Modelos Lineares Generalizados

Unidade I

Terezinha K. A. Ribeiro

terezinha.ribeiro@unb.br

Instituto de Ciências Exatas
Departamento de Estatística
Universidade de Brasília

Organização

- Introdução
- Família exponencial uniparamétrica e suas propriedades
- Casos particulares
- Funções de ligação
- Estimação dos parâmetros
- Procedimento iterativo

Introdução

O modelo de regressão linear normal (MRLN) amostral é definido por

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

em que

- Y_i é uma variável aleatória (observável) denominada de resposta;
- $x_{i1}, x_{i2}, \ldots, x_{ip}$ são $p < n$ variáveis não aleatórias (observáveis) chamadas de covariáveis; se $x_{i1} = 1, \forall i$, o modelo possui intercepto;
- $\beta_1, \beta_2, \ldots, \beta_p$, são os coeficientes/parâmetros da regressão desconhecidos que devem ser estimados;
- ε_i é uma variável aleatória (não observável) denominada de erro da regressão e supõe-se que $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma^2\right)$, com $\overset{\text{iid}}{\sim}$ denotando "se distribuem de forma independente e são identicamente distribuídas".

Introdução

Veja que

$$\mathrm{E}(Y_i) = \mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$
$$\mathrm{Var}(Y_i) = \sigma^2, \quad \forall i.$$

Note que este modelo é homoscedástico, ou seja, apesar da média de Y_i variar com as observações, a variância de Y_i é constante.

Para interpretar o coeficiente β_j, $j = 1, 2, \ldots, p$, considere o aumento de uma unidade em x_{ij} e que as demais covariáveis estão fixadas:

$$\mu_i(x'_{ij}) - \mu_i(x_{ij}) = \beta_j, \quad x'_{ij} = x_{ij} + 1.$$

Observa-se que β_j é a variação (aumento ou diminuição) na resposta média quando é adicionada uma unidade em x_{ij}. A variação na resposta média para um aumento de b unidades em x_{ij} é mensurada a partir de $b \times \beta_j$.

Introdução

Veja que

$$\mathrm{E}(Y_i) = \mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$
$$\mathrm{Var}(Y_i) = \sigma^2, \quad \forall i.$$

Note que este modelo é homoscedástico, ou seja, apesar da média de Y_i variar com as observações, a variância de Y_i é constante.

Para interpretar o coeficiente β_j, $j = 1, 2, \ldots, p$, considere o aumento de uma unidade em x_{ij} e que as demais covariáveis estão fixadas:

$$\mu_i(x'_{ij}) - \mu_i(x_{ij}) = \beta_j, \quad x'_{ij} = x_{ij} + 1.$$

Observa-se que β_j é a variação (aumento ou diminuição) na resposta média quando é adicionada uma unidade em x_{ij}. A variação na resposta média para um aumento de b unidades em x_{ij} é mensurada a partir de $b \times \beta_j$.

Introdução

Forma equivalente de definir o MRLN:

i) $Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \mathrm{N}\left(\mu_i, \sigma^2\right)$ – componente aleatória;

ii) $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ – componente determinística ou sistemática

em que $\boldsymbol{x}_i^\top = \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} \end{pmatrix} \in \mathbb{R}^p$ é o vetor de valores conhecidos das p covariáveis para a i-ésima observação, $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \end{pmatrix}^\top \in \mathbb{R}^p$ é o vetor de coeficientes da regressão, e $\overset{\text{ind}}{\sim}$ denota "se distribuem de forma independente".

Note que nesta última definição, o erro da regressão ε_i não está explicitado.

Introdução

Os modelos lineares generalizados (MLGs) são definidos por

i) $Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \mathsf{FE}\,(\mu_i, \phi)$

ii) $g\,(\mu_i) = \eta_i \Leftrightarrow \mu_i = g^{-1}\,(\eta_i)$ com $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$,

em que FE $(\cdot,\cdot)$ denota uma distribuição pertencente à família de distribuições exponencial uniparamétrica; $\phi > 0$ é um parâmetro de precisão (ϕ^{-1} é um parâmetro de dispersão); $g(\cdot)$ é uma função de ligação monótona e diferenciável; η_i é o preditor linear.

- Em i) ampliamos o número de opções para a distribuição de probabilidades da resposta Y_i.
- Em ii), ganhamos flexibilidade na modelagem da média μ_i por meio de uma função $g(\cdot)$.

Os modelos lineares generalizados (MLGs) são definidos por

i) $Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \text{FE}\,(\mu_i, \phi)$

ii) $g\,(\mu_i) = \eta_i \Leftrightarrow \mu_i = g^{-1}\,(\eta_i)$ com $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$,

em que FE $(\cdot, \cdot)$ denota uma distribuição pertencente à família de distribuições exponencial uniparamétrica; $\phi > 0$ é um parâmetro de precisão (ϕ^{-1} é um parâmetro de dispersão); $g(\cdot)$ é uma função de ligação monótona e diferenciável; η_i é o preditor linear.

- Em i) ampliamos o número de opções para a distribuição de probabilidades da resposta Y_i.
- Em ii), ganhamos flexibilidade na modelagem da média μ_i por meio de uma função $g(\cdot)$.

Introdução

Em situações em que Y_i possui uma distribuição assimétrica, o MRLN não é adequado pois a inferência dos parâmetros (intervalar e testes de hipóteses) é construída sob a suposição de normalidade de $Y_i \mid \boldsymbol{x}_i$.

Em cenários como este, pode-se transformar Y_i para atingir normalidade. Por exemplo, para $Y_i > 0$, pode-se utilizar a transformação logarítmica. Daí, o MRLN é ajustado considerando como variável resposta $Y_i^* = \log(Y_i)$:

$$Y_i^* = \log(Y_i) = \mu_i^* + \varepsilon_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i,$$

com $\mathrm{E}(Y_i^*) = \mu_i^*$.

Nesta abordagem, a interpretação de $\beta_1, \beta_2, \ldots, \beta_p$ é feita com relação à média do logaritmo da resposta. Neste caso, β_j é a variação na média de Y_i^* quando é adicionada uma unidade em x_{ij}.

Introdução

Aqui β_j não é interpretado (de forma exata) diretamente em termos da média de Y_i.

Observe que

$$
\begin{aligned}
Y_i = \mathrm{e}^{\log(Y_i)} &= \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i\} \\
&= \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\} \exp\{\varepsilon_i\}.
\end{aligned}
$$

Assim,

$$
\mathrm{E}(Y_i) = \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\} \, \mathrm{E}(\exp\{\varepsilon_i\}).
$$

Será que é possível calcular a média de Y_i?

Dado que $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, qual é a distribuição da variável aleatória $\mathrm{e}^{\varepsilon_i}$?

Revisão. Se $Z \sim \mathrm{N}(\mu, \sigma^2)$, então vale que

- $W = \mathrm{e}^Z \sim \mathrm{LN}(\mu, \sigma)$ com $\mathrm{LN}(\mu, \sigma)$ denotando a distribuição log-normal de parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$;
- $W = \mathrm{e}^Z \sim \mathrm{LN}(\mu, \sigma)$ com $\mathrm{Med}(W) = \mathrm{e}^\mu$,

$$\mathsf{E}(W) = \mathrm{e}^{\mu+\sigma^2/2}, \quad \mathsf{Var}(W) = (\mathrm{e}^{\sigma^2} - 1)\mathrm{e}^{2\mu+\sigma^2},$$

  em que $\mathrm{Med}(W)$ denota a mediana da variável W.

De $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, tem-se que $\exp\{\varepsilon_i\} \sim \mathrm{LN}(0, \sigma)$. Assim, obtemos

$$\mathsf{E}(\mathrm{e}^{\varepsilon_i}) = \mathrm{e}^{\sigma^2/2}, \quad \mathrm{Med}(\mathrm{e}^{\varepsilon_i}) = \mathrm{e}^0 = 1.$$

Daí, segue que

$$\mu_i = \mathrm{E}(Y_i) = \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\} e^{\sigma^2/2}.$$

Se σ^2 é um valor próximo a zero, então

$$\mu_i = \mathrm{E}(Y_i) \approx \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\}.$$

Alguns estudos mostram que σ^2 é um valor muito próximo a zero quando a transformação logarítmica é aplicada em Y.

Supondo que σ^2 é próximo a zero, como interpretamos β_0 e β_1?

Introdução

Considere o aumento de uma unidade em x_{ij} e que as demais covariáveis estão fixadas:

$$\frac{\mu_i(x'_{ij})}{\mu_i(x_{ij})} \approx \frac{\exp\{\beta_1 x_{i1} + \cdots + \beta_j(x_{ij}+1) + \cdots + \beta_p x_{ip}\}}{\exp\{\beta_1 x_{i1} + \cdots + \beta_j x_{ij} + \cdots + \beta_p x_{ip}\}} = \mathrm{e}^{\beta_j},$$

em que $x'_{ij} = x_{ij} + 1$.

Observa-se que e^{β_j} é a variação percentual **aproximada** na resposta média quando é adicionada uma unidade em x_{ij}.

Logo, a interpretação dos coeficientes do MRLN com resposta transformada através do logaritmo não é exata!

Introdução

De forma geral,

$$\mathrm{E}(g(Y)) \neq g(\mathrm{E}(Y)).$$

Por esta razão, ao aplicar uma transformação $g(\cdot)$ na resposta Y_i e, em seguida, ajustar um MRLN sob $g(Y_i)$, perdemos a fácil interpretação dos coeficientes da regressão em termos da média da resposta.

Esta é a grande desvantagem desta abordagem. Perde-se interpretação dos β's diretamente em termos da média de Y_i.

Uma grande diferença dos MLGs com relação ao MRLN é manter a distribuição assumida para a variável resposta intacta, e transformar o parâmetro μ_i flexibilizando o ajuste do modelo aos dados.

Introdução

A transformação $g(\cdot)$ é aplicada na média da resposta (parâmetro) através de

$$g\left(\mu_i\right) = \eta_i.$$

Além da flexibilidade, a ideia da função de ligação é mapear o espaço em que está o parâmetro de localização μ_i para a reta real, garantindo que os parâmetros de regressão sejam estimados sem restrições.

Por exemplo, suponha que Y_i segue uma distribuição de probabilidades positiva. Assim, $\mu_i = E\left(Y_i\right) > 0$. Seja $g(z) = \log(z)$, então $g^{-1}(z) = \mathrm{e}^z$. Consequentemente,

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \Rightarrow \log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \in \mathbb{R} \Rightarrow \mu_i = \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} > 0.$$

Família exponencial uniparamétrica

Se $Y \in \text{FE}(\mu, \phi)$, a função densidade (ou função de probabilidade) de Y possui a forma

$$f(y; \theta, \phi) = \exp\{\phi[\theta y - b(\theta)] + c(y; \phi)\}, \tag{1}$$

em que

- o suporte desta família de distribuições não depende de θ ou ϕ;
- $\theta = \theta(\mu)$ é uma função de μ e é chamado de parâmetro canônico;
- $b(\theta)$ é uma função diferenciável que depende apenas de θ;
- $c(y; \phi)$ é uma função que depende de y e ϕ e é diferenciável nos seus argumentos.

Note a função densidade em (1) é um caso particular da família exponencial de distribuições. A variável y aparece de forma linear em (1).

Família exponencial uniparamétrica

Para derivar algumas propriedades desta família de distribuições, considere

$$\ell(\theta, \phi) = \log[f(y; \theta, \phi)] = \phi[\theta y - b(\theta)] + c(y; \phi).$$

As funções escore para θ e ϕ são dadas por

$$U_\theta = \frac{\partial \ell(\theta, \phi)}{\partial \theta} = \phi \left[y - b'(\theta) \right],$$

$$U_\phi = \frac{\partial \ell(\theta, \phi)}{\partial \phi} = \theta y - b(\theta) + c'(y; \theta),$$

em que

$$b'(\theta) = \frac{d}{d\theta}[b(\theta)], \quad c'(y; \phi) = \frac{d}{d\phi}[c(y; \theta)].$$

Família exponencial uniparamétrica

Considere as seguintes condições de regularidade:

C_1) O suporte da distribuição de Y não depende de θ ou ϕ;

C_2) É válida a troca entre os sinais de derivação e integração sob a densidade $f(\cdot;\theta,\phi)$.

Sob estas condições é válido que

$$\mathrm{E}\left(U_\theta\right) = 0, \quad \mathrm{E}\left(U'_\theta\right) = -\,\mathrm{E}\left(U_\theta^2\right),$$

com

$$U'_\theta = \frac{\partial^2 \ell(\theta,\phi)}{\partial\theta^2}.$$

Daí, segue que

$$\mathrm{E}\left\{\phi\left[Y - b'(\theta)\right]\right\} = 0 \Rightarrow \mathrm{E}(Y) = \mu = b'(\theta).$$

Família exponencial uniparamétrica

Também, temos que

$$U_\theta^2 = \left\{\phi\left[y - b'(\theta)\right]\right\}^2 = \phi^2 (y-\mu)^2$$

$$\Rightarrow \mathrm{E}\left(U_\theta^2\right) = \phi^2\,\mathrm{E}\left[(Y-\mu)^2\right] = \phi^2\,\mathrm{Var}(Y).$$

Ainda, $U'_\theta = -\phi b''(\theta)$ com $b''(\theta) = \frac{d^2}{d\theta^2}[b(\theta)]$. Então,

$$\mathrm{E}\left(U'_\theta\right) = -\,\mathrm{E}\left(U_\theta^2\right) \Rightarrow -\phi b''(\theta) = -\phi^2\,\mathrm{Var}(Y)$$

$$\Rightarrow \mathrm{Var}(Y) = \phi^{-1} b''(\theta) = \phi^{-1}\,\mathrm{V}(\mu),$$

em que $\mathrm{V}(\mu) = b''(\theta)$ é conhecida como função de variância.

Família exponencial uniparamétrica

A função de variância caracteriza a distribuição da família exponencial uniparamétrica. Por exemplo, a função de variância

$$V(\mu) = \mu(1-\mu), 0 < \mu < 1,$$

caracteriza a classe de distribuições binomiais de sucesso μ ou $1-\mu$.

Note que para $V(\mu)$ fixada, $\mathrm{Var}(Y) = \phi^{-1} V(\mu)$ aumenta quando ϕ diminui, e, diminui quando ϕ cresce. Por esta razão, ϕ é considerado um parâmetro de precisão.

Resultados aproximados

Sejam $Y_1, Y_2, \ldots, Y_n$ variáveis aleatórias independentes e identicamente distribuídas (iid) com $\mathsf{E}\left(Y_i\right) = \mu$ e $\mathsf{Var}\left(Y_i\right) = \sigma^2$. Pelo Teorema do Limite Central, para n grande,

$$\overline{Y} \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

em que $\overset{a}{\sim}$ denota "segue distribuição aproximada".

Nos MLGs temos apenas independência entre $Y_1, Y_2, \ldots, Y_n$. Para ϕ grande,

$$Y \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\mathsf{V}(\mu)}{\phi}\right).$$

Note que este último resultado é válido para $n = 1$ e precisão ϕ grande.

Casos particulares - Distribuição normal

Seja Y uma variável aleatória com distribuição normal de média μ e variância σ^2. A função densidade de Y é dada por

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}, \qquad \begin{array}{l} y \in \mathbb{R}, \\ \mu \in \mathbb{R}, \\ \sigma^2 > 0. \end{array}$$

Denotaremos por $Y \sim \mathrm{N}\left(\mu,\sigma^2\right)$.

Veja que

$$\begin{aligned} f(y;\mu,\sigma^2) &= \exp\left\{-\frac{1}{2\sigma^2}\left(y^2 - 2\mu y + \mu^2\right) - \log\left[\left(2\pi\sigma^2\right)^{1/2}\right]\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2\sigma^2}y^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2}\left[\log\left(2\pi\sigma^2\right) + \frac{y^2}{\sigma^2}\right]\right\}. \end{aligned}$$

Casos particulares - Distribuição normal

Tome $\phi = 1/\sigma^2$, $\theta = \mu$, $b(\theta) = \mu^2/2 = \theta^2/2$, e

$$\begin{aligned} c(y;\phi) &= -\frac{1}{2}\left[\log\left(2\pi\sigma^2\right) + \frac{y^2}{\sigma^2}\right] \\ &= -\frac{1}{2}\left[\log\left(2\pi\phi^{-1}\right) + \phi y^2\right] \\ &= -\frac{1}{2}\left[\log\left(\frac{2\pi}{\phi}\right) + \phi y^2\right] \\ &= \frac{1}{2}\left[\log\left(\frac{\phi}{2\pi}\right) - \phi y^2\right]. \end{aligned}$$

Também, o suporte da distribuição normal não depende de θ ou ϕ pois $y \in \mathbb{R}$.

Portanto, $Y \in \text{FE}(\mu, \phi)$.

Casos particulares - Distribuição normal

Pelas propriedades da família exponencial uniparamétrica, temos que

$$\mathrm{E}(Y) = b'(\theta) = \frac{2\theta}{2} = \theta = \mu;$$

$$\mathrm{V}(\mu) = b''(\theta) = 1 \Rightarrow \mathrm{Var}(Y) = \phi^{-1}\,\mathrm{V}(\mu) = \sigma^2.$$

Para $Y \sim \mathrm{N}\left(\mu, \sigma^2\right)$, temos

- média igual a μ;
- função de variância $\mathrm{V}(\mu)$ não depende da média μ;
- variância igual a σ^2 (homoscedasticidade);
- adequação para dados contínuos reais;
- simetria em torno da média μ.

Seja Y uma variável aleatória com distribuição de Poisson de parâmetro μ. A função de probabilidade de Y é dada por

$$f(y;\mu)=\frac{e^{-\mu}\mu^{y}}{y!},\quad y\in\{0,1,2,\ldots\},\qquad \mu>0.$$

Denotaremos por $Y\sim \text{Poisson}(\mu)$.

Veja que

$$f(y;\mu)=\exp\left\{y\log(\mu)-\mu-\log(y!)\right\}.$$

Tome $\theta=\log(\mu)\Rightarrow\mu=\mathrm{e}^{\theta}$, $b(\theta)=\mu=\mathrm{e}^{\theta}$, $\phi=1$ (conhecido), e $c(y;\phi)=-\log\left(y!\right)$.

Além disso, o suporte da distribuição Poisson não depende de θ ou ϕ desde que $y\in\{0,1,2,\ldots\}$.

Logo, $Y\in \text{FE}(\mu,\phi)$.

Casos particulares - Distribuição Poisson

Segue que

$$\mathsf{E}(Y) = b'(\theta) = e^{\theta} = \mu;$$

$$\mathsf{V}(\mu) = b''(\theta) = e^{\theta} = \mu \Rightarrow \mathsf{Var}(Y) = \phi^{-1}\,\mathsf{V}(\mu) = \mu.$$

Para $Y \sim \mathsf{Poisson}(\mu)$, temos

- média igual a μ;
- variância igual a média (heteroscedasticidade);
- adequação para dados de contagem.

Dizemos que o modelo de Poisson acomoda equidispersão, isto é, acomoda dados que satisfazem $\mathsf{E}(Y) = \mathsf{Var}(Y)$. Esta é uma restrição forte desta distribuição para modelagem de dados de contagem.

Casos particulares - Distribuição gama

Seja Y uma variável aleatória com distribuição gama de média μ e coeficiente de variação $\phi^{-1/2}$. A função densidade de Y é dada por

$$f(y;\mu,\phi) = \frac{1}{\Gamma(\phi)}\left(\frac{\phi y}{\mu}\right)^{\phi} \exp\left\{-\frac{\phi y}{\mu}\right\}\frac{1}{y}, \qquad \begin{array}{l} y > 0, \\ \mu > 0, \\ \phi > 0, \end{array}$$

em que $\Gamma(\cdot)$ é a função gama definida por

$$\Gamma(\phi) = \int_0^{\infty} t^{\phi-1} \mathrm{e}^{-t} dt.$$

Esta função densidade está reparametrizada de tal forma que $\mu = \mathrm{E}(Y)$.

Denotaremos por $Y \sim \mathrm{gama}(\mu,\phi)$.

Se $\phi = 1$ temos a distribuição exponencial. Se $\phi = k/2$ e $\mu = k$ temos a distribuição qui-quadrado com k graus de liberdade.

Casos particulares - Distribuição gama

Veja que

$$\begin{aligned} f(y;\mu,\phi) &= \exp\left\{-\log\Gamma(\phi)+\phi\log\left(\frac{\phi y}{\mu}\right)-\frac{\phi y}{\mu}-\log(y)\right\} \\ &= \exp\left\{-\log\Gamma(\phi)+\phi[\log(\phi y)-\log(\mu)]-\frac{\phi y}{\mu}-\log(y)\right\} \\ &= \exp\{\phi[-y/\mu-\log(\mu)]-\log\Gamma(\phi)+\phi\log(\phi y)-\log(y)\}. \end{aligned}$$

Tome $\phi = \phi$, $\theta = -\frac{1}{\mu} \Rightarrow \mu = -\theta^{-1}$, $b(\theta) = \log(\mu) = -\log(-\theta)$, com $\theta < 0$,

$$\begin{aligned} c(y;\phi) &= \phi\log(\phi y)-\log\Gamma(\phi)-\log(y) \\ &= \phi\log(\phi)+\phi\log(y)-\log\Gamma(\phi)-\log(y) \\ &= (\phi-1)\log(y)+\phi\log(\phi)-\log\Gamma(\phi). \end{aligned}$$

Além disto, o suporte da distribuição gama não depende de θ ou ϕ desde que $y > 0$. Logo, $Y \in \text{FE}(\mu,\phi)$.

Casos particulares - Distribuição gama

Segue que

$$E(Y) = b'(\theta) = -\frac{1}{-\theta}(-1) = -\frac{1}{\theta} = \mu;$$

$$V(\mu) = b''(\theta) = \frac{1}{\theta^2} = \left(-\frac{1}{\theta}\right)^2 = \mu^2 \Rightarrow \text{Var}(Y) = \phi^{-1}V(\mu) = \phi^{-1}\mu^2.$$

Note que

$$\text{CV}(Y) = \frac{\text{DP}(Y)}{\text{E}(Y)} = \frac{(\phi^{-1}\mu^2)^{1/2}}{\mu} = \phi^{-1/2},$$

com $\text{CV}(Y)$ e $\text{DP}(Y)$ denotando coeficiente de variação e desvio-padrão de Y, respectivamente.

Casos particulares - Distribuição gama

Para $Y \sim \text{gama}(\mu, \phi)$, temos

- média igual a μ;
- variância dependente da média através de μ^2 (heteroscedasticidade);
- coeficiente de variação igual a $\phi^{-1/2}$;
- adequação para dados contínuos positivos assimétricos.

Adicionalmente, temos que para ϕ grande

$$Y \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\mu^2}{\phi}\right).$$

Seja Y uma variável aleatória com distribuição normal inversa de média μ e parâmetro de precisão ϕ. A função densidade de Y é dada por

$$f(y;\mu,\phi) = \sqrt{\frac{\phi}{2\pi y^3}} \exp\left\{-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right\}, \qquad \begin{array}{l} y > 0, \\ \mu > 0, \\ \phi > 0. \end{array}$$

Denotaremos por $Y \sim \text{NI}(\mu,\phi)$.

Veja que

$$\begin{aligned} f(y;\mu,\phi) &= \exp\left\{-\frac{\phi(y^2-2\mu y+\mu^2)}{2\mu^2 y} + \frac{1}{2}\log\left(\frac{\phi}{2\pi y^3}\right)\right\} \\ &= \exp\left\{\phi\left[-\frac{y}{2\mu^2}+\frac{1}{\mu}\right] - \frac{\phi}{2y} + \frac{1}{2}\log\left(\frac{\phi}{2\pi y^3}\right)\right\}. \end{aligned}$$

Casos particulares - Distribuição normal inversa

Tome $\phi = \phi$, $\theta = -\frac{1}{2\mu^2} \Rightarrow \mu = (-2\theta)^{-1/2}$, $b(\theta) = -\frac{1}{\mu} = -(-2\theta)^{1/2}$, com $\theta < 0$,

$$c(y;\phi) = -\frac{\phi}{2y} + \frac{1}{2}\log\left(\frac{\phi}{2\pi y^3}\right).$$

Além disto, o suporte da distribuição normal inversa não depende de θ ou ϕ desde que $y > 0$. Logo, $Y \in \mathrm{FE}(\mu, \phi)$.

Segue que

$$\mathrm{E}(Y) = b'(\theta) = -\frac{1}{2}(-2\theta)^{-1/2}(-2) = (-2\theta)^{-1/2} = \mu;$$

$$\mathrm{V}(\mu) = b''(\theta) = -\frac{1}{2}(-2\theta)^{-3/2}(-2) = \left[(-2\theta)^{-1/2}\right]^3 = \mu^3$$

$$\Rightarrow \mathrm{Var}(Y) = \phi^{-1}\mathrm{V}(\mu) = \phi^{-1}\mu^3.$$

Casos particulares - Distribuição normal inversa

Para $Y \sim \text{NI}(\mu, \phi)$, temos

- média igual a μ;
- variância dependente da média através de μ^3 (heteroscedasticidade);
- variância cresce mais rapidamente com μ quando comparado ao modelo gama;
- adequação para dados contínuos positivos assimétricos;
- alternativa ao modelo gama.

Adicionalmente, temos que para ϕ grande

$$Y \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\mu^3}{\phi}\right).$$

Casos particulares - Distribuição binomial

Seja Y uma variável aleatória com distribuição binomial de parâmetros n e $0 < \mu < 1$. A função de probabilidade de Y é dada por

$$f(y;\mu) = \binom{n}{y} \mu^y (1-\mu)^{n-y}, \quad y \in \{0,1,2,\ldots,n\}.$$

Denotaremos por $Y \sim \text{binomial}(n,\mu)$. O parâmetro n é conhecido.

Observe que

$$\begin{aligned} f(y;\mu) &= \exp\left\{\log\binom{n}{y} + y\log(\mu) + (n-y)\log(1-\mu)\right\} \\ &= \exp\left\{y\log\left(\frac{\mu}{1-\mu}\right) + n\log(1-\mu) + \log\binom{n}{y}\right\}. \end{aligned}$$

Casos particulares - Distribuição binomial

Considere $\phi = 1$,

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) \Rightarrow \mu = \frac{e^{\theta}}{1+e^{\theta}}, \quad \theta \in \mathbb{R};$$

$$b(\theta) = -n\log(1-\mu) = -n\log\left(\frac{1}{1+e^{\theta}}\right) = n\log\left(1+e^{\theta}\right);$$

$$c(y;\phi) = \log\binom{n}{y}.$$

Além disto, o suporte da distribuição binomial é o conjunto $\{0,1,2,\ldots,n\}$ que não depende de parâmetros desconhecidos. Logo, $Y \in \mathrm{FE}(\mu,\phi)$.

Note que o suporte depende de n, entretanto o valor n sempre será conhecido.

Casos particulares - Distribuição binomial

Segue que

$$\mathrm{E}(Y) = b'(\theta) = \frac{n\mathrm{e}^{\theta}}{1+\mathrm{e}^{\theta}} = n\mu;$$

$$\mathrm{V}(\mu) = b''(\theta) = \frac{n\left[\mathrm{e}^{\theta}\left(1+\mathrm{e}^{\theta}\right) - \mathrm{e}^{\theta}\mathrm{e}^{\theta}\right]}{\left(1+\mathrm{e}^{\theta}\right)^2} = \frac{n\mathrm{e}^{\theta}}{\left(1+\mathrm{e}^{\theta}\right)^2} = n\mu(1-\mu);$$

$$\Rightarrow \mathrm{Var}(Y) = \phi^{-1}\,\mathrm{V}(\mu) = n\mu(1-\mu).$$

Assim, a função de variância $\mathrm{V}(\mu)$ depende do tamanho amostral n. Gostaríamos que $\mathrm{V}(\mu)$ dependa apenas da média μ.

Casos particulares - Distribuição binomial

Para resolver este "problema", realizaremos uma mudança de escala através da transformação $Y^* = \frac{Y}{n}$. Note que Y^* representa a proporção de sucessos em n ensaios independentes de Bernoulli, cada um com probabilidade de ocorrência μ com $nY^* \sim \text{binomial}(n, \mu)$.

Daí, a função de probabilidade de Y^* fica dada por

$$
\begin{aligned}
f\left(y^*; \mu\right) &= \binom{n}{ny^*} \mu^{ny^*} \left(1-\mu\right)^{n-ny^*} \\
&= \exp\left\{ n \left[y^* \log\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu) \right] + \log\binom{n}{ny^*} \right\}.
\end{aligned}
$$

com $y^* \in \left\{0, \frac{1}{n}, \ldots, 1\right\}$.

Casos particulares - Distribuição binomial

Tome $\phi = n$ (conhecido),

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) \Rightarrow \mu = \frac{e^{\theta}}{1+e^{\theta}}, \quad \theta \in \mathbb{R};$$

$$b(\theta) = -\log(1-\mu) = -\log\left(\frac{1}{1+e^{\theta}}\right) = \log\left(1+e^{\theta}\right);$$

$$c(y;\phi) = \log\binom{n}{ny^*} = \log\binom{\phi}{\phi y^*}.$$

Além disto, o suporte da distribuição binomial é o conjunto $\left\{0, \frac{1}{n}, \ldots, 1\right\}$ que não depende de parâmetros desconhecidos. Logo, $Y^* \in \text{FE}(\mu,\phi)$.

Casos particulares - Distribuição binomial

Segue que

$$\mathrm{E}(Y^*) = b'(\theta) = \frac{e^\theta}{1+e^\theta} = \mu;$$

$$\mathrm{V}(\mu) = b''(\theta) = \frac{e^\theta\left(1+e^\theta\right) - e^\theta e^\theta}{\left(1+e^\theta\right)^2} = \frac{e^\theta}{\left(1+e^\theta\right)^2} = \mu(1-\mu);$$

$$\Rightarrow \mathrm{Var}(Y^*) = \phi^{-1}\,\mathrm{V}(\mu) = n^{-1}\mu(1-\mu).$$

Casos particulares - Distribuição binomial

Para $Y^* = \frac{Y}{n}$ com $nY^* \sim \text{binomial}(n, \mu)$, temos

- média igual a μ;
- variância dependente da média através de $\mu(1-\mu)$ (heteroscedasticidade);
- variância assume valor menor nos extremos (μ próximo a zero ou μ próximo a um);
- adequação para modelagem de probabilidade de sucesso;
- quando $n = 1$ obtém-se a regressão logística.

Adicionalmente, temos que para n grande

$$Y \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\mu(1-\mu)}{n}\right).$$

Casos particulares

Na tabela abaixo estão apresentadas as principais distribuições pertencentes à família exponencial uniparamétrica.

Tabela 1: Principais distribuições pertencentes à FE(μ, ϕ).

Distribuição	θ	ϕ	$b(\theta)$	$V(\mu)$
Normal	μ	$1/\sigma^2$	$\theta^2/2$	1
Poisson	$\log(\mu)$	1	e^θ	μ
Gama	$-\frac{1}{\mu}$	$1/\,CV^2$	$-\log(-\theta)$	μ^2
Normal Inversa	$-\frac{1}{2\mu^2}$	ϕ	$-(-2\theta)^{1/2}$	μ^3
Binomial	$\log\left(\frac{\mu}{1-\mu}\right)$	n	$\log(1+e^\theta)$	$\mu(1-\mu)$

Funções de ligação

Como escolher uma função de ligação $g(\cdot)$ adequada?

A princípio qualquer função monótona e diferenciável pode ser utilizada. Entretanto, existem algumas funções de ligação que são mais interessantes.

Dizemos que uma função de ligação $g(\cdot)$ é canônica quando esta é obtida de $\theta_i = \eta_i$.

Para a distribuição normal sabemos que o parâmetro canônico é $\theta_i = \mu_i$. Assim, a igualdade $\theta_i = \eta_i$ implica em $\mu_i = \eta_i$. Portanto, a função de ligação canônica para a distribuição normal é a ligação identidade

$$g(z) = z, \quad z \in \mathbb{R}.$$

Funções de ligação

Como escolher uma função de ligação $g(\cdot)$ adequada?

A princípio qualquer função monótona e diferenciável pode ser utilizada. Entretanto, existem algumas funções de ligação que são mais interessantes.

Dizemos que uma função de ligação $g(\cdot)$ é canônica quando esta é obtida de $\theta_i = \eta_i$.

Para a distribuição normal sabemos que o parâmetro canônico é $\theta_i = \mu_i$. Assim, a igualdade $\theta_i = \eta_i$ implica em $\mu_i = \eta_i$. Portanto, a função de ligação canônica para a distribuição normal é a ligação identidade

$$g(z) = z, \quad z \in \mathbb{R}.$$

Funções de ligação

Para a distribuição Poisson temos que $\theta_i = \log(\mu_i)$. Assim, $\theta_i = \eta_i$ implica em $\log(\mu_i) = \eta_i$. Daí, a função de ligação canônica para a distribuição Poisson é a ligação logarítmica

$$g(z) = \log(z), \quad z > 0.$$

Para a distribuição gama temos que $\theta_i = -1/\mu_i$. Assim, $\theta_i = \eta_i$ implica em $-1/\mu_i = \eta_i$. Por conveniência, a função de ligação canônica para a distribuição gama é definida como

$$g(z) = \frac{1}{z}, \quad z \in \mathbb{R}$$

e esta é chamada de ligação recíproca.

Funções de ligação

Para a distribuição Poisson temos que $\theta_i = \log(\mu_i)$. Assim, $\theta_i = \eta_i$ implica em $\log(\mu_i) = \eta_i$. Daí, a função de ligação canônica para a distribuição Poisson é a ligação logarítmica

$$g(z) = \log(z), \quad z > 0.$$

Para a distribuição gama temos que $\theta_i = -1/\mu_i$. Assim, $\theta_i = \eta_i$ implica em $-1/\mu_i = \eta_i$. Por conveniência, a função de ligação canônica para a distribuição gama é definida como

$$g(z) = \frac{1}{z}, \quad z \in \mathbb{R}$$

e esta é chamada de ligação recíproca.

Funções de ligação

Para a distribuição normal inversa temos que $\theta_i = -1/2\mu_i^2$. Assim, $\theta_i = \eta_i$ implica em $-1/2\mu_i^2 = \eta_i$. Por conveniência, a função de ligação canônica para a distribuição normal inversa é definida como

$$g(z) = \frac{1}{z^2}, \quad z \in \mathbb{R}$$

e esta é chamada de ligação recíproca ao quadrado.

Para a distribuição binomial temos que $\theta_i = \log(\mu_i/(1-\mu_i))$. Assim, $\theta_i = \eta_i$ implica em $\log(\mu_i/(1-\mu_i)) = \eta_i$. Daí, a função de ligação canônica para a distribuição binomial, denominada de logito, é

$$g(z) = \text{logito}(z) = \log\left(\frac{z}{1-z}\right), \quad 0 < z < 1.$$

Neste caso, a função de ligação é inversa da função de distribuição acumulada (FDA) da distribuição logística.

Funções de ligação

Para a distribuição normal inversa temos que $\theta_i = -1/2\mu_i^2$. Assim, $\theta_i = \eta_i$ implica em $-1/2\mu_i^2 = \eta_i$. Por conveniência, a função de ligação canônica para a distribuição normal inversa é definida como

$$g(z) = \frac{1}{z^2}, \quad z \in \mathbb{R}$$

e esta é chamada de ligação recíproca ao quadrado.

Para a distribuição binomial temos que $\theta_i = \log(\mu_i/(1-\mu_i))$. Assim, $\theta_i = \eta_i$ implica em $\log(\mu_i/(1-\mu_i)) = \eta_i$. Daí, a função de ligação canônica para a distribuição binomial, denominada de logito, é

$$g(z) = \text{logito}(z) = \log\left(\frac{z}{1-z}\right), \quad 0 < z < 1.$$

Neste caso, a função de ligação é inversa da função de distribuição acumulada (FDA) da distribuição logística.

Funções de ligação

Na tabela abaixo temos as funções de ligação canônicas das principais distribuições pertencentes à família exponencial uniparamétrica.

Tabela 2: Ligações canônicas das principais distribuições pertencentes à FE(μ, ϕ).

Distribuição	$g(z)$	Nome
Normal	z	identidade
Poisson	$\log(z)$	logarítmica
Gama	$1/z$	recíproca
Normal Inversa	$1/z^2$	recíproca ao quadrado
Binomial	$\text{logito}(z)$	logito

Funções de ligação

Existem outras funções de ligação que são muito utilizadas na prática. Em especial, quando o parâmetro a ser modelado pertence ao intervalo $(0, 1)$.

- Ligação probito:

$$g(z) = \text{probito}(z) = \Phi^{-1}(z), \quad 0 < z < 1,$$

  em que $\Phi^{-1}(\cdot)$ denota a inversa da FDA da distribuição normal padrão.

- Ligação complemento log-log:

$$g(z) = \log(-\log(1 - z)), \quad 0 < z < 1.$$

  Neste caso, $g(z)$ corresponde a inversa da FDA da distribuição Gumbel padrão.

Funções de ligação

Existem outras funções de ligação que são muito utilizadas na prática. Em especial, quando o parâmetro a ser modelado pertence ao intervalo $(0, 1)$.

- Ligação probito:

$$g(z) = \text{probito}(z) = \Phi^{-1}(z), \quad 0 < z < 1,$$

  em que $\Phi^{-1}(\cdot)$ denota a inversa da FDA da distribuição normal padrão.

- Ligação complemento log-log:

$$g(z) = \log(-\log(1 - z)), \quad 0 < z < 1.$$

  Neste caso, $g(z)$ corresponde a inversa da FDA da distribuição Gumbel padrão.

Funções de ligação

Existem outras funções de ligação que são muito utilizadas na prática. Em especial, quando o parâmetro a ser modelado pertence ao intervalo $(0, 1)$.

- Ligação probito:

$$g(z) = \text{probito}(z) = \Phi^{-1}(z), \quad 0 < z < 1,$$

em que $\Phi^{-1}(\cdot)$ denota a inversa da FDA da distribuição normal padrão.

- Ligação complemento log-log:

$$g(z) = \log(-\log(1 - z)), \quad 0 < z < 1.$$

Neste caso, $g(z)$ corresponde a inversa da FDA da distribuição Gumbel padrão.

Funções de ligação

- Ligação cauchit:

$$g(z) = \mathtt{tan}(\pi(z - 0{,}5)), \quad 0 < z < 1,$$

em que $\mathtt{tan}(\cdot)$ denota a função tangente. Neste caso, $g(z)$ é a inversa da FDA da distribuição Cauchy padrão.

As ligações probito, complemento log-log e cauchit são alternativas a ligação logito para modelar parâmetros θ tais que $0 < \theta < 1$. Por esta razão, estas ligações podem ser utilizadas para modelar a média da proporção de sucessos μ do modelo binomial.

Funções de ligação

- Ligação cauchit:

$$g(z) = \mathtt{tan}(\pi(z - 0{,}5)), \quad 0 < z < 1,$$

em que $\mathtt{tan}(\cdot)$ denota a função tangente. Neste caso, $g(z)$ é a inversa da FDA da distribuição Cauchy padrão.

As ligações probito, complemento log-log e cauchit são alternativas a ligação logito para modelar parâmetros θ tais que $0 < \theta < 1$. Por esta razão, estas ligações podem ser utilizadas para modelar a média da proporção de sucessos μ do modelo binomial.

Funções de ligação

As ligações logito, probito e cauchit são simétricas, isto é,

$$g(z) = -g(1 - z).$$

Isto é uma consequência das ligações serem baseadas em distribuições simétricas. O mesmo não pode ser dito sobre a ligação complemento log-log que é assimétrica.

Várias das ligações discutidas foram obtidas a partir de inversas de FDAs de distribuições de probabilidade contínuas padronizadas.

De forma geral, considerando $F(z)$ uma FDA contínua em sua forma padronizada, pode-se definir uma função de ligação a partir de

$$\mu_i = F(\eta_i) \Leftrightarrow F^{-1}(\mu_i) = \eta_i, \quad 0 < \mu_i < 1, \quad \eta_i \in \mathbb{R}.$$

Daí, define-se a ligação como sendo $g(z) = F^{-1}(z)$ com $0 < z < 1$.

Funções de ligação

As ligações logito, probito e cauchit são simétricas, isto é,

$$g(z) = -g(1-z).$$

Isto é uma consequência das ligações serem baseadas em distribuições simétricas. O mesmo não pode ser dito sobre a ligação complemento log-log que é assimétrica.

Várias das ligações discutidas foram obtidas a partir de inversas de FDAs de distribuições de probabilidade contínuas padronizadas.

De forma geral, considerando $F(z)$ uma FDA contínua em sua forma padronizada, pode-se definir uma função de ligação a partir de

$$\mu_i = F(\eta_i) \Leftrightarrow F^{-1}(\mu_i) = \eta_i, \;\; 0 < \mu_i < 1, \;\; \eta_i \in \mathbb{R}.$$

Daí, define-se a ligação como sendo $g(z) = F^{-1}(z)$ com $0 < z < 1$.

Funções de ligação

Para ajustar MLGs no R utiliza-se a função `glm` do pacote `stats`.

- Para o modelo normal, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo Poisson, as ligações disponíveis são identidade, logarítmica, raiz quadrada;
- Para o modelo gama, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo normal inversa, as ligações disponíveis são identidade, logarítmica, recíproca e recíproca ao quadrado;
- Para o modelo binomial, as ligações disponíveis são logito, probito, complemento log-log, cauchit, logarítmica.

O padrão para cada modelo é a ligação canônica. Entretanto, a ligação canônica não é necessariamente a mais utilizada.

Funções de ligação

Para ajustar MLGs no R utiliza-se a função `glm` do pacote `stats`.

- Para o modelo normal, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo Poisson, as ligações disponíveis são identidade, logarítmica, raiz quadrada;
- Para o modelo gama, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo normal inversa, as ligações disponíveis são identidade, logarítmica, recíproca e recíproca ao quadrado;
- Para o modelo binomial, as ligações disponíveis são logito, probito, complemento log-log, cauchit, logarítmica.

O padrão para cada modelo é a ligação canônica. Entretanto, a ligação canônica não é necessariamente a mais utilizada.

Funções de ligação

Para ajustar MLGs no R utiliza-se a função `glm` do pacote `stats`.

- Para o modelo normal, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo Poisson, as ligações disponíveis são identidade, logarítmica, raiz quadrada;
- Para o modelo gama, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo normal inversa, as ligações disponíveis são identidade, logarítmica, recíproca e recíproca ao quadrado;
- Para o modelo binomial, as ligações disponíveis são logito, probito, complemento log-log, cauchit, logarítmica.

O padrão para cada modelo é a ligação canônica. Entretanto, a ligação canônica não é necessariamente a mais utilizada.

Funções de ligação

Para ajustar MLGs no R utiliza-se a função `glm` do pacote `stats`.

- Para o modelo normal, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo Poisson, as ligações disponíveis são identidade, logarítmica, raiz quadrada;
- Para o modelo gama, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo normal inversa, as ligações disponíveis são identidade, logarítmica, recíproca e recíproca ao quadrado;
- Para o modelo binomial, as ligações disponíveis são logito, probito, complemento log-log, cauchit, logarítmica.

O padrão para cada modelo é a ligação canônica. Entretanto, a ligação canônica não é necessariamente a mais utilizada.

Funções de ligação

Para ajustar MLGs no R utiliza-se a função `glm` do pacote `stats`.

- Para o modelo normal, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo Poisson, as ligações disponíveis são identidade, logarítmica, raiz quadrada;
- Para o modelo gama, as ligações disponíveis são identidade, logarítmica e recíproca;
- Para o modelo normal inversa, as ligações disponíveis são identidade, logarítmica, recíproca e recíproca ao quadrado;
- Para o modelo binomial, as ligações disponíveis são logito, probito, complemento log-log, cauchit, logarítmica.

O padrão para cada modelo é a ligação canônica. Entretanto, a ligação canônica não é necessariamente a mais utilizada.

Funções de ligação - Interpretações

Para realizar interpretação de algumas funções de ligação, considere a estrutura de regressão:

$$g(\mu) = \eta = \alpha + \beta x,$$

em que x é uma covariável contínua. Seja $x' = x + 1$.

Se $g(z) = z$, então

$$\mu' = \alpha + \beta(x + 1) = \alpha + \beta x + \beta = \mu + \beta.$$

Assim, β é a variação na média da resposta quando x é acrescido de uma unidade.

A interpretação é análoga para o caso com duas ou mais covariáveis desde que as demais covariáveis estejam fixadas.

Funções de ligação - Interpretações

Se $g(z) = \log(z)$, então $\log(\mu) = \alpha + \beta x \Rightarrow \mu = \mathrm{e}^{\alpha+\beta x}$. Daí, segue que

$$\mu' = \mathrm{e}^{\alpha+\beta(x+1)} = \mathrm{e}^{\alpha+\beta x}\mathrm{e}^{\beta} = \mu\mathrm{e}^{\beta} \Rightarrow \mathrm{e}^{\beta} = \frac{\mu'}{\mu}.$$

Logo, e^{β} é a variação percentual na média da resposta quando x é acrescido de uma unidade.

Se $\mathrm{e}^{\beta} > 1$, então tem-se aumento percentual na média quando x aumenta. Se $0 < \mathrm{e}^{\beta} < 1$, então tem-se redução percentual na média quando x aumenta.

Novamente, a interpretação é análoga para o caso com duas ou mais covariáveis desde que as demais covariáveis estejam fixadas.

Funções de ligação - Interpretações

Se $g(z) = \text{logito}(z)$, então $\text{logito}(z) = \alpha + \beta x$.

Veja que

$$\log\left(\frac{\mu}{1-\mu}\right) = \alpha + \beta x \Rightarrow \frac{\mu}{1-\mu} = e^{\alpha+\beta x},$$

em que $0 < \mu < 1$ é a probabilidade de sucesso. A quantidade $\mu/(1-\mu)$ é denominada de chance do sucesso.

- Se a chance é igual a 1, sucesso e fracasso são igualmente prováveis;
- Se a chance é maior do que 1, o sucesso é mais provável do que o fracasso;
- Se a chance é menor do que 1, o fracasso é mais provável do que o sucesso.

Funções de ligação - Interpretações

Adicionando uma unidade em x, obtemos que a chance do sucesso fica dada por

$$\frac{\mu'}{1-\mu'} = \mathrm{e}^{\alpha+\beta(x+1)} = \mathrm{e}^{\alpha+\beta x}\mathrm{e}^{\beta} = \frac{\mu}{1-\mu}\mathrm{e}^{\beta}.$$

Consequentemente,

$$\mathrm{e}^{\beta} = \frac{\mu'/(1-\mu')}{\mu/(1-\mu)}.$$

Portanto, e^{β} é a razão de chances do sucesso quando adiciona-se uma unidade em x.

Por exemplo, considere Y é a ocorrência de uma certa doença ($Y = 1$ se doente, caso contrário, $Y = 0$) e x é a idade do indivíduo. Neste caso, e^{β} é a variação percentual na chance do indivíduo desenvolver a doença quando aumenta-se um ano na idade.

Estimação dos parâmetros

Por suposição $Y_i \mid \boldsymbol{x}_i \stackrel{\text{ind}}{\sim}$ FE (μ_i, ϕ) com $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$.

O vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \ \phi)^\top$ é desconhecido e deve ser estimado com base em dados amostrais. Sob os MLGs, o procedimento de estimação baseia-se no método de máxima verossimilhança.

Considere a função de verossimilhança $L(\boldsymbol{\theta})$ como sendo a função densidade de probabilidade conjunta de $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$.

O estimador de máxima verossimilhança (EMV) para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \ \phi)^\top$ é o valor de $\boldsymbol{\theta}$ que maximiza a função de verossimilhança $L(\boldsymbol{\theta})$, isto é,

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmax}} \ [L(\boldsymbol{\theta})],$$

em que $\widehat{\boldsymbol{\theta}}$ é o EMV para $\boldsymbol{\theta}$. Ao maximizar $L(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$ estaremos determinando o valor de $\boldsymbol{\theta}$ com maior plausibilidade de ter gerado a amostra $\boldsymbol{Y}$.

Estimação dos parâmetros

Pela independência de $Y_1, \ldots, Y_n$, a função de verossimilhança fica dada por

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(y_i; \theta_i, \phi) \\
&= \prod_{i=1}^{n} \exp\{\phi[\theta_i y_i - b(\theta_i)] + c(y_i; \phi)\} \\
&= \exp\left\{\sum_{i=1}^{n} [\phi[\theta_i y_i - b(\theta_i)] + c(y_i; \phi)]\right\} \\
&= \exp\left\{\phi \sum_{i=1}^{n} [\theta_i y_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i; \phi)\right\}.
\end{aligned}
$$

Estimação dos parâmetros

O valor de $\boldsymbol{\theta}$ que maximiza $L(\boldsymbol{\theta})$ é o mesmo que maximiza $\ell(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$ pois a função logaritmo é monótona crescente. Assim, o logaritmo da função de verossimilhança para $\boldsymbol{\theta}$ é

$$\ell(\boldsymbol{\theta}) = \phi \sum_{i=1}^{n} [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i; \phi).$$

Para encontrar o valor de $\boldsymbol{\theta}$ que maximiza $\ell(\boldsymbol{\theta})$, podemos derivar $\ell(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$, e posteriormente, igualar a derivada obtida ao vetor de zeros de dimensão $(p+1)$.

O vetor escore para $\boldsymbol{\theta}$ é definido por

$$U(\boldsymbol{\theta}) = \begin{bmatrix} U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ U_{\phi}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} \end{bmatrix}.$$

Estimação dos parâmetros

Daí, para $j = 1, 2, \ldots, p$, temos

$$U_{\beta_j} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \phi \sum_{i=1}^{n} \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\}.$$

Sabemos que

$$\mu_i = b'(\theta_i) = \frac{db(\theta_i)}{d\theta_i}; \quad \frac{\partial \eta_i}{\partial \beta_j} = x_{ij};$$

$$V_i = V(\mu_i) = b''(\theta_i) = \frac{d^2 b(\theta_i)}{d\theta_i^2} = \frac{d\mu_i}{d\theta_i} = \frac{1}{\frac{d\theta_i}{d\mu_i}};$$

$$\frac{d\mu_i}{d\eta_i} = \frac{d[g^{-1}(\eta_i)]}{d\eta_i} = \frac{1}{g'(g^{-1}(\eta_i))} = \frac{1}{g'(\mu_i)}.$$

Estimação dos parâmetros

Assim, ficamos com

$$U_{\beta_j} = \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij} = \sum_{i=1}^{n} \phi \left[\sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) x_{ij} \right],$$

em que

$$\omega_i = \frac{(d\mu_i/d\eta_i)^2}{V_i} > 0$$

é visto com um peso.

Note que o numerador de ω_i depende da função de ligação e o denominador de ω_i caracteriza a distribuição de probabilidades da FE.

Daí, o vetor escore para $\boldsymbol{\beta}$ é

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \phi X^\top W^{1/2} V^{-1/2} (\boldsymbol{y} - \boldsymbol{\mu}),$$

Estimação dos parâmetros

em que $W = \mathrm{diag}\{\omega_1, \omega_2, \ldots, \omega_n\}$, $V = \mathrm{diag}\{V_1, V_2, \ldots, V_n\}$,

$$X = \begin{bmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{bmatrix}, \quad \boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Por fim, o escore para ϕ é dado por

$$U_\phi(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} = \sum_{i=1}^{n} [y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n} c'(y_i; \phi),$$

em que $c'(y_i; \phi) = dc(y_i; \phi)/d\phi$.

Estimação dos parâmetros

Note que se $g(\cdot)$ é a ligação canônica ($\theta_i = \eta_i$),

$$\frac{d\mu_i}{d\eta_i} = \frac{d\mu_i}{d\theta_i}\frac{d\theta_i}{d\eta_i} = V_i,$$

pois $d\theta_i/d\eta_i = 1$.

Assim, sob ligação canônica tem-se que os pesos se reduzem a

$$\omega_i = \frac{(d\mu_i/d\eta_i)^2}{V_i} = \frac{V_i^2}{V_i} = V_i,$$

e o vetor escore para $\boldsymbol{\beta}$ fica expresso por

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \phi X^\top V^{1/2} V^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}) = \phi X^\top (\boldsymbol{y} - \boldsymbol{\mu}).$$

Estimação dos parâmetros

Para encontrar $\widehat{\theta}$, fazemos

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{0}, \quad U_{\phi}(\boldsymbol{\theta}) = 0.$$

Mas,

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{0} \Leftrightarrow X^{\top}\widehat{W}^{1/2}\widehat{V}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}) = \mathbf{0}.$$

Note que o sistema de equações acima não depende de $\widehat{\phi}$. Assim, o parâmetro ϕ não precisa ser estimado conjuntamente com $\boldsymbol{\beta}$.

De forma geral, $\widehat{\boldsymbol{\beta}}$ não possui forma fechada. Por esta razão, estudaremos um processo iterativo para encontrar o valor de $\widehat{\boldsymbol{\beta}}$.

Estimação dos parâmetros

De $U_\phi(\boldsymbol{\theta}) = 0$, obtemos que

$$-\sum_{i=1}^{n}[y_i\widehat{\theta}_i - b(\widehat{\theta}_i)] = \sum_{i=1}^{n} c'(y_i; \widehat{\phi}).$$

Quando estivermos sob os modelos normal e normal inversa, veremos que $\widehat{\phi}$ possui forma fechada.

Lembre-se que há situações em que não será necessário estimar ϕ pois este será um valor conhecido.

Estimação dos parâmetros

A matriz de informação de Fisher é obtida a partir de

$$K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} K_{\boldsymbol{\beta}\boldsymbol{\beta}} & K_{\boldsymbol{\beta}\phi} \\ K_{\phi\boldsymbol{\beta}} & K_{\phi\phi} \end{bmatrix} = \begin{bmatrix} -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right) & -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \phi}\right) \\ -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \phi \partial \boldsymbol{\beta}^\top}\right) & -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \phi \partial \phi}\right) \end{bmatrix}.$$

Para $j, l = 1, 2, \ldots, p$, tem-se que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} &= \frac{\partial U_{\beta_j}}{\partial \beta_l} = \frac{\partial}{\partial \beta_l}\left\{\phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij}\right\} \\ &= \phi \sum_{i=1}^{n} \frac{\partial[(y_i - \mu_i)]}{\partial \beta_l} V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} + \phi \sum_{i=1}^{n} (y_i - \mu_i) x_{ij} \frac{\partial}{\partial \beta_l}\left[V_i^{-1} \frac{d\mu_i}{d\eta_i}\right] \end{aligned}$$

Estimação dos parâmetros

A matriz de informação de Fisher é obtida a partir de

$$K_{\boldsymbol{\theta\theta}} = \begin{bmatrix} K_{\boldsymbol{\beta\beta}} & K_{\boldsymbol{\beta}\phi} \\ K_{\phi\boldsymbol{\beta}} & K_{\phi\phi} \end{bmatrix} = \begin{bmatrix} -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right) & -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\phi}\right) \\ -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\phi\partial\boldsymbol{\beta}^\top}\right) & -\operatorname{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\phi\partial\phi}\right) \end{bmatrix}.$$

Para $j, l = 1, 2, \ldots, p$, tem-se que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_j\partial\beta_l} &= \frac{\partial U_{\beta_j}}{\partial\beta_l} = \frac{\partial}{\partial\beta_l}\left\{\phi\sum_{i=1}^{n}(y_i-\mu_i)\frac{1}{V_i}\frac{d\mu_i}{d\eta_i}x_{ij}\right\} \\ &= \phi\sum_{i=1}^{n}\frac{\partial[(y_i-\mu_i)]}{\partial\beta_l}V_i^{-1}\frac{d\mu_i}{d\eta_i}x_{ij} + \phi\sum_{i=1}^{n}(y_i-\mu_i)x_{ij}\frac{\partial}{\partial\beta_l}\left[V_i^{-1}\frac{d\mu_i}{d\eta_i}\right] \end{aligned}$$

Estimação dos parâmetros

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} &= \phi \sum_{i=1}^{n} \left\{ -\frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_l} V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\
&+ \phi \sum_{i=1}^{n} (y_i - \mu_i)\, x_{ij} \left\{ \frac{\partial}{\partial \beta_l} \left[V_i^{-1} \right] \frac{d\mu_i}{d\eta_i} + V_i^{-1} \frac{d^2\mu_i}{d\eta_i^2} \frac{\partial \eta_i}{\partial \beta_l} \right\} \\
&= -\phi \sum_{i=1}^{n} V_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il} + \phi \sum_{i=1}^{n} (y_i - \mu_i)\, x_{ij} \frac{d^2\theta_i}{d\mu_i^2} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_l} \frac{d\mu_i}{d\eta_i} \\
&+ \phi \sum_{i=1}^{n} (y_i - \mu_i)\, x_{ij} V_i^{-1} \frac{d^2\mu_i}{d\eta_i^2} x_{il},
\end{aligned}
$$

desde que

$$
V_i^{-1} = \frac{d\theta_i}{d\mu_i}.
$$

Segue que

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} = -\phi \sum_{i=1}^{n} V_i^{-1} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{il} + \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{d^2\theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{il}x_{ij}$$
$$+ \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{d^2\mu_i}{d\eta_i^2} V_i^{-1} x_{ij}x_{il}.$$

Portanto,

$$-\,\mathrm{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l}\right\} = \phi \sum_{i=1}^{n} V_i^{-1} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{il}$$
$$= \phi \sum_{i=1}^{n} \omega_i x_{ij}x_{il},$$

desde que

$$\mathrm{E}\,(Y_i - \mu_i) = 0, \quad \omega_i = \frac{(d\mu_i|d\eta_i)^2}{V_i}.$$

Matricialmente, a matriz de informação de Fisher para $\boldsymbol{\beta}$ é

$$K_{\boldsymbol{\beta\beta}} = -\,\mathrm{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right\} = \phi X^\top W X.$$

Se W é uma matriz positiva definida, então $r(X^\top W X) = r(X)$. Se X é de posto completo, então $X^\top W X$ é invertível e positiva definida.

Se a ligação é a canônica ($\theta_i = \eta_i$), tem-se $\omega_i = V_i$, e consequentemente,

$$K_{\boldsymbol{\beta\beta}} = \phi X^\top V X.$$

Estimação dos parâmetros

Com relação ao parâmetro ϕ,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \phi^2} = \frac{\partial}{\partial \phi}\left\{\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n} c'(y_i;\phi)\right\} = \sum_{i=1}^{n} c''(y_i;\phi),$$

em que

$$c''(y_i;\phi) = \frac{d^2 c(y_i;\phi)}{d\phi^2}.$$

Daí, a informação de Fisher para ϕ é

$$K_{\phi\phi} = -\operatorname{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \phi^2}\right\} = -\operatorname{E}\left\{\sum_{i=1}^{n} c''(Y_i;\phi)\right\} = -\sum_{i=1}^{n}\operatorname{E}[c''(Y_i;\phi)].$$

Finalmente, para $j = 1, 2, \ldots, p$,

$$\frac{\partial^2 \ell(\theta)}{\partial \beta_j \partial \phi} = \frac{\partial}{\partial \phi} \left\{ \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} = \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij}.$$

Implicando que

$$K_{\boldsymbol{\beta}\phi} = K_{\phi\boldsymbol{\beta}}^\top = -\,\mathrm{E}\left\{ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \phi} \right\} = \mathbf{0}_{p\times 1}.$$

A matriz de informação de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \ \phi)^\top$ é dada por

$$K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathrm{diag}\{K_{\boldsymbol{\beta}\boldsymbol{\beta}}, K_{\phi\phi}\} = \begin{bmatrix} \phi X^\top W X & \mathbf{0} \\ \mathbf{0} & -\sum_{i=1}^{n} \mathrm{E}[c''(Y_i; \phi)] \end{bmatrix}.$$

Como a matriz $K_{\boldsymbol{\theta}\boldsymbol{\theta}}$ é bloco diagonal, concluímos que os parâmetros $\boldsymbol{\beta}$ e ϕ são ortogonais.

Normalidade assintótica do EMV

Para n grande, pode-se mostrar que

$$\widehat{\boldsymbol{\theta}} \sim \mathrm{N}_{p+1}\left(\boldsymbol{\theta}, K_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\right),$$

em que

$$K_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} = \mathrm{diag}\{K_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}, K_{\phi\phi}^{-1}\} = \begin{bmatrix} \phi^{-1}(X^\top WX)^{-1} & \mathbf{0} \\ \mathbf{0} & \{-\sum_{i=1}^n \mathrm{E}[c''(Y_i;\phi)]\}^{-1} \end{bmatrix}.$$

Particularmente, para n grande,

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_p\left(\boldsymbol{\beta}, \phi^{-1}(X^\top WX)^{-1}\right),$$

$$\widehat{\phi} \sim \mathrm{N}\left(\phi, \left\{-\sum_{i=1}^n \mathrm{E}[c''(Y_i;\phi)]\right\}^{-1}\right).$$

Casos particulares - Distribuição normal

Sabemos que $\theta_i = \mu_i$, $\phi = 1/\sigma^2$ e $V_i = 1$. Daí, os pesos ficam expressos por

$$\omega_i = \frac{(d\mu_i | d\eta_i)^2}{V_i} = \left(\frac{d\mu_i}{d\theta_i}\frac{d\theta_i}{d\eta_i}\right)^2 = \left(\frac{d\theta_i}{d\eta_i}\right)^2.$$

O vetor escore para $\boldsymbol{\beta}$ fica expresso por

$$U_{\boldsymbol{\beta}} = \frac{1}{\sigma^2}X^\top W^{1/2}(\boldsymbol{y} - \boldsymbol{\mu}), \quad K_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{1}{\sigma^2}X^\top W X.$$

Se a ligação é canônica ($\theta_i = \eta_i$) obtemos que $\omega_i = 1, \forall i$, reduzindo as expressões para

$$U_{\boldsymbol{\beta}} = \frac{1}{\sigma^2}X^\top(\boldsymbol{y} - \boldsymbol{\mu}), \quad K_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{1}{\sigma^2}X^\top X.$$

Casos particulares - Distribuição normal

Sabemos que $\widehat{\phi}$ é obtido tal que $U_{\phi} = 0$, isto é,

$$-\sum_{i=1}^{n}[y_i\widehat{\theta}_i - b(\widehat{\theta}_i)] = \sum_{i=1}^{n} c'(y_i;\widehat{\phi}).$$

Neste caso,

$$c\,(y_i;\phi) = \frac{1}{2}\left[\log\left(\frac{\phi}{2\pi}\right) - \phi y_i^2\right].$$

Daí, obtemos que

$$c'\,(y_i;\phi) = \frac{1}{2}\left[\frac{2\pi}{\phi}\frac{1}{2\pi} - y_i^2\right] = \frac{1}{2\phi} - \frac{{y_i}^2}{2}.$$

Também, $\widehat{\theta}_i = \widehat{\mu}_i$ e $b(\widehat{\theta}_i) = \widehat{\mu}_i^2/2$.

Assim, $\widehat{\phi}$ é tal que

$$-\sum_{i=1}^{n}\left[y_i\widehat{\mu}_i - \frac{\widehat{\mu}_i^2}{2}\right] = \sum_{i=1}^{n}\left[\frac{1}{2\widehat{\phi}} - \frac{{y_i}^2}{2}\right] \Leftrightarrow$$

$$-\sum_{i=1}^{n} y_i\widehat{\mu}_i + \frac{1}{2}\sum_{i=1}^{n}\widehat{\mu}_i^2 = \frac{n}{2\widehat{\phi}} - \frac{1}{2}\sum_{i=1}^{n} y_i^2 \Leftrightarrow$$

$$\frac{n}{\widehat{\phi}} = \sum_{i=1}^{n}(y_i^2 - 2y_i\widehat{\mu}_i + \widehat{\mu}_i^2) \Leftrightarrow$$

$$\widehat{\phi} = \frac{n}{D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})},$$

em que

$$D(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n}(y_i - \widehat{\mu}_i)^2.$$

Casos particulares - Distribuição normal

Também, temos que

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

e portanto, a matriz de informação de Fisher para ϕ fica dada por

$$K_{\phi\phi} = -\sum_{i=1}^{n} \mathrm{E}\left[c''(Y_i;\phi)\right] = \sum_{i=1}^{n} \frac{1}{2\phi^2} = \frac{n}{2\phi^2}.$$

Daí, a variância assintótica de $\widehat{\phi}$ é dada por

$$\mathrm{Var}(\widehat{\phi}) = \frac{2\phi^2}{n} \quad \Rightarrow \quad \widehat{\mathrm{Var}}(\widehat{\phi}) = \frac{2\widehat{\phi}^2}{n}.$$

Casos particulares - Distribuição Poisson

Sabemos que $\phi = 1$ (constante), $\theta_i = \log(\mu_i)$ e $V_i = \mu_i$. Daí, temos que

$$\omega_i = \frac{1}{V_i}\left(\frac{d\mu_i}{d\theta_i}\frac{d\theta_i}{d\eta_i}\right)^2 = \frac{1}{\mu_i}\left(\mu_i\frac{d\theta_i}{d\eta_i}\right)^2 = \mu_i\left(\frac{d\theta_i}{d\eta_i}\right)^2.$$

De forma geral,

$$U_{\boldsymbol{\beta}} = X^\top W^{1/2} V^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}), \quad K_{\boldsymbol{\beta}\boldsymbol{\beta}} = X^\top W X.$$

Em particular, se $g(\cdot)$ é a ligação canônica $(\theta_i = \eta_i)$, tem-se que $\omega_i = V_i = \mu_i, \forall i$. Neste caso, os pesos são as próprias médias.

Casos particulares - Distribuição Poisson

Se $g(\mu_i) = \sqrt{\mu_i}$, tem-se $\sqrt{\mu_i} = \eta_i \Rightarrow \mu_i = {\eta_i}^2$. Como $\theta_i = \log(\mu_i) = \log\left(\eta_i^2\right)$. Segue que

$$\frac{d\theta_i}{d\eta_i} = \frac{1}{\eta_i^2} 2\eta_i = \frac{2}{\eta_i} = \frac{2}{\sqrt{\mu_i}}.$$

$$\Rightarrow \omega_i = \mu_i \left(\frac{2}{\sqrt{\mu_i}}\right)^2 = 4, \quad \forall i.$$

Assim, se a ligação é a raiz quadrada, os pesos ω_i são constantes.

Nesse caso,

$$U_{\boldsymbol{\beta}} = 2X^\top V^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}), \quad K_{\boldsymbol{\beta}\boldsymbol{\beta}} = 4X^\top X.$$

Casos particulares - Distribuição binomial

Sabemos que $\phi = n$ (conhecido), $\theta_i = \log\left(\mu_i / \left(1 - \mu_i\right)\right)$ e $V_i = \mu_i\left(1 - \mu_i\right)$.

Daí,

$$
\begin{aligned}
\omega_i &= \frac{1}{V_i}\left(\frac{d\mu_i}{d\theta_i}\frac{d\theta_i}{d\eta_i}\right)^2 \\
&= \frac{1}{\mu_i(1-\mu_i)}\left[\mu_i(1-\mu_i)\frac{d\theta_i}{d\eta_i}\right]^2 \\
&= \mu_i(1-\mu_i)\left(\frac{d\theta_i}{d\eta_i}\right)^2.
\end{aligned}
$$

Se a ligação é canônica, então $\omega_i = \mu_i(1 - \mu_i)$.

Casos particulares - Distribuição binomial

Entretanto, pode-se encontrar na literatura (e nos *softwares*) os pesos definidos através de

$$\omega_i = n\mu_i(1-\mu_i)\left(\frac{d\theta_i}{d\eta_i}\right)^2.$$

Para ligação canônica, $\omega_i = n\mu_i(1-\mu_i)$.

Isto se deve ao fato que o MLG binomial pode ser definido sem fazer a mudança de escala $Y^* = Y/n$.

Casos particulares - Distribuição gama

Neste caso, tem-se que $V_i = \mu_i^2$ e $\theta_i = -1/\mu_i$. Então,

$$\omega_i = \frac{1}{\mu_i^2}\left[\mu_i^2 \frac{d\theta_i}{d\eta_i}\right]^2 = \mu_i^2 \left(\frac{d\theta_i}{d\eta_i}\right)^2.$$

Se a ligação é a canônica, então $\omega_i = \mu_i^2 \quad \forall i$.

Quando a ligação escolhida é a logarítmica, tem-se

$$\log(\mu_i) = \eta_i \;\Rightarrow\; \mu_i = \mathrm{e}^{\eta_i} \;\Rightarrow\; -\frac{1}{\theta_i} = \mathrm{e}^{\eta_i} \;\Rightarrow\; \theta_i = -\mathrm{e}^{-\eta_i}$$

$$\Rightarrow \frac{d\theta_i}{d\eta_i} = -\mathrm{e}^{-\eta_i} \times (-1) = \mathrm{e}^{-\eta_i} = \mathrm{e}^{-\log(\mu_i)} = \mathrm{e}^{\log\left(\mu_i^{-1}\right)} = \frac{1}{\mu_i}$$

$$\Rightarrow \omega_i = \mu_i^2 \left(\frac{1}{\mu_i}\right)^2 = 1, \forall i.$$

Casos particulares - Distribuição gama

Também, sabemos que $\widehat{\phi}$ é obtido tal que $U_\phi = 0$, isto é,

$$-\sum_{i=1}^{n}[y_i\widehat{\theta}_i - b(\widehat{\theta}_i)] = \sum_{i=1}^{n} c'(y_i; \widehat{\phi}).$$

Neste caso,

$$c\,(y_i; \phi) = (\phi - 1)\log(y_i) + \phi\log(\phi) - \log(\Gamma(\phi)).$$

Daí,

$$\begin{aligned} c'(y_i; \phi) &= \log(y_i) + \log(\phi) + \frac{\phi}{\phi} - \psi(\phi) \\ &= \log(y_i) + \log(\phi) + 1 - \psi(\phi). \end{aligned}$$

em que $\psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$ é a função digama.

Casos particulares - Distribuição gama

Sabendo que $\theta_i = -1/\mu_i$ e $b(\theta_i) = \log(\mu_i)$, ficamos com

$$\begin{aligned}
-\sum_{i=1}^{n}\left[-\frac{y_i}{\widehat{\mu}_i} - \log(\widehat{\mu}_i)\right] &= \sum_{i=1}^{n}[\log(y_i) + \log(\widehat{\phi}) + 1 - \psi(\widehat{\phi})] \\
\sum_{i=1}^{n}\left[\frac{y_i}{\widehat{\mu}_i} - \log\left(\frac{y_i}{\widehat{\mu}_i}\right)\right] &= n[\log(\widehat{\phi}) + 1 - \psi(\widehat{\phi})].
\end{aligned}$$

Note que não é possível isolar $\widehat{\phi}$ na equação acima, e portanto, $\widehat{\phi}$ não possui forma fechada.

Casos particulares - Distribuição gama

Também,

$$c''(y_i;\phi) = \frac{1}{\phi} - \psi'(\phi), \quad \psi'(\phi) = \frac{d\psi(\phi)}{d\phi},$$

com $\psi'(\cdot)$ sendo a função trigama.

Daí, segue que

$$K_{\phi\phi} = -\sum_{i=1}^{n}\left[\frac{1}{\phi} - \psi'(\phi)\right] = -\frac{n\left[1-\phi\psi'(\phi)\right]}{\phi} = \frac{n\left[\phi\psi'(\phi)-1\right]}{\phi}.$$

Logo, a variância assintótica de $\widehat{\phi}$ é dada por

$$\mathrm{Var}(\widehat{\phi}) = \frac{\phi}{n\left[\phi\psi'(\phi)-1\right]} \quad \Rightarrow \quad \widehat{\mathrm{Var}}(\widehat{\phi}) = \frac{\widehat{\phi}}{n\left[\widehat{\phi}\psi'(\widehat{\phi})-1\right]}.$$

Casos particulares - Distribuição normal inversa

Neste caso, $V_i = \mu_i^3$ e $\theta_i = -1/2\mu_i^2$. Segue que

$$\omega_i = \frac{1}{\mu_i^3}\left[\mu_i^3 \frac{d\theta_i}{d\eta_i}\right]^2 = \mu_i^3 \left(\frac{d\theta_i}{d\eta_i}\right)^2.$$

Se a ligação é a canônica, então $\omega_i = \mu_i^3, \forall i$.

Se $g(\mu_i) = \log(\mu_i)$, tem-se $\mu_i = \mathrm{e}^{\eta_i}$. Daí, a relação entre θ_i e η_i é obtida de

$$\mu_i^2 = -\frac{1}{2\theta_i} \;\Rightarrow\; \mu_i = \left(-\frac{1}{2\theta_i}\right)^{1/2} \;\Rightarrow\; \mathrm{e}^{\eta_i} = \left(-\frac{1}{2\theta_i}\right)^{1/2}$$

$$\mathrm{e}^{2\eta_i} = -\frac{1}{2\theta_i} \;\Rightarrow\; \theta_i = -\frac{1}{2\mathrm{e}^{2\eta_i}} \;\Rightarrow\; \theta_i = \frac{-\mathrm{e}^{-2\eta_i}}{2}.$$

Casos particulares - Distribuição normal inversa

Assim, obtemos que

$$\frac{d\theta_i}{d\eta_i} = -\frac{1}{2}\mathrm{e}^{-2\eta_i}(-2) = \mathrm{e}^{-2\eta_i} = \mathrm{e}^{-2\log(\mu_i)} = \mu_i^{-2} = \frac{1}{\mu_i^2}.$$

Portanto, se a ligação é logarítmica, os pesos são dados por

$$\omega_i = \mu_i^3\left(\frac{1}{\mu_i^2}\right)^2 = \frac{1}{\mu_i}, \forall i.$$

Ainda, sabemos que

$$c(y_i;\phi) = \frac{1}{2}\log\left(\frac{\phi}{2\pi y_i^3}\right) - \frac{\phi}{2y_i}.$$

Casos particulares - Distribuição normal inversa

Assim, obtemos que

$$c'(y_i;\phi) = \frac{1}{2}\frac{2\pi y_i^3}{\phi}\frac{1}{2\pi y_i^3} - \frac{1}{2y_i} = \frac{1}{2\phi} - \frac{1}{2y_i}.$$

Ainda, sabendo que $b(\widehat{\theta}_i) = -1/\widehat{\mu}_i$, temos $\widehat{\phi}$ é obtido tal que

$$-\sum_{i=1}^{n}[y_i\widehat{\theta}_i - b(\widehat{\theta}_i)] = \sum_{i=1}^{n} c'(y_i;\widehat{\phi})$$

$$-\sum_{i=1}^{n}\left[-\frac{y_i}{2\widehat{\mu}_i^2} + \frac{1}{\widehat{\mu}_i}\right] = \sum_{i=1}^{n}\left[\frac{1}{2\widehat{\phi}} - \frac{1}{2y_i}\right] \Leftrightarrow$$

$$\sum_{i=1}^{n}\left[\frac{(y_i^2 - 2\widehat{\mu}_i y_i + \widehat{\mu}_i^2)}{2y_i\widehat{\mu}_i^2}\right] = \frac{n}{2\widehat{\phi}}.$$

Casos particulares - Distribuição normal inversa

Isolando $\widehat{\phi}$, obtemos que

$$\widehat{\phi} = \frac{n}{D(\boldsymbol{y};\widehat{\boldsymbol{\mu}})},$$

com

$$D(\boldsymbol{y};\widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu}_i)^2}{y_i \widehat{\mu}_i^2}.$$

Por fim, veja que

$$c''(y_i;\phi) = -\frac{1}{2\phi^2} \Rightarrow K_{\phi\phi} = -\sum_{i=1}^{n} -\frac{1}{2\phi^2} = \frac{n}{2\phi^2}.$$

Consequentemente,

$$\mathrm{Var}(\widehat{\phi}) = \frac{2\phi^2}{n} \Rightarrow \widehat{\mathrm{Var}}(\widehat{\phi}) = \frac{2\widehat{\phi}^2}{n}.$$

Processo iterativo para estimar β

Vimos que o EMV para $\boldsymbol{\beta}$ é obtido encontrando o valor de $\boldsymbol{\beta}$ tal que

$$U_{\boldsymbol{\beta}} = \mathbf{0} \Leftrightarrow X^\top \widehat{W}^{1/2} \widehat{V}^{-1/2} (\mathbf{y} - \widehat{\boldsymbol{\mu}}) = \mathbf{0}, \tag{2}$$

que não depende de $\widehat{\phi}$.

Para encontrar o valor de $\boldsymbol{\beta}$ que resolve o sistema de equações não linear (2), considere a expansão do vetor escore $U_{\boldsymbol{\beta}}$ em série de Taylor em torno de $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$:

$$U_{\boldsymbol{\beta}} \cong U_{\boldsymbol{\beta}}^{(0)} + J_{\boldsymbol{\beta\beta}}{}^{(0)} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}), \tag{3}$$

em que $J_{\boldsymbol{\beta\beta}} = U'_{\boldsymbol{\beta}} = \partial U_{\boldsymbol{\beta}} / \partial \boldsymbol{\beta}^\top$ é a matriz hessiana para $\boldsymbol{\beta}$, $U_{\boldsymbol{\beta}}^{(0)}$ e $J_{\boldsymbol{\beta\beta}}{}^{(0)}$ são avaliadas em $\boldsymbol{\beta}^{(0)}$ que é um valor inicial ("chute") preestabelecido.

Processo iterativo para estimar β

Seja $\beta^{(1)}$ o valor β atualizado no passo 1 do procedimento iterativo após substituir $\beta^{(0)}$ em (3) e isolar β. Assim, segue que

$$U_{\boldsymbol{\beta}}^{(0)} + {J_{\boldsymbol{\beta\beta}}}^{(0)}(\beta^{(1)} - \beta^{(0)}) = \mathbf{0} \quad \Leftrightarrow$$

$$U_{\boldsymbol{\beta}}^{(0)} + {J_{\boldsymbol{\beta\beta}}}^{(0)}\beta^{(1)} - {J_{\boldsymbol{\beta\beta}}}^{(0)}\beta^{(0)} = \mathbf{0} \Leftrightarrow$$

$${J_{\boldsymbol{\beta\beta}}}^{(0)}\beta^{(1)} = {J_{\boldsymbol{\beta\beta}}}^{(0)}\beta^{(0)} - U_{\boldsymbol{\beta}}^{(0)} \Leftrightarrow$$

$$\beta^{(1)} = \left\{J_{\boldsymbol{\beta\beta}}^{-1}\right\}^{(0)} {J_{\boldsymbol{\beta\beta}}}^{(0)}\beta^{(0)} - \left\{J_{\boldsymbol{\beta\beta}}^{-1}\right\}^{(0)} U_{\boldsymbol{\beta}}^{(0)} \Leftrightarrow$$

$$\beta^{(1)} = \beta^{(0)} - \left\{J_{\boldsymbol{\beta\beta}}^{-1}\right\}^{(0)} U_{\boldsymbol{\beta}}^{(0)},$$

desde que $J_{\boldsymbol{\beta\beta}}^{-1}$ exista.

Processo iterativo para estimar β

Repetindo o procedimento m vezes, obtemos:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left\{ J^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}} \right\}^{(m)} U^{(m)}_{\boldsymbol{\beta}}, \quad m = 0, 1, 2, \ldots,$$

em que $\boldsymbol{\beta}^{(m+1)}$ denota o valor de $\boldsymbol{\beta}$ no passo $(m+1)$ do processo iterativo.

Observações:

- A matriz hessiana $J_{\boldsymbol{\beta}\boldsymbol{\beta}}$ pode não ser invertível;
- $\ell(\boldsymbol{\beta}^{(m+1)}) > \ell(\boldsymbol{\beta}^{(m)}), \forall m$, pode não ocorrer.

Estes possíveis problemas podem ser resolvidos substituindo a matriz $-J_{\boldsymbol{\beta}\boldsymbol{\beta}}$ (Fisher observada) pela matriz $K_{\boldsymbol{\beta}\boldsymbol{\beta}}$ (Fisher esperada) que sempre será positiva definida.

Processo iterativo para estimar β

Ao fazer esta modificação no procedimento iterativo estaremos utilizando o procedimento Escore de Fisher definido por

$$\beta^{(m+1)} = \beta^{(m)} + \left\{K_{\beta\beta}^{-1}\right\}^{(m)} U_{\beta}^{(m)}, \quad m = 0, 1, 2, \ldots, \tag{4}$$

em que

$$U_{\beta}^{(m)} = \phi X^{\top} \left\{W^{1/2}\right\}^{(m)} \left\{V^{-1/2}\right\}^{(m)} (\boldsymbol{y} - \boldsymbol{\mu}^{(m)}),$$

$$\left\{K_{\beta\beta}^{-1}\right\}^{(m)} = \phi^{-1} \left\{(X^{\top} W X)^{-1}\right\}^{(m)}.$$

Processo iterativo para estimar β

Ainda, observe que

$$\begin{aligned}\beta^{(m)} &= \left\{(X^\top W X)^{-1}\right\}^{(m)} \left[X^\top W^{(m)} X\right] \beta^{(m)} \\ &= \left\{(X^\top W X)^{-1}\right\}^{(m)} X^\top W^{(m)} \eta^{(m)},\end{aligned}$$

em que $\eta^{(m)} = X\beta^{(m)}$ é preditor linear no passo m do processo iterativo Escore de Fisher.

Substituindo a expressão acima de $\beta^{(m)}$ no processo iterativo em (4), ficaremos com

Processo iterativo para estimar β

$$
\begin{aligned}
\beta^{(m+1)} &= \beta^{(m)} + \left\{K_{\beta\beta}^{-1}\right\}^{(m)} U_{\beta}^{(m)} \\
&= \left\{(X^\top WX)^{-1}\right\}^{(m)} X^\top W^{(m)} \boldsymbol{\eta}^{(m)} \\
&+ \phi^{-1}\left\{(X^\top WX)^{-1}\right\}^{(m)} \phi X^\top \left\{W^{1/2}\right\}^{(m)} \left\{V^{-1/2}\right\}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\
&= \left\{(X^\top WX)^{-1}\right\}^{(m)} X^\top W^{(m)} \left\{\boldsymbol{\eta}^{(m)}\right. \\
&+ \left.\left\{W^{-1/2}\right\}^{(m)} \left\{V^{-1/2}\right\}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})\right\} \\
&= \left\{(X^\top WX)^{-1}\right\}^{(m)} X^\top W^{(m)} \mathbf{z}^{(m)},
\end{aligned}
$$

em que

$$
\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \left\{W^{-1/2}\right\}^{(m)} \left\{V^{-1/2}\right\}^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)}).
$$

Processo iterativo para estimar β

O vetor de variáveis $\boldsymbol{z}$ é vista como uma variável resposta modificada. As entradas do vetor $\boldsymbol{z} = (z_1 \ z_2 \ \cdots \ z_n)^\top$ são dadas por

$$z_i = \eta_i + \frac{y_i - \mu_i}{\sqrt{\omega_i V_i}}, \quad i = 1, 2, \ldots, n,$$

com $\omega_i = (d\mu_i / d\eta_i)^2 / V_i$.

Revisão. O estimador de mínimos quadrados ponderados para $\boldsymbol{\beta}$ no modelo linear $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$ e $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 V$ com V conhecida é

$$\widehat{\boldsymbol{\beta}}_G = (X^\top V^{-1} X)^{-1} X^\top V^{-1} \boldsymbol{Y}.$$

Processo iterativo para estimar β

Observe que o processo Escore de Fisher é atualizado por meio da expressão

$$\beta^{(m+1)} = \left\{ (X^\top W X)^{-1} \right\}^{(m)} X^\top W^{(m)} \boldsymbol{z}^{(m)} \tag{5}$$

que se assemelha com $\widehat{\beta}_G$.

Por esta razão que o processo iterativo de Escore de Fisher pode ser visto como um processo iterativo de mínimos quadrados reponderados.

Note que a matriz W possui o papel de reponderar $\widehat{\beta}$, ou seja, W é vista como uma matriz de ponderações ou "pesos" que muda a cada passo do processo iterativo.

Processo iterativo para estimar β

Para estabelecer um valor inicial para $\boldsymbol{\beta}^{(0)}$, considere que

$$\boldsymbol{\mu}^{(0)} = \boldsymbol{y} \;\;\Rightarrow\;\; \boldsymbol{\eta}^{(0)} = g(\boldsymbol{y}).$$

Daí, o i-ésimo valor da variável transformada z_i no passo 0 é calculado por meio de

$$z_i^{(0)} = \eta_i^{(0)} + \frac{y_i - \mu_i^{(0)}}{\sqrt{\omega_i^{(0)} V_i^{(0)}}},$$

em que

$$\eta_i^{(0)} = g(y_i), \quad \mu_i^{(0)} = y_i, \quad V_i^{(0)} = V_i\left(\mu_i^{(0)}\right), \quad \omega_i^{(0)} = \frac{[1/g'(\mu_i^{(0)})]^2}{V_i^{(0)}}.$$

Calculado $z_i^{(0)}, \forall i$ e $W^{(0)}$, obtemos $\boldsymbol{\beta}^{(1)}$ através de (5).

Processo iterativo para estimar $\boldsymbol{\beta}$

Consequentemente, com o valor atualizado $\boldsymbol{\beta}^{(1)}$ calcula-se $\boldsymbol{\eta}^{(1)}$ e obtém-se $\boldsymbol{\beta}^{(2)}$. O procedimento segue até que não há mudanças significativas na estimativa de $\boldsymbol{\beta}$.

Quando devemos parar o procedimento iterativo?

Dizemos que há convergência do procedimento iterativo se

$$\left| \frac{\beta_j^{(m+1)} - \beta_j^{(m)}}{\beta_j^{(m)}} \right| < \epsilon, \quad \forall j = 1, 2, \ldots, p, \tag{6}$$

em que $\epsilon > 0$ é um valor pequeno preestabelecido.

Processo iterativo para estimar β

Usualmente, fixam-se dois valores de ϵ. Sejam ϵ_1 e ϵ_2 os valores de ϵ fixados tais que $\epsilon_1 < \epsilon_2$.

Se a condição em (6) é verificada para ϵ_1, dizemos que houve "convergência forte". Note que se a condição é satisfeita para ϵ_1, então é satisfeita para ϵ_2.

Se a condição em (6) é verificada apenas para ϵ_2, dizemos que houve "convergência fraca".

Se a condição em (6) não é satisfeita para ϵ_2, dizemos que o procedimento iterativo não convergiu.

Para o caso de possível não convergência do procedimento iterativo, deve-se fixar um número máximo de iterações. Usualmente fixa-se 200 iterações.

Processo iterativo para estimar ϕ

De forma análoga, o procedimento iterativo para ϕ é

$$\phi^{(m+1)} = \phi^{(m)} - \left\{ {J_{\phi\phi}}^{-1} \right\}^{(m)} U_{\phi}^{(m)}, \quad m = 0, 1, 2, \ldots,$$

em que $\phi^{(m+1)}$ denota o valor de ϕ no passo $(m+1)$.

Também pode-se substituir $-J_{\phi\phi}$ por $K_{\phi\phi}$.

O valor inicial $\phi^{(0)}$ pode ser definido como sendo o estimador de momentos para ϕ.

Por exemplo, sob o modelo gama, temos que

$$\mathrm{E}\left[\frac{(Y_i - \mu_i)^2}{\mu_i^2} \right] = \phi^{-1}.$$

Processo iterativo para estimar ϕ

Daí, segue que um estimador de momentos para ϕ^{-1} sob o modelo gama é

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

Mostra-se que $\widehat{\phi}_M$ é um estimador consistente para ϕ.

Assim, no procedimento iterativo utiliza-se $\phi^{(0)} = \widehat{\phi}_M$.

Perceba que o procedimento iterativo para obtenção da estimativa de ϕ depende da estimativa de $\boldsymbol{\beta}$.

Portanto, primeiro encontra-se o valor $\widehat{\boldsymbol{\beta}}$, em seguida, obtém-se $\widehat{\phi}$.