

Introdução geral aos MLG: parte 2

Prof. Caio Azevedo

Estimação paramétrica (função de ligação geral)

- Lembremos que $\theta_i = h(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ em que $\eta_i = \sum_{j=1}^p x_{ji}\beta_j$.
- Logo $\theta_i = h(g^{-1}(\eta_i))$.
- Log-verossimilhança

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

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

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

- Depois, devemos resolver o sistema de equações $\begin{cases} \mathbf{S}(\beta) = \mathbf{0}_{(p \times 1)} \\ S(\phi) = 0 \end{cases}$, 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 β , como feito para o caso da função de ligação canônica.

- Vetor escore para β (usando a regra da cadeia)

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

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

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

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

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

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \quad (\text{matriz de planejamento$$

considerando-se todos os indivíduos), $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ e

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'.$$

- Função escore para ϕ

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

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

- Logo

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

Informação de Fisher

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

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

$$\begin{aligned}
 -\mathcal{E} \left(\frac{\partial l(\beta, \phi)}{\partial \beta_j \partial \beta_r} \right) &= -\mathcal{E} \left\{ \phi \left[\sum_{i=1}^n \frac{1}{V_i^{1/2}} (Y_i - \mu_i) X_{ji} \frac{\partial \omega_i^{1/2}}{\partial \beta_r} \right. \right. \\
 &+ \left. \left. \omega_i^{1/2} (Y_i - \mu_i) X_{ji} \frac{\partial V_i^{-1/2}}{\partial \beta_r} - \sqrt{\frac{\omega_i}{V_i}} X_{ji} \frac{\partial \mu_i}{\partial \beta_r} \right] \right\} \\
 &= -\left\{ \phi \left[\sum_{i=1}^n \frac{1}{V_i^{1/2}} \underbrace{\mathcal{E}(Y_i - \mu_i)}_0 X_{ji} \frac{\partial \omega_i^{1/2}}{\partial \beta_r} \right. \right. \\
 &+ \left. \left. \omega_i^{1/2} \underbrace{\mathcal{E}(Y_i - \mu_i)}_0 X_{ji} \frac{\partial V_i^{-1/2}}{\partial \beta_r} - \sqrt{\frac{\omega_i}{V_i}} X_{ji} \frac{\partial \mu_i}{\partial \beta_r} \right] \right\} \\
 &= \phi \sum_{i=1}^n \sqrt{\frac{\omega_i}{V_i}} X_{ji} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r}
 \end{aligned}$$

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

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

- Matricialmente,

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

- Componente $I_{12}(\beta, \phi)$ (Informação de Fisher).

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

- Informação de Fisher para ϕ

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

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

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

$$I(\beta, \phi) = \begin{bmatrix} \phi \mathbf{X}' \mathbf{W} \mathbf{X} & \mathbf{0}_{(p \times 1)} \\ \mathbf{0}_{(1 \times p)} & -\mathcal{E}(\sum_{i=1} c''(Y_i, \phi)) \end{bmatrix}$$

- O algoritmo Escore de Fisher (AEF) é definido como:

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

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

$t = 0, 1, 2, \dots$, até que algum critério de convergência seja satisfeito.

■ Em que

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

■ E

$$\mathbf{W}^{(t)} = \text{diag} \left(\left(\left(\frac{\partial \mu_1}{\partial \eta_1} \right)^{(t)} \right)^2 / V_1^{(t)}, \left(\left(\frac{\partial \mu_2}{\partial \eta_2} \right)^{(t)} \right)^2 / V_2^{(t)}, \dots, \left(\left(\frac{\partial \mu_n}{\partial \eta_n} \right)^{(t)} \right)^2 / V_n^{(t)} \right)$$

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

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

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

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

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

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

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

$t=0,1,2,\dots$

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

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

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

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

- 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:

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

para o modelo gama.

Inferência

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

$$V(\hat{\beta}) \approx \phi^{-1} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1}$$

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, seja $l(\boldsymbol{\mu}, \mathbf{y}) \equiv l(\boldsymbol{\beta}, \phi)$ a logverossimilhança associada ao modelo em estudo e $l(\boldsymbol{\mu}^0, \mathbf{y})$ 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 μ_i é dado por $\hat{\mu}_i = y_i$.

■ Defina

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

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

■ Assim, o desvio escalonado e não escalonado, respectivamente, por

$D^*(\mathbf{y}, \boldsymbol{\mu}) = \phi D(\mathbf{y}, \boldsymbol{\mu})$ e $D(\mathbf{y}, \boldsymbol{\mu})$, em que:

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

Desvio (ou função desvio)

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

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

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

Desvio (ou função desvio)

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

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

Desvio (ou função desvio)

- Logo, para o modelo saturado, vem que

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

- 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, $l(\tilde{\mu}_i^{(0)}, 0) = l(\tilde{\mu}_i^{(0)}, 1) = 0$.

Desvio (ou função desvio)

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

$$D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = -2 \sum_{i=1}^n \{ \ln(1 - \tilde{\mu}_i) \mathbb{1}_{\{0\}}(y_i) + \ln(\tilde{\mu}_i) \mathbb{1}_{\{1\}}(y_i) \}$$

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = -2 \sum_{i=1}^n \{ \ln(1 - \hat{\mu}_i) \mathbb{1}_{\{0\}}(y_i) + \ln(\hat{\mu}_i) \mathbb{1}_{\{1\}}(y_i) \}$$

Desvio: outros exemplos

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

Desvio: outros exemplos

- Binomial:

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

- Poisson:

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

Desvio: comportamento assintótico

- Em geral $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ ou $D^*(\mathbf{y}, \hat{\boldsymbol{\mu}})$ não seguem (mesmo assintoticamente) uma distribuição $\chi^2_{(n-p)}$, sob a hipótese de que o modelo em questão é adequado. Tal convergência ($D(\mathbf{Y}; \hat{\boldsymbol{\mu}}) \xrightarrow{D} \chi^2_{(n-p)}$) ocorre (sob a hipótese de que o modelo em questão é adequado):
 - Binomial: se $m_i \rightarrow \infty, \forall i$ 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 $\mu_i \rightarrow \infty, \forall i$.
 - Nos casos em que $D^*(\mathbf{Y}, \hat{\boldsymbol{\mu}})$ depende do parâmetro de precisão, se $\phi \rightarrow \infty$.

Mais sobre inferência

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

$$Q_t = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{M})' (\mathbf{C}\hat{\mathcal{V}}(\hat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{M}).$$

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

Mais sobre inferência

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

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

então sob H_0 e para n suficientemente grande, $Q_t^* \approx \chi_c^2$, em que c é o número de linhas de \mathbf{C} . Além disso, $\widehat{\mathbf{W}} \xrightarrow[n \rightarrow \infty]{P} \mathbf{W}$ o que implica que $\widehat{\mathcal{V}}(\widehat{\beta}) \xrightarrow[n \rightarrow \infty]{P} \mathcal{V}(\widehat{\beta})$, pois cada componente de $\widehat{\mathcal{V}}(\widehat{\beta})$ é uma função contínua da respectiva componente de $\widehat{\mathbf{W}}$ e também por Slutsky pois $\widehat{\phi} \xrightarrow[n \rightarrow \infty]{P} \phi$.

Mais sobre inferência

- Portanto, sob H_0 e para n suficientemente grande, pelos resultados anteriores $(Q_t^*, \widehat{\mathbf{W}})$, $Q_t \approx \chi_c^2$.
- Assim, rejeita-se H_0 se p -valor $\leq \alpha$, em que p -valor $\approx P(X \geq q_t | H_0)$, em que $X \sim \chi_c^2$
 $q_t = (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M})' (\mathbf{C}\tilde{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M})$.
- Sob H_1 , temos que $Q_t \approx \chi_{(c,\delta)}^2$ (qui-quadrado não central com c graus de liberdade e parâmetro de assimetria δ), em que $\delta = (\mathbf{C}\boldsymbol{\beta} - \mathbf{M})' (\mathbf{C}\mathbf{V}(\widehat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\boldsymbol{\beta} - \mathbf{M})$. Uma estimativa de δ é dada por $\tilde{\delta} = (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M})' (\mathbf{C}\tilde{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{C}')^{-1} (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{M})$.