Modelos Lineares Generalizados

Unidade IV

Terezinha K. A. Ribeiro

terezinha.ribeiro@unb.br

Instituto de Ciências Exatas
Departamento de Estatística
Universidade de Brasília

Organização

- Introdução
- Família exponencial
- Casos particulares
- Estimação dos parâmetros
- Procedimento iterativo
- Função desvio
- Resíduos
- Alavancagem
- Influência

Introdução

Os modelos lineares generalizados duplos (MLGDs) são extensões dos MLGs que admitem modelar a média μ e o parâmetro de precisão ϕ simultaneamente. Assim, a precisão ϕ também terá associada uma estrutura de regressão linear.

Ao admitir que a precisão da distribuição varie com a amostra, esteramos flexibilizando o modelo, e consequentemente, melhorando a qualidade do ajuste global do modelo suposto aos dados.

A partir dos MLGDs além de identificar quais covariáveis afetam o comportamento médio da reposta, também identificaremos quais covariáveis afetam a precisão da resposta.

MLGs Duplos

Os MLGDs podem ser definidos por

i) $Y_i | \boldsymbol{x}_i \overset{\text{ind}}{\sim} \text{FE}(\mu_i, \phi_i)$;

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

ii) $h(\phi_i) = \lambda_i$ com $\lambda_i = \boldsymbol{z}_i^\top \gamma \quad \Leftrightarrow \quad \phi_i = h^{-1}(\lambda_i)$;

- Y_i é a i-ésima variável resposta;
- $\boldsymbol{x}_i^\top = \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} \end{pmatrix} \in \mathbb{R}^p$ é o vetor de valores fixados das p covariáveis para a i-ésima observação no submodelo da média;
- $\boldsymbol{z}_i^\top = \begin{pmatrix} z_{i1} & z_{i2} & \cdots & z_{iq} \end{pmatrix} \in \mathbb{R}^q$ é o vetor de valores fixados das k covariáveis para a i-ésima observação no submodelo da precisão;

MLGs Duplos

- $g(\cdot)$ e $h(\cdot)$ são as respectivas funções de ligação (monótonas e diferenciáveis) para os submodelos da média e da precisão;
- η_i e λ_i são os respectivos preditores lineares dos submodelos da média e da precisão para a i-ésima observação;
- $\mathrm{FE}(\mu_i, \phi_i)$ denota a família exponencial uniparamétrica de média μ_i e parâmetro de precisão ϕ_i com função densidade de probabilidades dada por

$$f(y_i; \theta_i, \phi_i) = \exp\left\{\phi_i\left[y_i\theta_i - b(\theta_i)\right] + c(y_i; \phi_i)\right\},$$

em que

$$c(y_i; \phi_i) = d(\phi_i) + \phi_i a(y_i) + u(y_i),$$

é a decomposição da função $c(y_i; \phi_i)$ válida para as distribuições normal, gama e normal inversa.

MLGs Duplos - Normal

Se $Y_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu_i, \sigma_i^2)$, a função densidade de probabilidades de Y_i é

$$
\begin{aligned}
f(y_i; \mu_i, \sigma_i^2) &= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_i - \mu_i)^2\right\} \\
&= \exp\left\{\frac{1}{\sigma_i^2}\left(\mu_i y_i - \frac{\mu_i^2}{2}\right) - \frac{1}{2}\left[\log(2\pi\sigma_i^2) + \frac{y_i^2}{\sigma_i^2}\right]\right\}.
\end{aligned}
$$

Tome $\phi_i = 1/\sigma_i^2$, $\theta_i = \mu_i$, $b(\theta_i) = \mu_i^2/2 = \theta_i^2/2$, e

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

Para o modelo normal temos

$$
d(\phi_i) = \frac{1}{2}\log(\phi_i), \quad a(y_i) = -\frac{1}{2}y_i^2, \quad u(y_i) = -\frac{1}{2}\log(2\pi)
$$

Se $Y_i \stackrel{\text{ind}}{\sim} \text{gama}(\mu_i, \phi_i)$, a função densidade de probabilidades de Y_i é

$$
\begin{aligned}
f(y_i;\mu_i,\phi_i) &= \frac{1}{\Gamma(\phi_i)}\left(\frac{\phi_i y_i}{\mu_i}\right)^{\phi_i} \exp\left\{-\frac{\phi_i y_i}{\mu_i}\right\}\frac{1}{y_i} \\
&= \exp\{\phi[-y_i/\mu_i - \log(\mu_i)] - \log\Gamma(\phi_i) + \phi_i\log(\phi_i y_i) - \log(y_i)\}.
\end{aligned}
$$

Tome $\phi_i = \phi_i$, $\theta_i = -\frac{1}{\mu_i} \Rightarrow \mu_i = -\theta_i^{-1}$, $b(\theta_i) = \log(\mu_i) = -\log(-\theta_i)$, com $\theta_i < 0$,

$$
\begin{aligned}
c(y_i;\phi_i) &= \phi_i\log(\phi_i y_i) - \log\Gamma(\phi_i) - \log(y_i) \\
&= \phi_i\log(\phi_i) - \log\Gamma(\phi_i) + \phi_i\log(y_i) - \log(y_i).
\end{aligned}
$$

Para o modelo gama temos

$$
d(\phi_i) = \phi_i\log(\phi_i) - \log\Gamma(\phi_i), \quad a(y_i) = \log(y_i), \quad u(y_i) = -\log(y_i).
$$

Se $Y_i \overset{\text{ind}}{\sim} \text{NI}(\mu_i, \phi_i)$, a função densidade de probabilidades de Y_i é

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

Tome $\phi_i = \phi_i$, $\theta_i = -\frac{1}{2\mu_i^2} \Rightarrow \mu_i = (-2\theta_i)^{-1/2}$, $b(\theta_i) = -\frac{1}{\mu_i} = -(-2\theta_i)^{1/2}$, com $\theta_i < 0$,

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

Para o modelo NI temos

$$d(\phi_i) = \frac{1}{2} \log(\phi_i), \quad a(y_i) = -\frac{1}{2y_i}, \quad u(y_i) = -\frac{1}{2} \log(2\pi y_i^3).$$

Família exponencial

Assim, para os modelos normal, gama e NI, vale que

$$
\begin{aligned}
f(y_i;\theta_i,\phi_i) &= \exp\left\{\phi_i\left[y_i\theta_i - b(\theta_i)\right] + d(\phi_i) + \phi_i a(y_i) + u(y_i)\right\} \\
&= \exp\left\{\phi_i\left[y_i\theta_i - b(\theta_i) + a(y_i)\right] + d(\phi_i) + u(y_i)\right\} \\
&= \exp\left\{\phi_i t_i + d(\phi_i) + u(y_i)\right\},
\end{aligned}
$$

em que

$$t_i = y_i\theta_i - b(\theta_i) + a(y_i).$$

Note que t_i não depende de ϕ_i.

Se θ_i for fixado, $f\left(y_i;\theta_i,\phi_i\right)$ é a densidade da família exponencial uniparamétrica de uma variável aleatória T_i com $\theta_i = \phi_i$, $b\left(\theta_i\right) = -d\left(\phi_i\right)$, $\phi = 1$, $c\left(y_i,\phi\right) = u\left(y_i\right)$.

Em outras palavras, $T_1, T_2, \ldots, T_n$ com θ_i fixado são variáveis aleatórias independentes com densidade na forma da família exponencial uniparamétrica.

Família exponencial

Pelas propriedades da família exponencial uniparamétrica, temos que

$$\mu_{T_i} = \mathrm{E}\left(T_i\right) = b'\left(\theta_i\right) = -d'\left(\phi_i\right) = -\frac{d}{d\phi_i}\left[d\left(\phi_i\right)\right];$$

$$\mathrm{Var}\left(T_i\right) = -d''\left(\phi_i\right) = -\frac{d^2}{d\phi_i^2}\left[d\left(\phi_i\right)\right].$$

Para o modelo normal, temos

$$d\left(\phi_i\right) = \frac{1}{2}\log\left(\phi_i\right) \Rightarrow d'\left(\phi_i\right) = \frac{1}{2\phi_i}, \quad d''\left(\phi_i\right) \quad = -\frac{1}{2\phi_i^2}.$$

Família exponencial

Para o modelo gama, temos

$$
\begin{aligned}
d(\phi_i) &= \phi_i \log(\phi_i) - \log\Gamma(\phi_i) \Rightarrow \\
d'(\phi_i) &= \log(\phi_i) + \phi_i \frac{1}{\phi_i} - \psi(\phi_i) \;= 1 + \log(\phi_i) - \psi(\phi_i) \\
d''(\phi_i) &= \frac{1}{\phi_i} - \psi'(\phi_i),
\end{aligned}
$$

em que

$$\psi'(\phi) = \frac{d}{d\phi}\psi(\phi).$$

Para o modelo NI, temos

$$d(\phi_i) = \frac{1}{2}\log(\phi_i) \Rightarrow d'(\phi_i) = \frac{1}{2\phi_i}, \quad d''(\phi_i) \quad = -\frac{1}{2\phi_i^2}.$$

Estimação dos parâmetros

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \ \boldsymbol{\gamma}^\top)^\top \in \mathbb{R}^{p+q}$ o vetor de parâmetros a ser estimado. O logaritmo da função de verossimilhança fica dado por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[f\left(y_i; \theta_i, \phi_i\right)\right] = \sum_{i=1}^{n} \left\{\phi_i t_i + d\left(\phi_i\right) + u\left(y_i\right)\right\},$$

em que $t_i = y_i\theta_i - b\left(\theta_i\right) + a\left(y_i\right)$.

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

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

em que $\widehat{\boldsymbol{\theta}}$ é o EMV para $\boldsymbol{\theta}$.

Estimação dos parâmetros

O vetor escore para θ é definido por

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

Para $j = 1, 2, \ldots, p$, temos

$$\begin{aligned} U_{\beta_j} &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \left\{ \phi_i t_i + d\left(\phi_i\right) + u\left(y_i\right) \right\} \\ &= \sum_{i=1}^{n} \phi_i \frac{\partial}{\partial \beta_j} \left\{ y_i \theta_i - b\left(\theta_i\right) + a\left(y_i\right) \right\} \\ &= \sum_{i=1}^{n} \phi_i \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\}. \end{aligned}$$

Estimação dos parâmetros

Sabemos que

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

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

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

Ficamos com

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

em que

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

Estimação dos parâmetros

O vetor escore para β fica dado por

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

em que $\Phi = \mathrm{diag}\{\phi_1, \phi_2, \ldots, \phi_n\}$, $W = \mathrm{diag}\{\omega_1, \omega_2, \ldots, \omega_n\}$, $V = \mathrm{diag}\{V_1, V_2, \ldots, V_n\}$,

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

Estimação dos parâmetros

Para $j = 1, 2, \ldots, q$, temos

$$\begin{aligned} U_{\gamma_j} &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \gamma_j} \{\phi_i t_i + d(\phi_i) + u(y_i)\} \\ &= \sum_{i=1}^{n} \left\{ t_i \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} + d'(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} \right\}. \end{aligned}$$

Sabemos que

$$\mu_{T_i} = -d'(\phi_i); \quad \frac{\partial \lambda_i}{\partial \gamma_j} = z_{ij};$$

$$\frac{d\phi_i}{d\lambda_i} = \frac{d[h^{-1}(\lambda_i)]}{d\lambda_i} = \frac{1}{h'(h^{-1}(\lambda_i))} = \frac{1}{h'(\phi_i)}.$$

Estimação dos parâmetros

Assim, ficamos com

$$U_{\gamma_j} \quad = \quad \sum_{i=1}^{n} \{t_i - \mathrm{E}(T_i)\} \, \frac{z_{ij}}{h'(\phi_i)}.$$

O vetor escore para γ é

$$U_\gamma(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} = Z^\top H_\gamma^{-1}(\boldsymbol{t} - \boldsymbol{\mu}_T),$$

em que $H_\gamma = \mathrm{diag}\{h'(\phi_1), h'(\phi_2), \ldots, h'(\phi_n)\}$

$$Z = \begin{bmatrix} \boldsymbol{z}_1^\top \\ \boldsymbol{z}_2^\top \\ \vdots \\ \boldsymbol{z}_n^\top \end{bmatrix}, \quad \boldsymbol{z}_i = \begin{bmatrix} z_{i1} \\ z_{i2} \\ \vdots \\ z_{iq} \end{bmatrix}, \quad \boldsymbol{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \quad \boldsymbol{\mu}_T = \begin{bmatrix} \mathrm{E}(T_1) \\ \mathrm{E}(T_2) \\ \vdots \\ \mathrm{E}(T_n) \end{bmatrix}.$$

Estimação dos parâmetros

A matriz de informação de Fisher é obtida a partir de

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

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

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

Estimação dos parâmetros

Portanto,

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

$$= \sum_{i=1}^{n}\phi_i\omega_i x_{ij}x_{il},$$

desde que

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

A matriz de informação de Fisher para $\boldsymbol{\beta}$ é

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

Estimação dos parâmetros

Também, temos que

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_l} &= \frac{\partial U_{\gamma_j}}{\partial \gamma_l} \\
&= \frac{\partial}{\partial \gamma_l} \left\{ \sum_{i=1}^{n} [t_i + d'(\phi_i)] \frac{z_{ij}}{h'(\phi_i)} \right\} \\
&= \sum_{i=1}^{n} \frac{z_{ij}}{\{h'(\phi_i)\}^2} \left\{ d''(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_l} h'(\phi_i) - [t_i + d'(\phi_i)]\, h''(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_l} \right\} \\
&= \sum_{i=1}^{n} \frac{z_{ij} z_{il}}{\{h'(\phi_i)\}^2} \left\{ \frac{d''(\phi_i)}{h'(\phi_i)} h'(\phi_i) - [t_i + d'(\phi_i)] \frac{h''(\phi_i)}{h'(\phi_i)} \right\} \\
&= \sum_{i=1}^{n} \frac{z_{ij} z_{il}}{\{h'(\phi_i)\}^2} \left\{ d''(\phi_i) - [t_i + d'(\phi_i)] \frac{h''(\phi_i)}{h'(\phi_i)} \right\}.
\end{aligned}
$$

Portanto,

$$-\,\mathrm{E}\left\{\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_l}\right\} = -\sum_{i=1}^{n}\frac{z_{ij}z_{il}d''\left(\phi_i\right)}{\left\{h'\left(\phi_i\right)\right\}^2} = \sum_{i=1}^{n} p_i z_{ij} z_{il},$$

desde que

$$\mathrm{E}\left(T_i + d'\left(\phi_i\right)\right) = 0, \quad p_i = -\frac{d''\left(\phi_i\right)}{\left\{h'\left(\phi_i\right)\right\}^2}.$$

A matriz de informação de Fisher para γ é

$$K_{\gamma\gamma} = -\,\mathrm{E}\left\{\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right\} = Z^\top PZ,$$

em que $P = \mathrm{diag}\{p_1, p_2, \ldots, p_n\}$. Ainda, $P = V_\gamma H_\gamma^{-2}$ com $V_\gamma = \mathrm{diag}\{-d''(\phi_1), \ldots, -d''(\phi_n)\}$.

Estimação dos parâmetros

Finalmente,

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_l} &= \frac{\partial U_{\beta_j}}{\partial \gamma_l} \\
&= \frac{\partial}{\partial \gamma_l} \left\{ \sum_{i=1}^{n} \phi_i (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\
&= \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij} \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_l} \\
&= \sum_{i=1}^{n} (y_i - \mu_i) \frac{1}{V_i} \frac{d\mu_i}{d\eta_i} x_{ij} \frac{z_{il}}{h'(\phi_i)}.
\end{aligned}
$$

Observe que

$$
-\operatorname{E}\left\{ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_l} \right\} = 0,
$$

pois $\operatorname{E}(Y_i - \mu_i) = 0$.

Estimação dos parâmetros

Isto implica que

$$K_{\boldsymbol{\beta}\boldsymbol{\gamma}} = K_{\boldsymbol{\gamma}\boldsymbol{\beta}}^{\top} = -\operatorname{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}}\right\} = 0_{p\times q}.$$

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

$$K_{\boldsymbol{\theta}\boldsymbol{\theta}} = \operatorname{diag}\{K_{\boldsymbol{\beta}\boldsymbol{\beta}}, K_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\} = \begin{bmatrix} X^{\top}\Phi W X & 0_{p\times q} \\ 0_{q\times p} & Z^{\top} P Z \end{bmatrix}.$$

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

Normalidade assintótica do EMV

Para n grande, pode-se mostrar que

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

em que

$$K_{\boldsymbol{\theta\theta}}^{-1} = \mathrm{diag}\{K_{\boldsymbol{\beta\beta}}^{-1}, K_{\boldsymbol{\gamma\gamma}}^{-1}\} = \begin{bmatrix} (X^\top \Phi W X)^{-1} & 0_{p\times q} \\ 0_{q\times p} & (Z^\top P Z)^{-1} \end{bmatrix}.$$

Particularmente, para n grande,

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

$$\widehat{\boldsymbol{\gamma}} \sim \mathrm{N}_q\left(\boldsymbol{\gamma},\ (Z^\top P Z)^{-1}\right).$$

Processo iterativo para estimar β e γ

Os estimadores de máxima verossimilhança para β e γ devem ser obtidos a partir de

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

$$U_{\gamma}(\boldsymbol{\theta}) = \mathbf{0} \Leftrightarrow Z^{\top}H_{\gamma}^{-1}(\boldsymbol{t} - \boldsymbol{\mu}_T) = \mathbf{0},$$

respectivamente.

Observe que $\widehat{\boldsymbol{\beta}}$ e $\widehat{\gamma}$ não possuem forma fechada.

Por esta razão, desenvolve-se o processo iterativo escore de Fisher sob os MLGDs.

Processo iterativo para estimar β e γ

O procedimento iterativo Escore de Fisher para $\boldsymbol{\beta}$ fica definido por

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

com

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

Analogamente, para γ o procedimento iterativo fixa expresso por

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

com

$$U_{\gamma}(\boldsymbol{\theta}) = Z^\top H_{\gamma}^{-1}(\boldsymbol{t} - \boldsymbol{\mu}_T), \quad K_{\gamma\gamma}^{-1} = (Z^\top P Z)^{-1}.$$

Processo iterativo para estimar β e γ

Ainda, veja que

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

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

De forma similar, temos

$$\begin{aligned}\gamma^{(m)} &= \left\{(Z^\top P Z)^{-1}\right\}^{(m)} \left[Z^\top P^{(m)} Z\right] \gamma^{(m)} \\ &= \left\{(Z^\top P Z)^{-1}\right\}^{(m)} Z^\top P^{(m)} \lambda^{(m)},\end{aligned}$$

em que $\lambda^{(m)} = Z\gamma^{(m)}$ é o preditor linear associado ao submodelo da precisão no passo m do processo iterativo Escore de Fisher.

Processo iterativo para estimar β e γ

Substituindo a expressão de $\beta^{(m)}$ em (1) ficamos com

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

em que

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

Processo iterativo para estimar β e γ

Substituindo a expressão de $\gamma^{(m)}$ em (2) ficamos com

$$
\begin{aligned}
\gamma^{(m+1)} &= \gamma^{(m)} + \left\{K_{\gamma\gamma}^{-1}\right\}^{(m)} U_{\gamma}^{(m)} \\
&= \left\{(Z^{\top}PZ)^{-1}\right\}^{(m)} Z^{\top}P^{(m)}\boldsymbol{\lambda}^{(m)} \\
&+ \left\{(Z^{\top}PZ)^{-1}\right\}^{(m)} Z^{\top}\left\{H_{\gamma}^{-1}\right\}^{(m)}(\boldsymbol{t}^{(m)} - \boldsymbol{\mu}_T^{(m)}) \\
&= \left\{(Z^{\top}PZ)^{-1}\right\}^{(m)} Z^{\top}P^{(m)}\left\{\boldsymbol{\lambda}^{(m)} + \left\{V_{\gamma}^{-1}\right\}^{(m)} H_{\gamma}^{(m)}(\boldsymbol{t}^{(m)} - \boldsymbol{\mu}_T^{(m)})\right\} \\
&= \left\{(Z^{\top}PZ)^{-1}\right\}^{(m)} Z^{\top}P^{(m)}\boldsymbol{z}^{*(m)},
\end{aligned}
$$

em que

$$\boldsymbol{z}^{*(m)} = \boldsymbol{\lambda}^{(m)} + \left\{V_{\gamma}^{-1}\right\}^{(m)} H_{\gamma}^{(m)}(\boldsymbol{t}^{(m)} - \boldsymbol{\mu}_T^{(m)}).$$

Lembre-se que $P = V_{\gamma}H_{\gamma}^{-2}$.

Processo iterativo para estimar β e γ

Novamente, observa-se que o processo iterativo de Escore de Fisher pode ser visto como um processo iterativo de mínimos quadrados reponderados.

A matriz ΦW possui o papel de reponderar $\widehat{\beta}$, enquanto que a matriz P possui o papel de reponderar $\widehat{\gamma}$. Por esta razão, ΦW e P são vistas como matrizes de ponderações ou "pesos" que mudam a cada passo do processo iterativo.

O processo iterativo pode ser resolvido alternando-se as duas equações obtidas para $\beta^{(m+1)}$ e $\gamma^{(m+1)}$ até a convergência.

O processo pode ser iniciado com a equação obtida para $\beta^{(m+1)}$ avaliada sob as estimativas de um MLG com $\phi_i = \phi$, $\forall i$ (ou ainda $\Phi = \phi I_n$), isto é, os valores iniciais $\beta^{(0)}$ e $\gamma^{(0)}$ são as estimativas de máxima verossimilhança sob um MLG com precisão constante.

MLG normal duplo

O MLGD normal é definido por

i) $Y_i|\boldsymbol{x}_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu_i, \phi_i^{-1})$ com $\phi_i = \sigma_i^2$;

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

ii) $h(\phi_i) = \boldsymbol{z}_i^\top \gamma \;\;\Leftrightarrow\;\; \phi_i = h^{-1}(\boldsymbol{z}_i^\top \gamma)$.

As opções de função de ligação que podem ser utilizadas para $g(\cdot)$ são as mesmas vistas sobre os MLGs.

As opções de função de ligação que podem ser utilizadas para $h(\cdot)$ são as mesmas utilizadas para modelar parâmetros estritamente positivos. Em particular, a ligação logarítmica é a ligação vastamente utilizada.

MLG normal duplo

Um modelo de regressão linear normal heteroscedástico pode ser obtido como caso especial dos MLGDs normais.

Considere o modelo

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind}}{\sim} \mathrm{N}(0, \phi_i^{-1}),$$

com estrutura de regressão para a precisão ϕ dada por

$$\log(\phi_i) = \boldsymbol{z}_i^\top \gamma \;\; \Rightarrow \;\; \log(\sigma_i^2) = -\boldsymbol{z}_i^\top \gamma.$$

Sob este modelo,

$$\mu_i = \mathrm{E}(Y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad \mathrm{Var}(Y_i) = \phi_i^{-1} = \sigma_i^2 = \mathrm{e}^{-\boldsymbol{z}_i^\top \gamma}.$$

Assim, a média e a variância da resposta estão sendo modeladas simultaneamente. Neste caso, $g(x) = x$ e $h(x) = \log(x)$.

MLG normal duplo

Lembre-se que sob ligação $g(\cdot)$ canônica, $W = V$. Daí, o vetor escore para $\boldsymbol{\beta}$ fica dado por

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

em que $\Phi = \mathrm{diag}\{\sigma_1^{-2}, \sigma_2^{-2}, \ldots, \sigma_n^{-2}\}$ sob o modelo normal.

O vetor escore para γ fica dado por

$$U_{\gamma}(\boldsymbol{\theta}) = Z^{\top}H_{\gamma}^{-1}(\boldsymbol{t}-\boldsymbol{\mu}_T),$$

em que $H_{\gamma} = \mathrm{diag}\{h'(\phi_1), \ldots, h'(\phi_n)\}$ com $h'(\phi_i) = \phi_i^{-1}$ desde que $h(x) = \log(x)$.

Assim, $H_{\gamma} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$.

MLG normal duplo

Também, $\boldsymbol{t} = (t_1 \ t_2 \cdots t_n)^\top$ com $t_i = y_i\theta_i - b(\theta_i) + a(y_i)$.

Para este modelo, ficamos com

$$\begin{aligned} t_i &= y_i\mu_i - \frac{\mu_i^2}{2} - \frac{1}{2}y_i^2 \\ &= \frac{2y_i\mu_i - \mu_i^2 - y_i^2}{2} \\ &= -\frac{(y_i - \mu_i)^2}{2}. \end{aligned}$$

Assim,

$$\boldsymbol{t}^\top = \left[-\tfrac{(y_1-\mu_1)^2}{2} \quad -\tfrac{(y_2-\mu_2)^2}{2} \quad \ldots \quad -\tfrac{(y_n-\mu_n)^2}{2} \right]$$

MLG normal duplo

O vetor $\boldsymbol{\mu}_T$ fica expresso por

$$\boldsymbol{\mu}_T^\top = \begin{bmatrix} -d'(\phi_1) & -d'(\phi_2) & \cdots & -d'(\phi_n) \end{bmatrix}.$$

Sob o modelo normal, $d(\phi_i) = \frac{1}{2}\log(\phi_i)$ que implica em $d'(\phi_i) = \frac{1}{2\phi_i}$. Assim, ficamos com

$$\begin{aligned} \boldsymbol{\mu}_T^\top &= \begin{bmatrix} -\frac{1}{2\phi_1} & -\frac{1}{2\phi_2} & \cdots & -\frac{1}{2\phi_n} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{\sigma_1^2}{2} & -\frac{\sigma_2^2}{2} & \cdots & -\frac{\sigma_n^2}{2} \end{bmatrix}. \end{aligned}$$

A matriz de Fisher para $\boldsymbol{\beta}$ é

$$K_{\boldsymbol{\beta}\boldsymbol{\beta}} = X^\top \Phi W X = X^\top \Phi X,$$

desde que $W = V = I_n$ pois $V_i = 1,\ \forall i$.

MLG normal duplo

A matriz de Fisher para γ é

$$K_{\gamma\gamma} = Z^\top P Z,$$

em que $P = \mathrm{diag}\{p_1, p_2, \ldots, p_n\}$ com

$$p_i = -\frac{d''(\phi_i)}{\{h'(\phi_i)\}^2} = -\frac{[-1/(2\phi_i^2)]}{\{\phi_i^{-1}\}^2} = \frac{1}{2}.$$

No modelo normal linear heteroscedástico as ponderações no processo iterativo associado ao submodelo da precisão são constantes.

Daí, ficamos com

$$K_{\gamma\gamma} = \frac{1}{2} Z^\top Z.$$

Para n grande, temos que

$$\widehat{\beta} \sim \mathrm{N}_p\left(\beta,\ (X^\top \Phi X)^{-1}\right), \quad \widehat{\gamma} \sim \mathrm{N}_q\left(\gamma,\ 2(Z^\top Z)^{-1}\right).$$

Função desvio - submodelo da média

Sob os MLGDs pode-se definir as respectivas funções desvio para cada submodelo.

Supondo ϕ_i conhecido $\forall i$, o desvio para o submodelo da média é

$$D^*_\mu(\boldsymbol{y};\widehat{\boldsymbol{\mu}},\boldsymbol{\phi}) = \sum_{i=1}^{n} {d^*_\mu}^2(y_i;\widehat{\mu}_i,\phi_i) = 2\left\{\ell(\widetilde{\boldsymbol{\mu}}) - \ell(\widehat{\boldsymbol{\mu}})\right\},$$

em que $\boldsymbol{\phi} = (\phi_1 \ \phi_2 \ \cdots \ \phi_n)^\top$, $\ell(\boldsymbol{\mu})$ é o logaritmo da função de verossimilhança avaliado em $\boldsymbol{\mu}$, e

$${d^*_\mu}^2(y_i;\widehat{\mu}_i,\phi_i) = 2\phi_i\left\{y_i(\widetilde{\theta}_i - \widehat{\theta}_i) + b(\widehat{\theta}_i) - b(\widetilde{\theta}_i)\right\}.$$

Na prática ϕ_i é desconhecido, assim deve-se substituir ϕ_i por $\widehat{\phi}_i = h^{-1}(\widehat{\lambda}_i)$.

Para ϕ_i grande para todo i, mostra-se que

$$D^*(\boldsymbol{y};\widehat{\boldsymbol{\mu}},\boldsymbol{\phi}) \sim \chi^2_{n-p}.$$

Função desvio - submodelo da precisão

Supondo μ_i conhecido $\forall i$, o desvio para o submodelo da precisão é

$$D_\phi^*(\boldsymbol{y}; \widehat{\boldsymbol{\phi}}, \boldsymbol{\mu}) = \sum_{i=1}^{n} {d_\phi^*}^2(y_i; \widehat{\phi}_i, \mu_i) = 2\left\{\ell(\widetilde{\boldsymbol{\phi}}) - \ell(\widehat{\boldsymbol{\phi}})\right\},$$

em que $\ell(\boldsymbol{\phi})$ é o logaritmo da função de verossimilhança avaliado em $\boldsymbol{\phi}$.

Sob o MLGD, supondo μ_i conhecido,

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{n} \ell(\phi_i),$$

em que

$$\ell(\phi_i) = \phi_i t_i + d(\phi_i) + u(y_i).$$

Função desvio - submodelo da precisão

$$
\begin{aligned}
D^*_\phi(\boldsymbol{y};\widehat{\boldsymbol{\phi}},\boldsymbol{\mu}) &= 2\left\{\ell(\widetilde{\boldsymbol{\phi}})-\ell(\widehat{\boldsymbol{\phi}})\right\}\\
&= 2\sum_{i=1}^{n}[\ell(\widetilde{\phi}_i)-\ell(\widehat{\phi}_i)]\\
&= 2\sum_{i=1}^{n}\left\{\widetilde{\phi}_i t_i + d(\widetilde{\phi}_i) + u(y_i) - \widehat{\phi}_i t_i - d(\widehat{\phi}_i) - u(y_i)\right\}\\
&= 2\sum_{i=1}^{n}\left\{t_i(\widetilde{\phi}_i-\widehat{\phi}_i) + d(\widetilde{\phi}_i) - d(\widehat{\phi}_i)\right\}.
\end{aligned}
$$

Assim, obtemos que

$$
{d^*_\phi}^2(y_i;\widehat{\phi}_i,\mu_i) = 2\left\{t_i(\widetilde{\phi}_i-\widehat{\phi}_i)+d(\widetilde{\phi}_i)-d(\widehat{\phi}_i)\right\},
$$

em que $\widetilde{\phi}_i$ é a estimativa de máxima verossimilhança para ϕ_i sob o modelo saturado.

Modelo saturado

Na prática μ_i (e, portanto, t_i) é desconhecido. Então, deve-se substituir μ_i (t_i) por $\widehat{\mu}_i = g^{-1}(\widehat{\eta}_i)$ ($\widehat{t}_i = y_i\widehat{\theta}_i - b(\widehat{\theta}_i) + a(y_i)$).

Qual é a estimativa de máxima verossimilhança para ϕ_i sob o modelo saturado?

Considere que $Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \text{FE}\,(\mu_i, \phi_i)$ com μ_i fixado, e $h\,(\phi_i) = \gamma_i$, para $i = 1, 2, \ldots, n$.

O modelo suposto para os dados é um MLG que possui n parâmetros, $\gamma_1, \gamma_2, \ldots, \gamma_n$. Mais especificamente,

$$\phi_1 = h^{-1}(\gamma_1), \;\; \phi_2 = h^{-1}(\gamma_2), \quad \ldots, \quad \phi_n = h^{-1}(\gamma_n).$$

Função desvio - submodelo da precisão

Neste caso, a matriz associada ao submodelo da precisão é a matriz identidade de ordem n, isto é, $Z = I_n$.

Note que a relação entre as precisões e os γ's é única e expressa por $h(\phi_i) = \gamma_i$. Portanto, sob o modelo saturado podemos estimar diretamente as precisões $\phi_1, \phi_2, \ldots, \phi_n$, e posteriormente, estabelecer as estimativas de $\gamma_1, \gamma_2, \ldots, \gamma_n$.

O escore para ϕ_j, $j = 1, 2, \ldots, n$, sob o modelo saturado é

$$U_{\phi_j} = \sum_{i=1}^{n} \frac{\partial}{\partial \phi_j} \{\phi_i t_i + d(\phi_i) + u(y_i)\} = t_j + d'(\phi_j).$$

O vetor escore para $\boldsymbol{\phi} = (\phi_1 \ \phi_2 \ \ldots \ \phi_n)^\top$ é

$$U_\phi = (\boldsymbol{t} - \boldsymbol{\mu}_T).$$

Função desvio - submodelo da precisão

O EMV para o vetor de parâmetros $\boldsymbol{\phi}$ é obtido de

$$(\boldsymbol{t} - \widetilde{\boldsymbol{\mu}}_T) = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{t} = \widetilde{\boldsymbol{\mu}}_T,$$

em que $\widetilde{\boldsymbol{\mu}}_T$ denota o EMV de $\boldsymbol{\mu}_T$ sob o modelo saturado.

Portanto, $\widetilde{\phi}_i$ é obtido tal que forma que $t_i = -d'(\widetilde{\phi}_i)$, ou ainda,

$$d'(\widetilde{\phi}_i) = -t_i, \;\; \forall i.$$

Para o modelo normal, temos que $d'(\widetilde{\phi}_i) = 1/(2\widetilde{\phi}_i)$, e portanto,

$$\frac{1}{2\widetilde{\phi}_i} = -t_i \;\Rightarrow\; \widetilde{\phi}_i = -\frac{1}{2t_i},$$

em que

$$t_i = -\frac{(y_i - \mu_i)^2}{2}.$$

Função desvio - submodelo da precisão

Na prática, troca-se t_i por $\widehat{t}_i$ na expressão de $\widetilde{\phi}_i$. Daí, ficamos com

$$\widetilde{\phi}_i = -\frac{1}{2\widehat{t}_i} = \frac{1}{(y_i - \widehat{\mu}_i)^2}.$$

Para ϕ_i grande para todo i, mostra-se que

$$D^*(\boldsymbol{y}; \widehat{\boldsymbol{\phi}}, \boldsymbol{\mu}) \sim \chi^2_{n-q}.$$

Para o modelo normal inversa, também temos que $d'(\widetilde{\phi}_i) = 1/(2\widetilde{\phi}_i)$, e portanto,

$$\frac{1}{2\widetilde{\phi}_i} = -t_i \;\Rightarrow\; \widetilde{\phi}_i = -\frac{1}{2t_i},$$

porém t_i é definido diferente.

Função desvio - submodelo da precisão

Temos que $\theta_i = -\frac{1}{2\mu_i^2}$, $b(\theta_i) = -1/\mu_i$, e $a(y_i) = -1/(2y_i)$. Assim, ficamos com

$$\begin{aligned} t_i &= y_i\theta_i - b(\theta_i) + a(y_i) \\ &= -\frac{y_i}{2\mu_i^2} + \frac{1}{\mu_i} - \frac{1}{2y_i} \\ &= -\frac{y_i^2 - 2\mu_i y_i + \mu_i^2}{2\mu_i^2 y_i} \\ &= -\frac{(y_i - \mu_i)^2}{2\mu_i^2 y_i}. \end{aligned}$$

Função desvio - submodelo da precisão

Para o modelo gama, temos

$$d'(\phi_i) = 1 + \log(\phi_i) - \psi(\phi_i).$$

Sabemos que $\widetilde{\phi}_i$ é obtido tal que forma que

$$\begin{aligned} d'(\widetilde{\phi}_i) &= -t_i, \ \ \forall i \Rightarrow \\ 1 + \log(\widetilde{\phi}_i) - \psi(\widetilde{\phi}_i) &= -t_i \end{aligned}$$

Observe que, neste caso, $\widetilde{\phi}_i$ não possui forma fechada. Assim, $\widetilde{\phi}_i$ deve ser obtido através da solução de uma equação não linear.

Resíduos - submodelo da média

O resíduo componente do desvio para o submodelo da média é definido por

$$\widehat{t_{D_\mu}}_i = \frac{d^*_\mu(y_i; \widehat{\mu}_i, \widehat{\phi}_i)}{\sqrt{1 - \widehat{h}_{ii}}},$$

em que

$$d^*_\mu(y_i; \widehat{\mu}_i, \widehat{\phi}_i) = \pm\sqrt{2\widehat{\phi}_i}\left\{y_i(\widetilde{\theta}_i - \widehat{\theta}_i) + b(\widehat{\theta}_i) - b(\widetilde{\theta}_i)\right\}^{1/2},$$

$\pm$ corresponde ao sinal de $(y_i - \widehat{\mu}_i)$, e

$$\widehat{h}_{ii} = \widehat{\phi}_i\widehat{\omega}_i\boldsymbol{x}_i^\top(X^\top\widehat{\Phi}\widehat{W}X)^{-1}\boldsymbol{x}_i.$$

Alavancagem - submodelo da média

Sob os MLGDs, a medida $\widehat{h}_{ii}$ pode ser vista como a variação causada em $\widehat{\eta}_i$ quando a resposta modificada associada ao submodelo da média $\widehat{y}_i^*$ é acrescida de uma contaminação infinitesimal, isto é,

$$\widehat{h}_{ii} = \frac{\partial \widehat{\eta}_i}{\partial \widehat{y}_i^*}.$$

Assim, $\widehat{h}_{ii}$ é a medida de alavancagem associada ao submodelo da média obtida como sendo o i-ésimo elemento da diagonal principal de

$$\widehat{H} = \widehat{\Phi}^{1/2}\widehat{W}^{1/2}X(X^\top\widehat{\Phi}\widehat{W}X)^{-1}X^\top\widehat{\Phi}^{1/2}\widehat{W}^{1/2}.$$

Resíduos - submodelo da precisão

O resíduo componente do desvio para o submodelo da precisão é definido por

$$\widehat{t_{D_{\phi}}}_i = \frac{d_{\phi}^*(y_i; \widehat{\phi}_i, \widehat{\mu}_i)}{\sqrt{1 - \widehat{r}_{ii}}},$$

em que

$$d_{\phi}^*(y_i; \widehat{\phi}_i, \widehat{\mu}_i) = \pm\sqrt{2}\left\{\widehat{t}_i(\widetilde{\phi}_i - \widehat{\phi}_i) + d(\widetilde{\phi}_i) - d(\widehat{\phi}_i)\right\}^{1/2},$$

$\pm$ corresponde ao sinal de $\left(\widehat{t}_i + d'(\widehat{\phi}_i)\right)$, e

$$\widehat{r}_{ii} = \widehat{p}_i \boldsymbol{z}_i^{\top}(Z^{\top}\widehat{P}Z)^{-1}\boldsymbol{z}_i.$$

Alavancagem - submodelo da precisão

Sob os MLGDs, a medida $\widehat{r}_{ii}$ pode ser vista como a variação causada em $\widehat{\lambda}_i$ quando a resposta modificada associada ao submodelo da precisão $\widehat{z}_i^*$ é acrescida de uma contaminação infinitesimal, isto é,

$$\widehat{r}_{ii} = \frac{\partial \widehat{\lambda}_i}{\partial \widehat{z}_i^*}.$$

Assim, $\widehat{r}_{ii}$ é a medida de alavancagem associada ao submodelo da precisão obtida como sendo o i-ésimo elemento da diagonal principal de

$$\widehat{R} = \widehat{P}^{1/2} Z (Z^\top \widehat{P} Z)^{-1} Z^\top \widehat{P}^{1/2}.$$

Resíduos - Ajuste global

O resíduo quantílico sob os MLGDs é definido por

$$r_{q_i} = \Phi^{-1}(F(y_i; \widehat{\mu}_i, \widehat{\phi}_i)), \quad i = 1, 2, \ldots, n.$$

Observe que, para n finito, os resíduos quantílicos não são normais padrão independentes desde que calcula-se r_{q_i} sob as estimativas $\widehat{\mu}_i, \widehat{\phi}_i$.

Entretanto, se $\widehat{\mu}_i$ e $\widehat{\phi}_i$ são estimadores consistentes de μ_i e ϕ, respectivamente, então $r_{q_i} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$ para n grande.

Se y_i for um valor discrepante com relação ao modelo postulado F, o valor de $F(y_i; \widehat{\mu}_i, \widehat{\phi}_i)$ será próximo de zero ou próximo de um. Consequentemente, $|r_{q_i}|$ assumirá um valor maior ou igual que 3, indicando assim y_i como aberrante.

Influência - submodelo da média

A influência da i-ésima observação em $\widehat{\beta}$ pode ser mensurada através do afastamento pela verossimilhança definido por

$$LD_{\mu_i} = 2\{\ell(\widehat{\beta}) - \ell(\widehat{\beta}_{(i)})\},$$

em que $\widehat{\beta}_{(i)}$ é o EMV de β sem a i-ésima observação.

A distância de Cook aproximada para o submodelo da média é

$$LD_{\mu_i} = \widehat{ts}_{\mu_i}^{2} \frac{\widehat{h}_{ii}}{1 - \widehat{h}_{ii}},$$

em que

$$\widehat{ts}_{\mu_i} = \frac{\sqrt{\widehat{\phi}_i}(y_i - \widehat{\mu}_i)}{\sqrt{\widehat{V}_i(1 - \widehat{h}_{ii})}}$$

é o resíduo de Pearson studentizado para o submodelo da média.

Influência - submodelo da precisão

A influência da i-ésima observação em $\widehat{\gamma}$ pode ser mensurada através do afastamento pela verossimilhança definido por

$$LD_{\phi_i} = 2\{\ell(\widehat{\gamma}) - \ell(\widehat{\gamma}_{(i)})\},$$

em que $\widehat{\gamma}_{(i)}$ é o EMV de γ sem a i-ésima observação.

A distância de Cook aproximada para o submodelo da precisão é

$$LD_{\phi_i} \quad = \quad \widehat{ts}^{\,2}_{\phi_i} \frac{\widehat{r}_{ii}}{1-\widehat{r}_{ii}},$$

em que

$$\widehat{ts}_{\phi_i} = \frac{\widehat{t}_i + d'(\widehat{\phi}_i)}{\sqrt{-d''(\widehat{\phi}_i)(1-\widehat{r}_{ii})}}$$

é o resíduo de Pearson studentizado para o submodelo da precisão.