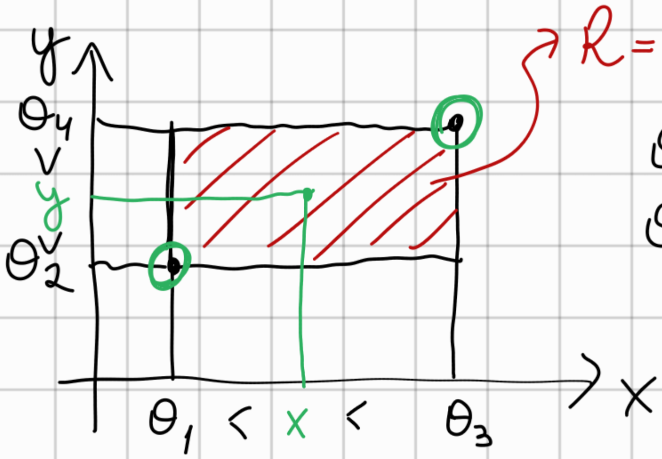
Monitoria 11/07/2022

Introdução ao Conceito de Suficiência

→ pdf constante

6.7 Let $\boxed{f(x,y}|\theta_1,\theta_2,\theta_3,\theta_4)$ be the $\boxed{\text{bivariate}}$ pdf for the $\boxed{\text{uniform distribution}}$ on the rectangle with lower left corner (θ_1,θ_2) and upper right corner (θ_3,θ_4) in $\Re^2$. The parameters satisfy $\theta_1<\theta_3$ and $\theta_2<\theta_4$. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from this pdf. Find a four-dimensional sufficient statistic for $(\theta_1,\theta_2,\theta_3,\theta_4)$.



→ R = suporte $f = \{x : f(x) > 0\}$

$\theta_1 < \theta_3$
$\theta_2 < \theta_4$

Denoto $f(x,y|\theta_1,\theta_2,\theta_3,\theta_4) = f_\theta(x,y)$

$$f_\theta(x,y) = \begin{cases} c, & (x,y)\in R \\ 0, & (x,y)\notin R \end{cases}$$

Queremos que $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\theta(x,y)\,\boxed{dx\,dy} = 1$

↳ medida de Lebesguem em $\mathbb{R}^2$

$$\| $$

$$\int_{\theta_2}^{\theta_4}\int_{\theta_1}^{\theta_3} c\,dx\,dy$$

$$\|$$

$$(\theta_4-\theta_2)(\theta_3-\theta_1)\,c = 1$$

$\Rightarrow c = (\theta_4-\theta_2)^{-1}(\theta_3-\theta_1)^{-1}.$

Portanto,

$$f_\theta(x,y) = \frac{1}{(\theta_3-\theta_1)(\theta_4-\theta_2)} \mathbb{1}_R(x,y) \quad = \begin{cases} 1, & (x,y)\in R \\ 0, & \text{c.c.} \end{cases}$$

independência

$$f_\theta(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} f_\theta(x_i,y_i) \rightarrow \text{identicamente distribuídos}$$

$$= (\theta_4-\theta_2)^{-n}(\theta_3-\theta_1)^{-n}\prod_{i=1}^{n}\mathbb{1}_R(x_i,y_i)$$

Proporcional

$$\propto \begin{cases} 1, & (x_i,y_i)\in R, \ \forall i \\ 0, & \text{c.c.} \end{cases}$$

Isso é equivalente a

$$\theta_1 \le x_i \le \theta_3 \ \text{ e } \ \theta_2 \le y_i \le \theta_4, \quad \forall i \in \{1,\dots,n\}.$$

$$\Updownarrow$$

$$\theta_1 \le \min\{x_i\}, \ \max\{x_i\} \le \theta_3, \ \theta_2 \le \min\{y_i\}, \ \max\{y_i\} \le \theta_4,$$

Com isso, podemos escrever

$$\prod_{i=1}^{n}\mathbb{1}_R(x_i,y_i) = \underbrace{\mathbb{1}\{\min\{x_i\}\ge\theta_1\}\,\mathbb{1}\{\min\{y_i\}\ge\theta_2\}\times \mathbb{1}\{\max\{x_i\}\le\theta_3\}\,\mathbb{1}\{\max\{y_i\}\le\theta_4\}}$$

Pelo Teorema da Fatorização

$$T(x,y) = (\min x_i, \min y_i, \max x_i, \max y_i)$$

é estatística suficiente, bastando fazer

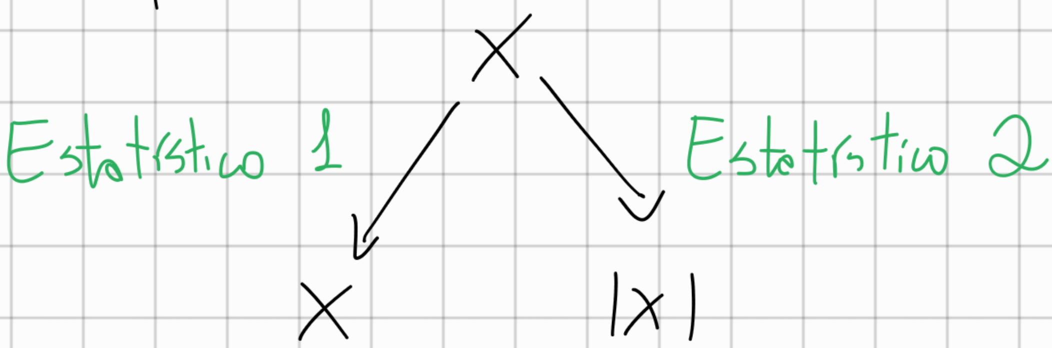
$$h(x) = 1$$

$$g(T(x)|\theta) = (\theta_4-\theta_2)^{-n}(\theta_3-\theta_1)^{-n}\times E$$

$f(x,y \mid T(x) = (t_1,\dots,t_4))$ não depende de θ.

6.1 $X \sim N(0, \sigma^2)$, $T(X) = |X|$

Suponha que um termômetro mede uma temperatura e tem erro X. Assim $T = \mu + X$. O fabricante deve informar σ^2. Será que o sinal de X é importante?



Pelo Teorema da Fatorização, como $|x|^2 = x^2$, $T(x)$ é suficiente. Calcule

$$\mathbb{P}(X = x \mid T(x) = t),$$

que é discreta, visto que $X = t$ ou $X = -t$.

O que acontece se $X \sim N(1, \sigma^2)$?

$X_1, \dots, X_n \sim P$, $\mathbb{P}(X_1 = i) = p_i$. Note que $\mathbb{P}(X_1 = 3) = 1 - p_1 - p_2$. Com isso, temos só dois valores desconhecidos.

Vamos comparar com o caso Bernoulli

$$X_1, \dots, X_n \sim \text{Bernoulli}(\theta),$$
$$\mathbb{P}(X_1 = 1) = \theta$$
$$\mathbb{P}(X_1 = 0) = 1 - \theta.$$

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

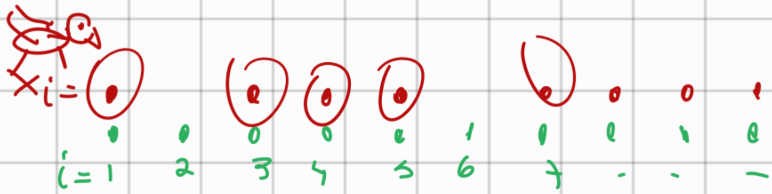
$$= \theta^{\sum x_i} (1-\theta)^{\sum(1-x_i)}$$

$$= \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}$$

$$T(x) = \sum x_i = \sum_{i=1}^{n} \mathbb{1}_1(x_i) = \begin{cases} 1, & x_i = 1 \\ 0, & x_i = 0 \end{cases}$$

contagem

$$= \#\{i : x_i = 1\}$$



Extenda essa ideia para o problema.

Corolário B.55 Schervish

Seja (S, Σ, P) espaço de probabilidade,

$X : S \to \mathbb{R}$

$g : \mathbb{R} \to \mathbb{R}$ mensurável / contínua.

$Y = g(X)$

E suponha que a distribuição de X tenha densidade $f_{\mathbf{X}}(x)$

Então

$$f_{X \mid Y}(x \mid y) = \frac{f_X(x)}{f_Y(y)} \mathbb{1}_{\{g(x)\}}(y)$$

mensurável,

Na (4), $X_1, \dots, X_n \sim \mathcal{U}nif(0, \theta)$, $T(x) = \max(x_i)$

$$f_{X \mid T}(x \mid t) = \frac{\prod_{i=1}^{n} f_{\mathbf{X}}(x_i \mid \theta)}{f_T(t \mid \theta)}$$