

# Introdução geral aos MLG: parte 1

Prof. Caio Azevedo

# Modelo linear generalizado

- Consideraremos que

$$Y \sim \text{FE}(\theta, \phi),$$

e que temos uma amostra aleatória de  $Y$ , ou seja

$$Y_i \stackrel{\text{ind.}}{\sim} \text{FE}(\theta_i, \phi), i = 1, 2, \dots, n,$$

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 (invertível 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

- Temos, assim, que um MLG (na sua forma mais simples) é dado por:

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

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

# Escolhas naturais para funções de ligação

Distribuição	$g(\mu)$
Normal	$\mu$ (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  $\mu$  a proporção de sucessos de uma binomial ( $\mu \in (0, 1)$ ).
- 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  $\mu$  ( $\mu \in (0, 1)$ ) 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:

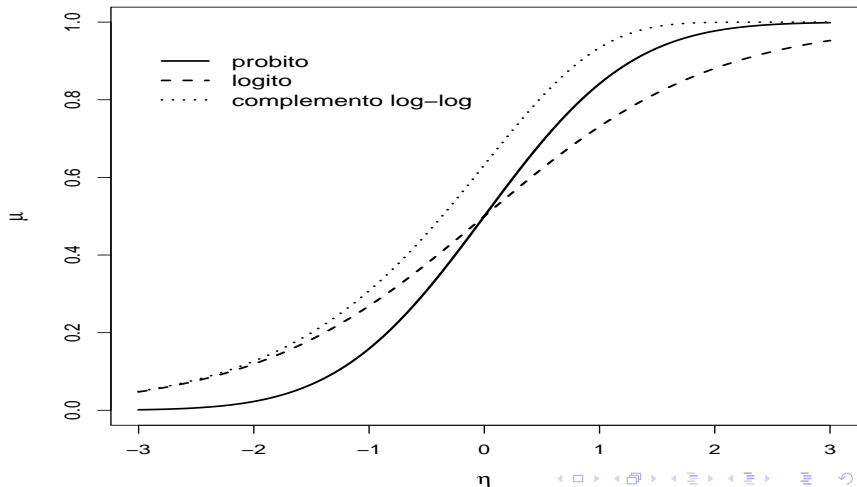
$$F(x) = 1 - \exp\{-\exp(x)\}$$

- 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)





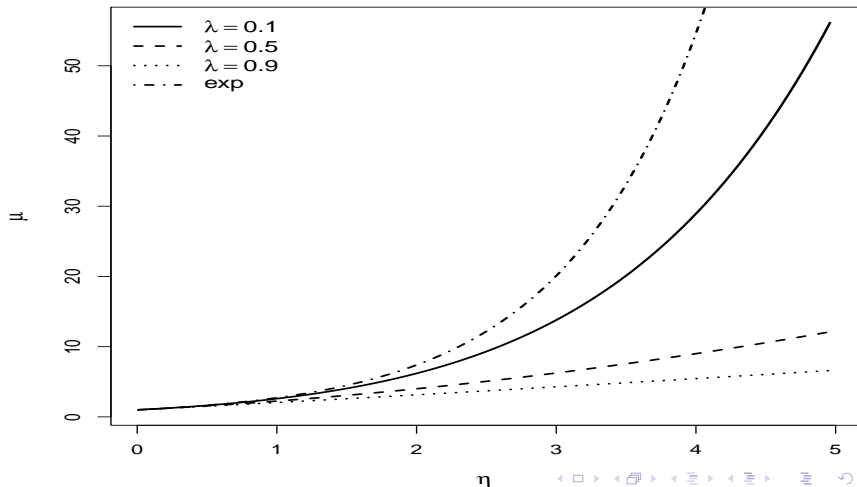
# Funções de ligação Box-Cox

- 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  $\lambda$ , ajustando o modelo para cada um deles, e utilizar algum critério de seleção de modelos para escolher o valor mais apropriado.
- Podemos também estimar  $\lambda$  em concomitância com os outros parâmetros (veja também [aqui](#)).

# Funções de ligação da família Box-Cox



## Resumo: principais funções de ligação

Distribuição	$g(\mu)$
Normal	$\mu, (\text{canônica}), 1/\mu$ e $\ln \mu$ (se $\mu > 0$ )
Poisson	$\ln \mu$ (canônica), $\sqrt{\mu}$
Binomial	$\ln(\mu/(1 - \mu))$ (canônica), $\Phi^{-1}(\mu)$ , $\ln(-\ln(1 - \mu))$
Gama	$1/\mu$ (canônica), $\ln \mu$
N.Inversa	$1/(2\mu^2)$ (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 penalizada, 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:  $\hat{\boldsymbol{\beta}}(\hat{\phi})$  estimador,  $\tilde{\boldsymbol{\beta}}(\tilde{\phi})$  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:

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

# Derivadas matriciais úteis

- Logo

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

# Derivadas matriciais úteis

- Alguns resultados:

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

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

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

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



# Estimação paramétrica (ligação canônica)

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

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

- Temos que encontrar o vetor de derivadas de  $l(., .)$  com relação à  $\boldsymbol{\beta}$ ,  $\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}}$  e a derivada com relação à  $\phi$ ,  $S(\phi) = \frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \phi}$ .

# Estimação paramétrica (ligação canônica)

- (Cont.) Em que:

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

## Cont.

- Depois, devemos resolver o sistema de equações

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

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).

## Cont.

- Calcular  $S(\phi)$  é relativamente simples.
- Para obtermos  $\mathbf{S}(\beta)$  temos duas opções:
  - Derivar  $\mathbf{S}(\beta)$  diretamente com relação ao vetor  $\beta$  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.

## Cont.

- Temos que:

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

## Cont.

- 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:

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

## Cont.

■ 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

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

## Cont.

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



## Cont.

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

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

# Cont

- Com relação à  $\phi$ , 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  $c'(y_i, \phi) = \frac{\partial c(y_i, \phi)}{\partial \phi}$ .

- Observações com relação à estimação de  $\phi$ .
  - Não é necessário estimar  $\phi$  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.

## Cont.

- Podemos notar que o sistema de equações (2) não tem, em geral, solução explícita pois, via de regra,  $\mu_i$  é uma função não linear de  $\beta$  (quando se considera a função de ligação identidade e a distribuição Normal, 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 (aqui).
- Por simplicidade e por apresentar, em geral, um excelente desempenho, consideraremos o algoritmo EF.

## Cont.

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

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

## Cont.

- 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

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

$$\mathbf{H}_{12}(\boldsymbol{\beta}, \phi) = \left[ \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_1 \partial \phi} \quad \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_2 \partial \phi} \quad \cdots \quad \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \beta_p \partial \phi} \right]', \quad \mathbf{H}_{22}(\boldsymbol{\beta}, \phi) = \left[ \frac{\partial^2 I(\boldsymbol{\beta}, \phi)}{\partial \phi^2} \right].$$

## Cont.

- Por outro lado, temos que

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

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

## Cont.

- 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:

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



## Cont.

- 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

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

## Cont.

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

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

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

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

## Cont.

- O algoritmo Escore de Fisher é definido como: Sejam  $\beta^{(0)}$  e  $\phi^{(0)}$  estimativas iniciais de  $\beta$  e  $\phi$  (chute inicial), respectivamente, então faça (até que algum critério de convergência seja satisfeito):

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

## Cont.

- Em que

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

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

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

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

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

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

## Cont.

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

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

## Cont.

- Sendo assim pode-se conduzir o processo iterativo, primeiramente, para  $\beta$ , ou seja,

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

- É possível ainda provar (exercício) 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  $\mathbf{z} = \boldsymbol{\eta} + \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ ,  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ .

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

## Cont.

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

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

para o modelo gama.

## Cont.

- 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

$$\beta^{(0)} = (\mathbf{X}'\mathbf{V}^{(0)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{(0)}\eta^{(0)}.$$



## Cont.

- Depois de obtida uma estimativa para  $\beta$ , digamos  $\tilde{\beta}$ , uma estimativa inicial para  $\phi$  é dada por:

$$\phi^{(0)} = \frac{n - p}{\sum_{i=1}^n \frac{(y_i - \tilde{\mu}_i)^2}{V(\tilde{\mu}_i)}},$$

(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  $\epsilon > 0$ .

## Cont.

- Pode-se, alternativamente ao EMV, utilizar o seguinte estimador para  $\phi$ :

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

em que  $\hat{\mu}_i = g^{-1}(\mathbf{X}_i' \hat{\boldsymbol{\beta}})$  e  $\hat{\boldsymbol{\beta}}$  é o estimador de MV de  $\boldsymbol{\beta}$ .

- A justificativa da utilização do estimador é:

1 Temos que  $\mathcal{V}(Y_i) = \frac{V(\mu_i)}{\phi} \rightarrow \phi^{-1} = \frac{\mathcal{V}(Y_i)}{V(\mu_i)}$ .

2 Se  $\hat{\beta} \xrightarrow[n \rightarrow \infty]{P} \beta$ , 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  $\hat{\beta}$ ).

3 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  $\phi$  pode ser utilizado ao invés do seu respectivo EMV (exercício, comparar os dois via simulação). Em geral consideraremos o emv.

## Cont.

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

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

para  $n$  suficientemente grande.

## Cont.

- Uma outra forma de escrever os resultados anteriores é:

$$\begin{aligned}\sqrt{\widehat{\phi}}\sqrt{n}(\widehat{\beta} - \beta) &\xrightarrow[n \rightarrow \infty]{D} N_p(\mathbf{0}, \Psi_{\widehat{\beta}}), \\ \sqrt{n}(\widehat{\phi} - \phi) &\xrightarrow[n \rightarrow \infty]{D} N(0, \psi_{\widehat{\phi}}),\end{aligned}$$

em que  $\Psi_{\widehat{\beta}} = \lim_{n \rightarrow \infty} \left[ \frac{1}{n} (\mathbf{X}'_n \mathbf{V}_n \mathbf{X}_n)^{-1} \right]$ , é uma matriz positiva

definida  $\psi_{\widehat{\phi}} = - \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \phi) \right) \right]^{-1} \right\} > 0$ .

## Cont.

- (Cont.) Em que:  $\mathbf{X}_n \equiv \mathbf{X}$  e  $\mathbf{V}_n \equiv \mathbf{V}$  (o subíndice  $n$  apenas ressalta que  $\mathbf{X}$  depende de  $n$ ).
- De modo equivalente, o resultados anteriores podem ser escritos como:

$$(\mathbf{X}'_n \mathbf{V}_n \mathbf{X}_n)^{1/2} \sqrt{\widehat{\phi}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[n \rightarrow \infty]{D} N_p(\mathbf{0}, \mathbf{I}_p), \quad (5)$$

$$\left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \phi) \right) \right]^{1/2} (\widehat{\phi} - \phi) \xrightarrow[n \rightarrow \infty]{D} N(0, 1), \quad (6)$$

- em que se  $\mathbf{A}^{1/2}$  é raiz quadrada da matriz  $\mathbf{A}$ , então  $\mathbf{A} = \mathbf{A}^{1/2} \mathbf{A}^{1/2}$ .

## Cont.

- Contudo, como  $\beta$  e  $\phi$  são desconhecidos, os resultados (5) e (6) não são completamente úteis. Por outro lado, temos que:

$$(\mathbf{X}'_n \mathbf{V}_n \mathbf{X}_n)^{-1} (\mathbf{X}'_n \widehat{\mathbf{V}}_n \mathbf{X}_n) \xrightarrow[n \rightarrow \infty]{P} \mathbf{I}_p \quad (7)$$

$$\frac{\left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \phi) \right) \right] \Big|_{\beta=\widehat{\beta}, \phi=\widehat{\phi}}}{\mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \phi) \right)} \xrightarrow[n \rightarrow \infty]{P} 1, \quad (8)$$

## Cont.

- (Cont.) em que:

- $\mathcal{E}(Y, \phi)|_{\beta=\hat{\beta}, \phi=\hat{\phi}}$ , significa primeiro calcular a esperança e, posteriormente, avaliar o respectivo resultado em  $\hat{\beta}$  e  $\hat{\phi}$ , os respectivos EMV de  $\beta$  e  $\phi$ .

- $\hat{\mathbf{V}}_n \equiv \hat{\mathbf{V}} = \text{diag} \left( \left. \frac{\partial \mu_1}{\partial \eta_1} \right|_{\beta=\hat{\beta}}, \left. \frac{\partial \mu_2}{\partial \eta_2} \right|_{\beta=\hat{\beta}}, \dots, \left. \frac{\partial \mu_n}{\partial \eta_n} \right|_{\beta=\hat{\beta}} \right)$ , por exemplo,

$$\text{se } \mu_i = \exp(\eta_i) \text{ então } \left. \frac{\partial \mu_i}{\partial \eta_i} \right|_{\beta=\hat{\beta}} = \exp(\hat{\eta}_i) = \exp \left( \sum_{r=1}^p X_{ji} \hat{\beta}_j \right).$$



cont.

- Assim, usando (7) em (5) e (8) em (6), vem que:

$$\begin{aligned} & \left( \mathbf{X}'_n \widehat{\mathbf{V}}_n \mathbf{X}_n \right)^{1/2} \sqrt{\widehat{\phi}} \left( \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) \xrightarrow[n \rightarrow \infty]{D} N_p(\mathbf{0}, \mathbf{I}_p), \\ & \left\{ \left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \phi) \right) \right] \Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \phi=\widehat{\phi}} \right\}^{1/2} \left( \widehat{\phi} - \phi \right) \xrightarrow[n \rightarrow \infty]{D} N(0, 1). \end{aligned}$$

- Adicionalmente, pelo fato da Informação de Fisher ser bloco diagonal,  $\widehat{\boldsymbol{\beta}} \perp \widehat{\phi}$ , quando  $n \rightarrow \infty$ .
- Essencialmente, os resultados acima dão origem a todos os demais resultados assintóticos (ao menos os principais) relativos aos MLG.

## Cont.

- Para mais detalhes veja : [Sen and Singer \(1994\)](#), [Sen et al \(2009\)](#), [Fahrmeir and Kaufmann \(1985\)](#)
- 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  $\psi_j$  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  $\frac{\hat{\phi} - \phi}{\sqrt{\hat{\mathcal{V}}(\hat{\phi})}} \approx N(0, 1)$ , em que  $\hat{\mathcal{V}}(\hat{\phi}) = - \left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \hat{\phi}) \right) \right]^{-1} \equiv \left[ \mathcal{E} \left( \sum_{i=1}^n c''(Y_i, \hat{\phi}) \right) \Big|_{\beta=\hat{\beta}, \phi=\hat{\phi}} \right]^{-1}$ .
- Portanto (considerando-se  $P(X \leq z_{\frac{1+\gamma}{2}}) \approx \frac{1+\gamma}{2}$ ,  $X \sim N(0, 1)$ ), temos que

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

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

- 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  $\beta_{j0}$  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  $\approx 2P(X \geq |z_t| | H_0), X \sim N(0, 1)$ .

# Condições de regularidade (CR)

- As condições de regularidade mencionadas até o momento, dizem respeito à uma amostra iid (independentes e identicamente distribuída).
- No caso dos MLG a amostra continua sendo ind. (independente mas não é mais identicamente distribuída). Assim, faz-se mister adaptar tais CR ao nosso caso.
- Nos focaremos na CR para o  $\beta$  sendo aquelas para o  $\phi$  definíveis de forma análoga (pesquisar).
- Seja o MLG apresentado nestes slides página 04.

## CR para o MLG

1  $\lim_{n \rightarrow \infty} \left( \frac{1}{n} \mathbf{I}_{n11}(\boldsymbol{\beta}, \phi) \right) = \boldsymbol{\Psi}_{\hat{\boldsymbol{\beta}}}$ , em que  $\boldsymbol{\Psi}_{\hat{\boldsymbol{\beta}}}$  é positiva definida e  $\mathbf{I}_{n11}(\boldsymbol{\beta}, \phi) \equiv \mathbf{I}_{11}(\boldsymbol{\beta}, \phi)$ .

2  $\lim_{n \rightarrow \infty} \left[ \frac{1}{n^2} \sum_{i=1}^n V_i(\mu_i(\boldsymbol{\beta})) \text{tr}(\mathbf{G}_i \mathbf{G}_i') \right] = 0$ , em que  $(V_i(\mu_i(\boldsymbol{\beta}))) \equiv V_i(\boldsymbol{\beta}), \mathbf{X}'_{ni} \equiv \mathbf{X}_i, i = 1, 2, \dots)$

$$\mathbf{G}_i = \left[ \frac{g''(\mu_i(\boldsymbol{\beta}))}{[g''(\mu_i(\boldsymbol{\beta}))]^2} + \frac{b'''(h(\mathbf{X}'_{ni}\boldsymbol{\beta}))}{[g'(\mu_i(\boldsymbol{\beta}))]^2 [V_i(\boldsymbol{\beta})]^3} \right] \mathbf{X}_{ni} \mathbf{X}'_{ni},$$

3 As funções  $[g'(\mu_i(\boldsymbol{\beta}^*))]^{-2}$ ,  $[V_i(\boldsymbol{\beta}^*)]^{-3}$ ,  $g''(\mu_i(\boldsymbol{\beta}^*))$  e  $b'''(h(\mathbf{X}'_{ni}\boldsymbol{\beta}^*))$ ,  $i=1,2,\dots,n$  são uniformemente contínuas no conjunto

$$B(\delta) = \{\boldsymbol{\beta}^* \in \mathcal{R}^p : \|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| < \delta\}, \delta \downarrow 0,$$

## CR para o MLG (cont.)

### 4 Definindo

$$\begin{aligned}w_{1i} &= \{g'(\mu_i(\beta))\}^{-2} [V_i(\beta)]^{-1}, \\w_{2i} &= \mu_i(\beta) \left\{ g''(\mu_i(\beta)) [g'(\mu_i(\beta))]^{-2} \right. \\&\quad \left. + b'''(h(\mathbf{X}'_{ni}\beta)) [g'(\mu_i(\beta)) [V_i(\beta)]^{-3}] \right\}, \text{ então}\end{aligned}$$

1 Para  $k = 1, 2$ , à medida que  $\delta \downarrow 0$  tem-se que

$$\sup_{\beta^* \in B(\delta)} \left[ \frac{1}{n} \sum_{i=1}^n \| [w_{ki}(\beta^*) - w_{ki}(\beta)] \mathbf{X}_{in} \mathbf{X}'_{in} \| \right] \rightarrow 0$$

2 À medida que  $\delta \downarrow 0$ , tem-se que

$$\mathcal{E}_\beta \left\{ \sup_{\beta^* \in B(\delta)} \left[ \frac{1}{n} \sum_{i=1}^n |Y_i| \| [w_{2i}(\beta^*) - w_{2i}(\beta)] \mathbf{X}_{in} \mathbf{X}'_{in} \| \right] \right\} \rightarrow 0$$



## CR (Cont.)

- Sob as condições de 1 a 4.2, temos a validade do resultado (5).
- Para provar a convergência em probabilidade, ou seja, para provar:

1 Que  $\hat{\beta} \xrightarrow[n \rightarrow \infty]{P} \beta$ ,

2 e o resultado (7),

podemos usar o resultado de convergência em distribuição (Equação (5)) e o Teorema de Slutsky.

- Exercício: pesquisar sobre os outros resultados.