Universidade Federal do Paraná - Departamento de Estatística
CE225: Modelos lineares generalizados
Prof. Cesar Augusto Taconeli
Material complementar

Propriedades- família exponencial

Seja y uma obervação da variável aleatória Y, cuja função (densidade) de probabilidade pode ser expressa na forma:

$$f(y|\theta,\phi) = \exp\left\{\frac{\theta y - b(\theta)}{\phi} + c(y;\phi)\right\}, \tag{1}$$

sendo θ e ϕ denominados parâmetros canônico e de dispersão da distribuição, respectivamente.

Vamos derivar a média (esperança) de Y. Para isso, vamos considerar a função de log-verossimilhança fica dada por:

$$l(\theta,\phi|y) = \frac{\theta y - b(\theta)}{\phi} + c(y;\phi) \tag{2}$$

A função escore é definida pelas derivadas da log-verossimilhança com relação aos parâmetros do modelo. No caso:

$$S(\theta) = \frac{\partial l(\theta,\phi|y)}{\partial\theta}, \tag{3}$$

e em suas propriedades, como detalhado na sequência.

$$\mathrm{E}(S(\theta)) = 0 \tag{4}$$

$$\begin{aligned}\mathrm{Var}(S(\theta)) =& \mathrm{E}\left[S(\theta)^2\right] - \{\mathrm{E}\left[S(\theta)\right]\}^2 = \\ & \mathrm{E}\left[S(\theta)^2\right] = \\ & \left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right)^2,\end{aligned}$$

e, conforme iremos verificar adiante,

$$\left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right)^2 = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} l(\theta,\phi|y)\right] \tag{5}$$

Essas propriedades da função escore, que valem em geral sob certas condições de regularidade, também serão serão demonstradas adiante.

Média e variância das distribuições pertencentes à família exponencial

Para o caso de Y com distribuição pertencente à família familia exponencial, com base na função (densidade) de probabilidade definida em (1), e na definição da função escore (3) e de suas propriedades apresentadas em (4) e (5), vamos derivar, inicialmente, a média de Y:

$$\begin{aligned} \mathrm{E}(S(\theta)) =& \mathrm{E}\left[\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right] = \\ & \mathrm{E}\left[\frac{\partial}{\partial\theta}\frac{\theta Y - b(\theta)}{\phi}\right] = \\ & \mathrm{E}\left[\frac{Y - b'(\theta)}{\phi}\right]. \end{aligned} \tag{6}$$

Lembrando que $\mathrm{E}(S(\theta)) = 0$, podemos deduzir, temos que:

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

de onde podemos extrair que:

$$\mathrm{E}(Y) = b'(\theta) = \frac{\partial}{\partial\theta} b(\theta) \tag{7}$$

Agora, vamos obter a variância de Y. Neste caso:

$$\mathrm{Var}(S(\theta)) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right)^2\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} l(\theta,\phi|y)\right] \tag{8}$$

Vamos trabalhar cada membro dessa igualdade. Primeiramente:

$$\begin{aligned} \mathrm{E}\left[\left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right)^2\right] =& \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\frac{Y - b'(\theta)}{\phi}\right)^2\right] = \\ & \frac{\mathrm{E}Y^2 - 2\mathrm{E}(Y)b'(\theta) + [b'(\theta)]^2}{\phi^2} = \\ & \frac{\mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2}{\phi^2} = \frac{\mathrm{Var}(Y)}{\phi^2} \end{aligned} \tag{9}$$

Agora, tomando o membro à direta em (8):

$$\begin{aligned} -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} l(\theta,\phi|y)\right] =& -\mathrm{E}\left[\frac{\partial}{\partial\theta}\left(\frac{Y - b'(\theta)}{\phi}\right)\right] \\ & -\mathrm{E}\left[\frac{-b''(\theta)}{\phi}\right] = \frac{b''(\theta)}{\phi} \end{aligned}$$

Agora, retornando à igualdade em (8), temos:

$$\mathrm{E}\left(\frac{\partial}{\partial\theta}l(\theta,\phi|y)\right)^2 = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}l(\theta,\phi|y)\right]$$

$$\frac{\mathrm{Var}(Y)}{\phi^2} = \frac{b''(\theta)}{\phi}$$

De onde se extrai:

$$\mathrm{Var}(Y) = \phi b''(\theta)$$

Note que:

$$\begin{aligned}
\mathrm{Var}(Y) = \phi b''(\theta) = \\
\phi\frac{\partial}{\partial\theta}b'(\theta) = \\
\phi\frac{\partial}{\partial\theta}\mu = \\
\phi\mathrm{V}(\mu),
\end{aligned}$$

onde $\mathrm{V}(\mu)$ é chamada função de variância, que descreve a relação entre média e variância para as distribuições pertencentes à família exponencia.

Complemento- Propriedades gerais da função escore

- Média (esperança) da função escore.

$$\begin{aligned}
\mathrm{E}(S(\theta)) = \int_Y S(\theta)f(y|\theta,\phi)dy = \\
\int_Y \left[\frac{\partial}{\partial\theta}\log(f(y|\theta,\phi))\right]f(y|\theta,\phi)dy = \\
\int_Y \frac{1}{f(y|\theta,\phi)}\left[\frac{\partial}{\partial\theta}f(y|\theta,\phi)\right]f(y|\theta,\phi)dy = \\
\int_Y \frac{\partial}{\partial\theta}f(y|\theta,\phi)dy
\end{aligned}$$

As distribuições que pertencem à família exponencial atendem a condições de regularidade que permitem intercambiar operações de derivação e integração. Desta forma:

$$
\begin{aligned}
\mathrm{E}(S(\theta)) = \int_Y \frac{\partial f(y|\theta,\phi)}{\theta} dy = \\
\frac{\partial}{\partial\theta}\int_Y f(y|\theta,\phi)dy = \\
\frac{\partial}{\partial\theta} 1 = 0
\end{aligned}
$$

- Variância da função escore.

$$
\mathrm{Var}(S(\theta)) = \mathrm{E}\left[S(\theta)^2\right] - \{E\left[S(\theta)\right]\}^2
$$

Como verificado anteriormente, $E\left[S(\theta)\right] = 0$, de maneira que $\{E\left[S(\theta)\right]\}^2 = 0$ e, consequentemente:

$$
\begin{aligned}
\mathrm{Var}(S(\theta)) =& \mathrm{E}\left[S(\theta)^2\right] = \\
& \mathrm{E}\left[\left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)^2\right)^2\right] = \\
& \int_Y \left(\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right)^2 f(y|\theta,\phi)dy,
\end{aligned}
$$

que é a informação de Fisher relativa a θ, que pode ser expressa como:

$$
I(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} l(\theta,\phi|y)\right]
$$

uma vez que

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta^2} l(\theta,\phi|y) =& \frac{\partial}{\partial\theta}\left[\frac{f'(y;\theta,\phi)}{f(y;\theta,\phi)}\right] = \\
& \frac{1}{f(y;\theta,\phi)} f''(y;\theta,\phi) - \left[\frac{f(y;\theta,\phi)}{\left[f(y;\theta,\phi)\right]^2 f'(y;\theta,\phi)} + \frac{1}{f(y;\theta,\phi)} f''(y;\theta,\phi)\right] = \\
& \left[\frac{f''(y;\theta,\phi)}{f(y;\theta\phi)}\right]^2 = \\
& \left[\frac{\partial}{\partial\theta} l(\theta,\phi|y)\right]^2
\end{aligned}
$$

Logo,

$$
I(\theta) = \mathrm{E}\left(\frac{\partial}{\partial\theta}\log f(y|\theta,\phi)\right)^2 = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(y|\theta,\phi)\right].
$$

Principais membros da família exponencial

- **Distribuição binomial**

Seja $X \sim \text{binomial}(m, \pi)$, com função de probabilidades dada por:

$$f_X(x|m,\pi) = \binom{m}{x}\pi^x(1-\pi)^{m-x}, \quad m \in \mathbb{N}; 0 < \pi < 1; x = 0, 1, ..., m$$

O modelo apresentado representa o número de secessos y em m ensaios independentes de Bernoulli, cada um com probabilidade de sucesso p. De maneira alternativa, podemos definir a fração de sucessos $Y = X/m$, de forma que a função de probabilidades fica dada por:

$$f_Y(y|m,\pi) = \binom{m}{my}p^{my}(1-\pi)^{m-my} =$$

$$\exp\left\{my\log\left(\frac{\pi}{1-\pi}\right) + m\log(1-\pi) + \log\binom{m}{my}\right\} =$$

$$\exp\left\{m\left[y\log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right] + \log\binom{m}{my}\right\},$$

de onde podemos extrair:

$$\theta = \log\frac{\pi}{1-\pi} \to \pi = \frac{e^\theta}{e^\theta + 1}$$

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

$$\log\left(1 + e^\theta\right)$$

$$\phi = \frac{1}{m}$$

$$c(y;\phi) = \log\binom{m}{my}$$

$$E(Y) = \mu = b'(\theta) = \frac{\partial}{\partial\theta}\log\left(1+e^\theta\right) = \frac{e^\theta}{e^\theta+1} = \pi$$

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

$$\frac{e^{\theta}}{\left(e^{\theta}+1\right)^2} =$$

$$\mu(1-\mu),$$

de tal forma que:

$$\mathrm{V}(\mu) = b''(\theta) = \mu(1-\mu),$$

e

$$\mathrm{Var}(Y) = \phi b''(\theta) = \frac{\mu(1-\mu)}{m}.$$

- **Distribuição de Poisson**

Seja $Y \sim \text{Poisson}(\mu)$, com fumção de probabilidades dada por:

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

Logo,

$$\begin{aligned} f_Y(y;\mu) = & \frac{e^{-\mu}\mu^y}{y!} = \\ & \exp\left\{\log\left(\frac{e^{-\mu}\mu^y}{y!}\right)\right\} = \\ & \exp\left\{y\log(\mu) - \mu - \log y!\right\}, \end{aligned}$$

de tal forma que:

$$\theta = \log(\mu)$$

$$b(\theta) = \mu = e^\theta$$

$$\phi = 1$$

$$c(\phi, y) = -\log(y!)$$

$$E(Y) = b'(\theta) = \mu = e^\theta$$

$$\text{V}(\mu) = b''(\theta) = e^\theta = \mu$$

$$\text{Var}(Y) = \phi b''(\theta) = e^\theta = \mu$$

- **Distribuição normal**

Seja $Y \sim \text{normal}(\mu, \sigma^2)$, com fumção densidade de probabilidades dada por:

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - mu)^2\right\}, \quad y \in \mathbb{R}; \mu \in \mathbb{R}; \sigma^2 > 0.$$

Desta forma,

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

tal que:

$$\theta = \mu$$

$$b(\theta) = \frac{\mu^2}{2} = \frac{\theta^2}{2}$$

$$\phi = \sigma^2$$

$$c(\phi, y) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$

$$E(Y) = b'(\theta) = \theta = \mu$$

$$\text{V}(\mu) = b''(\theta) = 1$$

$$\text{Var}(Y) = \phi b''(\theta) = \sigma^2 \times 1 = \sigma^2$$

- **Distribuição gama**

Seja $Y \sim \text{gama}(\mu, \nu)$, com fumção densidade de probabilidades dada por:

$$f_Y(y; \mu, \nu) = \frac{\left(\frac{\nu}{\mu}\right)^{\nu}}{\Gamma(\nu)} y^{\nu-1} \exp\left\{\left(-\frac{y\nu}{\nu}\right)\right\}, \quad y > 0, \mu > 0, \nu > 0.$$

Desta forma,

$$\begin{aligned}
f_Y(y; \mu, \nu) =& \frac{\left(\frac{\nu}{\mu}\right)^{\nu}}{\Gamma(\nu)} y^{\nu-1} \exp\left\{\left(-\frac{y\nu}{\nu}\right)\right\} = \\
& \exp\left\{\log\left(\frac{\left(\frac{\nu}{\mu}\right)^{\nu}}{\Gamma(\nu)} y^{\nu-1} \exp\left\{\left(-\frac{y\nu}{\nu}\right)\right\}\right)\right\} = \\
& \exp\left\{\nu \log\left((\frac{\nu}{\mu}\right) - log\left(\Gamma(\nu)\right) + (\nu - 1)\log(y) - \frac{y\nu}{\mu}\right\} = \\
& \exp\left\{\nu\left(-\frac{y}{\mu} - \log(\mu) + \log(\nu) - \log\left(\Gamma(\nu) + (\nu-1)\log(y)\right)\right)\right\},
\end{aligned}$$

podendo-se deduzir que:

$$\theta = -\frac{1}{\mu}$$

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

$$\phi = \frac{1}{\nu}$$

$$c(\phi, y) = \log(\nu) - \log\left(\Gamma(\nu)\right) + (\nu - 1)\log(y)$$

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

$$\mathrm{V}(\mu) = b''(\theta) = \frac{1}{\theta^2} = \mu^2$$

$$\mathrm{Var}(Y) = \phi b''(\theta) = \frac{\mu^2}{\nu}$$

- **Distribuição normal inversa**

Seja $Y \sim$ normal inversa(μ, τ), com fumção densidade de probabilidades dada por:

$$f_Y(y; \mu, \tau) = \frac{\tau^{\frac{1}{2}}}{\sqrt{2\pi y^3}} \exp\left\{-\frac{\tau(y-\mu)^2}{2\mu^2 y}\right\}, \quad y > 0; \mu > 0; \tau > 0$$

Desta forma,

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

de onde podemos extrair que:

$$\theta = -\frac{1}{2\mu^2}$$

$$b(\theta) = -\frac{1}{\mu} = -\sqrt{-2\theta}$$

$$\phi = \frac{1}{\tau}$$

$$c(\tau, y) = -\frac{1}{2}\left[\log\left(\frac{2\pi y^3}{\tau} + \frac{\tau}{y}\right)\right]$$

$$E(Y) = b'(\theta) = 2 \times \frac{1}{2}(-2\theta)^{-\frac{1}{2}} = \left(\frac{2}{2\mu^2}\right)^{-\frac{1}{2}} = \mu$$

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

$$\mathrm{Var}(Y) = \phi b''(\theta) = \frac{\mu^3}{\tau}$$

Algoritmo de estimação para o modelo de regressão logística

- Sem perda de generalidade, vamos considerar $m_i = 1$, para todo $i = 1, 2, ..., n$, ou seja, a situação de dados não agrupados (distribuição de Bernoulli).
- Neste caso:

$$Y|\boldsymbol{x} \sim \text{binomial}(1, \mu_{\mathbf{x}})$$

$$\log\left(\frac{\mu}{1-\mu}\right) = \eta_{\boldsymbol{x}} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k,$$

uma vez que, neste caso, $\mu_{\boldsymbol{x}} = \pi_{\boldsymbol{x}} = P(Y = 1|\boldsymbol{x})$.

Voltando ao algoritmo de estimação, a variável dependente ajustada, dada por:

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}$$

para o caso do modelo de regressão logística fica dado por:

$$z_i = \log\left(\frac{\mu_i}{1-\mu_i}\right) + \frac{y_i - \mu_i}{\mu_i(1-\mu_i)},$$

uma vez que, para o modelo de regressão logística:

$$\eta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right)$$

e

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}\log\left(\frac{\mu_i}{1-\mu_i}\right) =$$

$$\frac{1}{\mu_i(1-\mu_i)}.$$

- A matriz de pesos, como visto anteriormente, é uma matriz diagonal com elementos:

$$\omega_i = \frac{(\partial \mu_i/\partial \eta_i)^2}{\mathrm{V}(\mu_i)}.$$

- No caso do modelo de regressão logística, temos:

$$\omega_i = \mu_i(1-\mu_i),$$

uma vez que:

$$\frac{\partial \mu_i}{\partial \eta_i} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^{-1} = \left(\frac{\partial \log\left(\frac{\mu_i}{1-\mu_i}\right)}{\partial \mu_i}\right)^{-1} = \left(\frac{1}{\mu_i(1-\mu_i)}\right)^{-1} = \mu_i(1-\mu_i),$$

e $\mathrm{V}(\mu_i) = \mu_i(1-\mu_i)$. Logo,

$$\omega_i = \frac{(\partial\mu_i/\partial\eta_i)^2}{\mathrm{V}(\mu_i)} = \frac{[\mu_i(1-\mu_i)]^2}{\mu_i(1-\mu_i)} = \mu_i(1-\mu_i),$$

de tal forma que o estimador de β via mínimos quadrados ponderados fica dado por:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}},$$

em que $\hat{\mathbf{W}}$ é a matriz diagonal $n \times n$ com entradas $\hat{\omega}_i = \hat{\mu}_i(1-\hat{\mu}_i)$, e $\hat{\mathbf{z}}$ é o vetor de tamanho n com entradas $\hat{z}_i = \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) + \frac{y_i-\hat{\mu}_i}{\hat{\mu}_i(1-\hat{\mu}_i)}$, para $i = 1, 2, ..., n$.

- Devemos relembrar que tanto $\hat{\beta}$ quanto $\mathbf{z}$ e $\mathbf{W}$ dependem das estimativas dos parâmetros, o que justifica a necessidade de um processo iterativo para estimar os $\beta'\mathbf{s}$. Detalhes adicionais desse processo podem ser vistos nos scripts em R referentes a esta aula.
- Além disso, a matriz de variância estimada dos estimadores dos $\beta'\mathbf{s}$ fica dada por:

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}.$$

Algoritmo de estimação para o modelo de regressão Poisson com ligação logaritmica

- Vamos considerar:

$$Y|\boldsymbol{x} \sim \text{Poisson}(\mu_i)$$

$$g(\mu_{\mathbf{x}}) = \log(\mu_{\mathbf{x}}) = \eta_{\boldsymbol{x}} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

A variável dependente ajustada, dada por:

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}$$

que, para o caso do modelo de regressão Poisson com ligação logarítmica, fica dada por:

$$z_i = \log(\mu_i) + \frac{y_i - \mu_i}{\mu_i}$$

uma vez que, para o modelo de regressão Poisson com ligação logarítmica:

$$\eta_i = \log(\mu_i)$$

e

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}\log(\mu_i) = \frac{1}{\mu_i}.$$

- A matriz de pesos, como visto anteriormente, é uma matriz diagonal com elementos:

$$\omega_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\text{V}(\mu_i)}.$$

- No caso do modelo de regressão gama com ligação inversa, temos:

$$\omega_i = \mu_i,$$

uma vez que:

$$\frac{\partial \mu_i}{\partial \eta_i} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^{-1} = \left(\frac{\partial \log(\mu_i)}{\partial \mu_i}\right)^{-1} = \mu_i,$$

e $\text{V}(\mu_i) = \mu_i$. Logo,

$$\omega_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\text{V}(\mu_i)} = \frac{\mu_i^2}{\mu_i} = \mu_i,$$

de tal forma que o estimador de β via mínimos quadrados ponderados fica dado por:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}},$$

em que $\hat{\mathbf{W}}$ é a matriz diagonal $n \times n$ com entradas $\hat{\omega}_i = \hat{\mu}_i$ e $\hat{\mathbf{z}}$ é o vetor de tamanho n com entradas $z_i = \log(\mu_i) + \frac{y_i - \mu_i}{\mu_i}$, para $i = 1, 2, ..., n$.

- A matriz de variância estimada dos estimadores dos $\beta'\mathbf{s}$ fica dada por:

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}.$$

Algoritmo de estimação para o modelo de regressão gama com ligação inversa

- Vamos considerar:

$$Y|\boldsymbol{x} \sim \text{gamma}(\mu_i, \nu)$$

$$g(\mu_{\mathbf{x}}) = \frac{1}{\mu_{\mathbf{x}}} = \eta_{\boldsymbol{x}} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

A variável dependente ajustada, dada por:

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}$$

para o caso do modelo de regressão gama com ligação inversa fica dada por:

$$z_i = \frac{1}{\mu_i} - \frac{y_i - \mu_i}{\mu_i^2} = \frac{1}{\mu_i} + \frac{\mu_i - y_i}{\mu_i^2}$$

uma vez que, para o modelo de regressão gama com ligação inversa:

$$\eta_i = \frac{1}{\mu_i}$$

e

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}\frac{1}{\mu_i} = -1/\mu_i^2.$$

- A matriz de pesos, como visto anteriormente, é uma matriz diagonal com elementos:

$$\omega_i = \frac{\left(\partial \mu_i / \partial \eta_i\right)^2}{\mathrm{V}(\mu_i)}.$$

- No caso do modelo de regressão gama com ligação inversa, temos:

$$\omega_i = \mu_i^2,$$

uma vez que:

$$\frac{\partial \mu_i}{\partial \eta_i} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^{-1} = \left(\frac{\partial\left(\frac{1}{\mu_i}\right)}{\partial \mu_i}\right)^2 = \mu_i^2,$$

e $\mathrm{V}(\mu_i) = \mu_i^2$. Logo,

$$\omega_i = \frac{\left(\partial \mu_i / \partial \eta_i\right)^2}{\mathrm{V}(\mu_i)} = \frac{\mu_i^4}{\mu_i^2} = \mu_i^2,$$

de tal forma que o estimador de β via mínimos quadrados ponderados fica dado por:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}},$$

em que $\hat{\mathbf{W}}$ é a matriz diagonal $n \times n$ com entradas $\hat{\omega}_i$ e $\hat{\mathbf{z}}$ é o vetor de tamanho n com entradas $\hat{z}_i = \frac{1}{\hat{\mu}_i} + \frac{y_i - \hat{\mu}_i}{\mu_i^2}$, para $i = 1, 2, ..., n$.

- A matriz de variância estimada dos estimadores dos $\beta'\mathbf{s}$ fica dada por:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}.$$

Função desvio para as pricipais distribuições pertencentes à família exponencial

- **Distribuição binomial**

Para a distribuição binomial, assumimos $Y_i \sim \text{binomial}(\mu_i, \pi_i)$, de tal forma que, como visto anteriormente, $\tilde{\theta}_i = \log\{y_i/(m_i - y_i)\}$ e $\hat{\theta}_i = \log\{\hat{\mu}_i/(m_i - \hat{\mu}_i)\}$; $\text{b}(\tilde{\theta}_i) = -\log(1 - y_i)$, $\text{b}(\hat{\theta}_i) = -\log(1 - \hat{\mu}_i)$, produzindo:

$$\text{D}(\mathbf{y}, \hat{\mu}) = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{m_i\hat{\mu}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i/m_i}{1 - \hat{\mu}_i}\right)\right]$$

- **Distribuição Poisson**

Para a distribuição Poisson, assumimos $Y_i \sim \text{Poisson}(\mu_i)$, de tal forma que, como visto anteriormente, $\tilde{\theta}_i = \log(y_i)$ e $\hat{\theta}_i = \log(\hat{\mu}_i)$; $\text{b}(\tilde{\theta}_i) = y_i$, $\text{b}(\hat{\theta}_i) = \hat{\mu}_i$ produzindo:

$$\text{D}(\mathbf{y}, \hat{\mu}) = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]$$

- **Distribuição normal**

Para a distribuição normal, assumimos $Y_i \sim \text{normal}(\mu_i, \sigma^2)$, de tal forma que, como visto anteriormente, $\tilde{\theta}_i = y_i$ e $\hat{\theta}_i = \hat{\mu}_i$; $\text{b}(\tilde{\theta}_i) = y_i^2/2$ e $\text{b}(\hat{\theta}_i) = \hat{\mu}_i^2$, produzindo:

$$\text{D}(\mathbf{y}, \hat{\mu}) = 2\sum_{i=1}^{n}\left[y_i(y_i - \hat{\mu}_i) + \frac{\hat{\mu}_i^2 - y_i^2}{2}\right] = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2/2,$$

que é a soma de quadrados dos resíduos.

- **Distribuição gama**

Para a distribuição gama, assumimos $Y_i \sim \text{gama}(\mu_i, \nu)$, de tal forma que, como visto anteriormente, $\tilde{\theta}_i = -1/y_i$ e $\hat{\theta}_i = -1/\hat{\mu}_i$; $\text{b}(\tilde{\theta}_i) = \log(y_i)$ e $\text{b}(\hat{\theta}_i) = \log(\hat{\mu}_i)$, produzindo:

$$\text{D}(\mathbf{y}, \hat{\mu}) = 2\sum_{i=1}^{n}\left[-\log\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right]$$

- **Distribuição normal inversa**

Para a distribuição normal inversa, assumimos $Y_i \sim \text{normal inversa}(\mu_i, \tau)$, de tal forma que, como visto anteriormente, $\tilde{\theta}_i = -1/2y_i^2$ e $\hat{\theta}_i = -1/2\hat{\mu}_i^2$; $\text{b}(\tilde{\theta}_i) = -1/y_i$ e $\text{b}(\hat{\theta}_i) = -1/\hat{\mu}_i$, produzindo:

$$\text{D}(\mathbf{y}, \hat{\mu}) = 2\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{y_i\hat{\mu}_i^2}.$$