

3a. Prova - Estat para CC

Renato Assunção - DCC, UFMG

Maio de 2014

1. Suponha que $X_1, \dots, X_n$ forme uma amostra aleatória de v.a.'s i.i.d. com a seguinte densidade para $x \in (0, \infty)$ e $\theta > 0$:

$$f(x; \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

- Encontre o MLE de θ .
- Obtenha a estatística suficiente para θ .
- Esta distribuição é um membro da família exponencial?
- Encontre a informação de Fisher usando a segunda derivada da log-verossimilhança e usando o resultado de que $\mathbb{E}(X_i) = \theta$.
- Se $\sum_i x_i = 20$ e $n = 10$, obtenha um IC de 95% para θ .

Solução: A densidade conjunta das observações é igual a

$$f(\mathbf{x}; \theta) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_i x_i\right).$$

Pelo Teorema da fatoração, $T(\mathbf{X}) = \sum_i X_i$ é estatística suficiente para estimar θ . A distribuição encaixa-se na definição de família exponencial.

A log-verossimilhança é dada por

$$\ell(\theta) = -n \log(\theta) - \frac{1}{\theta} \sum_{i=1}^n x_i$$

e o escore por

$$\frac{\partial \ell}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i$$

Igualando a zero, encontramos o MLE: $\hat{\theta} = \sum_i X_i / n$, a média aritmética.

Derivando uma segunda vez, temos

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{n}{\theta^3} - \frac{2}{\theta^3} \sum_{i=1}^n x_i$$

e portanto encontramos a informação de Fisher igual a

$$\begin{aligned} I(\theta) &= -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) \\ &= \frac{n}{\theta^3} - \frac{2}{\theta^3} \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \frac{n}{\theta^3} - \frac{2}{\theta^3} n\theta \\ &= \frac{n}{\theta^2} \end{aligned}$$

Usando a aproximação normal para o MLE, podemos construir um IC de 95%:

$$\left(\hat{\theta} \pm 1.95 \sqrt{I^{-1}(\hat{\theta})} \right) = \left(\hat{\theta} \pm 1.95 \frac{\hat{\theta}}{\sqrt{n}} \right) = \left(\frac{20}{10} \pm 1.95 \frac{20/10}{\sqrt{10}} \right)$$

2. $Y_1, \dots, Y_n$ são v.a.'s independentes. Cada Y_i é uma variável binária, com valores 0 ou 1, e com probabilidade de sucesso p_i que varia com i . Esta probabilidade de sucesso depende de uma variável explicativa x_i (tomada como constante) da seguinte forma:

$$p_i = p_i(\beta) = \beta_0 + \beta_1 x_i.$$

Este modelo se parece com o de regressão logística, apenas não usa a função logística para representar p_i . Suponha que os dados sejam semelhantes aos abaixo:

i	x_i	y_i
1	2.3	0
2	4.3	1
3	1.7	0
4	3.2	1
$\vdots$	$\vdots$	$\vdots$
n	5.4	1

- Obtenha a equação de verossimilhança para encontrar o MLE de $\beta = (\beta_0, \beta_1)$. (um sistema com duas equações, não tente resolvê-las).
- Este sistema deve ser resolvido numericamente. Especifique o algoritmo de Newton-Raphson para obter a estimativa MLE de θ . Explique como o MLE é obtido numericamente a partir do valor inicial $\beta^{(0)} = (0.5, 0)$.
- Encontre a matriz 2×2 de informação de Fisher.

Solução: A verossimilhança é dada por

$$L(\beta) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} = \prod_{i=1}^n (\beta_0 + \beta_1 x_i)^{y_i} (1 - \beta_0 - \beta_1 x_i)^{1-y_i}.$$

A log-verossimilhança é igual a

$$\ell(\theta) = \sum_i y_i \log(\beta_0 + \beta_1 x_i) + \sum_i (1 - y_i) \log(1 - \beta_0 - \beta_1 x_i)$$

e a equação de verossimilhança é

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_0} &= \sum_i \frac{y_i}{\beta_0 + \beta_1 x_i} + \sum_i \frac{(1 - y_i)(-1)}{1 - \beta_0 - \beta_1 x_i} = 0 \\ \frac{\partial \ell}{\partial \beta_1} &= \sum_i \frac{y_i x_i}{\beta_0 + \beta_1 x_i} - \sum_i \frac{(1 - y_i)x_i}{1 - \beta_0 - \beta_1 x_i} = 0 \end{aligned}$$

Isto é equivalente a

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_0} &= \sum_i \frac{2y_i - 1}{\beta_0 + \beta_1 x_i} = 0 \\ \frac{\partial \ell}{\partial \beta_1} &= \sum_i \frac{(2y_i - 1)x_i}{\beta_0 + \beta_1 x_i} = 0 \end{aligned}$$

O k -ésimo passo iterativo do algoritmo Newton é dado por

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)} - [D^2]^{-1} \frac{\partial \ell}{\partial \boldsymbol{\beta}}$$

onde

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \ell}{\partial \beta_0}, \frac{\partial \ell}{\partial \beta_1} \right)^t$$

é o vetor gradiente e D^2 é a matriz hessiana (ou matriz de derivadas parciais de segunda ordem) da função $\ell(\boldsymbol{\theta})$, com

$$D^2 = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \beta_0^2} & \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \ell}{\partial \beta_1^2} \end{bmatrix}$$

Tomando cuidado com a multiplicação de vários sinais negativos, encontramos

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta_0^2} &= - \sum_i \frac{y_i}{(\beta_0 + \beta_1 x_i)^2} - \sum_i \frac{(1 - y_i)(-1)(-1)}{(1 - \beta_0 - \beta_1 x_i)^2} \\ &= - \sum_i \frac{y_i}{(\beta_0 + \beta_1 x_i)^2} - \sum_i \frac{(1 - y_i)}{(1 - \beta_0 - \beta_1 x_i)^2} \\ &= - \sum_i \frac{y_i + (1 - y_i)}{(1 - \beta_0 - \beta_1 x_i)^2} \\ &= - \sum_i \frac{1}{(1 - \beta_0 - \beta_1 x_i)^2} \end{aligned}$$

De forma similar, obtemos as demais derivadas parciais de segunda ordem e portanto

$$D^2 = - \sum_i \frac{1}{(\beta_0 + \beta_1 x_i)^2} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix}$$

O vetor gradiente e a matriz D^2 são avaliadas no valor corrente $\boldsymbol{\beta}^{(k-1)}$ do parâmetro.

Supondo que temos apenas os $n = 5$ dados mostrados na tabela no enunciado no problema, se o valor inicial for $\boldsymbol{\beta}^{(0)} = (0.5, 0)$, temos na primeira iteração

$$\begin{aligned} \boldsymbol{\beta}^{(1)} &= \boldsymbol{\beta}^{(0)} - [D^2]^{-1} \frac{\partial \ell}{\partial \boldsymbol{\beta}} \\ &= \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} - \left(- \frac{1}{(\beta_0 + \beta_1 2.3)^2} \begin{bmatrix} 1 & 2.3 \\ 2.3 & 2.3^2 \end{bmatrix} - \dots - \frac{1}{(\beta_0 + \beta_1 5.4)^2} \begin{bmatrix} 1 & 5.4 \\ 5.4 & 5.4^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{\partial \ell}{\partial \beta_0} \\ \frac{\partial \ell}{\partial \beta_1} \end{bmatrix} \end{aligned}$$

Fazendo o cálculo numericamente, encontramos

$$D^2 = \begin{bmatrix} -10.0 & -33.80 \\ -33.80 & -132.14 \end{bmatrix} \quad \text{e portanto,} \quad [D^2]^{-1} = \begin{bmatrix} -0.74 & 0.19 \\ 0.19 & -0.06 \end{bmatrix}$$

Precisamos também do vetor gradiente avaliado em $\boldsymbol{\beta}^{(0)} = (0.5, 0)$. Temos

$$\begin{bmatrix} \frac{\partial \ell}{\partial \beta_0} \\ \frac{\partial \ell}{\partial \beta_1} \end{bmatrix} = \begin{bmatrix} \sum_i \frac{2y_i - 1}{\beta_0 + \beta_1 x_i} \\ \sum_i \frac{(2y_i - 1)x_i}{\beta_0 + \beta_1 x_i} \end{bmatrix} = \begin{bmatrix} \frac{2*0 - 1}{\beta_0 + \beta_1 * 2.3} + \dots + \frac{2*1 - 1}{\beta_0 + \beta_1 * 5.4} \\ \frac{(2*0 - 1)*2.3}{\beta_0 + \beta_1 * 2.3} + \dots + \frac{(2*1 - 1)*5.4}{\beta_0 + \beta_1 * 5.4} \end{bmatrix} = \begin{bmatrix} 2.0 \\ 17.8 \end{bmatrix}$$

onde tomamos $\beta_0 = 0.5$ e $\beta_1 = 0$.

Assim,

$$\boldsymbol{\beta}^{(1)} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} - \begin{bmatrix} -0.74 & 0.19 \\ 0.19 & -0.06 \end{bmatrix} \begin{bmatrix} 2.0 \\ 17.8 \end{bmatrix} = \begin{bmatrix} -1.39 \\ 0.62 \end{bmatrix}$$

Usamos $\boldsymbol{\beta}^{(1)} = (-1.39, 0.62)^t$ para obter $\boldsymbol{\beta}^{(2)}$, etc.

Como D^2 não envolve as variáveis aleatórias $Y_1, \dots, Y_n$, a matriz de informação de Fisher é dada por

$$I(\boldsymbol{\beta}) = -\mathbb{E}(D^2) = -D^2.$$

Esta matriz de informação é avaliada usando a estimativa $\hat{\boldsymbol{\beta}}$ resultante do algoritmo de Newton-Raphson.