E}(c''(Y_i, \phi^{(t)}))} \end{bmatrix}
 \end{aligned}$$

- Sendo assim pode-se conduzir o processo iterativo, primeiramente, para β , ou seja,

$$\beta^{(t+1)} = \beta^{(t)} + \left(\mathbf{X}'\mathbf{V}^{(t)}\mathbf{X}\right)^{-1} \mathbf{X}' \left(\mathbf{y} - \boldsymbol{\mu}^{(t)}\right), t = 0, 1, 2, \dots$$

- É possível ainda provar que o processo acima pode ser escrito como

$$\beta^{(t+1)} = \left(\mathbf{X}'\mathbf{V}^{(t)}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{(t)}\mathbf{z}^{(t)}, t = 0, 1, 2, \dots \quad (4)$$

em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = \mathbf{X}\beta$.

- Devido à equação (4), o algoritmo Escore de Fisher para os MLG é chamado também de mínimos quadrados ponderados.

- Depois de obtida uma estimativa para β , digamos $\tilde{\beta}$, estimamos ϕ (caso necessário), através da forma analítica (normal e normal inversa) ou através do processo iterativo:

$$\phi^{(t+1)} = \phi^{(t)} - \frac{\sum_{i=1}^n (y_i \eta_i^{(t)} - b(\eta_i^{(t)})) + \sum_{i=1}^n c'(y_i, \phi^{(t)})}{\sum_{i=1}^n \mathcal{E}(c''(Y_i, \phi^{(t)}))} \quad t = 0, 1, 2, \dots$$

para o modelo gama.

- Estimativas iniciais: $\beta^{(0)} = (\mathbf{X}'\mathbf{V}^{(0)}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{(0)}\mathbf{z}^{(0)}$, com $\eta^{(0)} = g(\mathbf{y})$ ($g(\cdot)$ é a função de ligação), o que implica que $\beta^{(0)} = (\mathbf{X}'\mathbf{V}^{(0)}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{(0)}\eta^{(0)}$. Depois de obtida uma estimativa para β , digamos $\tilde{\beta}$, uma estimativa inicial para ϕ é dada por:

$$\phi^{(0)} = \frac{n-p}{\sum_{i=1}^n \frac{(y_i - \tilde{\mu}_i)^2}{V(\tilde{\mu}_i)}} \quad (\text{estimativa "do tipo " método dos$$

momentos), em que $\tilde{\mu}_i = g^{-1}(\mathbf{X}'_i\tilde{\beta})$.

- Critérios de convergência: $\|\beta^{(t+1)} - \beta^{(t)}\| < \epsilon$ e $|\phi^{(t+1)} - \phi^{(t)}| < \epsilon$, para algum $\epsilon > 0$.

- A justificativa da utilização do estimador $\hat{\phi} = \frac{n-p}{\sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}}$, em que $\hat{\mu}_i = g^{-1}(\mathbf{X}'_i \hat{\boldsymbol{\beta}})$ e $\hat{\boldsymbol{\beta}}$ é o estimador de MV de $\boldsymbol{\beta}$.
- Em outras palavras temos que $\mathcal{V}(Y_i) = \frac{V(\mu_i)}{\phi} \rightarrow \phi^{-1} = \frac{\mathcal{V}(Y_i)}{V(\mu_i)}$.
- Se $\hat{\boldsymbol{\beta}} \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\beta}$, então $\hat{\mu}_i \xrightarrow[n \rightarrow \infty]{P} \mu_i$ e $V(\hat{\mu}_i) \xrightarrow[n \rightarrow \infty]{P} V(\mu_i)$ ($\hat{\mu}_i$ e $V(\hat{\mu}_i)$ são funções contínuas de $\hat{\boldsymbol{\beta}}$).
- Assim, por Slutsky, $\hat{\phi}^{-1} \xrightarrow[n \rightarrow \infty]{P} \phi^{-1}$ e, como $\hat{\phi}$ é uma função contínua de $1/\hat{\phi}$ então $\hat{\phi} \xrightarrow[n \rightarrow \infty]{P} \phi$.
- Esse estimador de ϕ pode ser utilizado ao invés do seu respectivo EMV (exercício, comparar os dois via simulação). Em geral consideraremos o emv.

- Sob as condições de regularidade e para n suficientemente grande, $\hat{\beta}$ e $\hat{\phi}$ (os respectivos estimadores de MV de β e ϕ) são mutuamente independentes e

$$\hat{\beta} \approx N_p \left(\beta, \phi^{-1} (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1} \right); \hat{\phi} \approx N \left(\phi, - \left[\mathcal{E} \left(\sum_{i=1} c''(Y_i, \phi) \right) \right]^{-1} \right).$$

- Se $\hat{\beta}_j$ é a j -ésima componente do vetor $\hat{\beta}$ então $\hat{\beta}_j \approx N(\beta_j, \phi^{-1}\psi_j)$ em que ψ_j é o j -ésimo elemento da diagonal principal de $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$.
- Temos ainda que como $\hat{\beta}$ e $\hat{\phi}$ são estimadores consistentes, então $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\phi}^{-1}\hat{\psi}_j}} \approx N(0, 1)$, em que $\hat{\psi}_j$ é o j -ésimo elemento da diagonal principal de $(\mathbf{X}'\hat{\mathbf{V}}\mathbf{X})^{-1}$.

Inferência

- Analogamente, temos que $\frac{\hat{\phi} - \phi}{\sqrt{\widehat{\mathcal{V}}(\hat{\phi})}} \approx N(0, 1)$, em que

$$\widehat{\mathcal{V}}(\hat{\phi}) = - \left[\mathcal{E} \left(\sum_{i=1} c''(Y_i, \hat{\phi}) \right) \right]^{-1}.$$

- Portanto (considerando-se $P(X \leq z_{\frac{1+\gamma}{2}}) \approx \frac{1+\gamma}{2}$, $X \sim N(0, 1)$), temos que

$$IC(\beta_j, \gamma) \approx \left[\hat{\beta}_j - z_{\frac{1+\gamma}{2}} \sqrt{\hat{\phi}^{-1} \hat{\psi}_j}; \hat{\beta}_j + z_{\frac{1+\gamma}{2}} \sqrt{\hat{\phi}^{-1} \hat{\psi}_j} \right]$$

$$IC(\phi, \gamma) \approx \left[\hat{\phi} - z_{\frac{1+\gamma}{2}} \sqrt{\widehat{\mathcal{V}}(\hat{\phi})}; \hat{\phi} + z_{\frac{1+\gamma}{2}} \sqrt{\widehat{\mathcal{V}}(\hat{\phi})} \right]$$

- Os intervalos de confiança numéricos são obtidos substituindo-se os estimadores pelas respectivas estimativas.

Testes de hipóteses

- Suponha que queremos testar $H_0 : \beta_j = \beta_{j0}$ vs $H_1 : \beta_j \neq \beta_{j0}$, para algum j , em que β_{j0} é um valor fixado.

- Estatística do teste $Z_t = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\hat{\phi}^{-1}\hat{\psi}_j}}$.

- Assim, rejeita-se H_0 se $|z_t| \geq z_c$, em que $z_t = \frac{\tilde{\beta}_j - \beta_{j0}}{\sqrt{\tilde{\phi}^{-1}\tilde{\psi}_j}}$ e $P(X \geq z_c | H_0) \approx \alpha/2$, $X \sim N(0, 1)$, $\tilde{\beta}_j$, $\tilde{\psi}_j$ e $\tilde{\phi}$ são estimativas.

- De modo equivalente, rejeita-se H_0 se p-valor $\leq \alpha$, em que p-valor $\approx 2P(X \geq |z_t| | H_0)$, $X \sim N(0, 1)$.

Testes de hipóteses

- Suponha que queremos testar $H_0 : \phi = \phi_0$ vs $H_1 : \phi \neq \phi_0$.
- Estatística do teste $Z_t = \frac{\hat{\phi} - \phi_0}{\sqrt{\hat{\mathcal{V}}(\hat{\phi})}}$.
- Assim, rejeita-se H_0 se $|z_t| \geq z_c$, em que $z_t = \frac{\tilde{\phi} - \phi_0}{\sqrt{\tilde{\mathcal{V}}(\hat{\phi})}}$ e $P(X \geq z_c | H_0) \approx \alpha/2$, $X \sim N(0, 1)$, $\tilde{\beta}_j$, $\tilde{\psi}_j$ e $\tilde{\phi}$ são estimativas.
- De modo equivalente, rejeita-se H_0 se p-valor $\leq \alpha$, em que p-valor $\approx 2P(X \geq |z_t| | H_0)$, $X \sim N(0, 1)$.