Resumo sobre modelos lineares generalizados

Prof. Caio Azevedo

Família exponencial bi-paramétrica

- Dizemos que uma v.a. Y (discreta ou contínua) pertence à família exponencial biparamétrica se sua fdp é dada por:

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

em que $\theta \in \Theta \subset \Re$ (espaço paramétrico), $b(\theta), c(y,\phi) : \Re \rightarrow \Re$, A é um conjunto que não depende de θ e ϕ e, por sua vez, $\phi(\phi > 0)$ é o parâmetro de precisão.

Família exponencial bi-paramétrica

- Consideraremos que $Y \sim \text{FE}(\theta, \phi)$ e que temos $Y_i \overset{ind.}{\sim} FE(\theta_i, \phi), i = 1, 2, ..., n$, ou seja

$$f(y_i; \theta) = exp\left\{\phi\left[y\theta_i - b(\theta_i)\right] + c(y_i, \phi)\right\} \mathbb{1}_A(y_i)$$

Modelo linear generalizado

- $Y_i \sim \text{FE}(\theta_i, \phi), i = 1, 2, ..., n$, $\theta_i = h(\mu_i)$.
- $g(\mu_i) = \eta_i$, $\eta_i = \sum_{j=1}^{p} x_{ji}\beta_j$, $\mathcal{E}(Y_i) = \mu_i = b'(\theta_i)$, x_{ji} : covariável j associada ao indivíduo i (fixa e conhecida) e $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)', \phi$: parâmetros desconhecidos.
- $\mathcal{V}(Y) = \phi^{-1}V(\mu_i)$, em que $V(\mu_i) = \dfrac{d\mu_i}{d\theta_i}$.
- $g(.)$ é uma função de ligação (inversível e duplamente diferenciável). Quando $\theta = g(.)$ temos a função de ligação canônica.

Principais distribuições pertencentes à FE

Distribuição	$b(\theta)$	θ	ϕ	$V(\mu)$
Normal	$\theta^2/2$	μ	σ^{-2}	1
Poisson	e^{θ}	$\ln \mu$	1	μ
Binomial	$\ln(1+e^{\theta})$	$\ln(\mu/(1-\mu))$	n	$\mu(1-\mu)$
Gama	$-\ln(-\theta)$	$-1/\mu$	$1/(CV^2)$	μ^2
N.Inversa	$-\sqrt{-2\theta}$	$-1/2\mu^2$	ϕ	μ^3

Desvio (ou função desvio)

- Sem perda de generalidade, seja $l(\boldsymbol{\mu}, \mathbf{y}) \equiv l(\boldsymbol{\beta}, \phi)$ a logverossimilhança associada ao modelo em estudo e $l(\boldsymbol{\mu}^0, \mathbf{y})$ a logverossimilhança do modelo saturado (n=p), ou seja, em que cada média é representada por ela mesma.
- $l(\boldsymbol{\mu}, \mathbf{y}) = \phi\left[\sum_{i=1}^{n} y_i\theta_i - \sum_{i=1}^{n} b(\theta_i)\right] + \sum_{i=1}^{n} c(y_i, \phi)$
- $l(\boldsymbol{\mu}^{(0)}, \mathbf{y}) = \phi\left[\sum_{i=1}^{n} y_i\theta_i^{(0)} - \sum_{i=1}^{n} b(\theta_i^{(0)})\right] + \sum_{i=1}^{n} c(y_i, \phi)$
  em que $\theta_i^{(0)} = h(\mu_i^{(0)})$.
- Assim, $D^*(\mathbf{y}, \boldsymbol{\mu}) = 2(l(\boldsymbol{\mu}^{(0)}, \mathbf{y}) - l(\boldsymbol{\mu}, \mathbf{y})) = \phi\sum_{i=1}^{n}\left[y_i\left(\theta_i^{(0)} - \theta_i\right) + b(\theta_i) - b(\theta_i^{(0)})\right] = \phi D(\mathbf{y}, \boldsymbol{\mu})$

Desvio (ou função desvio)

- Sejam $\widehat{\boldsymbol{\mu}}^{(0)} = \mathbf{Y}$ e $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}}) = g^{-1}(\mathbf{X}\widehat{\boldsymbol{\beta}})$ os respectivos estimadores de MV e defina $\widehat{\theta}_i^{(0)} = h(\widehat{\mu}_i^{(0)})$ e $\widehat{\theta}_i = h(\widehat{\mu}_i)$.
- Assim, o desvio não escalonado é dado por $D(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \left[y_i \left(\widehat{\theta}_i^{(0)} - \widehat{\theta}_i \right) + b(\widehat{\theta}_i) - b(\widehat{\theta}_i^{(0)}) \right]$.
- O desvio escalonado é dado por $D^*(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}, \widehat{\boldsymbol{\mu}})$. Na prática, substituimos ϕ por algum estimador consistente.

Desvio: exemplos

- Já vimos como ficam os desvios no caso dos modelos Poisson e binomial (lembrando que $\phi = 1$) para esses modelos.
- Normal: $D(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2$ e $D^*(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2$.
- Gama: $D(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^{n} \{-\ln(y_i/\widehat{\mu}_i) + (y_i - \widehat{\mu}_i)/\widehat{\mu}_i\}$ e $D^*(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}, \widehat{\boldsymbol{\mu}})$.

Desvio: comportamento assintótico

- Já vimos o comportamento assintótico para os modelos Poisson e binomial.
- Para os outros modelos, $D^*(\mathbf{y}, \widehat{\boldsymbol{\mu}})$ depende de ϕ. Então, nesses casos, utilizamos uma estimativa consistente de ϕ. Além disso, temos que

$$D^*(\mathbf{y}, \widehat{\boldsymbol{\mu}}) \xrightarrow[\phi \to \infty]{D} \chi^2_{n-p}$$

- Assim, podemos comparar o valor do desvio com os quantis de uma distribuição qui-quadrado, para verificar a qualidade de ajuste do modelo, quando a variabilidade dos dados for pequena.

Análise do desvio

- Podemos definir um procedimento para testar as hipóteses $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ vs $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$.
- A estatística $Q_{AD} = \dfrac{\left(D(\mathbf{y}; \widehat{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \widehat{\boldsymbol{\mu}})\right)/q}{D(\mathbf{y}; \widehat{\boldsymbol{\mu}})/(n-p)}$ sob H_0 e para n suficientemente grande, é tal que $Q_{AD} \approx F_{(q,n-p)}$
- Note que só podemos utilizar esta abordagem para modelos não saturados (n $>$ p).
- Assim, rejeita-se H_0 se $p - valor \leq \alpha$, em que $p - valor \approx P(X \geq q_{AD} | H_0)$, e $X \sim F_{(q,n-p)}$

$$q_{AD} = \frac{\left(D(\mathbf{y}; \widetilde{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \widetilde{\boldsymbol{\mu}})\right)/q}{D(\mathbf{y}; \widetilde{\boldsymbol{\mu}})/(n-p)}.$$

Outras funções de ligação

- Seja μ a proporção de sucessos de uma binomial.
- A ligação probito é dada por

$$\Phi^{-1}(\mu) = \boldsymbol{\eta}$$

ou, de modo equivalente, $\mu = \Phi(\eta)$, em que $\Phi(.)$ é a fda de uma distribuição normal padrão.

Outras funções de ligação

- Novamente, seja μ a proporção de sucessos de uma binomial.
- A fda de uma distribuição do valor extremo padrão (ou Gumbell padrão, a qual corresponde ao logaritmo natural de uma distribuição exponencial com seu parâmetro igual a 1) é dada por:

$$F(x) = 1 - \exp\{-\exp(x)\}$$

- Assim, o modelo binomial com ligação log-log é dado por

$$\mu = 1 - \exp\{-\exp(\eta)\}$$

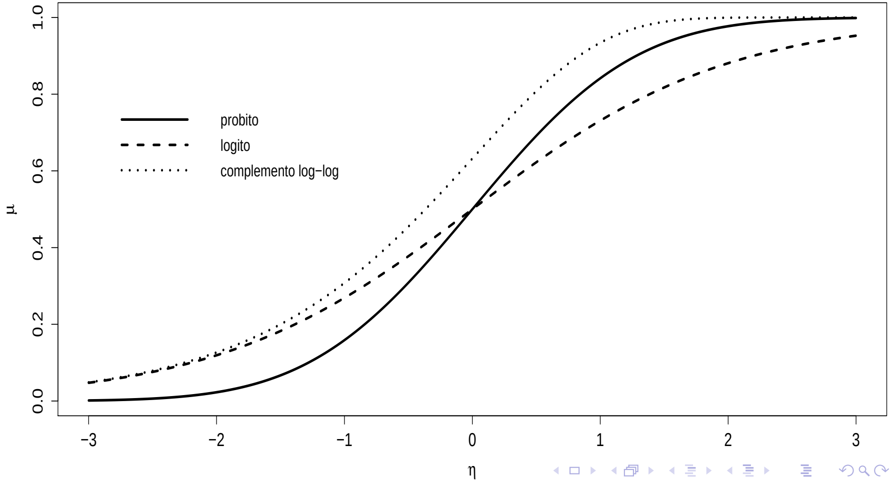
ou de modo equivalente,

$\ln(-\ln(1-\mu)) = \eta.$

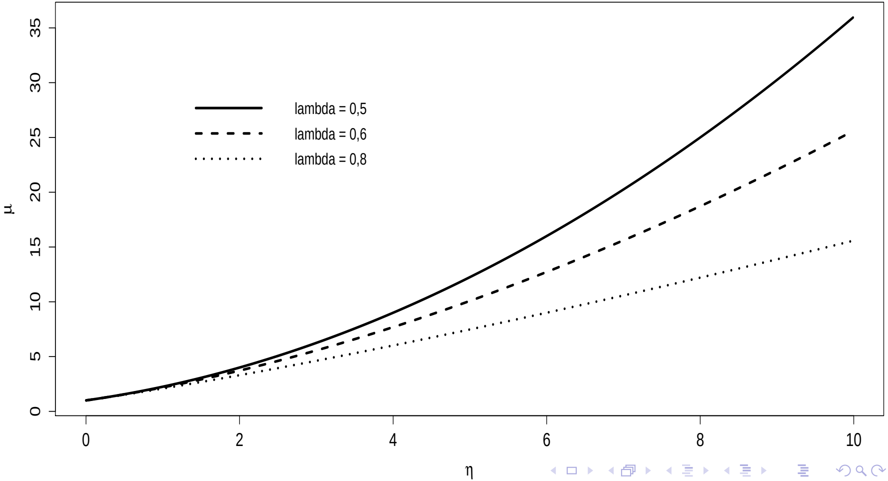
Resumo: principais funções de ligação

Distribuição	$g(\theta)$
Normal	μ,(canônica), $1/\mu$ e $\ln \mu$ (se $\mu > 0$)
Poisson	$\ln \mu$(canônica), $\sqrt{\mu}$
Binomial	$\ln(\mu/(1-\mu))$ (canônica), $\Phi^{-1}(\mu)$, $\ln(-\ln(1-\mu))$
Gama	$1/\mu$(canônica), $\ln \mu$
N.Inversa	$1/(2\mu^2)$ (canônica), $\ln \mu$, $1/\mu$

Funções de ligação para médias no intervalo (0,1)



Funções de ligação da família Box-Cox



Estimação

- Lembremos que $\theta_i = h(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ em que $\eta_i = \sum_{j=1}^{p} x_{ji}\beta_j$.
- Logo $\theta_i = h(g^{-1}(\eta_i))$. Se $g(.)$ form uma função de ligação canônica, então $\theta_i = \eta_i$.
- Verossimilhança

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

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

Estimação

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

$$
\begin{aligned}
S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\boldsymbol{\beta}} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\boldsymbol{\beta}} \right\} \\
&= \phi \sum_{i=1}^{n} \left\{ \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) \mathbf{X}_i \right\} = \phi \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})
\end{aligned}
$$

em que $\mathbf{W} = \text{diag}(\omega_1, ..., \omega_n), \omega_i = (d\mu_i/d\eta_i)^2/V_i, \; V_i = V(\mu_i)$ e $\mathbf{V} = \text{diag}(V_1, ..., V_n)'$. As outras quantidades são como definidas para os modelos de regressão de Poisson e logístico.

Estimação

- Vetor Escore para ϕ

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

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

Estimação

- Informação de Fisher (usando a regra da cadeia)

$$
\begin{aligned}
I(\boldsymbol{\beta},\boldsymbol{\beta}) &= -\mathcal{E}\left\{\phi\left\{\sum_{i=1}^{n}\frac{1}{V_i^{1/2}}(Y_i-\mu_i)\mathbf{X}_i\frac{d\omega_i^{1/2}}{d\boldsymbol{\beta}'}\right.\right.\\
&+ \left.\left.\omega_i^{1/2}(Y_i-\mu_i)\mathbf{X}_i\frac{dV_i^{-1/2}}{d\boldsymbol{\beta}'}-\sqrt{\frac{\omega_i}{V_i}}\mathbf{X}_i\frac{d\mu_i}{d\boldsymbol{\beta}'}\right\}\right\} = \phi\mathbf{X}'\mathbf{W}\mathbf{X}
\end{aligned}
$$

$$
I(\boldsymbol{\beta},\phi) = -\mathcal{E}\left(\left\{\sqrt{\frac{\omega_i}{V_i}}(Y_i-\mu_i)x_{ji}\right\}\right) = \mathbf{0}
$$

Estimação

- Informação de Fisher para ϕ

$$I(\phi,\phi) = -\mathcal{E}\left(\sum_{i=1}^{n} c''(Y_i,\phi)\right)$$

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

Inferência para o modelo

- O sistema de equações $\mathbf{S}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}, S(\widehat{\phi}) = 0$ não tem solução explícita e algum método de otimização numérica, como o algoritmo escore de Fisher, deve ser utilizado para obter-se as estimativas de MV.
- Contudo, como os parâmetros são ortogonais $I(\boldsymbol{\beta}, \phi) = -\mathcal{E}\left(\left\{\sqrt{\frac{\omega_i}{V_i}}(Y_i - \mu_i)x_{ji}\right\}\right) = \mathbf{0}$ e o produto matricial $\mathbf{I}^{-1}(\boldsymbol{\beta})\mathbf{S}(\boldsymbol{\beta})$ não depende de ϕ a estimação é feita em duas etapas.

Estimação de β: Algoritmo escore de Fisher

- Seja $\beta^{(0)}$ uma estimativa inicial de β (chute inicial), então faça

$$\beta^{(t+1)} = \beta^{(t)} + \mathbf{I}^{-1}(\beta^{(t)})\mathbf{S}(\beta^{(t)}), t = 1, 2, ... \tag{1}$$

até que algum critério de convergência seja satisfeito, como

$$|l(\beta^{(t+1)}) - l(\beta^{(t)})| < \epsilon, \epsilon > 0,$$

em que $l(.)$ é a logverossimilhança.

Estimação de ϕ

- Com as estimativas de $\boldsymbol{\beta}$, digamos $\widetilde{\boldsymbol{\beta}}$, obtidas no passo anterior, obtenha as estimativas de ϕ através de $\widehat{\phi} = \frac{n-p}{\sum_{i=1}^{n}\left(\frac{(Y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)}\right)}$ que é o estimador do método dos momentos (e consistente) de ϕ.
- O R fornece a estimativa associada ao estimador $\widehat{\phi} = \frac{n-p}{D(\mathbf{y},\widehat{\boldsymbol{\mu}})}$ que não é consistente para ϕ.
- Para os modelos Poission e Binomial, $\phi = 1$.

Análise de diagnóstico

- O procedimento para verificação do ajuste do modelo segue o mesmo padrão anteriormente apresentado para os modelos Poisson e binomial.
- As fórmulas para o resíduo componente do desvio e da variável $\mathbf{z}$ já foram apresentadas para os dois modelos acima.
- Para os outros modelos, veja o livro do Prof. Gilberto.