# Introdução geral aos MLG: parte 2

Prof. Caio Azevedo

# Estimação paramétrica (função de ligação geral)

- Vamos estender alguns resultados obtidos/apresentados aqui, considerando uma função de ligação que não, necessariamente, é a canônica mas que ainda é inversível e duplamente diferenciável.
- Lembremos que $\theta_i = h(\mu_i)$ e que $\mu_i = g^{-1}(\eta_i)$, em que $\eta_i = \sum_{j=1}^{p} X_{ji}\beta_j$.
- Logo $\theta_i = h(g^{-1}(\eta_i))$.
- Log-verossimilhança

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

# Cont.

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

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

# Cont.

- Depois, devemos resolver o sistema de equações $\left\{ \begin{array}{c} \boldsymbol{S}(\boldsymbol{\beta}) = \mathbf{0}_{(p \times 1)} \\ S(\phi) = 0 \end{array} \right.$ , em que $\mathbf{0}_{(p \times 1)}$ é um vetor de zeros de dimensão $(p \times 1)$.
- Vamos derivar com relação à cada componente do vetor $\boldsymbol{\beta}$, como feito para o caso da função de ligação canônica.

# Cont.

- Vetor escore para $\boldsymbol{\beta}$ (usando a regra da cadeia)

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

# Cont.

- (Cont.) Em que
  - $\dfrac{\partial \mu_i}{\partial \theta_i} = V(\mu_i) \equiv V_i \rightarrow \dfrac{\partial \theta_i}{\partial \mu_i} = V_i^{-1}$,
  - $E(Y_i) = \mu_i = \dfrac{\partial b(\theta_i)}{\partial \theta_i}$,
  - $\dfrac{\partial \eta_i}{\partial \beta_j} = \dfrac{\partial \sum_{s=1}^{p} X_{si}\beta_s}{\partial \beta_j} = X_{ji}$, $\omega_i = \dfrac{(\partial \mu_i / \partial \eta_i)^2}{V_i}$.
  - Obs: sob a função de ligação canônica, $\omega_i = V_i$.

# Cont.

- Assim, o vetor escore, definido pela concatenação de cada uma das $p$ derivadas anteriores, é dado em sua forma matricial por:

$$S(\boldsymbol{\beta}) = \phi \boldsymbol{X}' \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}),$$

em que $\boldsymbol{W} = \text{diag}(\omega_1, ..., \omega_n)$, $\boldsymbol{V} = \text{diag}(V_1, ..., V_n)$,

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix}$$ (matriz de planejamento considerando-se todos os indivíduos), $\boldsymbol{y} = (y_1, y_2, ..., y_n)'$ e $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_n)'$.

# Cont.

- Função escore para $\phi$

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

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

- Logo

$$\boldsymbol{S}(\boldsymbol{\beta}, \phi) = \left[ \begin{array}{c} \phi \boldsymbol{X}' \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}) \\ \sum_{i=1}^{n} \{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c'(y_i, \phi) \end{array} \right].$$

# Informação de Fisher

$$
\begin{aligned}
\boldsymbol{I}(\boldsymbol{\beta},\phi) &= -\mathcal{E}(\boldsymbol{H}(\boldsymbol{\beta},\phi)) = \left[\begin{array}{cc} -\mathcal{E}\left(\boldsymbol{H}_{11}(\boldsymbol{\beta},\phi)\right) & -\mathcal{E}\left(\boldsymbol{H}_{12}(\boldsymbol{\beta},\phi)\right) \\ -\mathcal{E}\left(\boldsymbol{H}_{21}(\boldsymbol{\beta},\phi)\right) & -\mathcal{E}\left(H_{22}(\boldsymbol{\beta},\phi)\right) \end{array}\right] \\
&= \left[\begin{array}{cc} -\mathcal{E}\left(\dfrac{\partial^2 l(\boldsymbol{\beta},\phi)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) & -\mathcal{E}\left(\dfrac{\partial^2 l(\boldsymbol{\beta},\phi)}{\partial\boldsymbol{\beta}\partial\phi}\right) \\ -\mathcal{E}\left(\dfrac{\partial^2 l(\boldsymbol{\beta},\phi)}{\partial\phi\partial\boldsymbol{\beta}'}\right) & -\mathcal{E}\left(\dfrac{\partial^2 l(\boldsymbol{\beta},\phi)}{\partial\phi^2}\right) \end{array}\right] \\
&= \left[\begin{array}{cc} \boldsymbol{I}_{11}(\boldsymbol{\beta},\phi) & \boldsymbol{I}_{12}(\boldsymbol{\beta},\phi) \\ \boldsymbol{I}_{21}(\boldsymbol{\beta},\phi) & I_{22}(\boldsymbol{\beta},\phi) \end{array}\right].
\end{aligned}
$$

# Cont.

- Componente de ordem $(j,r)$ de $\boldsymbol{I}_{11}(\boldsymbol{\beta},\boldsymbol{\beta})$ (Informação de Fisher).

$$
\begin{aligned}
-\mathcal{E}\left(\frac{\partial l(\boldsymbol{\beta},\phi)}{\partial\beta_j\partial\beta_r}\right) &= -\mathcal{E}\left\{\phi\sum_{i=1}^{n}\left[\frac{1}{V_i^{1/2}}(Y_i-\mu_i)X_{ji}\frac{\partial\omega_i^{1/2}}{\partial\beta_r}\right.\right.\\
&+ \left.\left.\omega_i^{1/2}(Y_i-\mu_i)X_{ji}\frac{\partial V_i^{-1/2}}{\partial\beta_r}-\sqrt{\frac{\omega_i}{V_i}}X_{ji}\frac{\partial\mu_i}{\partial\beta_r}\right]\right\}\\
&= -\left\{\phi\sum_{i=1}^{n}\left[\frac{1}{V_i^{1/2}}\underbrace{\mathcal{E}(Y_i-\mu_i)}_{0}X_{ji}\frac{\partial\omega_i^{1/2}}{\partial\beta_r}\right.\right.\\
&+ \left.\left.\omega_i^{1/2}\underbrace{\mathcal{E}(Y_i-\mu_i)}_{0}X_{ji}\frac{dV_i^{-1/2}}{\partial\beta_r}-\sqrt{\frac{\omega_i}{V_i}}X_{ji}\frac{\partial\mu_i}{\partial\beta_r}\right]\right\}\\
&= \phi\sum_{i=1}^{n}\sqrt{\frac{\omega_i}{V_i}}X_{ji}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_r}.
\end{aligned}
$$

# Cont.

- Como $\omega_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 / V_i$, então $\frac{\partial \mu_i}{\partial \eta_i} = \sqrt{V_i \omega_i}$. Portanto:

$$
\begin{aligned}
-\mathcal{E}\left(\frac{\partial l(\boldsymbol{\beta},\phi)}{\partial \beta_j \partial \beta_r}\right) &= \phi \sum_{i=1}^{n} \sqrt{\frac{\omega_i}{V_i}} X_{ji} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_r} \\
&= \phi \sum_{i=1}^{n} \omega_i X_{ji} X_{ri}.
\end{aligned}
$$

- Matricialmente,

$$
\boldsymbol{I}_{11}(\boldsymbol{\beta},\boldsymbol{\beta}) = \phi \boldsymbol{X}' \boldsymbol{W} \boldsymbol{X}.
$$

# Cont.

- Componente $\boldsymbol{I}_{12}(\boldsymbol{\beta}, \phi)$ (Informação de Fisher):

$$
\begin{aligned}
-\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \beta \partial \phi}\right) &= -\mathcal{E}\left\{\frac{\partial}{\partial \phi}\left(\phi \boldsymbol{X}' \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu})\right)\right\} \\
&= -\boldsymbol{X}' \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} \mathcal{E}\left(\boldsymbol{Y} - \boldsymbol{\mu}\right) = \boldsymbol{0}.
\end{aligned}
$$

- Informação de Fisher para $\phi$:

$$
-\mathcal{E}\left(\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi^2}\right) = -\mathcal{E}\left(\sum_{i=1}^{n} c''(Y_i, \phi)\right)
$$

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

## Cont.

- Assim, temos que a Informação de Fisher é dada por:

$$\boldsymbol{I}(\boldsymbol{\beta},\phi) = \left[\begin{array}{cc} \phi \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{0}_{(p\times 1)} \\ \boldsymbol{0}_{(1\times p)} & -\mathcal{E}\left(\sum_{i=1} c''(Y_i,\phi)\right) \end{array}\right].$$

- O algoritmo Escore de Fisher (AEF) é definido como: Sejam $\boldsymbol{\beta}^{(0)}$ e $\phi^{(0)}$ estimativas iniciais de $\boldsymbol{\beta}$ e $\phi$ (chutes iniciais), respectivamente, então faça ($t = 0, 1, 2, ....$, até que algum critério de convergência seja satisfeito.) (próximo slide)

# Cont.

- (Cont.):

$$
\begin{bmatrix} \boldsymbol{\beta}^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix}
+ \begin{bmatrix} \phi^{(t)} \boldsymbol{X}' \boldsymbol{W}^{(t)} \boldsymbol{X} & \mathbf{0}_{(p \times 1)} \\ \mathbf{0}_{(1 \times p)} & -\mathcal{E}\left(\sum_{i=1} c''(Y_i, \phi^{(t)})\right) \end{bmatrix}^{-1}
\times \begin{bmatrix} \phi^{(t)} \boldsymbol{X}' \left(\boldsymbol{W}^{(t)}\right)^{1/2} \left(\boldsymbol{V}^{(t)}\right)^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}^{(t)}) \\ \sum_{i=1}^{n} \left\{ y_i \theta_i^{(t)} - b(\theta_i^{(t)}) \right\} + \sum_{i=1}^{n} c'(y_i, \phi^{(t)}) \end{bmatrix}.
$$

# Cont.)

- (Cont.) em que:

$$
\begin{aligned}
\boldsymbol{\mu}^{(t)} &= g^{-1}\left(\boldsymbol{\eta}^{(t)}\right), \eta^{(t)} = \boldsymbol{X}\beta^{(t)}, \theta_i^{(t)} = h(\mu_i^{(t)}), \\
\boldsymbol{V}^{(t)} &= \mathrm{diag}(V_1^{(t)}, V_2^{(t)}, ..., V_n^{(t)}) \\
&= \mathrm{diag}\left(\left(\frac{\partial \mu_1}{\partial \theta_1}\right)^{(t)}, \left(\frac{\partial \mu_2}{\partial \theta_2}\right)^{(t)}, ..., \left(\frac{\partial \mu_n}{\partial \theta_n}^{(t)}\right)\right).
\end{aligned}
$$

- (Cont.) e:

$$\begin{aligned} \boldsymbol{W}^{(t)} &= \text{diag}\left[\left(\left(\frac{\partial \mu_1}{\partial \eta_1}\right)^{(t)}\right)^2 / V_1^{(t)}, \left(\left(\frac{\partial \mu_2}{\partial \eta_2}\right)^{(t)}\right)^2 / V_2^{(t)} \right. \\ &, ..., \left. \left(\left(\frac{\partial \mu_n}{\partial \eta_n}\right)^{(t)}\right)^2 / V_n^{(t)}\right]. \\ \left(\frac{\partial \mu_j}{\partial \theta_j}\right)^{(t)} &= \left.\frac{\partial \mu_j}{\partial \theta_j}\right|_{\theta_j = \theta_j^{(t)}}, \left(\frac{\partial \mu_j}{\partial \eta_j}\right)^{(t)} = \left.\frac{\partial \mu_j}{\partial \eta_j}\right|_{\eta_j = \eta_j^{(t)}}, j = 1, 2, ..., n \end{aligned}$$

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

Analogamente para $\left(\frac{\partial \mu_j}{\partial \theta_j}\right)^{(t)}$.

# Cont.

- Como a informação de Fisher é bloco diagonal, temos que (AEF) ($t = 0, 1, 2, ...$, até que algum critério de parada seja alcançado)

$$
\begin{bmatrix} \boldsymbol{\beta}^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix}
+ \begin{bmatrix} \left(\phi^{(t)}\right)^{-1} \left(\boldsymbol{X}'\boldsymbol{W}^{(t)}\boldsymbol{X}\right)^{-1} \phi^{(t)}\boldsymbol{X}' \left(\boldsymbol{W}^{(t)}\right)^{1/2} \left(\boldsymbol{V}^{(t)}\right)^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}^{(t)}) \\ -\dfrac{\sum_{i=1}^{n}(y_i\theta_i^{(t)} - b(\theta_i^{(t)})) + \sum_{i=1}^{n} c'(y_i, \phi^{(t)})}{\sum_{i=1}^{n} \mathcal{E}\left(c''(Y_i, \phi^{(t)})\right)} \end{bmatrix},
$$

# Cont.

- (Cont.)

$$
\rightarrow \begin{bmatrix} \boldsymbol{\beta}^{(t+1)} \\ \phi^{(t+1)} \end{bmatrix}
$$

$$
= \begin{bmatrix} \boldsymbol{\beta}^{(t)} \\ \phi^{(t)} \end{bmatrix} + \begin{bmatrix} \left(\boldsymbol{X}'\boldsymbol{W}^{(t)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{W}^{(t)}\right)^{1/2}\left(\boldsymbol{V}^{(t)}\right)^{-1/2}(\boldsymbol{y}-\boldsymbol{\mu}^{(t)}) \\ -\dfrac{\sum_{i=1}^{n}(y_i\theta_i^{(t)} - b(\theta_i^{(t)})) + \sum_{i=1}^{n} c'(y_i, \phi^{(t)})}{\sum_{i=1}^{n}\mathcal{E}\left(c''(Y_i, \phi^{(t)})\right)} \end{bmatrix}.
$$

# Cont.

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

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

t=0,1,2,...

# Cont.

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

$$\boldsymbol{\beta}^{(t+1)} = \left(\boldsymbol{X}'\boldsymbol{W}^{(t)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}^{(t)}\boldsymbol{z}^{(t)}, t = 0, 1, 2, ... \qquad (1)$$

em que $\boldsymbol{z} = \boldsymbol{\eta} + \boldsymbol{W}^{-1/2}\boldsymbol{V}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$.

- Devido à equação (1), o algoritmo Escore de Fisher para os MLG é chamado também de mínimos quadrados reponderados.

# Cont.

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

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

para o modelo gama.

# Inferência

- Os comentários feitos anteriormente (considerando-se a função de ligação canônica) para os chutes iniciais, critério de convergência e estimação do $\phi$ continuam valendo, utilizando agora o fato de que $\theta_i = h(g^{-1}(\eta_i))$.
- As distribuições assintóticas de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\phi}$, construção de intervalos de confiança e testes de hipótese são equivalentes àquelas apresentadas anteriormente (quando se considerou a função de ligação canônica) utilizando a matriz de covariâncias apropriada para $\widehat{\boldsymbol{\beta}}$ (para $\widehat{\phi}$ tal matriz não muda), ou seja:

$$\mathcal{V}(\widehat{\boldsymbol{\beta}}) \approx \phi^{-1} \left( \boldsymbol{X}' \boldsymbol{W} \boldsymbol{X} \right)^{-1}.$$

# Mais sobre inferência

- Hipóteses do tipo $H_0 : \boldsymbol{C\beta} = \boldsymbol{M}$ vs $H_1 : \boldsymbol{C\beta} \neq \boldsymbol{M}$ podem ser testadas através da estatística (do tipo Wald)

$$Q_t = \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right),$$

em que $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) = \widehat{\phi}^{-1}\left(\boldsymbol{X}'\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}$, e $\widehat{\boldsymbol{W}}$ corresponde à matriz $\boldsymbol{W}$ na qual as quantidades desconhecidas são substituídas pelos respectivos estimadores (em geral, de MV, os quais são consistentes).

# Mais sobre inferência

- Note que $\widehat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, \phi \left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1})$, portanto, devido à esse resultado e algumas propriedades da normal multivariada (aqui e aqui), temos que se

$$Q_t^* = \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\mathcal{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{M}\right),$$

então sob $H_0$ e para $n$ suficientemente grande, $Q_t^* \approx \chi_c^2$, em que $c$ é o número de linhas de $\boldsymbol{C}$.

# Mais sobre inferência

- Além disso, $\widehat{\boldsymbol{W}} \xrightarrow[n \to \infty]{P} \boldsymbol{W}$ o que implica que $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}}) \xrightarrow[n \to \infty]{P} \mathcal{V}(\widehat{\boldsymbol{\beta}})$, pois cada componente de $\widehat{\mathcal{V}}(\widehat{\boldsymbol{\beta}})$ é uma função contínua da respectiva componente de $\widehat{\boldsymbol{W}}$ e também por Slutsky, pois $\widehat{\phi} \xrightarrow[n \to \infty]{P} \phi$.

# Mais sobre inferência

- Portanto, sob $H_0$ e para $n$ suficientemente grande, pelos resultados anteriores $(Q_t^*, \widehat{\boldsymbol{W}})$, temos que:

$$Q_t \approx \chi^2_c.$$

- Assim, rejeita-se $H_0$ se $p - valor \leq \alpha$, em que $p - valor \approx P(X \geq q_t | H_0)$, em que $X \sim \chi^2_c$

$$q_t = \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right)' \left(\boldsymbol{C}\widetilde{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}'\right)^{-1} \left(\boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M}\right).$$

# Mais sobre inferência

- Sob $H_1$, temos que $Q_t \approx \chi^2_{(c,\delta)}$ (qui-quadrado não central com $c$ graus de liberdade e parâmetro de não centralidade $\delta$), em que

$$\delta = (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M})' \left( \boldsymbol{C}\mathcal{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}' \right)^{-1} (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M}).$$

- Uma estimativa de $\delta$ é dada por

$$\widetilde{\delta} = \left( \boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M} \right)' \left( \boldsymbol{C}\widetilde{\mathcal{V}}(\widehat{\boldsymbol{\beta}})\boldsymbol{C}' \right)^{-1} \left( \boldsymbol{C}\widetilde{\boldsymbol{\beta}} - \boldsymbol{M} \right).$$

# Cont.

- Resumidamente, temos que:
  - $\widehat{\boldsymbol{\beta}} \xrightarrow[n\to\infty]{P} \boldsymbol{\beta}$, $\widehat{\phi} \xrightarrow[n\to\infty]{P} \phi$.
  - Sob $H_0(\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{M})$, $Q_t \xrightarrow[n\to\infty']{P} \chi^2_c$.
  - $\left(\boldsymbol{X}_n'\widehat{\boldsymbol{W}}_n\boldsymbol{X}_n\right)^{1/2}\sqrt{\widehat{\phi}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow[n\to\infty]{D} N_p\left(\boldsymbol{0}, \boldsymbol{I}_p\right)$, em que $\widehat{\boldsymbol{W}}_n \equiv \widehat{\boldsymbol{W}}$,
  - $\left\{\left[\mathcal{E}\left(\sum_{i=1}^{n} c''(Y_i,\phi)\right)\right]\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}},\phi=\widehat{\phi}}\right\}^{1/2}\left(\widehat{\phi} - \phi\right) \xrightarrow[n\to\infty]{D} N\left(0,1\right)$.
- Mais detalhes podem ser obtidos aqui,Sen and Singer (1994), Sen et al (2009), Fahrmeir and Kaufmann (1985) .

# Desvio (ou função desvio)

- Tal função é útil para termos uma idéia da adequabilidade do modelo (verificação da qualidade do ajuste).
- Sem perda de generalidade, sejam:
  - $l(\boldsymbol{\mu}, \boldsymbol{y}) \equiv l(\boldsymbol{\beta}, \phi)$ a logverossimilhança associada ao modelo em estudo e
  - $l(\boldsymbol{\mu}^0, \boldsymbol{y})$ a logverossimilhança do modelo saturado (n=p), ou seja, em que cada média é representada por ela mesma.
- Para o modelo saturado o emv de cada $\mu_i$ é dado por $\widehat{\mu}_i = y_i$ (exercício).

# Cont.

- Defina
  - Modelo em estudo:

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

  - Modelo saturado:

$$l(\boldsymbol{\mu}^{(0)}, \boldsymbol{y}) = \sum_{i=1}^{n} l(\mu_i^{(0)}, y_i) = \phi\left[\sum_{i=1}^{n} y_i\theta_i^{(0)} - \sum_{i=1}^{n} b(\theta_i^{(0)})\right] + \sum_{i=1}^{n} c(y_i, \phi).$$

em que $\theta_i^{(0)} = h(\mu_i^{(0)})$.

# Cont.

- Assim, o desvio escalonado e não escalonado, respectivamente, por $D^*(\boldsymbol{y}, \boldsymbol{\mu}) = \phi D(\boldsymbol{y}, \boldsymbol{\mu})$ e $D(\boldsymbol{y}, \boldsymbol{\mu})$, em que:

$$
\begin{aligned}
D(\boldsymbol{y}, \boldsymbol{\mu}) &= \frac{2}{\phi}(l(\boldsymbol{\mu}^{(0)}, \boldsymbol{y}) - l(\boldsymbol{\mu}, \boldsymbol{y})) = \frac{2}{\phi}\sum_{i=1}^{n}\left(l(\mu_i^{(0)}, y_i) - l(\mu_i, y_i)\right) \\
&= 2\sum_{i=1}^{n} D(y_i, \mu_i) = 2\sum_{i=1}^{n}\left[y_i\left(\theta_i^{(0)} - \theta_i\right) + b(\theta_i) - b(\theta_i^{(0)})\right],
\end{aligned}
$$

em que $D(y_i, \mu_i)$ é chamado de componente do desvio não escalonado.

# Desvio (ou função desvio)

- Sejam $\widehat{\boldsymbol{\mu}}^{(0)} = \boldsymbol{Y}$ e $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}}) = g^{-1}(\boldsymbol{X}\widehat{\boldsymbol{\beta}})$ os respectivos estimadores de MV e defina $\widehat{\theta}_i^{(0)} = h(\widehat{\mu}_i^{(0)}) = h(Y_i)$ e $\widehat{\theta}_i = h(\widehat{\mu}_i)$.
- Portanto, o desvio não escalonado é estimado por

$$D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left[y_i\left(\widehat{\theta}_i^{(0)} - \widehat{\theta}_i\right) + b(\widehat{\theta}_i) - b(\widehat{\theta}_i^{(0)})\right].$$

- O desvio escalonado é dado por $D^*(\boldsymbol{y}, \boldsymbol{\mu}) = \phi D(\boldsymbol{y}, \boldsymbol{\mu})$ e estimado por $D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \widehat{\phi} D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})$, em que $\widehat{\phi}$ é algum estimador consistente.

# Desvio (ou função desvio)

- De uma forma grosseira, o desvio lembra a estatística do teste da razão de verossimilhanças (TRV) para testar $H_0 : \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ vs $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}$.
- As respectivas estimativas, $D(\boldsymbol{y}, \widetilde{\boldsymbol{\mu}})$ e $D^*(\boldsymbol{y}, \widetilde{\mu})$ são obtidas substituindo-se os estimadores pelas respectivas estimativas e as variáveis aleatórias pelos seus respectivos valores observados.
- Exemplo: Bernoulli. Temos que

$$l(\boldsymbol{\mu}, \boldsymbol{y}) = \sum_{i=1}^{n} [y_i \ln(\mu_i) + (1 - y_i) \ln(1 - \mu_i)] = \sum_{i=1}^{n} D(y_i, \mu_i).$$

# Desvio (ou função desvio)

- Logo, para o modelo saturado, vem que

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\mu}^{(0)}, \boldsymbol{y})}{\partial \mu_i^{(0)}} &= \frac{\partial}{\partial \mu_i^{(0)}} \sum_{i=1}^{n} \left[ y_i \ln(\mu_i^{(0)}) + (1 - y_i) \ln(1 - \mu_i^{(0)}) \right] \\
&= \frac{y_i}{\mu_i^{(0)}} - \frac{1 - y_i}{1 - \mu_i^{(0)}} \rightarrow \widetilde{\mu}_i^{(0)} = y_i.
\end{aligned}
$$

- Dessa forma, se $y_i = 0$, $l(\mu_i^{(0)}, y_i) = \ln(1 - \mu_i^{(0)})$ e se $y_i = 1$, $l(\mu_i^{(0)}, y_i) = \ln(\mu_i^{(0)})$. Logo, $l(\widetilde{\mu}_i^{(0)}, 0) = l(\widetilde{\mu}_i^{(0)}, 1) = 0$.

# Desvio (ou função desvio)

- Portanto, se $y_i = 0$, $l(\widetilde{\mu}_i^{(0)}, y_i) - l(\widetilde{\mu}_i, y_i) = -\ln(1 - \widetilde{\mu}_i)$ e, se $y_i = 1$, $l(\widetilde{\mu}_i^{(0)}, y_i) - l(\widetilde{\mu}_i, y_i) = -\ln(\widetilde{\mu}_i)$.
- Segue-se que a estimativa e o estimador do desvio, nesse caso, são, dados, respectivamente, por:

$$
\begin{aligned}
D(\boldsymbol{y}, \widetilde{\boldsymbol{\mu}}) &= -2\sum_{i=1}^{n} \left\{\ln(1 - \widetilde{\mu}_i)\mathbb{1}_{\{0\}}(y_i) + \ln(\widetilde{\mu}_i)\mathbb{1}_{\{1\}}(y_i)\right] , \\
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= -2\sum_{i=1}^{n} \left\{\ln(1 - \widehat{\mu}_i)\mathbb{1}_{\{0\}}(y_i) + \ln(\widehat{\mu}_i)\mathbb{1}_{\{1\}}(y_i)\right] .
\end{aligned}
$$

# Desvio: outros exemplos

- Normal: $D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2$ e $D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \frac{1}{\widehat{\sigma}^2} \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2$.
- Gama: $D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^{n} \{-\ln(y_i/\widehat{\mu}_i) + (y_i - \widehat{\mu}_i)/\widehat{\mu}_i\}$ e $D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \widehat{\phi} D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})$.

# Desvio: outros exemplos

- Binomial:

$$
\begin{aligned}
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) &= 2\sum_{i=1}^{n}\left\{\left[y_i \ln[y_i/(m_i\widehat{\mu}_i)]\right.\right. \\
&+ \left.(m_i - y_i)\ln\left[(1 - y_i/m_i)/(1 - \widehat{\mu}_i)\right]\right] \times \mathbb{1}_{\{1,\ldots,(m_i-1)\}}(y_i) \\
&- \left. 2[m_i \ln(1 - \widehat{\mu}_i)]\mathbb{1}_{\{0\}}(y_i) - 2[m_i \ln \widehat{\mu}_i]\mathbb{1}_{\{m_i\}}(y_i)\right\}.
\end{aligned}
$$

- Poisson:

$$
D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left\{[y_i \ln(y_i/\widehat{\mu}_i) - (y_i - \widehat{\mu}_i)]\, I_{\{1,2,\ldots\}}(y_i) + \widehat{\mu}_i \mathbb{1}_{\{0\}}(y_i)\right\}.
$$

# Desvio: comportamento assintótico

- Em geral $D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})$ ou $D^{*}(\boldsymbol{y},\widehat{\boldsymbol{\mu}})$ não seguem (mesmo assintoticamente) uma distribuição $\chi^2_{(n-p)}$, sob a hipótese de que o modelo em questão é adequado, como se esperaria de um TRV usual (veja aqui e aqui).
- Tal convergência ($D(\boldsymbol{Y};\widehat{\boldsymbol{\mu}}) \xrightarrow{D} \chi^2_{(n-p)}$) ocorre (sob a hipótese de que o modelo em questão é adequado) e se alguma das seguintes condições ocorrerem (próximo slide):

# Desvio: comportamento assintótico

- Binomial: se $m_i \to \infty, \forall i$ e $n$ (tamanho da amostra) é fixo. Assim, em geral, para o modelo Bernoulli ($m_i = 1, \forall i$), tal resultado não é válido.
- Poisson: se $\mu_i \to \infty, \forall i$.
- Nos casos em que $D^*(\boldsymbol{Y}, \widehat{\boldsymbol{\mu}})$ depende do parâmetro de precisão, se $\phi \to \infty$.
- Obs: além das referências mencionadas, também sugere-se essa referência.