Procedimento: Máxima Verossimilhança

- procedimento aplicável (faz sentido) em modelos paramétricos regulares.

- Seja $p(x/\theta) \in \mathcal{P} = \{P_\theta, \theta \in \Theta\}$; $p(x/\theta)$: densidade de $X = x$ dado q' θ é o verdadeiro valor ; $\theta \in \Theta$
$\Theta \subset \mathbb{R}^d$

$L_x(\theta) \doteq p(x/\theta)$ ← verossimilhança de θ.
↳ medida de quão provável é q' θ tenha produzido a observação x.

Método: Achar $\hat{\theta}(x)$ q' maximize $L_x(\theta)$ como função de θ.

$x \in \mathcal{X}$; $L_x(\cdot) : \Theta \to \mathbb{R}$.

$\hat{\theta}(x)$ verifica. ie $\hat{\theta}(x) = \arg\max_{\theta \in \Theta} \{p(x;\theta)\}$

$$L_x(\hat{\theta}(x)) = p(x/\hat{\theta}(x)) = \max_\theta \{p(x,\theta), \theta \in \Theta\}$$

Exemplo: Suponho q' $\theta \in \Theta = \{0, 1/2\}$ e $p(x/\theta)$ é dada pela seguinte tabela:

$x \backslash \theta$	0	1/2
1	0	0.10
2	1	0.90

$x = 1 \to L_1$ é maximizado em $\hat{\theta}(1) = 1/2$

$x = 2 \to \hat{\theta}(2) = 0$

a maximização depende de x

Exemplo: Distribuição normal com variância conhecida.

$X \sim \mathcal{N}(\theta, \sigma^2)$; σ^2 conhecida ; φ: fdp da normal padrão.

parâmetro de interesse $= \theta$.

x : valor observado

$$L_x(\theta) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-\theta}{\sigma}\right)$$

então $\hat{\theta}(x) = x$ ← estimador de Máximo-Verossimilhança de θ (MLE)

Em geral, se $X_1, \dots, X_n$ é a a $N(\theta, \sigma^2)$ σ^2 conhecida

então, se $x_1, \dots, x_n$ são os elementos observados,

$$\hat{\theta}(x_1, \dots, x_n) = \bar{x} \qquad \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

da q' $$L_{(x_1, \dots, x_n)}(\theta) = \prod_{i=1}^{n} L_{x_i}(\theta) = \frac{1}{\sigma^n} \prod_{i=1}^{n} \varphi\left(\frac{\theta - x_i}{\sigma}\right)$$

$$\frac{d}{d\theta} L_{\underline{x}}(\theta) = 0 \text{ sse } \frac{d}{d\theta}\left\{\sum_{i=1}^{n} (\theta - x_i)^2\right\} = 0 \quad \text{sse } \sum_{i=1}^{n} (\theta - x_i) = 0$$

versão 7.1

sse $\hat{\theta}(x_1, \dots, x_n) = \bar{x}$. → est suf

<u>Exemplo</u> (Tamanho populacional) $X_1, \dots, X_n$ iid $\mathcal{U}\{1, 2, \dots, \theta\}$

$\underline{x} = (x_1, \dots, x_n)$

ie $L_{\underline{x}}(\theta) = \frac{1}{\theta^n} \mathbb{1}_{\{X_{(n)} \leq \theta\}}$ ie • $L_{\underline{x}}(\theta) = 0 \quad \forall \theta \in \{1, 2, \dots, X_{(n)} - 1\}$

$$L_{\underline{x}}(\theta) = \left[\frac{1}{X_{(n)}}\right]^n \text{ se } \theta = X_{(n)}$$

$$\left[\frac{1}{X_{(n)}+1}\right]^n \text{ se } \theta = X_{(n)}+1$$

$$\vdots$$

crescente

ou seja o valor $\hat{\theta}(\underline{x})$ q' maximiza $L_{\underline{x}}(\theta)$ é $\underbrace{\hat{\theta}(\underline{x}) = X_{(n)}}$

- MLE
- est suf

• $\hat{\theta}(\underline{x}) = X_{(n)}$: estimativa ruim se $\theta > X_{(n)}$

Verdadeiro.

$$\frac{1}{\sigma^n}\prod_{i=1}^{n}\varphi\left(\frac{\theta-x_i}{\sigma}\right) = \frac{1}{\sigma^n}\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(\theta-x_i)^2}{2\sigma}\right\}$$

$$= \underbrace{\frac{1}{\sigma^n}\cdot\frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{n}(x_i-\theta)^2\right\}}$$

$$\frac{d}{d\theta} = \frac{1}{\sigma^n(2\pi)^{n/2}}e^{-\frac{1}{2\sigma}\sum_{i=1}^{n}(x_i-\theta)^2}\cdot\left(-\frac{1}{2\sigma}\right)\cdot\frac{d}{d\theta}\left\{\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

$= 0$ see $\frac{d}{d\theta}\left\{\sum_{i=1}^{n}(\theta-x_i)^2\right\} = 0$

$$\sum_{i=1}^{n}2(\theta-x_i) = 0.$$

$$n\hat{\theta} = \sum_{i=1}^{n}x_i$$

$$\hat{\theta} = \bar{x}$$

Contraste Mínimo

Seja $\ell_x(\theta) \doteq \ln(L_x(\theta)) = \ln(p(x/\theta))$; $p(\cdot/\theta) \in \mathcal{P} \doteq \{P_\theta, \theta \in \Theta\}$

- MLE $\hat{\theta}(x)$ minimiza $-\ln(p(x/\theta))$. ou seja minimiza $-\ell_x(\theta)$
- ln permite associar o critério de MV com função contraste.

Se $X_1, \ldots, X_n$ são iid $p(\cdot/\theta)$ então

$$\ell_{\underline{x}}(\theta) = \ln(p(\underline{x}/\theta)) = \ln\left(\prod_{i=1}^{n} p(x_i/\theta)\right) = \sum_{i=1}^{n} \ln(p(x_i/\theta))$$

$$\ell_{\underline{x}}(\theta) = \sum_{i=1}^{n} \ell_{x_i}(\theta) \qquad \underline{x} = (x_1, \ldots, x_n)$$

contraste

definimos $\rho(x,\theta) \doteq -\ell_x(\theta)$ neste contexto $\hat{\theta}^{MLE}(x)$

q' minimiza $-\ell_x(\theta)$ oferece o menor contraste

$$D(\theta_0,\theta) = -E_{\theta_0} \ln[p(X/\theta)]$$

θ_0: verdadeiro valor de $\theta \in \Theta$

D é minimizado qdo $P_\theta = P_{\theta_0}$. equivalentemente (exercício)

$$D(\theta_0,\theta) - D(\theta_0,\theta_0) > 0.$$

- Note q' $D(\theta_0,\theta_0) = -E_{\theta_0} \ln[p(X/\theta_0)]$ recebe o nome de entropia de X medida de coos.

- mutua entropia ou divergencia de Kullback leibler entre p_0 e p_1

$$K(p_0,p_1) = -E_0\left(\ln\left(\frac{p_1}{p_0}(X)\right)\right)$$

$$= -\sum_x p_0(x) \cdot \ln \frac{p_1}{p_0}(x)$$

se X assume valores discretos.

ou $K(f_0,f_1) = -E_0\left[\ln \frac{f_1}{f_0}(x)\right]$

$$= -\int f_0(x) \cdot \ln\left(\frac{f_1}{f_0}(x)\right) dx$$

X contínua.

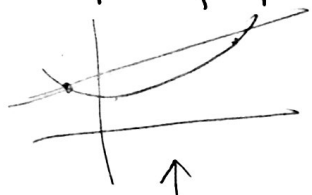
Para a definição K-L por convenção $\overbrace{\frac{0}{0}=0}^{a}$; $\overbrace{0\times\infty=0}^{b}$

. ou seja se $p_0(x)=0$ então $\xrightarrow{\overset{0}{\pi}} \overbrace{p_0(x)}^{0} \overbrace{\ln\left(\frac{p_1(x)}{p_0(x)}\right)}^{\infty} \overset{b}{=} 0$

$p_1(x)=0$ $\quad \overbrace{p_0(x)}^{0} \cdot \overbrace{\ln\left(\frac{p_1(x)}{p_0(x)}\right)}^{-\infty} \overset{b}{=} 0$ $\quad \frac{p_1(x)}{p_0(x)} \neq 0$

(*)

<u>lema</u> A entropia mútua $K(p_0,p_1)$ é sempre bem definida e $K(p_0,p_1) \geq 0$ com = sse $\{x : p_0(x) = p_1(x)\}$ possui probabilidade 1 relativa a P_0 e P_1.



dem: (caso discreto)

Desigualdade de Jensen: se $g: \mathbb{R} \to \mathbb{R}$ é convexa estrita e Z v.a. tq $E(Z)$ é definida, então $E(g(Z))$ é bem definido e $E(g(Z)) \geq g(E(Z))$ (Jensen)

Valendo a = sse $P(Z=c_0)=1$; c_0 = cte

Assumimos q' $X \sim P_0$; seja $Z \doteq \frac{p_1(X)}{p_0(X)}$ $\quad p_1$ e p_0 massas.

$$g(Z) \doteq -\ln(Z)$$

Como $g''(z) = z^{-2} > 0 \Rightarrow g$ é convexa estrita (Exercício)

Então $E(g(Z)) = -\sum_x p_0(x) \cdot \ln\left(\frac{p_1(x)}{p_0(x)}\right) \geq g(E(Z)) = -\ln\left[\sum_x p_0(x) \cdot \frac{p_1(x)}{p_0(x)}\right]$

$X \sim P_0$

assumindo (*) para controlar a boa definição, temos $E(g(Z)) \geq g(E(Z)) \geq -\underbrace{\ln(1)}_{0}$

logo $\underbrace{E(g(Z))}_{K(P_0,P_1)} = -\sum_x p_0(x) \ln\left(\frac{p_1(x)}{p_0(x)}\right) = K(P_0,P_1) \geq 0.$

$K(P_0,P_1) = 0 \iff E(g(Z)) = g(E(Z))$ sse $P(Z=c_0)=1$ sse

Nota $p_1(x)>0 \Rightarrow p_0(x)>0$

$P\left(\frac{p_1(X)}{p_0(X)} = c_0\right) = 1$ sse $P(p_1(X) = c_0 p_0(X)) = 1$ então $1 = \sum_x p_1(x) = \sum_x c_0 p_0(x) = c_0$

ie $p_0(x) = p_1(x)\ \forall x$.

Nota. suponho q' $p_1(x)>0 \not\Rightarrow p_0(x)>0$

ie $\exists x$ tq $p_1(x)>0$; $p_0(x)=0$ pela desigualdade de Jensen $\frac{p_1(x)}{p_0(x)} = c_0$ ← cte

↓

impossível!

Exemplo Ensaio multinomial $\theta_j \doteq P(X=j)$ $j=1,\dots,k$.

$X_1,\dots,X_n$ a.a de $X \sim$ Multinomial k categorias

$$N_j = \sum_{i=1}^n 1_{(X_i=j)}$$

$p(\underline{x},\theta) = \prod_{j=1}^k \theta_j^{n_j}$ n_j valor assumido por N_j em $x_1,\dots,x_n$

$$\ell_{\underline{x}}(\theta) = \ln\left(\prod_{j=1}^k \theta_j^{n_j}\right) = \sum_{j=1}^k \ln(\theta_j^{n_j}) = \sum_{j=1}^k n_j \ln(\theta_j)$$

como $\sum_{j=1}^{k} \theta_j = 1 \Rightarrow \theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$

$$\frac{\partial \ell_{\underline{x}}(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\left\{\sum_{s=1}^{k} n_s \ln(\theta_s)\right\} = \sum_{s=1}^{k} \frac{n_s}{\theta_s} \cdot \frac{\partial \theta_s}{\partial \theta_j} = 0 \qquad j = 1, \ldots, k-1$$

equação de MV

$$\frac{\partial \theta_k}{\partial \theta_j} = -1 \quad \forall\, j \in \{1, \ldots, k-1\}.$$

se $j = 1$

$$\frac{\partial \ell_{\underline{x}}(\theta)}{\partial \theta_1} = \frac{n_1}{\theta_1}\overset{1}{\left(\frac{\partial \theta_1}{\partial \theta_1}\right)} + \cancel{\frac{n_2}{\theta_2} \cdot \frac{\partial \theta_2}{\partial \theta_1}}^{\,0} + \cdots + \frac{n_k}{\theta_k}\overset{-1}{\left(\frac{\partial \theta_k}{\partial \theta_1}\right)}$$

$$= \frac{n_1}{\theta_1} + (-1) \cdot \frac{n_k}{\theta_k} = 0 \text{ sse } \frac{n_1}{\theta_1} = \frac{n_k}{\theta_k}$$

$$\frac{n_1}{\theta_1} = \frac{n_k}{1 - \sum_{j=1}^{k-1} \theta_j} \quad \text{ie} \quad \frac{\theta_1}{n_1} = \frac{1 - \sum_{j=1}^{k-1} \theta_j}{n_k} = \frac{1 - \theta_1 - \sum_{j=2}^{k-1} \theta_j}{n_k}$$

$$\frac{\theta_1}{n_1} + \frac{\theta_1}{n_k} = \frac{1 - \sum_{j=2}^{k-1} \theta_j}{n_k} \quad \text{sse} \quad \left(\frac{1}{n_1} + \frac{1}{n_k}\right)\theta_1 = \frac{1 - \sum_{j=2}^{k-1} \theta_j}{n_k}.$$

$$\text{ie } \forall\, j \in \{1, \ldots, k-1\} \quad \left(\frac{1}{n_j} + \frac{1}{n_k}\right)\theta_j = \frac{1 - \sum_{\substack{s=1 \\ s \neq j}}^{k-1} \theta_s}{n_k}.$$

sistema de equações.

$k = 3$

$$\left(\frac{1}{n_j} + \frac{1}{n_3}\right)\theta_j = \frac{1 - \sum_{\substack{s=1 \\ s \neq j}}^{2} \theta_s}{n_3} \qquad j = 1, 2. \longrightarrow \begin{cases} \left(\frac{1}{n_1} + \frac{1}{n_3}\right)\theta_1 = \frac{1 - \theta_2}{n_3} \\ \left(\frac{1}{n_2} + \frac{1}{n_3}\right)\theta_2 = \frac{1 - \theta_1}{n_3} \end{cases}$$

$$\frac{(n_1+n_3)}{n_1 \cdot n_3} \cdot \theta_1 = \frac{1-\theta_2}{n_3} \quad \& \quad \frac{n_2+n_3}{n_2 \cdot n_3} \cdot \theta_2 = \frac{1-\theta_1}{n_3}$$

$$\underbrace{\frac{n_1+n_3}{n_1} \cdot \theta_1 = 1-\theta_2}_{a} \quad \& \quad \underbrace{\frac{n_2+n_3}{n_2} \cdot \theta_2 = 1-\theta_1}_{b}$$

$$b \Rightarrow \quad \theta_2 = (1-\theta_1) \cdot \frac{n_2}{n_2+n_3} = \frac{n_2}{n_2+n_3} - \theta_1 \cdot \frac{n_2}{n_2+n_3} \quad \text{em } a$$

$$\frac{(n_1+n_3)}{n_1} \cdot \theta_1 = 1 - \frac{n_2}{n_2+n_3} + \theta_1 \cdot \frac{n_2}{n_2+n_3} \Rightarrow \left[\frac{(n_1+n_3)}{n_1} - \frac{n_2}{n_2+n_3}\right] \theta_1 = 1 - \frac{n_2}{n_2+n_3}$$

$$\text{ser} \quad \left\{\frac{(n_1+n_3)(n_2+n_3) - n_2 \cdot n_1}{n_1 \cdot (n_2+n_3)}\right\} \theta_1 = \frac{\cancel{n_2} + n_3 - \cancel{n_2}}{n_2+n_3} \quad \text{ser}$$

$$\frac{(n_1+n_3)(n_2+n_3) - n_2 \cdot n_1}{n_1} \cdot \theta_1 = n_3$$

$$\hat{\theta}_1 = \frac{n_3 \cdot n_1}{(n_1+n_3)(n_2+n_3) - n_1 \cdot n_2} = \frac{n_1 \cdot \cancel{n_3}}{\cancel{n_1 n_2} + n_1 \cancel{n_3} + n_2 \cancel{n_3} + n_3 \cdot \cancel{n_3} - \cancel{n_1 n_2}} = \frac{n_1}{n_1+n_2+n}$$

$$\hat{\theta}_1 = \frac{n_1}{n} \quad e \quad \hat{\theta}_2 = \frac{n_2}{n} \quad \checkmark$$

<u>Exercício</u> demonstre q' $(\hat{\theta}_1, \hat{\theta}_2)$ maximiza $\ell_{\underline{x}}(\theta)$

$$= \ln\left(\prod_{j=1}^{k} \theta_j^{n_j}\right)$$

quem é $\quad D(\theta_0,\theta) - D(\theta_0,\theta_0) = ?$

$$D(\theta_0,\theta) = -E_{\theta_0}\left[\ln p(X,\theta)\right]$$

$$D(\theta_0,\theta) - \overbrace{D(\theta_0,\theta_0)}^{\text{entropia do medido } P_{\theta_0}} = -E_{\theta_0}\left[\ln p(X,\theta)\right] + E_{\theta_0}\left[\ln (p(X,\theta_0)\right]$$

$$= -\left\{E_{\theta_0}\left[\ln p(X,\theta)\right] - E_{\theta_0}\left[\ln p(X,\theta_0)\right]\right\}$$

$$= -E_{\theta_0}\left\{\ln p(X,\theta) - \ln (p(X,\theta_0)\right\}$$

$$= -E_{\theta_0}\left\{\ln\left(\frac{p(X,\theta)}{p(X,\theta_0)}\right)\right\} = K(P_{\theta_0}, P_\theta).$$

= divergencia K-L entre P_{θ_0} e P_θ.

Ou seja, se achamos θ q' minimize $K(P_{\theta_0}, P_\theta)$, achamos θ q' se aproxima a θ_0 no sentido do medido de K-L.