Modelos Lineares Generalizados - Família Exponencial

Erica Castilho Rodrigues

10 de Maio de 2017

Introdução

Família Exponencial

- No modelo de Regressão Linear temos que

$$Y_i \sim N(\mu, \sigma^2) \quad E(Y_i) = \mu_i = x_i^T \boldsymbol{\beta}$$

onde

  - Y_i são independentes;
  - x_i^T representa a i-ésima linha da matriz $\mathbf{X}$, correspondente ao i-ésimo indivíduo.

- Podemos estar interessados em situações mais genéricas.
- A variável resposta tem uma distribuição diferente da normal.
- A relação entre o valor esperado da variável resposta e as explicativas pode ter uma relação diferente de

$$E(Y_i) = \mu_i = x_i^T \boldsymbol{\beta}$$

podemos ter

$$E(Y_i) = \mu_i = g(x_i^T \beta)$$

onde $g(.)$ é uma função genérica.

- A variável Y_i não pode ter QUALQUER distribuição.
- Precisamos garantir certas propriedades para:
  - estimar os parâmetros,
  - fazer testes de hipóteses,
  - tirar conclusões sobre o modelo.
- Uma classe de distribuições garante essas propriedades.
- Essa classe é conhecida como **família exponencial**
- O que é a família exponencial?

Família

Conjunto de distribuições com características similares.

Família Exponencial

- O que a família exponencial?

Família

Conjunto de distribuições com características similares.

Família Exponencial

- O que a família exponencial?
- É uma família de distribuições cuja função densidade pode ser escrita na seguinte forma

$$f(y;\theta) = s(\mathbf{y})t(\theta)e^{a(\mathbf{y})b(\theta)}$$

onde $s(.)$, $t(.)$, $a(.)$ e $b(.)$ são funções não negativas.

- Essa expressão pode ser reescrita como

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\mathbf{y})\right]$$

onde

$$s(y) =$$

Família

Conjunto de distribuições com características similares.

Família Exponencial

- O que a família exponencial?
- É uma família de distribuições cuja função densidade pode ser escrita na seguinte forma

$$f(y;\theta) = s(\mathbf{y})t(\theta)e^{a(\mathbf{y})b(\theta)}$$

onde $s(.)$, $t(.)$, $a(.)$ e $b(.)$ são funções não negativas.
- Essa expressão pode ser reescrita como

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\mathbf{y})\right]$$

onde

$$s(y) = e^{d(y)} \quad t(\theta) = e^{c(\theta)} \, .$$

Exemplo:

- Seja Y uma variável tal que

$$Y \sim Poisson(\theta) \ .$$

- Vamos verificar que essa distribuição pertence à família exponencial

$$f(y, \theta) =$$

Exemplo:

- Seja Y uma variável tal que

$$Y \sim Poisson(\theta)\ .$$

- Vamos verificar que essa distribuição pertence à família exponencial

$$f(y,\theta) = \frac{e^{-\theta}\theta^{y}}{y!} = (y!)^{-1}e^{-\theta}\exp\{\log(\theta^{y})\}$$

$$= (y!)^{-1}e^{-\theta}\exp\{y\log(\theta)\} = \underbrace{(y!)^{-1}}_{s(y)}\underbrace{e^{-\theta}}_{t(\theta)}\exp\{\underbrace{y}_{a(y)}\underbrace{\log(\theta)}_{b(\theta)}\}\ .$$

- Notamos então que a distribuição de Poisson pertence a família exponencial.

Exemplo:

- A distribuição Normal pertence à Família Exponencial?

Exemplo:

- A distribuição Normal pertence à Família Exponencial? Sim.
- Vejamo porque isso é verdade.
- Seja

$$Y \sim N(\mu, \sigma^2)$$

$$f(y, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Exemplo:

- A distribuição Normal pertence à Família Exponencial? Sim.
- Vejamo porque isso é verdade.
- Seja

$$Y \sim N(\mu, \sigma^2)$$

$$f(y, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\}$$

$$= \exp\left\{-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)\right\}$$

Exemplo: (continuação)

$$= \exp\left\{\frac{\mu y}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)\right\}$$

$$= \exp\left\{ y \underbrace{\frac{\mu}{\sigma^2}}_{b(\mu)} \underbrace{-\frac{y^2}{2\sigma^2}}_{d(y)} \underbrace{-\frac{\mu^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)}_{c(\mu)} \right\}$$

- Portanto a distribuição normal pertence à família exponencial.

- Observe que no exemplo anterior consideramos $\theta = \mu$.
- Tratamos σ^2 como uma constante conhecida
- Ele é chamado parâmetro de ruído (nuisance parameter).
- Na prática, precisamos estimá-lo.
- Porém o nosso interesse está em μ e não em σ^2.

Parâmetro Canônico

- Veremos mais a frente que um tipo específico de parâmetro será de grande importância.
- Ele é chamado parâmetro canônico.
- Considere a função densidade escrita na forma

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right] \ .$$

- Se $a(y) = y$, $b(\theta)$ é chamado paramêtro canônico da distribuição.

Exemplo:

- Vamos considerar o exemplo da Poisson.
- Vimos que

$$f(y;\theta) = \underbrace{(y!)^{-1}}_{s(y)} \underbrace{e^{-\theta}}_{t(\theta)} \exp\{\underbrace{y}_{a(y)} \underbrace{\log(\theta)}_{b(\theta)}\} .$$

- Qual é o parâmetro canônico nesse caso?

Exemplo:

- Vamos considerar o exemplo da Poisson.
- Vimos que

$$f(y;\theta) = \underbrace{(y!)^{-1}}_{s(y)} \underbrace{e^{-\theta}}_{t(\theta)} \exp\{\underbrace{y}_{a(y)} \underbrace{\log(\theta)}_{b(\theta)}\} \, .$$

- Qual é o parâmetro canônico nesse caso?
- Como $a(y) = y$, o parâmetro canônico é dado por

$$b(\theta) = \log(\theta) \, .$$

Exemplo:

- Vamos considerar o exemplo da Normal.
- Vimos que

$$f(y,\mu) = \exp\left\{ y \underbrace{\frac{\mu}{\sigma^2}}_{b(\mu)} \underbrace{-\frac{y^2}{2\sigma^2}}_{d(y)} \underbrace{-\frac{\mu^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)}_{c(\mu)} \right\}$$

- Qual é o parâmetro canônico nesse caso?

Exemplo:

- Vamos considerar o exemplo da Normal.
- Vimos que

$$f(y,\mu) = \exp\left\{ y \underbrace{\frac{\mu}{\sigma^2}}_{b(\mu)} \underbrace{-\frac{y^2}{2\sigma^2}}_{d(y)} \underbrace{-\frac{\mu^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)}_{c(\mu)} \right\}$$

- Qual é o parâmetro canônico nesse caso?
- Como $a(y) = y$, o parâmetro canônico é dado por

$$b(\mu) = \frac{\mu}{\sigma^2}\,.$$

- Veremos agora algumas propriedades da Família Exponencial.
- Essas propriedades serão muito importantes para estimarmos os modelos.
- Elas nos fornecem maneiras diretas de calcularmos

$$E(a(y)) \quad \text{e} \quad Var(a(y))$$

em função de $a(.)$, $b(.)$, $c(.)$ e $d(.)$.

Teorema

- Se a função densidade pode ser escrita como

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right]$$

então

$$E(a(Y)) = -\frac{c'(\theta)}{b'(\theta)}$$

$$Var(a(Y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$$

- Vamos mostrar primeiro que

$$E(a(Y)) = -\frac{c'(\theta)}{b'(\theta)}\,.$$

- Vamos mostrar primeiro que

$$E(a(Y)) = -\frac{c'(\theta)}{b'(\theta)} .$$

- Temos que

$$\int f(y;\theta)dy = 1$$

derivando com relação a θ dos dois lados

$$\frac{d}{d\theta}\int f(y;\theta)dy = \frac{d}{d\theta}1 = 0 .$$

- Vamos supor que podemos inverter a ordem entre a derivada e a integral.
- Temos então que

$$\int \frac{d}{d\theta} f(y;\theta)dy = 0$$

- Temos, porém, que a função de densidade pode ser escrita como

$$f(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right] \,.$$

- Vamos derivar em relação a θ

$$\begin{aligned}
&\frac{d}{d\theta} f(y;\theta) \\
&= \left(a(y)b'(\theta) + c'(\theta)\right) \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right] \\
&= \left(a(y)b'(\theta) + c'(\theta)\right) f(y;\theta) \,.
\end{aligned}$$

- Mostramos que é verdade que

$$\int \frac{d}{d\theta} f(y;\theta)dy = 0$$

- Vamos agora integrar

$$\frac{d}{d\theta} f(y;\theta)$$

em relação a y.

- Temos que

$$\int \frac{d}{d\theta} f(y;\theta)dy = \int \left(a(y)b'(\theta) + c'(\theta)\right) f(y;\theta)dy$$

$$= \int a(y)b'(\theta)f(y;\theta) + c'(\theta)f(y;\theta)dy$$

$$= \int a(y)b'(\theta)f(y;\theta)dy + \int c'(\theta)f(y;\theta)dy$$

$$= b'(\theta) \int a(y) f(y;\theta) dy + c'(\theta) \int f(y;\theta) dy$$
$$= b'(\theta) E(a(y)) + c'(\theta) 1$$

pois $\int a(y) f(y;\theta) dy = E(a(y))$ e $\int f(y;\theta) dy = 1$.

- Temos então que

$$\int \frac{d}{d\theta} f(y;\theta) dy = b'(\theta) E(a(y)) + c'(\theta) = 0$$

- Isso implica que

$$b'(\theta) E(a(y)) = -c'(\theta) \Rightarrow E(a(y)) = -\frac{c'(\theta)}{b'(\theta)} .$$

- Vamos mostrar agora que

$$Var(a(Y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \,.$$

- Vimos que

$$\int \frac{d}{d\theta} f(y;\theta)dy = 0$$

derivando novamente em relação a θ

$$\frac{d}{d\theta} \int \frac{d}{d\theta} f(y;\theta)dy = 0$$

trocando a ordem entre a derivada e a integral

$$\int \frac{d^2}{d\theta^2} f(y;\theta)dy = 0 \,.$$

- Vejamos agora como fica
$$\frac{d^2}{d\theta^2} f(y;\theta) \,.$$
- Vimos que
$$\frac{d}{d\theta} f(y;\theta) =$$
$$\big(a(y)b'(\theta) + c'(\theta)\big) \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right]$$
- Derivando novamente em relação a θ e usando a regra do produto
$$\frac{d^2}{d\theta^2} f(y;\theta) =$$
$$\frac{d}{d\theta}\big(a(y)b'(\theta) + c'(\theta)\big) \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right]$$
$$+\big(a(y)b'(\theta) + c'(\theta)\big) \frac{d}{d\theta} \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right]$$

$$= (a(y)b''(\theta) + c''(\theta)) \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right] +$$
$$(a(y)b'(\theta)+c'(\theta))(a(y)b'(\theta)+c'(\theta)) \exp\left[a(y)b(\theta) + c(\theta) + d(\theta)\right]$$
$$= (a(y)b''(\theta) + c''(\theta))f(y;\theta)+$$
$$(a(y)b'(\theta) + c'(\theta))(a(y)b'(\theta) + c'(\theta))f(y;\theta)$$
$$= (a(y)b''(\theta) + c''(\theta))f(y;\theta)+$$
$$(a(y)b'(\theta) + c'(\theta))^2 f(y;\theta)$$

- Vamos olhar para o termo

$$[a(y)b'(\theta) + c'(\theta)]^2$$

colocando $b'(\theta)$ em evidência

$$\left[b'(\theta)\left(a(y) + \frac{c'(\theta)}{b'(\theta)}\right)\right]^2$$

$$= b'(\theta)^2 \left(a(y) + \frac{c'(\theta)}{b'(\theta)} \right)^2$$

- Vimos que

$$E(a(y)) = -\frac{c'(\theta)}{b'(\theta)} \Rightarrow \frac{c'(\theta)}{b'(\theta)} = -E(a(y))$$

- Substituindo na expressão acima

$$b'(\theta)^2 \left(a(y) + \frac{c'(\theta)}{b'(\theta)} \right)^2 = b'(\theta)^2 \left(a(y) - E(a(y))\right)^2$$

- Temos então que

$$\frac{d^2}{d\theta^2} f(y;\theta) =$$

$$= (a(y)b''(\theta) + c''(\theta))f(y;\theta) +$$

$$(a(y)b'(\theta) + c'(\theta))^2 f(y;\theta)$$

$$= (a(y)b''(\theta) + c''(\theta))f(y;\theta) +$$

$$b'(\theta)^2 \left(a(y) - E(a(y))\right)^2 f(y;\theta)$$

- Queremos usar o fato de que

$$\int \frac{d^2}{d\theta^2} f(y;\theta) dy = 0 \, .$$

então vamos integrar em y.

- Temos que

$$\int \frac{d^2}{d\theta^2} f(y;\theta)dy$$

$$= \int (a(y)b''(\theta)+c''(\theta))f(y;\theta)+b'(\theta)^2\left(a(y)-E(a(y))\right)^2 f(y;\theta)dy$$

$$= \int (a(y)b''(\theta)+c''(\theta))f(y;\theta)dy+\int b'(\theta)^2\left(a(y)-E(a(y))\right)^2 f(y;\theta)dy$$

$$= \int b''(\theta)a(y)f(y;\theta)dy + \int c''(\theta)f(y;\theta)dy$$

$$+ \int b'(\theta)^2\left(a(y)-E(a(y))\right)^2 f(y;\theta)dy$$

$$= b''(\theta)\int a(y)f(y;\theta)dy + c''(\theta)\int f(y;\theta)dy$$

$$+b'(\theta)^2\int \left(a(y) - E(a(y))\right)^2 f(y;\theta)dy$$

- Temos que

$$\int a(y)f(y;\theta) = E(a(y)) \quad \int f(y;\theta)dy = 1$$

$$\int \left(a(y) - E(a(y))\right)^2 f(y;\theta)dy = Var(a(y))$$

- Portanto

$$\int \frac{d^2}{d\theta^2} f(y;\theta)dy$$

$$= b''(\theta)\int a(y)f(y;\theta)dy + c''(\theta)\int f(y;\theta)dy$$

$$+b'(\theta)^2 \int (a(y) - E(a(y)))^2 f(y;\theta)dy$$

$$= b''(\theta)E(a(y)) + c''(\theta) + b'(\theta)^2 Var(a(y)) \,.$$

- Como

$$\int \frac{d^2}{d\theta^2} f(y;\theta)dy = 0$$

logo

$$b''(\theta)E(a(y)) + c''(\theta) + b'(\theta)^2 Var(a(y)) = 0$$

- Vamos agora isolar $Var(a(y))$.

- Temos que

$$b'(\theta)^2 Var(a(y)) = -b''(\theta)E(a(y)) - c''(\theta)$$

$$Var(a(y)) = \frac{-b''(\theta)E(a(y)) - c''(\theta)}{b'(\theta)^2}$$

mas sabemos que

$$E(a(y)) = -\frac{c'(\theta)}{b'(\theta)}$$

logo

$$Var(a(y)) = \frac{-b''(\theta)\left(-\frac{c'(\theta)}{b'(\theta)}\right) - c''(\theta)}{b'(\theta)^2}$$

- Multiplicando por $b'(\theta)$ no numerador e denominador

$$Var(a(y)) = \frac{-b''(\theta)\left(-\frac{c'(\theta)}{b'(\theta)}\right)b'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

$$= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}.$$

Exemplo:

- Vamos retomar o exemplo da Poisson.
- Vimos que

$$f(y;\theta) = \underbrace{(y!)^{-1}}_{s(y)} \underbrace{e^{-\theta}}_{t(\theta)} \exp\{\underbrace{y}_{a(y)} \underbrace{\log(\theta)}_{b(\theta)}\} .$$

que pode ser reescrita como

$$f(y;\theta) = \exp\{\underbrace{y}_{a(y)} \underbrace{\log(\theta)}_{b(\theta)} \underbrace{-\theta}_{c(\theta)} \underbrace{-\log(y!)}_{d(y)}\} .$$

Exemplo: (continuação)

- Temos que $a(y) = y$
- Podemos calcular a esperança e variância da Poisson usando os resultados anteriores.
- Temos que

$$E(y) = -\frac{c'(\theta)}{b'(\theta)}$$

mas

$$c'(\theta) = \frac{d}{d\theta}(-\theta) = -1$$

$$b'(\theta) = \frac{d}{d\theta}(\log(\theta)) = \frac{1}{\theta}$$

portanto

$$E(y) = -\frac{-1}{1/\theta} = \theta\,.$$

Exemplo: (continuação)

- Vimos ainda que

$$Var(a(y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3} \, .$$

- Mas temos que

$$b''(\theta) = \frac{d}{d\theta}(\frac{1}{\theta}) = -\frac{1}{\theta^2}$$

$$c''(\theta) = \frac{d}{d\theta}(-1) = 0$$

Exemplo: (continuação)

- Como

$$b''(\theta) = -\frac{1}{\theta^2} \quad c'(\theta) = -1 \quad c''(\theta) = 0 \quad b'(\theta) = \frac{1}{\theta}$$

temos que

$$Var(a(y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

$$= \frac{(-1/\theta^2)(-1) - 0}{(1/\theta)^3} = \frac{\theta^3}{\theta^2} = \theta \, .$$

Exemplo:

- Vamos retomar o exemplo da Normal.
- Vimos que

$$f(y,\theta) = \exp\left\{ y \underbrace{\frac{\theta}{\sigma^2}}_{b(\theta)} \underbrace{-\frac{y^2}{2\sigma^2}}_{d(y)} \underbrace{-\frac{\theta^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2)}_{c(\theta)} \right\}$$

Exemplo: (continuação)

- Temos que $a(y) = y$
- Podemos calcular a esperança e variância da Normal usando os resultados anteriores.
- Temos que

$$E(y) = -\frac{c'(\theta)}{b'(\theta)}$$

mas

$$c'(\theta) = \frac{d}{d\theta}\left(-\frac{\theta^2}{2\sigma^2} - 1/2\log(2\pi\sigma^2\right)) = -\frac{\theta}{\sigma^2}$$

$$b'(\theta) = \frac{d}{d\theta}(\frac{\theta}{\sigma^2}) = \frac{1}{\sigma^2}$$

portanto

$$E(y) = -\frac{(-\theta/\sigma^2)}{1/\sigma^2} = \mu\ .$$

Exemplo: (continuação)

- Vimos ainda que

$$Var(a(y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3} \, .$$

- Mas temos que

$$b''(\theta) = \frac{d}{d\theta}(\frac{1}{\sigma^2}) = 0$$

$$c''(\theta) = \frac{d}{d\theta}(-\frac{\theta}{\sigma^2}) = -\frac{1}{\sigma^2}$$

Exemplo: (continuação)

- Como

$$b''(\theta) = 0 \quad c'(\theta) = -\frac{\theta}{\sigma^2} \quad c''(\theta) = -\frac{1}{\sigma^2} \quad b'(\theta) = \frac{1}{\sigma^2}$$

temos que

$$Var(a(y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

$$= \frac{0 - \left(-\frac{1}{\sigma^2}\frac{1}{\sigma^2}\right)}{\left(\frac{1}{\sigma^2}\right)^3}$$

$$= \frac{(\sigma^2)^3}{(\sigma^2)^2} = \sigma^2 \,.$$