CC0288 - Inferência Estatística I

Segunda Chamada da Primeira Verificação de Aprendizagem - 19/10/2022.

Prof. Maurício

1. (Valor 1,5 pontos) Suponha que 200 observações independentes $X_1, X_2, \ldots, X_{200}$ sejam obtidas de variável aleatória X com

$$E(X) = \mu \quad e \quad V(X) = \sigma^2.$$

Sabe-se que

$$\sum_{i=1}^{200} x_i = 300 \quad e \quad \sum_{i=1}^{200} x_i^2 = 3754.$$

Usando esses valores calcule uma estimativa:

a. não tendenciosa de μ.
b. não tendenciosa de σ^2.
c. tendenciosa de σ.

Solução:

Um estimador não viciado para μ é :

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Assim nossa estimativa é dada por:

$$\bar{x} = \frac{\sum_{i=1}^{200} x_i}{200} = \frac{300}{200} = 1,5.$$

Um estimador não viciado para σ^2 é :

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2}{n-1}.$$

Assim nossa estimativa é dada por:

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}{n-1} = \frac{3754 - 200 \times 2,25}{199} = \frac{3304}{199} = 16,60302.$$

S é um estimador viciado de σ. Assim

$$s = \sqrt{s^2} = 4,07468.$$

Você também poderia ter usado:

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2}{n} = \frac{3304}{200} = 16,52.$$

Logo,

$$\widehat{\sigma} = \sqrt{16,52} = 4,06448.$$

```
>
> n=200
> SX=300
> xb=SX/n;xb
[1] 1.5
> SX2=3754
>
> num=SX2-n*xb^2;num
[1] 3304
> s2=num/(n-1);s2
[1] 16.60302
> s=sqrt(s2);s
[1] 4.07468
>
> sigma2_est=num/n;sigma2_est
[1] 16.52
>
> sigma_est=sqrt(sigma2_est);sigma_est
[1] 4.06448
>
```

2. (Valor 7,5 pontos) Seja X uma variável aleatória com a seguinte f.d.p.:

$$f(x|\theta) = (1+\theta)\,x^{\theta}, \;\; 0 < x < 1, \; \theta > -1.$$

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de X.

a. Identifique a variável aleatória especificando seu suporte, espaço paramétrico, sua média e variância.
b. Mostre que que ela pertence à família exponencial.
c. Qual a função escore e a informação de Fisher?
d. Encontre uma estatística suficiente e completa, S, para θ .
e. Seja T um estimador não viciado de $g(\theta) = \frac{\theta+1}{\theta+2}$. Ache um limite inferior para a variância de T?

f. Mostre que

$$T = -\frac{\sum_{i=1}^{n} \log(X_i)}{n}$$

é o estimador não viciado de variância mínima de $g(\theta) = \frac{1}{\theta+1}$.

g. Qual o estimador pelo método dos momentos para θ?
Uma amostra aleatória de tamanho $n = 6$ forneceu:

$$x_1 = 0,3; x_2 = 0,8; x_3 = 0,27; x_4 = 0,35; x_5 = 0,62, x_6 = 0,55.$$

Calcule a estimativa de θ por este método.

h. Qual o estimador pelo método de máxima verossimilhança para θ?
Calcule a estimativa de θ por este método usando a amostra do item **g**.

Solução:

A f.d.p. de $X \sim beta(a,b)$ é dada por:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, x^{a-1}(1-x)^{b-1}\, I_{(0,1)(x)}, a > 0, b > 0$$

Fazendo $a = \theta + 1$ e $b = 1$ temos:

$$\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} = \frac{\Gamma(\theta+2)}{\Gamma(\theta+1)\,\Gamma(1)} = \frac{\Gamma(\theta+2)}{\Gamma(\theta+1)} = \frac{(\theta+1)\Gamma(\theta+1)}{\Gamma(\theta+1)} = \theta+1.$$

e assim

$$X \sim Beta(a = \theta+1, b = 1)$$

O suporte é dado por:

$$A = (0,1).$$

O espaço paramétrico é dado por:

$$\Theta = (-1, \infty),$$

que independe de θ.
Além disso temos:

$$E(X) = \frac{a}{a+b} = \frac{\theta+1}{\theta+2}.$$

$$V(X) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\theta+1}{(\theta+2)^2(\theta+3)}.$$

Vamos mostrar agora que pertence à família exponencial:

$$\log\left(f(x|\theta)\right) = \theta\ \log(x) + \log(\theta+1) + 0.$$

Assim

$$c(\theta) = \theta\ \ ;\ T(x) = \log(x)\ ;\ d(\theta) = \log(\theta+1)\ \ ; h(x) = 0.$$

provando assim que pertence à família exponencial.
Além disso temos:

$$c'(\theta) = 1 \quad ; d'(\theta) = \frac{1}{\theta+1}$$

$$E(T(X)) = E(\log(X)) = -\frac{d'(\theta)}{c'(\theta)} = -\frac{1}{\theta+1}.$$

Lembrando que:

$$\log\left(f(X|\theta)\right) = \theta\ \log(X) + \log(\theta+1)$$

Derivando em relação a θ temos:

$$V = \frac{\partial \log f(X;\theta)}{\partial\theta} = \log(X) + \frac{1}{\theta+1}$$

é a função escore.
Seja

$$W = \frac{\partial 2 \log f(X;\theta)}{\partial\theta^2} = -\frac{1}{(\theta=1)^2}$$

$$I_F(\theta) = E(-W) = \frac{1}{(1+\theta)^2}.$$

Vamos responder ao item **d**:

$$S = \sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} \log(X_i)$$

é suficiente e completa para θ.
Seja T um estimador não viciado de $g(\theta) = \frac{\theta+1}{\theta+2}$
Derivando temos:

$$g'(\theta) = \frac{1}{(\theta+2)^2}.$$

Sabemos que o limite inferior de Cramer-Rao é dado por:

$$LICR = \frac{(g'(\theta))^2}{nI_F(\theta)} = \frac{\frac{1}{(\theta+2)^4}}{n\ \frac{1}{(1+\theta)^2}} = \frac{(1+\theta)^2}{n\ (\theta+2)^4}.$$

Vamos responder ao item **f**:

Vamos calcular $E(S)$:

$$E(S) = \sum_{i=1}^{n} E(\log(X_i)) = -n\,\frac{1}{\theta+1}$$

Logo,

$$E\left(-\frac{S}{n}\right) = \frac{1}{\theta+1}.$$

Logo

$$T = -\frac{S}{n} = \frac{-\sum_{i=1}^{n}\log(X_i)}{n},$$

é o nosso estimador ótimo procurado.

Vamos responder ao item **g**:

$$E_\theta(X) = \bar{X}.$$

$$\frac{\theta+1}{\theta+2} = \bar{X}.$$

Assim,

$$\theta + 1 = \theta\bar{X} + 2\bar{X}.$$

$$\theta - \theta\bar{X} = 2\,\bar{X} - 1.$$

$$(1-\bar{X})\,\theta = 2\,\bar{X} - 1$$

$$\widehat{\theta}_{MM} = \frac{2\,\bar{X}-1}{1-\bar{X}}.$$

Na nossa amostra temos $n = 6$ e

$$\sum_{i=1}^{n} x_i = 2,89.$$

A média amostral é dada por:

$$\bar{x} = \frac{2,89}{6} = 0,4817.$$

A estimativa de θ pelo método dos momentos:

$$\widehat{\theta}_{MM} = \frac{2\,\bar{x}-1}{1-\bar{x}} = \frac{0,9634-1}{1-0,4817} = -0,0707 > -1.$$

```
 X=c(0.3,0.8,0.27,0.35,0.62,0.55)
> n=length(X);n
[1] 6
>
> SX=sum(X);SX
[1] 2.89
> Xb=mean(X);SX/n;Xb
[1] 0.4816667
[1] 0.4816667
>
> tetaMM_est=(2*Xb-1)/(1-Xb);tetaMM_est
[1] -0.07073955
>
> tetaMM_est >-1
[1] TRUE
```

Vamos responder ao item **h**:

A função de verossimilhança é dada por:

$$L(\theta)=\prod_{i=1}^{n} f(x_i|\theta)=\prod_{i=1}^{n}(1+\theta)\ x_i^{\theta}=(1+\theta)^n\left(\prod_{i=1}^{n}\ x_i\right)^{\theta}$$

Aplicando logaritmo neperiano temos:

$$l(\theta)=n\ \log(1+\theta)+\theta\ \sum_{i=1}^{n}\ \log(x_i)$$

Derivando em relação a θ temos:

$$l^{'}(\theta)=n\ \frac{1}{1+\theta}+\ \sum_{i=1}^{n}\ \log(x_i)$$

A derivada segunda é dada por:

$$l^{''}(\theta)=-\ \frac{n}{(1+\theta)^2}<0.$$

Fazendo

$$l^{'}(\theta)=0,$$

temos:

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

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

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} -\log(X_i)} - 1,$$

é o nosso estimador de MV procurado.

A nossa estimativa de MV é dada por:

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} -\log(x_i)} - 1 = \frac{6}{4,86} - 1 = 0,234.$$

```
> X=c(0.3,0.8,0.27,0.35,0.62,0.55)
> n=length(X);n
[1] 6
> Y=-log(X);Y
[1] 1.2039728 0.2231436 1.3093333 1.0498221 0.4780358 0.5978370
> SY=sum(Y);SY
[1] 4.862145
>
>
> tetaMV_est=n/SY -1;tetaMV_est
[1] 0.2340234
>
```

3. (Valor 1 ponto) Seja $X \sim U[-\theta, \theta], \;\; \theta > 0.$

   Determine o estimador de de máxima verossimilhança para θ baseado em uma amostra aleatória $X_1, X_2, \ldots, X_n$ de X.

Solução: A f.d.p. de X é dada por:

$$f(x|\theta) = \frac{1}{2\theta}\, I_{[-\theta,\theta]}\,(x) = \frac{1}{2\theta}\, I_{[0,\theta]}\,(|x|).$$

A função de verossimilhança é dada por:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{2\theta}\, I_{[0,\theta]}\,(|x_i|).$$

$$L(\theta) = \frac{1}{2^n\theta^n}\, \prod_{i=1}^{n}\, I_{[0,\theta]}\,(|x_i|)$$

Seja

$$Y_n = max(|X_1|, |X_2|, \ldots, |X_n|).$$

Assim,

$$L(\theta) = \frac{1}{2^n \theta^n} \; I_{[y_n, \infty)} \, (\theta),$$

que é uma função decrescente de θ.

Logo

$$\widehat{\theta}_{MV} = Y_n = max(|X_1|, |X_2|, \ldots, |X_n|).$$

4. (Valor 1 ponto) Sejam $Y_1, Y_2, \ldots, Y_n$ variáveis aleatórias independentes com

$$Y_i \sim N(\beta \; x_i^2, \sigma^2), i = 1, 2, \ldots,$$

com x_i conhecidos.

Qual o estimador de mínimos quadrados de β?

Solução:

Note que

$$E(Y_i) = \beta \; x_i^2.$$

$$S(\beta) = \sum_{i=1}^{n} (Y_i - E(Y_i))^2 = \sum_{i=1}^{n} \left(Y_i - \beta x_i^2\right)^2.$$

Derivando em relação a β temos:

$$S^{'}(\beta) = -2 \; \sum_{i=1}^{n} \left(Y_i - \beta x_i^2\right) \; x_i^2 = -2 \left(\sum_{i=1}^{n} x_i^2 \; Y_i - \beta \; \sum_{i=1}^{n} x_i^4 \right)$$

A derivada segunda em relação a β é dada por:

$$S^{'}(\beta) = 2 \sum_{i=1}^{n} x_i^4 > 0.$$

Fazendo $S^{'}(\beta) = 0$ temos:

$$\widehat{\beta}_{MQ} = \frac{\displaystyle\sum_{i=1}^{n} x_i^2 \; Y_i}{\displaystyle\sum_{i=1}^{n} x_i^4}.$$