CC0288 - Inferência Estatística I

Quarta Verificação de Aprendizagem - 10/05/2023.

Prof. Maurício

1. (Valor 4 pontos) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de tamanho n da variável aleatória X, com função densidade de probabilidade dada por:

$$f(x|\theta) = \frac{x}{\theta^2}\ \exp(-\frac{x}{\theta})\ I_A(x),\ A = (0, \infty), \theta > 0.$$

a. Identifique a lei de X. Forneça sua média e variância.

b Qual o estimador pelo método dos momentos de θ? Calcule sua variância.

c. Encontre o estimador de máxima verossimilhança de θ e verifique se ele é eficiente.

d. Encontre o estimador de máxima verossimilhança de $g(\theta) = Var(X)$ e encontre sua distribuição aproximada em grandes amostras.

Solução:

Vamos responder ao item **a**:

Temos que

$$X \sim Gama\left(r = 2\ \ \lambda = \frac{1}{\theta}\right).$$

Assim

$$E(X) = \frac{2}{\lambda} = 2\theta;\ V(X) = \sigma^2 = \frac{2}{\lambda^2} = 2\theta^2.$$

Vamos responder ao item **b**:

A função de verossimilhança é dada por:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta^2}\ \exp(-\frac{x_i}{\theta})$$

$$L(\theta) = \theta^{-2n} \prod_{i=1}^{n} x_i\ \ \exp(-\frac{1}{\theta}\ \sum_{i=1}^{n} x_i)$$

Aplicando logaritmo neperiano temos:

$$l(\theta) = \log\left[L(\theta)\right] = -2n\ \log(\theta) + \sum_{i=1}^{n} \log(x_i) - \frac{1}{\theta}\ \sum_{i=1}^{n} x_i.$$

Derivando em relação à θ temos:

$$l'(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^2}\ \sum_{i=1}^{n} x_i.$$

.

A segunda derivada em relação à θ temos:

$$l^{''}(\theta) = \frac{2n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{n} x_i = \frac{2n}{\theta^2} - \frac{2n\bar{x}}{\theta^3} = 2n \frac{\theta - \bar{x}}{\theta^3}.$$

De

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

temos:

$$\frac{2n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta^2}.$$

Assim

$$\theta = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\bar{x}}{2}.$$

Como

$$\theta - \bar{x} = \frac{\bar{x}}{2} - \bar{x} = -\frac{\bar{x}}{2} < 0$$

Assim

$$T = \hat{\theta}_{MV} = \frac{\bar{X}}{2}.$$

Vamos verificar se ele é eficiente?

A variância de T ja foi obtida anteriormente.

Seja T um estimador não viciado de θ. O Limite inferior de Cramer-Rao é dado por:

$$LICR = \frac{1}{n\, I_F(\theta)}.$$

Sabemos que:
O suporte de X , $A = (0, \infty)$ não depende de θ.

$$\log(f(X|\theta)) = \log(X) - 2\log(\theta) - \frac{X}{\theta}$$

$$V = \frac{\partial \log f(X;\theta)}{\partial \theta} = -\frac{2}{\theta} + \frac{X}{\theta^2}.$$

A informação de Fisher é dada por:

$$Var(V) = I_F(\theta) = Var(-\frac{2}{\theta} + \frac{X}{\theta^2}) = \frac{1}{\theta^4} Var(X) = \frac{1}{\theta^4}\, \theta^2 = \frac{1}{\theta^2}.$$

$$LICR = \frac{1}{n\, I_F(\theta)} = \frac{\theta^2}{n} = Var(T).$$

Logo,

$$T = \frac{\bar{X}}{2}$$

é eficiente.

Vamos responder ao item **c**:

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

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

$$T = \theta_{MM} = \frac{\bar{X}}{2}.$$

A variância de T é dada por:

$$Var(T) = Var\left(\frac{\bar{X}}{2}\right) = \frac{1}{4}\, Var(\bar{X}).$$

$$Var(T) = \frac{1}{4}\, \frac{\sigma^2}{n} = \frac{1}{4}\, \frac{2\theta^2}{n} = \frac{\theta^2}{2n}.$$

Vamos responder ao item **d**:

Como

$$g(\theta) = Var(X) = 2\,\theta^2.$$

vamos utilizar a propriedade da invariância , isto é,

$$T_1 = \widehat{g(\theta)} = g\left(\hat{\theta}\right)$$

$$T_1 = 2\,\frac{\bar{X}^2}{4} = \frac{\bar{X}^2}{2}$$

é o nosso estimador de MV pedido.

Sabemos que se o tamanho da amostra é grande e as condições de regularidade estão satisfeitas temos:

$$\sqrt{n}\ \left(g\left(\hat{\theta}\right) - g(\theta)\right) \overset{a}{\sim} N\left(0, \frac{(g'(\theta))^2}{I_F(\theta)}\right).$$

Assim,

$$g'(\theta) = 4\,\theta \quad ; \; [g'(\theta)]^2 = 16\theta^2.$$

$$\frac{(g'(\theta))^2}{I_F(\theta)} = \frac{16\theta^2}{2/\theta^2} = 16\theta^4.$$

$$\sqrt{n}\, \left(g\left(\hat{\theta}\right) - g(\theta) \right) \overset{a}{\sim} N\left(0, 8\,\theta^4\right).$$

2. (Valor 2 pontos) Seja X variável aleatória com função densidade de probabilidade

$$f(x|\theta) = \frac{\theta}{x^2}, \quad \theta \leq x < \infty,$$

em que $\theta > 0$ é um parâmetro desconhecido. Seja $(X_1, ..., X_n)$ uma amostra aleatória de X.

a . Encontre o estimador de máxima verossimilhança para θ.

b. Obtenha uma estatística suficiente e completa para θ.

Solução:

Vamos resolver o item **a**:

A função de verossimilhança é dada por:

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

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

$$L(\theta) = \frac{1}{\prod_{i=1}^{n} x_i^2} \ \theta^n \ \ I_{(0,Y_1)} \ (\theta),$$

em que

$$Y_1 = min(X_1, X_2, \ldots, X_n).$$

Note que $L(\theta)$ é uma função crescente de θ . O Máximo de $L(\theta)$ é dado por:

$$\hat{\theta} = Y_1 = min(X_1, X_2, \ldots, X_n).$$

Vamos resolver o item **b**:

A distribuição conjunta da amostra é dada por:

$$F(x_1, x_2, \ldots, x_n|\theta) = \frac{1}{\prod_{i=1}^{n} x_i^2} \ \theta^n \ \ I_{(0,\infty)} \ (Y_1),$$

Fazendo

$$g(y_1, \theta) = \theta^n \ \ I_{(0,\infty)} \ (y_1) \quad e \quad h(x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^{n} x_i^2}.$$

Assim pelo critério da Fatoração temos que:

$$Y_1 = min(X_1, XX_2, \ldots, X_n)$$

é suficiente para θ.

Para mostrar que é completa vamos inicialmente achar a f.d.p. de Y_1.

A acumulada de X é dada por:

Seja $x > \theta$

$$F(x) = \int_\theta^x \frac{\theta}{t^2}\, dt = -\frac{\theta}{t}\Big|_\theta^x = 1 - \frac{\theta}{x}.$$

A sobrevivência de X, $x > \theta$, é dada por:

$$S(x) = 1 - F(x) = \frac{\theta}{x}.$$

A densidade de Y_1 é dada por:

$$g_{Y_1}(y) = n\,[1 - F(y)]^{n-1}\, f(y) = n\left[\frac{\theta}{y}\right]^{n-1} \frac{\theta}{y^2}\;\, I_{[\theta,\infty)}\,(y).$$

$$g_{Y_1}(y) = n\frac{\theta^n}{y^{n+1}}\; I_{[\theta,\infty)}\,(y).$$

Vamos mostrar que ela é completa:

$$E[h(Y_1)] = 0.\forall\theta > 0.$$

$$\int_\theta^\infty h(y)\; n\frac{\theta^n}{y^{n+1}}\; dy = 0$$

Manipulando as constantes temos:

$$\int_\theta^\infty h(y)\; \frac{1}{y^{n+1}}\; dy = 0$$

Diferenciando em relação a θ temos:

$$-h(\theta)\,\frac{1}{\theta^n} = 0.$$

Logo

$$h(\theta) = 0, \forall\theta > 0.$$

Assim h é a função nula e a Y_1 é completa.

3. (Valor 4 pontos) Considere o o modelo de regressão:

$$Y_i = \beta\,\sqrt{X_i} + u_i, i = 1, 2, n$$

com $u_i' s$ independentes

$$E(u_i) = 0 \quad e \quad Var(u_i){=}\sigma^2.$$

Responda ao que se pede:

a. Mostre que:

$$E(Y_i) = \beta\,\sqrt{X_i} \quad e \quad V(Y_i) = \sigma^2.$$

b. Mostre que o estimador de mínimos quadrados de β

$$B = \frac{\sum_{i=1}^{n} \sqrt{X_i}\,Y_i}{\sum_{i=1}^{n} X_i}.$$

c. Mostre que:

$$E(B) = \beta \quad e \quad Var(B) = \frac{\sigma^2}{\sum_{i=1}^{n} X_i}.$$

d. Considere a amostra de $n = 5$ pares de valores:

X	Y
4	5
9	7
16	11
16	5
25	9

Sendo Y a receita de um empresa comercial em certo intervalo de tempo e X é a quantidade vendida (em unidades físicas).

Calcule a estimativa b de mínimos quadrados.

Qual a equação de regressão?

$$\hat{Y}_i = b\,\sqrt{X_i}.$$

Qual a previsão da receita da empresa se vender 36 unidades?

você pode usar as seguintes informações:

```
>
> X=c(4,9,16,16,25);Y=c(5,7,11,5,9)
>
> n=length(X);n
[1] 5
>
> SX=sum(X);SX
[1] 70
>
> SY=sum(Y);SY
[1] 37
>
> S1=sum(X*Y);S1
[1] 564
> S2=sum(sqrt(X)*Y);S2
[1] 140
> S3=sum(sqrt(X));S3
[1] 18
>
```

Solução:

Vamos responder ao item **a**:

$$E(Y_i) = E\left(\beta\,\sqrt{X_i} + u_i\right) = \beta\,\sqrt{X_i} + E(u_i) = \beta\,\sqrt{X_i} + 0 = \beta\,\sqrt{X_i}.$$

$$Var(Y_i) = Var\left(\beta\,\sqrt{X_i} + u_i\right) = Var(u_i) = \sigma^2.$$

Vamos responder ao item **b**:

$$S(\beta) = \sum_{i=1}^{n} [(Y_i - E(Y_i))]^2 = \sum_{i=1}^{n} \left[(Y_i - \beta\sqrt{X_i})\right]^2$$

Elevando ao quadrado e distribuindo os somatórios temos:

$$S(\beta) = \sum_{i=1}^{n} Y_i^2 - 2\beta\,\sum_{i=1}^{n}\,\sqrt{X_i}\;\,Y_i + \beta^2\,\sum_{i=1}^{n} X_i.$$

Derivando em relação a β

temos

$$S^{'}(\beta) = -2\,\sum_{i=1}^{n}\,\sqrt{X_i}\;\,Y_i + 2\,\beta\,\sum_{i=1}^{n} X_i.$$

A derivada segunda é dada por:

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

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

$$-2 \sum_{i=1}^{n} \sqrt{X_i} \;\; Y_i + 2 \, \beta \, \sum_{i=1}^{n} X_i = 0$$

$$B = \hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n} \sqrt{X_i} \;\; Y_i}{\displaystyle\sum_{i=1}^{n} X_i}.$$

Vamos responder ao item **c**:

Vamos mostrar que $E(B) = \beta$.

Seja

$$Num = \sum_{i=1}^{n} \sqrt{X_i} \;\; Y_i$$

$$E(Num) = \sum_{i=1}^{n} \sqrt{X_i} \;\; E(Y_i) = \sum_{i=1}^{n} \sqrt{X_i} \;\; \sqrt{X_i} \; \beta$$

$$E(Num) = \beta \, \sum_{i=1}^{n} X_i$$

Logo

$$E\left(\frac{Num}{\displaystyle\sum_{i=1}^{n} X_i} \right) = E(B) = \beta.$$

Por outro lado

$$Var(B) = Var\left(\frac{Num}{\displaystyle\sum_{i=1}^{n} X_i} \right) = \frac{Var(Num)}{\left(\displaystyle\sum_{i=1}^{n} X_i\right)^2} .$$

$$Var(Num) = Var\left(\sum_{i=1}^{n} \sqrt{X_i} \;\; Y_i \right) = \sum_{i=1}^{n} X_i Var(Y_i) = \sum_{i=1}^{n} X_i \sigma^2.$$

Assim

$$Var(B) = \frac{\sigma^2 \displaystyle\sum_{i=1}^{n} X_i}{\left(\displaystyle\sum_{i=1}^{n} X_i\right)^2} = \frac{\sigma^2}{\displaystyle\sum_{i=1}^{n} X_i}.$$

Vamos responder ao item **d**:

A estimativa de mínimos quadrado b é dada por:

$$b = \frac{\sum_{i=1}^{n} \sqrt{X_i}\, Y_i}{\sum_{i=1}^{n} X_i} = \frac{S2}{S} = \frac{140}{70} = 2.$$

Assim

$$\hat{Y}_i = 2\sqrt{X_i}.$$

Quando $X = 36$

temos

$$\hat{Y} = 2 \times \sqrt{36} = 2 \times 6 = 12.$$