MAE5763 - Modelos Lineares Generalizados - Resolução da prova 2

Guilherme Marthe - 8661962

3/12/2020

1 Exercício 1

A variável Y_i, $i = 1, ..., n$, apresentada no enunciado tem a seguinte função de massa de probabilidades:

$$f(y_i, \psi_i) = \binom{m_i}{y_i} \left(\frac{\psi_i}{1+\psi_i}\right)^{y_i} \left(\frac{1}{1+\psi_i}\right)^{(m_i - y_i)}$$

Vamos testar a hipótese

$$H_0 : \psi_1 = \psi_2 = ... = \psi_n$$

contra

$$H_1 : \psi_i \neq \psi_l \quad \text{com} \quad i \neq l \;\text{ e }\; i, l = 1, ..., n$$

Através de um teste de razão de verossimilhanças.

O logaritmo da função de verossimilhança de $Y_1, ..., Y_n$ é:

$$\begin{aligned}
L(y_i, \psi_i) &= \sum_{i=1}^{n} y_i \log\left(\frac{\psi_i}{1+\psi_i}\right) + (m_i - y_i)\log\left(\frac{1}{1+\psi_i}\right) + \log\binom{m_i}{y_i} \\
&= \sum_{i=1}^{n} y_i \left(\log(\psi_i) - \log(1+\psi_i)\right) - (m_i - y_i)\log\left(1+\psi_i\right) + \log\binom{m_i}{y_i} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - y_i \log(1+\psi_i) + y_i \log\left(1+\psi_i\right) - m_i \log\left(1+\psi_i\right) + \log\binom{m_i}{y_i} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - m_i \log\left(1+\psi_i\right) + log\binom{m_i}{y_i} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \sum_{i=1}^{n} \log\binom{m_i}{y_i}
\end{aligned}$$

Sob H_0, podemos definir que $\psi_1 = \psi_2 = ... = \psi_n = \psi^0$, ou seja, os ψ_i têm um valor fixo chamado ψ^0. Dessa forma temos

$$L(y_i, \psi^0) = \sum_{i=1}^{n} y_i \log(\psi^0) - \sum_{i=1}^{n} m_i \log\left(1+\psi^0\right) + \sum_{i=1}^{n} \log\binom{m_i}{y_i}$$

$$
\begin{aligned}
&= \log(\psi^0)\sum_{i=1}^{n} y_i - \log\left(1+\psi^0\right)\sum_{i=1}^{n} m_i + \sum_{i=1}^{n}\log\binom{m_i}{y_i} \\
&= \log(\psi^0)n\bar{y} - \log\left(1+\psi^0\right)n\bar{m} + \sum_{i=1}^{n}\log\binom{m_i}{y_i}
\end{aligned}
$$

E sob a hipótese alternativa, não temos muito o que simplificar, uma vez que os ψ_i não são fixos. Então

$$
L(y_i,\psi_i) = \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \sum_{i=1}^{n}\log\binom{m_i}{y_i}
$$

A estatística do teste de razão de verossimilhança é:

$$
\xi_{RV} = 2[L(y_i;\psi_i) - L(y_i;\psi^0)]
$$

Assim, desenvolvendo a estatística temos:

$$
\begin{aligned}
\xi_{RV} &= \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \sum_{i=1}^{n} log\binom{m_i}{y_i} - \left(\log(\psi^0)n\bar{y} - \log\left(1+\psi^0\right)n\bar{m} + \sum_{i=1}^{n} log\binom{m_i}{y_i}\right) \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \sum_{i=1}^{n} log\binom{m_i}{y_i} - \log(\psi^0)n\bar{y} + \log\left(1+\psi^0\right)n\bar{m} - \sum_{i=1}^{n} log\binom{m_i}{y_i} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) - \log(\psi^0)n\bar{y} + \log\left(1+\psi^0\right)n\bar{m} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - \log(\psi^0)n\bar{y} - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \log\left(1+\psi^0\right)n\bar{m} \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} y_i \log(\psi^0) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + \sum_{i=1}^{n} m_i \log\left(1+\psi^0\right) \\
&= \sum_{i=1}^{n} y_i \log(\psi_i) - y_i \log(\psi^0) - \sum_{i=1}^{n} m_i \log\left(1+\psi_i\right) + m_i \log\left(1+\psi^0\right) \\
&= \sum_{i=1}^{n} y_i \log\left(\frac{\psi_i}{\psi^0}\right) - \sum_{i=1}^{n} m_i \log\left(\frac{1+\psi_i}{1+\psi^0}\right)
\end{aligned}
$$

1.1 Distribuição do teste de razão de verossimilhanças

Podemos pensar a hipótese $H_0 : \psi_1 = \psi_2 = ... = \psi_n$ como os pares $\psi_1 = \psi_2$ e $\psi_2 = \psi_3$, . . . , $\psi_{n-1} = \psi_n$. Essa formulação pode ser escrita como $H_0 : C\beta$, com $\beta = (\psi_1, \psi_2, ..., \psi_n)^T$ e

$$
C = \begin{bmatrix}
1 & -1 & 0 & \dots & 0 & 0 \\
0 & 1 & -1 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \dots & 1 & -1
\end{bmatrix}_{(k-1\times p)}
$$

Assim, como a matriz é construída com $k-1$ linhas, a distribuição assintótica da estatística do teste estudado é $\xi_{RV} \to \chi^2_{k-1}$.

1.2 Obtendo $\hat{\psi}_i$ e $\hat{\psi}^0$

Obtemos as estimativas de ψ_i através de sua função escore, que é a derivada do logaritmo da função de verossimilhança, e igualando à zero.

$$L(y_i, \psi_i) = \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log(1 + \psi_i) + \sum_{i=1}^{n} \log \binom{m_i}{y_i}$$

$$\begin{aligned}
\frac{\partial L}{\partial \psi_i} &= \frac{\partial}{\partial \psi_i} \sum_{i=1}^{n} y_i \log(\psi_i) - \sum_{i=1}^{n} m_i \log(1 + \psi_i) + \sum_{i=1}^{n} \log \binom{m_i}{y_i} \\
&= y_i \frac{1}{\psi_i} - m_i \frac{1}{1 + \psi_i}
\end{aligned}$$

Igualando a expressão anterior à zero temos chegamos na estimativa a seguir:

$$\begin{aligned}
y_i \frac{1}{\hat{\psi}_i} - m_i \frac{1}{1 + \hat{\psi}_i} &= 0 \\
y_i(1 + \hat{\psi}_i) &= m_i \hat{\psi}_i \\
y_i + y_i \hat{\psi}_i - m_i \hat{\psi}_i &= 0 \\
\hat{\psi}_i (y_i - m_i) &= -y_i \\
\hat{\psi}_i (m_i - y_i) &= y_i \\
\hat{\psi}_i &= \frac{y_i}{m_i - y_i}
\end{aligned}$$

Um procedimento similar chega em ψ^0.

$$\begin{aligned}
L(y_i, \psi^0) &= \log(\psi^0) \sum_{i=1}^{n} y_i - \log\left(1 + \psi^0\right) \sum_{i=1}^{n} m_i + \sum_{i=1}^{n} \log \binom{m_i}{y_i} \\
\frac{\partial L}{\partial \psi^0} &= \frac{1}{\psi^0} \sum_{i=1}^{n} y_i - \frac{1}{1 + \psi^0} \sum_{i=1}^{n} m_i
\end{aligned}$$

Igualando a expressão anterior à zero temos chegamos na estimativa a seguir:

$$\begin{aligned}
\frac{1}{\hat{\psi}^0} \sum_{i=1}^{n} y_i - \frac{1}{1 + \hat{\psi}^0} \sum_{i=1}^{n} m_i &= 0 \\
(1 + \hat{\psi}^0) \sum_{i=1}^{n} y_i - \hat{\psi}^0 \sum_{i=1}^{n} m_i &= 0 \\
\sum_{i=1}^{n} y_i + \hat{\psi}^0 \sum_{i=1}^{n} y_i - \hat{\psi}^0 \sum_{i=1}^{n} m_i &= 0 \\
-\sum_{i=1}^{n} y_i - \hat{\psi}^0 \sum_{i=1}^{n} y_i + \hat{\psi}^0 \sum_{i=1}^{n} m_i &= 0
\end{aligned}$$

$$\hat{\psi}^0(\sum_{i=1}^n m_i - \sum_{i=1}^n y_i) = \sum_{i=1}^n y_i$$
$$\hat{\psi}^0 = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i - \sum_{i=1}^n y_i}$$

Ou

$$\hat{\psi}^0 = \frac{n\bar{y}}{n\bar{m} - n\bar{y}} = \frac{\bar{y}}{\bar{m} - \bar{y}}$$

2 Exercício 2

É apresentado no enunciado $Y_i|Xi \overset{ind}{\sim} NI(\mu, \phi_i)$ com o o seguinte componente sistemático:

$$\log(\phi_i) = \lambda_i = \gamma_0 + \gamma_1(x_i - t_0)_+$$

Onde $(x_i - t_0)_+ = 0$ se i $x_i \leq t_0$ e $(x_i - t_0)_+ = x_i - t_0$ caso contrário. É pedido que calculemos Var($\hat{\gamma}_0$), Var($\hat{\gamma}_1$) e Cov($\hat{\gamma}_0, \hat{\gamma}_1$).

Iremos calcular os componentes necessários de maneira genérica uma vez que a variável Y_i pertence à família exponencial dupla. Vamos trabalhar com a verossimilhança restrita da variável $T_i = t_i$, onde, genericamente, $t_i = y_i\theta - b(\theta) + u(y_i)$ e, especificamente para uma variável normal inversa $t_i = -\left(\frac{y_i}{2\mu^2} - \frac{1}{\mu} + \frac{1}{2y_i}\right)$.

Assim, a derivação da função escore U_γ, $\gamma = (\gamma_0, \gamma_1)^T$ será com base na função de verossimilhança da variável transformada T_i.

$$L(t_i; \phi_i, \mu) = \sum_{i=1}^n \phi_i t_i + \mathrm{d}(\phi_i) + \mathrm{u}(y_i)$$

Para $k = 0, 1$:

$$\begin{aligned}
U_{\gamma_k} &= \frac{\partial L}{\partial \gamma_k} \\
&= \frac{\partial}{\partial \gamma_k} \sum_{i=1}^n \phi_i t_i + \mathrm{d}(\phi_i) + \mathrm{u}(y_i) \\
&= \sum_{i=1}^n \left[\frac{d\phi_i}{d\lambda_i}\frac{\partial \lambda_i}{\partial \gamma_k} t_i + \mathrm{d}'(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial \lambda_i}{\partial \gamma_k}\right] \quad (1) \\
&= \sum_{i=1}^n \left[\frac{d\phi_i}{d\lambda_i}\frac{\partial \lambda_i}{\partial \gamma_k}(t_i + \mathrm{d}'(\phi_i))\right] \quad (2)
\end{aligned}$$

Note que:

$$\frac{d\phi_i}{d\lambda_i} = \left(\frac{d\lambda_i}{d\phi_i}\right)^{-1} = \left(\frac{d\log\phi_i}{d\phi_i}\right)^{-1} = \phi_i$$

E

$$\frac{\partial \lambda_i}{\partial \gamma_k} = \frac{\partial}{\partial \gamma_k}\gamma_0 + \gamma_1(x_i - t_0)_+$$

Assim, $\frac{\partial\lambda_i}{\partial\gamma_k} = 1$ se $k = 0$ e $\frac{\partial\lambda_i}{\partial\gamma_k} = (x_i - t_0)_+$ se $k = 1$.

As funções escore são então:

$$U_{\gamma_0} = \sum_{i=1}^{n} \left[\phi_i(t_i + \mathrm{d}'(\phi_i))\right]$$

$$U_{\gamma_1} = \sum_{i=1}^{n} \left[\phi_i(x_i - t_0)_+(t_i + \mathrm{d}'(\phi_i))\right]$$

Para obter a Hessiana e em seguida os componentes da informação de Fisher,vamos derivar as funções escore:

$$\begin{aligned}
\frac{\partial U_{\gamma_0}}{\partial\gamma_0} &= \frac{\partial U_{\gamma_0}}{\partial\gamma_0}\sum_{i=1}^{n} \left[\phi_i(t_i + \mathrm{d}'(\phi_i))\right] \\
&= \sum_{i=1}^{n} \left[\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + \phi_i\frac{\partial\mathrm{d}'(\phi_i)}{\partial\gamma_0}\right] \\
&= \sum_{i=1}^{n} \left[\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + \phi_i\mathrm{d}''(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial\lambda_i}{\partial\gamma_0}\right] \\
&= \sum_{i=1}^{n} \left[\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + \phi_i^2\mathrm{d}''(\phi_1)\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial U_{\gamma_1}}{\partial\gamma_1} &= \frac{\partial U_{\gamma_1}}{\partial\gamma_1}\sum_{i=1}^{n} \left[(x_i - t_0)_+\phi_i(t_i + \mathrm{d}'(\phi_i))\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_1}(t_i + \mathrm{d}'(\phi_i)) + (x_i - t_0)_+\phi_i\frac{\partial\mathrm{d}'(\phi_i)}{\partial\gamma_1}\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_1}(t_i + \mathrm{d}'(\phi_i)) + (x_i - t_0)_+\phi_i\mathrm{d}''(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial\lambda_i}{\partial\gamma_1}\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_1}(t_i + \mathrm{d}'(\phi_i)) + \phi_i^2\left[(x_i - t_0)_+\right]^2\mathrm{d}''(\phi_i)\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial U_{\gamma_1}}{\partial\gamma_0} &= \frac{\partial U_{\gamma_1}}{\partial\gamma_0}\sum_{i=1}^{n} \left[(x_i - t_0)_+\phi_i(t_i + \mathrm{d}'(\phi_i))\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + (x_i - t_0)_+\phi_i\frac{\partial\mathrm{d}'(\phi_i)}{\partial\gamma_0}\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + (x_i - t_0)_+\phi_i\mathrm{d}''(\phi_i)\frac{d\phi_i}{d\lambda_i}\frac{\partial\lambda_i}{\partial\gamma_0}\right] \\
&= \sum_{i=1}^{n} \left[(x_i - t_0)_+\frac{\partial\phi_i}{\partial\gamma_0}(t_i + \mathrm{d}'(\phi_i)) + (x_i - t_0)_+\phi_i^2\mathrm{d}''(\phi_i)\right]
\end{aligned}$$

O valor esperado do negativo da das componentes da Hessiana são dados por:

$$\mathrm{K}\gamma_0\gamma_0 = E\left[-\frac{\partial U_{\gamma_0}}{\partial\gamma_0}\right] = -\sum_{i=1}^{n}\left[\phi_i^2\mathrm{d}''(\phi_i)\right]$$

$$\mathrm{K}\gamma_1\gamma_1 = E\left[-\frac{\partial U_{\gamma_1}}{\partial \gamma_1}\right] = -\sum_{i=1}^{n}\left[\phi_i^2\left[(x_i - t_0)_+\right]^2 \mathrm{d}''(\phi_i)\right]$$
$$\mathrm{K}\gamma_1\gamma_0 = E\left[-\frac{\partial U_{\gamma_1}}{\partial \gamma_0}\right] = -\sum_{i=1}^{n}\left[(x_i - t_0)_+\phi_i^2 \mathrm{d}''(\phi_i)\right]$$

Uma vez que $E(t_i) = -\mathrm{d}'(\phi)$. Outro ponto importante é, como a variável em questão é uma variável Normal inversa, seu d"$(\phi_i) = -(2\phi_i^2)^{-1}$. Nomeando como sugerido $z_i = (x_i - t_0)_+$ temos as seguinte expressões

$$\mathrm{K}_{\gamma_0\gamma_0} = \sum_{i=1}^{n}\frac{1}{2} = \frac{n}{2}$$
$$\mathrm{K}_{\gamma_1\gamma_1} = \frac{1}{2}\sum_{i=1}^{n} z_i^2 = \frac{1}{2}\sum_{i=r}^{n} z_i^2$$
$$\mathrm{K}_{\gamma_1\gamma_0} = \frac{1}{2}\sum_{i=1}^{n} z_i = \frac{1}{2}\sum_{i=r}^{n} z_i = \frac{1}{2}(n-r)\bar{z}_+$$

Onde, por consequência da definição de $z_i = (x_i - t_0)_+$, o r primeiros termos são nulos e $\bar{z}_+$ representa a média amostral dos $n - r$ valores não nulos ($x_i > t_0$). Então a matriz de Informação de Fisher é dada por:

$$\mathrm{K}_{\gamma\gamma} = \frac{1}{2}\begin{bmatrix} n & (n-r)\bar{z}_+ \\ (n-r)\bar{z}_+ & \sum_{i=r}^{n} z_i^2 \end{bmatrix}$$

As quantias Var$(\hat{\gamma}_0)$, Var$(\hat{\gamma}_1)$ e Cov$(\hat{\gamma}_0, \hat{\gamma}_1)$ são dados a partir do elemento correspondente em $\mathrm{K}_{\gamma\gamma}^{-1}$. O determinante de $\mathrm{K}_{\gamma\gamma}$ é:

$$\det \mathrm{K}_{\gamma\gamma} = n\sum_{i=r}^{n} z_i^2 - (n-r)^2\bar{z}_+^2$$

E

$$\mathrm{K}_{\gamma\gamma}^{-1} = \frac{2}{n\sum_{i=r}^{n} z_i^2 - (n-r)^2\bar{z}_+^2}\begin{bmatrix} \sum_{i=r}^{n} z_i^2 & -(n-r)\bar{z}_+ \\ -(n-r)\bar{z}_+ & n \end{bmatrix}$$

Em conclusão:

$$\mathrm{Var}(\hat{\gamma}_0) = \frac{2\sum_{i=r}^{n} z_i^2}{n\sum_{i=r}^{n} z_i^2 - (n-r)^2\bar{z}_+^2}$$

$$\mathrm{Var}(\hat{\gamma}_1) = \frac{2}{\sum_{i=r}^{n} z_i^2 - n^{-1}(n-r)^2\bar{z}_+^2}$$

$$\mathrm{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = -\frac{(n-r)\bar{z}_+}{n\sum_{i=r}^{n} z_i^2 - (n-r)^2\bar{z}_+^2}$$

2.1 Estatística do teste de Escore

Em seguida vamos calcular a estatística de escore para testar $H_0 : \gamma_1 = 0$ contra $H_1 : \gamma_1 \neq 0$. Sua definição aplicada ao modelo estudado é

$$\xi_{SR} = \left(U^0_{\gamma_1}\right)^2 \mathrm{Var}_0(\hat{\gamma}_1)$$

Vamos primeiro trabalhar a função escore que achamos anteriormente para que fique em uma versão mais amigável:

$$\begin{aligned}
U_{\gamma_1} &= \sum_{i=1}^{n} \left[\phi_i (x_i - t_0)_+ (t_i + \mathrm{d}'(\phi_i))\right] \\
&= \sum_{i=1}^{n} \left[\phi_i z_i \left(t_i + \frac{1}{2\phi_i}\right)\right] \\
&= \sum_{i=1}^{n} \left[\phi_i z_i \left(\frac{2\phi_i t_i + 1}{2\phi_i}\right)\right] \\
&= \sum_{i=1}^{n} \left[z_i \left(\frac{2\phi_i t_i + 1}{2}\right)\right] \\
&= \sum_{i=1}^{n} \left[z_i \left(\phi_i t_i + \frac{1}{2}\right)\right] \\
&= \sum_{i=r}^{n} \left[z_i \left(\phi_i t_i + \frac{1}{2}\right)\right] \quad \text{pela definição de } z_i \\
&= \sum_{i=r}^{n} \left[\left(z_i \phi_i t_i + \frac{z_i}{2}\right)\right] \\
&= \sum_{i=r}^{n} z_i \phi_i t_i + \sum_{i=r}^{n} \frac{z_i}{2} \\
&= \frac{1}{2}(n-r)\bar{z}_+ + \sum_{i=r}^{n} z_i \phi_i t_i
\end{aligned}$$

Sob H_0, $\phi_i \to \phi^0$ e :

$$U^0_{\gamma_1} = \frac{1}{2}(n-r)\bar{z}_+ + \phi^0 \sum_{i=r}^{n} z_i t_i$$

Para a variância, note que nenhum termo depende do parâmetro testado:

$$\mathrm{Var}(\hat{\gamma}_1) = \mathrm{Var}_0(\hat{\gamma}_1) = \frac{2}{\sum_{i=r}^{n} z_i^2 - n^{-1}(n-r)^2 \bar{z}_+^2}$$

Concluíndo a montagem da estatística de escore:

$$\begin{aligned}
\xi_{SR} &= \left(U^0_{\gamma_1}\right)^2 \mathrm{Var}_0(\hat{\gamma}_1) \\
&= \frac{\left((n-r)\bar{z}_+ + 2\phi^0 \sum_{i=r}^{n} z_i t_i\right)^2}{\sum_{i=r}^{n} z_i^2 - n^{-1}(n-r)^2 \bar{z}_+^2}
\end{aligned}$$

3 Exercício 3

A função de estimação parte da seguinte expressão:

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \sum_{i=1}^{n} E\left[\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\beta}}\right]^T \mathrm{Var}(\boldsymbol{u}_i)^{-1}\boldsymbol{u}_i$$

Nos foi fornecido que, para o caso de respostas marginais binomiais negativas, $\boldsymbol{u}_i = \boldsymbol{y_i} - \boldsymbol{\mu_i}$ onde $\boldsymbol{y_i} = (y_{i1}, \ldots, y_{ir_i})^T$ e $\boldsymbol{\mu_i} = (\mu_{i1}, \ldots, \mu_{ir_i})^T$. Assim, temos

$$\begin{aligned}
E\left[\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\beta}}\right]^T &= E\left[\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{y_i} - \boldsymbol{\mu_i})\right]^T \\
&= E\left[-\frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}}\right]^T \\
&= E\left[-\frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\eta_i}}\frac{\partial \boldsymbol{\eta_i}}{\partial \boldsymbol{\beta}}\right]^T \\
&= E\left[-\boldsymbol{F_i}\boldsymbol{X_i}\right]^T \\
&= (-\boldsymbol{F_i}\boldsymbol{X_i})^T \\
&= -\boldsymbol{X_i}^T\boldsymbol{F_i}
\end{aligned}$$

Onde $\boldsymbol{X_i}$ é a matriz de variáveis explicativas das observações no grupo i e $\boldsymbol{F_i} = \mathrm{diag}\left\{\frac{d\mu_{i1}}{d\eta_{i1}}, \ldots \frac{d\mu_{ir_i}}{d\eta_{ir_i}}\right\}$, ou seja, uma ponderação com relação à função de ligação usada. Seguindo a construção:

$$\mathrm{Var}(\boldsymbol{u_i}) = \mathrm{Var}(\boldsymbol{y_i}) = \boldsymbol{V_i}^{1/2}\boldsymbol{R}_i(\alpha)\boldsymbol{V_i}^{1/2}$$

No caso de respostas marginais binomias negativas, a variância delas é $\mathrm{Var}(Y_{ij}) = \mu_{ij} + \frac{\mu_{ij}}{\nu}$. Então, matriz de variâncias tem o seguinte formato:

$$\boldsymbol{V_i} = \mathrm{diag}\left\{\mu_{i1} + \frac{\mu_{i1}}{\nu}, ..., \mu_{ir_i} + \frac{\mu_{ir_i}}{\nu}\right\}$$

Então, munidos dos dois fatore necessários, a função de estimação $\boldsymbol{\Psi}(\boldsymbol{\beta})$ é

$$\begin{aligned}
\boldsymbol{\Psi}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} E\left[\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\beta}}\right]^T \mathrm{Var}(\boldsymbol{u}_i)^{-1}\boldsymbol{u}_i \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T\boldsymbol{F_i}\left[\boldsymbol{V_i}^{1/2}\boldsymbol{R}_i(\alpha)\boldsymbol{V_i}^{1/2}\right]^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu_i})
\end{aligned}$$

As passagens à seguir mostram que a função de estimação apresentada é não visada:

$$E\left[\boldsymbol{\Psi}(\boldsymbol{\beta})\right] = E\left[\sum_{i=1}^{n} -\boldsymbol{X_i}^T\boldsymbol{F_i}\left[\boldsymbol{V_i}^{1/2}\boldsymbol{R}_i(\alpha)\boldsymbol{V_i}^{1/2}\right]^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu_i})\right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} E\left[-\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \right] \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} E\left[(\boldsymbol{y_i} - \boldsymbol{\mu_i}) \right] \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (E\left[\boldsymbol{y_i} \right] - \boldsymbol{\mu_i}) \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (\boldsymbol{\mu_i} - \boldsymbol{\mu_i}) \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (0) \\
&= 0
\end{aligned}
$$

3.1 Processo iterativo para $\hat{\beta}$

No caso de funções de estimação, a estimativa é obtida a partir de suas raízes, fazendo $\boldsymbol{\Psi}(\hat{\boldsymbol{\beta}}) = 0$. Esta construção culmina assim no nome equações de estimação. O processo iterativo para se encontrar essas raízes é:

$$
\boldsymbol{\beta^{(m+1)}} = \boldsymbol{\beta^{(m)}} - \left(E\left[\boldsymbol{\Psi'}(\boldsymbol{\beta^{(m)}}) \right] \right)^{-1} \boldsymbol{\Psi}(\boldsymbol{\beta^{(m)}})
$$

Onde $m = 1, 2, \ldots$. Durante esse procedimento estamos supondo α, o parâmetro da matriz trabalho (correlação intragrupos) e ν, parâmetro associado com a dispersão da variável resposta, fixos. Note que é preciso determinar $\boldsymbol{\beta^{(0)}}$ para que possa iniciar o processo iterativo. O próximo passo então é encontrar uma forma para a derivada da função de estimação que estamos estudando. Os passos a seguir apresentam isso:

$$
\begin{aligned}
\boldsymbol{\Psi'}(\boldsymbol{\beta^{(m)}}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \\
&= \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \\
&= \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\mu_i} \\
&= \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i}
\end{aligned}
$$

Onde a última igualdade segue o mesmo raciocínio que usamos para montar a função de estimação $\boldsymbol{\Psi}(\boldsymbol{\beta})$ no início do exercício. Assim, o próximo passo é montar o processo iterativo usando as matrizes conhecidas:

$$
\begin{aligned}
\boldsymbol{\beta^{(m+1)}} &= \boldsymbol{\beta^{(m)}} - \left(E\left[\boldsymbol{\Psi'}(\boldsymbol{\beta^{(m)}}) \right] \right)^{-1} \boldsymbol{\Psi}(\boldsymbol{\beta^{(m)}}) \\
\boldsymbol{\beta^{(m+1)}} &= \boldsymbol{\beta^{(m)}} - \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} -\boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \\
\boldsymbol{\beta^{(m+1)}} &= \boldsymbol{\beta^{(m)}} + \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{F_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i})
\end{aligned}
$$

Onde a inserção pré multiplicada de $\boldsymbol{F_i} \boldsymbol{F_i}^{-1}$ está relacionada com a próxima passagem. Note que podemos multiplicar convenientemente $\boldsymbol{\beta^{(m)}}$ de tal forma que:

$$\beta^{(m)} = \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \boldsymbol{\beta}^{(m)}$$

E, "colocando em evidência" tudo menos o termo mais à direita $\boldsymbol{X_i}\boldsymbol{\beta}^{(m)}$, que nada mais é o preditor linear, chegamos à:

$$\begin{aligned} \beta^{(m+1)} &= \beta^{(m)} + \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{F_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \\ &= \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i} \left(\boldsymbol{X_i} \boldsymbol{\beta}^{(m)} + \boldsymbol{F_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \right) \end{aligned}$$

Note que as matrizes $\boldsymbol{F_i}$, $\boldsymbol{V_i}$, $\boldsymbol{R_i}(\alpha)$ e o vetor $\boldsymbol{\mu_i}$ devem ser recalculados à cada passo m do processo iterativo. Nomeando os seguintes fatores:

$$\boldsymbol{W_i} = \boldsymbol{F_i} \left[\boldsymbol{V_i}^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{V_i}^{1/2} \right]^{-1} \boldsymbol{F_i}$$

$$\boldsymbol{y_i^*} = \boldsymbol{X_i} \beta^{(m)} + \boldsymbol{F_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i})$$

Temos assim a matriz de pesos $\boldsymbol{W_i}$ e a variável resposta modificada $\boldsymbol{y_i^*}$. O processo iterativo por fim é dado por:

$$\beta^{(m+1)} = \left(\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{W_i}^{(m)} \boldsymbol{X_i} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{W_i}^{(m)} \boldsymbol{y_i^{*(m)}}$$

3.2 Estimador robusto para Var($\hat{\boldsymbol{\beta}}$)

Como a matriz trabalho de covariâncias intra grupos $R_i(\alpha)$ pode ser mal especificada, levando à estimadores inconsistentes das Var($\hat{\boldsymbol{\beta}}$), seguinte estimador é considerado robusto à esse tipo de falha. Nomeando as seguintes matrizes:

$$\boldsymbol{H_1}(\hat{\beta}) = \sum_{i=1}^{n} \boldsymbol{X_i}^T \hat{\boldsymbol{F}}_{\boldsymbol{i}} \left[\hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \boldsymbol{R}_i(\hat{\alpha}) \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \right]^{-1} \hat{\boldsymbol{F}}_{\boldsymbol{i}} \boldsymbol{X_i}$$

$$\boldsymbol{H_2}(\hat{\beta}) = \sum_{i=1}^{n} \boldsymbol{X_i}^T \hat{\boldsymbol{F}}_{\boldsymbol{i}} \left[\hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \boldsymbol{R}_i(\hat{\alpha}) \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \right]^{-1} (\boldsymbol{y_i} - \hat{\boldsymbol{\mu}}_{\boldsymbol{i}})(\boldsymbol{y_i} - \hat{\boldsymbol{\mu}}_{\boldsymbol{i}})^T \left[\hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \boldsymbol{R}_i(\hat{\alpha}) \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{1/2} \right]^{-1} \hat{\boldsymbol{F}}_{\boldsymbol{i}} \boldsymbol{X_i}$$

Com essas duas matrizes, podemos ter o seguinte estimador para a Var($\hat{\boldsymbol{\beta}}$):

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left[\boldsymbol{H_1}(\hat{\beta}) \right]^{-1} \boldsymbol{H_2}(\hat{\beta}) \left[\boldsymbol{H_1}(\hat{\beta}) \right]^{-1}$$

3.3 Estimando ν

A estimativa para o parâmetro ν, que está relacionada à dispersão da distribuição marginal de uma variável binomial negativa, não é clara dado o procedimento apresentado. Uma primeira ideia seria utilizar um estimador por método de momentos. Outra seria concluir que é natural de se pensar que a estimativa para um parâmetro de dispersão esteja relacionado com algum tipo de resíduo. Talvez a abordagem de equações de estimação generalizadas por meio da quasi-verossimelhança possibilite a criação de um quasi-desvio que culmine na estimação do parâmetro ν da mesma forma que a estimativa de máxima verossimilhança do parâmetro de ϕ no caso dos MLG se dá por meio do desvio.

Sobre em que etapa estimar o ν este pode acontecer durante o processo iterativo ou após. Para se entender melhor quando fazê-lo, considerações como complexidade computacional devem ser levadas em consideração.

4 Exercício 4

Supondo o modelo descrito no enunciado, vamos investigar o comportamento da medida de alavancagem $\hat{h}_{ii}$ de um ponto com relação à sua estimativa $\hat{\mu}_i$ e sua posição $x_i - \bar{x}$. A parte sistemática é dada por:

$$\eta_i = \beta(x_i - \bar{x})$$

Supondo a ligação canônica, vamos estudar a medida de alavancagem para 4 distribuições: binomial, poisson, gamma e normal inversa. A medida de alavancagem que usaremos é:

$$\mathrm{GL}_{ii} = \omega_i x_i^T (X^T W X)^{-1} x_i$$

Onde $W = \mathrm{diag}\{w_1, \ldots, w_n\}$ e $w_i = \left(\frac{dmu_i}{s\eta_i}\right)^2 V_i^{-1}$ e V_i^{-1} é a função de variância. A matriz X modelo é:

$$X = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

Desta forma temos

$$X^T W = \begin{bmatrix} (x_1 - \bar{x})\omega_1 & \ldots (x_n - \bar{x})\omega_n \end{bmatrix}$$

E

$$X^T W X = \begin{bmatrix} (x_1 - \bar{x})\omega_1 & \ldots (x_n - \bar{x})\omega_n \end{bmatrix} \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$
$$= \sum_{i=1}^{n} \omega_i (x_i - \bar{x})^2$$

E por fim a medida de alavanca

$$h_{ii} = \mathrm{GL}_{ii} = \frac{\omega_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} \omega_i (x_i - \bar{x})^2}$$

A suposição de ligação canônica faz com que $\omega_i = V_i = V(\mu_i)$, ou seja, a matriz diagonal de pesos resulta em apenas uma matriz diagonal com as funções de variância da distribuição estudada. Ou seja

$$h_{ii} = \frac{V(\mu_i)(x_i - \bar{x})^2}{\sum_{i=1}^{n} V(\mu_i)(x_i - \bar{x})^2}$$

Sua estimativa no caso dos MLG consiste:

$$\hat{h}_{ii} = \frac{V(\hat{\mu}_i)(x_i - \bar{x})^2}{\sum_{i=1}^{n} V(\hat{\mu}_i)(x_i - \bar{x})^2}$$

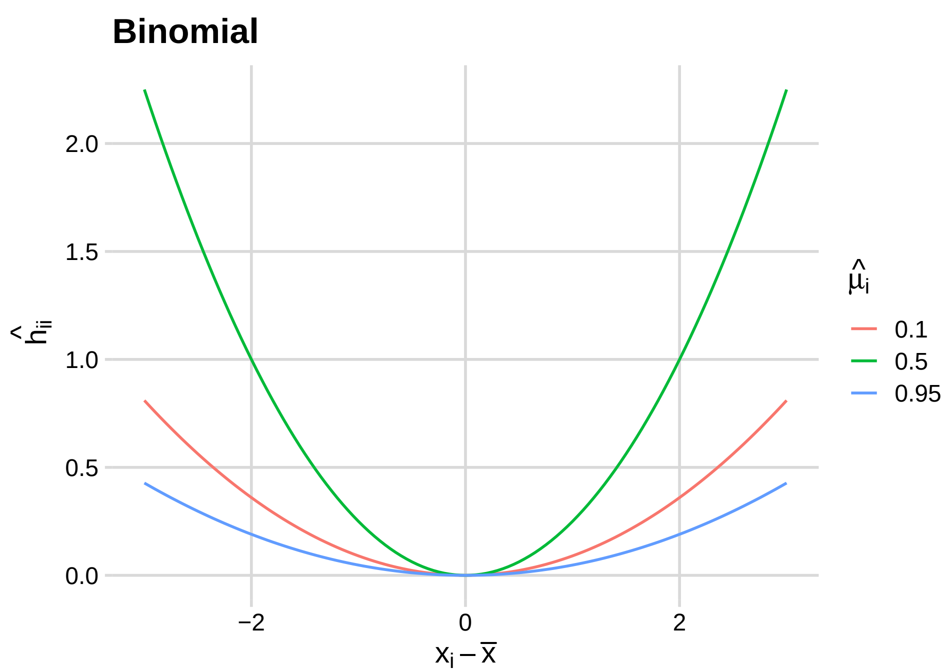
Assim, vamos estudar os casos de cada uma das distribuições requisitadas. No entanto, a suposição crucial para entender a relação entre $\hat{\mu}_i$, $(x_i - \bar{x})$ e $\hat{h}_{ii}$ é que a alteração via sua estimativa ou sua variável explicativa de um ponto não modifica o denominador de relevantemente. Isto é possível se a amostra for grande de tal forma que a alteração de um ponto i não altera $\sum_{i=1}^{n} V(\hat{\mu}_i)(x_i - \bar{x})^2$.

4.1 Binomial

A seguir temos a expressão de alavancagem para a distribuição binomial:

$$\hat{h}_{ii} = \frac{\hat{\mu}_i(1 - \hat{\mu}_i)(x_i - \bar{x})^2}{\sum_{i=1}^{n} \hat{\mu}_i(1 - \hat{\mu}_i)(x_i - \bar{x})^2}$$

E abaixo podemos ver a relação entre $\hat{h}_{ii}$ e alguns casos relevantes $\hat{\mu}_i$ ao longo de $x_i - \bar{x}$:



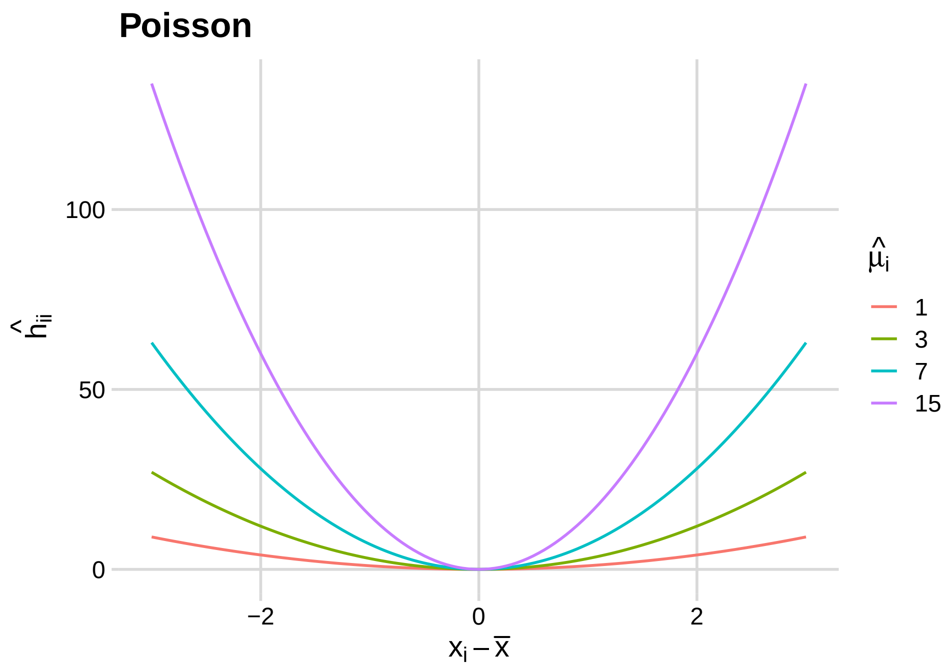
Podemos ver que o caso de maior impacto no valor da alavancagem ocorre quando a estimativa de localização está próxima de 0.50.

4.2 Poisson

Abaixo temos a medida de alavancagem para a distribuição de poisson:

$$\hat{h}_{ii} = \frac{\hat{\mu}_i(x_i - \bar{x})^2}{\sum_{i=1}^{n} \hat{\mu}_i(x_i - \bar{x})^2}$$

O gráfico da relação entre as grandezas e a alavancagem é



Podemos ver que quanto menor é a estimativa, menor será o impacto de grandes desvios da média. Todavia, diferentemente da distribuição binomial, aqui relação entre $\hat{\mu}_i$ e $\hat{h}_{ii}$ não é limitado.

4.3 Gamma e normal inversa

Vamos comparar as distribuições Gamma e Normal Inversa simultâneamente, uma vez que ambas modelam dados assimétricos positivos. Abaixo mostro as estimativas de h_{ii}.

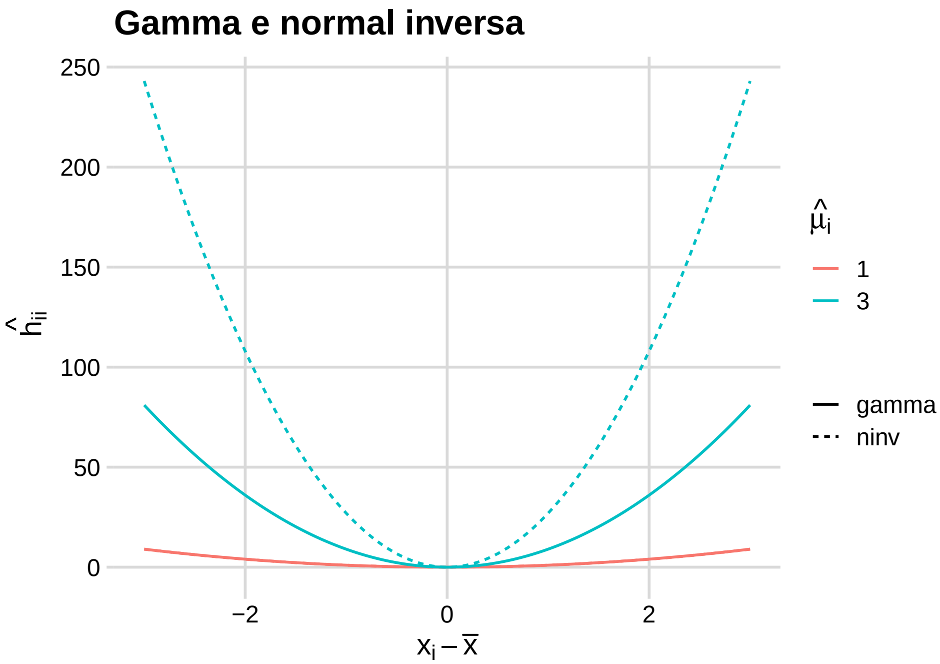
Caso da distribuição gamma:

$$\hat{h}_{ii} = \frac{\hat{\mu}_i^2 (x_i - \bar{x})^2}{\sum_{i=1}^{n} \hat{\mu}_i^2 (x_i - \bar{x})^2}$$

e para a distribuição normal inversa:

$$\hat{h}_{ii} = \frac{\hat{\mu}_i^3 (x_i - \bar{x})^2}{\sum_{i=1}^{n} \hat{\mu}_i^3 (x_i - \bar{x})^2}$$

Em seguida mostro a relação entre as quantidades. Note que, devido aos expoentes presentes nas funções de variância, vamos restringir $\hat{\mu}_i$ à valores menores na intenção de facilitar a comparação entre ambas as distribuições.



Assim, para valores menores de $\hat{\mu}_i$ as alavancagens são muito parecidas entre a normal inversa e gamma. Todavia, quanto maior é a estimativa, os desvios $x_i - \bar{x}$ acarretam em valores muito maiores de $\hat{h}_{ii}$ no caso da normal inversa.