Família exponencial

INTRODUÇÃO : DEFINIÇÃO DE GLM

- Nelder & Wedderburn mostraram , que a maioria dos problemas estatísticos resolvidos por modelos de regressão podem ser generalizados pelo denominado "modelo linear generalizado".

- Esses modelos envolvem uma **variável resposta** univariada , **variáveis explicativas** e uma amostra aleatória de n observações sendo que:

- Essa **variável resposta** , componente aleatório do modelo, tem uma distribuição pertencente à **família exponencial na forma canônica**

INTRODUÇÃO : FAMÍLIA EXPONENCIAL

As distribuições pertencentes a Família exponencial já são conhecidas

Quais distribuições poderemos aplicar aqui o GLM

Família exponencial de distribuições

Discreta

Contínuas

Independente do tipo dos dados Estaremos Aplicando o **Modelo linear Generalizado**

Não é necessário trabalhar com transformações, podemos utilizar apenas o componente aleatório e aplicar o GLM.

Família exponencial uniparamétrica:

É caracterizada por uma função de probabilidade ou densidade que depende de um parâmetro desconhecido $\theta \in \Theta$, especificada na forma:

$$\mathbf{f(x;\theta)} = \mathbf{h}(\boldsymbol{x}).\mathbf{exp}[\boldsymbol{\eta}\,(\boldsymbol{\theta}).\mathbf{t}(\boldsymbol{x}) - \mathbf{b}(\,\boldsymbol{\theta}\,)]$$

em que as funções $\eta\,(\theta), \mathrm{t}(\mathrm{x})\ e\ h(x)$ assumem valores no subconjunto dos reais e não são únicos.

Temos que pelo teorema da fatoração de Neyman, a estatística T(X) é suficiente para θ. Ou seja, a partir dessa estatística podemos construir estimadores ótimos (máxima verossimilhança)

Exemplo

A distribuição de Rayleigh, usada para análise de dados contínuos positivos, tem função densidade ($x > 0$, $\theta > 0$) dada por :

$$f(x;\theta) = \frac{x}{\theta^2} \cdot exp\left(-\frac{x^2}{2\theta^2}\right)$$

Essa função depende da família exponencial?

$$\mathbf{f(x;\,\theta)} = \mathbf{h}(\boldsymbol{x}).\mathbf{exp}[\boldsymbol{\eta}\,(\boldsymbol{\theta}).\mathbf{t}(\boldsymbol{x}) - \mathbf{b}(\,\boldsymbol{\theta}\,)]$$

Observa-se que:

$$\mathrm{f}(\mathrm{x};\,\theta) = x \cdot \exp\left[-\frac{1}{2\theta^2} \cdot x^2 - 2\,log(\theta)\right]$$

Exemplo

$$f(x;\theta) = x \cdot \exp\left[-\frac{1}{2\theta^2} \cdot x^2 - 2\,log(\theta)\right]$$

$$f(x;\theta) = h(x).\,exp[\eta\,(\theta).\,t(x) - b(\,\theta\,)]$$

Portanto a distribuição de Rayleigh pertence a família exponencial.

Distribuição Conjunta

Sejam $X_1, X_2, ..., X_n$ variáveis i.i.d que seguem distribuição pertencente a família exponencial. A distribuição conjunta de $X_1, X_2, ..., X_n$ é dada por:

$$f(x_1, x_2, ..., x_n;\ \theta) = \left[\prod_{i=1}^{n} h(x_i)\right] \cdot \exp\left[\eta(\theta) \sum_{i=1}^{n} t(x_i) - n\, b\,(\theta)\right]$$

- A distribuição conjunta de $X_1, X_2, ..., X_n$ também é um modelo que pertence à família exponencial.
- A estatística suficiente de um modelo da família exponencial tem distribuição, também, pertencente à família exponencial.

Por exemplo se $X_1, X_2, ..., X_n \sim Poisson\ (\theta)$ então a estatística suficiente T segue distribuição de Poisson, isto é, T$(X_i) \sim Poisson\ (n\theta)$

Família exponencial na forma canônica

Caso especial da família exponencial:

$$f(x; \theta) = h(x).\exp[\eta\ (\theta).t(x) - b(\ \theta\ )]$$

Onde:

$$\eta\ (\theta) = \theta \text{ e } t(x) = x$$ **Função identidade**

A família exponencial na forma canônica é definida considerando que as funções $\eta(\theta)$ e t(x) são iguais à funções identidade , na forma:

$$f(x; \theta) = h(x).\exp[\theta x - b(\theta)]$$

Família exponencial na forma canônica

Na parametrização, θ é denominado de parâmetro canônico. **O logaritmo da função de verossimilhança** corresponde a uma única observação no modelo é expresso como:

$$f(x;\theta) = h(x).\exp[\theta x - b(\theta)]$$

Log

$$\ell(\theta) = \theta x - b(\theta) + \log[h(x)]$$

Família exponencial na forma canônica

A função escore é dada por:

$$\ell(\theta) = \theta \mathrm{x} - \mathrm{b}(\theta) + \log[h(x)]$$

$$U = U(\theta) = \frac{d\,\ell(\theta)}{d\theta} \Longrightarrow U = x - b'(\theta)$$

Pelas propriedades da função escore, temos que:

$E(U) = E\left[\frac{d\,\ell(\theta)}{d\theta}\right] = 0$ e $E(U^2) = -E(U') = -E\left[\frac{d^2\,\ell(\theta)}{d\,\theta^2}\right] = 0$

Família exponencial na forma canônica

Além disso,

$Var(U) = E(U^2) – [E(U)]^2 = E(U^2) = - E(U') = b''(\theta)$

Portanto,

$E(X) = b'(\theta)$ e $Var(X) = b''(\theta)$

A importância da família exponencial da forma canônica na teoria dos GLM's está no fato de podermos calcular os dois primeiros momentos em termos da derivada da função $b(\theta)$ em relação ao parâmetro canônico θ.

PUC Minas
Virtual