Pasta 05

Nº cópias 72

NÃO VICIOSIDADE (Cap 2., TPE)

Definição Estimador Não Viciado

Def. Um estimador $\delta(x)$ de $g(\theta)$ é dito ser não viciado (não viesado) se

$$E_\theta[\delta(x)] = g(\theta), \text{ para todo } \theta \in \Omega.$$

Nem sempre existe estimador não-viciado.

Ex. (não existência de ENV)

$x \sim b(p, n)$ (binomial), $0 < p < 1$.

$g(p) = \frac{1}{p}$.

Para que $\delta(x)$ seja não viciado devemos ter

$$\sum_{k=0}^{n} \delta(k) \binom{n}{k} p^k (1-p)^{n-k} = g(p), \text{ para todo } 0 < p < 1.$$

Se $p \to 0$, o lado esquerdo tende a $\delta(0)$ e o direito a $+\infty$. Então, $\nexists$ $\delta(x)$ tq a igualdade acima ocorra, $\forall p$.

Se existe, dizemos que $g(\theta)$ é estimável (U-estimável)

Se existe um estimador não viciado de g, dizemos que g é estimável (ou U-estimável).

OBJETIVO: Procurar um estimador ~~com~~ não viciado com risco uniformemente mínimo (i.e, para todo $\theta \in \Omega$).

Lema: Se δ_0 é qualquer estimador não viciado de $g(\theta)$, a totalidade de estimadores não viciados é dada por

$$\delta = \delta_0 - U,$$

onde U representa qualquer estimador não viciado do "zero", ou seja, U satisfaz

$$E_\theta(U) = 0, \text{ para todo } \theta \in \Omega.$$

A ideia é procurar um estimador não viciado de $g(\theta)$ (δ_0 digamos) e buscar um estimador δ da forma acima que minimize o risco uniforme/ em θ.

Se a função de perda for quadrática,

$$R(\theta, \delta) = Var_\theta(\delta) = Var_\theta(\delta_0 - U) = E_\theta[(\delta_0 - U)^2] - g(\theta)^2.$$

Assim, o risco é minimizado, minimizando $E_\theta[(\delta_0 - U)^2]$.

Ex. X : v.a. tq
$P_p(X=-1)=p$, $P_p(X=k)=q^2p^k$, $k=0,1,\dots$,
$0<p<1$ e $q=1-p$.

a) Estimação de p

$$\delta_0 = \begin{cases} 1, & \text{se } X=-1 \\ 0, & \text{c.c.} \end{cases}$$

$E_p(\delta_0) = P_p(X=-1) = p$ (δ_0 é ENV de p).

Um estimador U é não viciado de zero sse

$U(k) = ak$, p/ $k=-1,0,1,\dots$, para algum a.
(mostraremos a seguir)

Perda quadrática: o risco (variância) é minimizado em um valor fixo p_0, minimizando

$$f(a) = \sum_{k=-1}^{+\infty} P_{p_0}(X=k)\left[\delta_0(k) - ak\right]^2$$

(buscamos a tq $f(a)$ seja mínimo).

Mas

$$f(a) = p_0(1+a)^2 + \sum_{k=0}^{+\infty} q_0^2 p_0^k a^2 k^2, \qquad q_0 = 1-p_0$$

$$f'(a) = 2p_0(1+a) + 2aq_0^2 \sum_{k=0}^{+\infty} p_0^k k^2$$

$$f''(a) = 2p_0 + 2q_0^2 \sum_{k=0}^{+\infty} p_0^k k^2 > 0$$

$$f'(a) = 0 \iff 2p_0(a+1) + 2aq_0^2 \sum_{k=0}^{+\infty} p_0^k k^2 = 0$$

$$\iff 2a\left[p_0 + q_0^2 \sum_{k=0}^{+\infty} p_0^k k^2\right] + 2p_0 = 0$$

$$\iff a = a_0 = -\frac{p_0}{p_0 + q_0^2 \sum_{k=0}^{+\infty} p_0^k k^2}.$$

Note que a_0 depende de p_0 e o estimador

$$\delta = \delta_0 - a_0 X$$

é não viciado e tem variância mínima em $p=p_0$.

Nota: U é um estimador não viciado de zero $\iff$

$$\sum_{k=-1}^{+\infty} U(k)P(X=k) = 0, \quad 0<p<1 \iff$$

$$U(-1)p + q^2 \sum_{k=0}^{+\infty} U(k)p^k = 0, \quad 0<p<1 \iff$$

$$U(-1)p + (1-2p+p^2)\sum_{k=0}^{+\infty} U(k)p^k = 0, \quad 0<p<1 \iff$$

$$U(-1)p + \sum_{k=0}^{+\infty} U(k)(p^k - 2p^{k+1} + p^{k+2}) = 0, \quad 0<p<1 \iff$$

$$\begin{aligned} U(-1)p + U(0) + U(1)p + U(2)p^2 + U(3)p^3 + U(4)p^4 + \dots \\ - 2U(0)p - 2U(1)p^2 - 2U(2)p^3 - 2U(3)p^4 + \dots \\ + U(0)p^2 + U(1)p^3 + U(2)p^4 + \dots = 0, \\ 0<p<1 \quad \iff \end{aligned}$$

$$U(0) + p\,[U(1) - 2U(0) + U(-1)] + p^2\,[U(2) - 2U(1) + U(0)] + p^3\,[U(3) - 2U(2) + U(1)] + p^4\,[U(4) - 2U(3) + U(2)] + \dots + p^k\,[U(k) - 2U(k-1) + U(k-2)] + \dots = 0, \quad 0 < p < 1 \iff$$

$$U(0) = 0$$
$$U(k) - 2U(k-1) + U(k-2) = 0, \text{ p/ } k = 1, 2, \dots \iff$$

$$U(k) = -k\,U(-1), \quad \text{para } k = 0, 1, 2, \dots \iff$$

$$U(k) = k\,a, \quad \text{p/ } k = -1, 0, 1, 2, \dots, \text{ para algum } a.$$

b) Estimação de q^2

$$\delta_1 = \begin{cases} 1, & \text{se } X = 0 \\ 0, & \text{c.c.} \end{cases}$$

$E_p(\delta_1) = P_p(X=0) = q^2$ (δ_1 é ENV de q^2).

Seja $\delta = \delta_1 - U = \delta_1 - aX$.

Procuramos a que minimiza

$$h(a) = \sum_{k=-1}^{+\infty} P_{p_0}(X=k)\,[\delta_1 - ak]^2.$$

$$h(a) = p_0\,(0 - a(-1))^2 + q_0^2\,(1 - a.0)^2 + \sum_{k=1}^{+\infty} q_0^2\, p_0^k\,(0 - ak)^2$$
$$= a^2 p_0 + q_0^2 + q_0^2\, a^2 \sum_{k=1}^{+\infty} p_0^k\, k^2.$$

$$h'(a) = 2a\,p_0 + 2a\,q_0^2 \sum_{k=1}^{+\infty} p_0^k\, k^2$$

$$h''(a) > 0.$$

$h'(a) = 0 \iff a = 0$ (não depende de p_0)

O estimador $\delta = \delta_1 - 0X = \delta_1$ tem variância uniformemente mínima entre todos os estimadores não viciados de q^2.

Def. Um estimador não viciado $\delta(X)$ de $g(\theta)$ é o estimador não viciado de variância uniformemente mínima (ENVVUM, UMVU) de $g(\theta)$ se

$$\text{var}_\theta(\delta(X)) \le \text{var}_\theta(\delta'(X)), \text{ para todo } \theta \in \Omega,$$

onde $\delta'(X)$ é qualquer outro estimador não viciado de $g(\theta)$.

O estimador é não viciado de variância localmente mínima (ENVVLM, LMVU) em $\theta = \theta_0$ se

$$\text{var}_{\theta_0}(\delta(X)) \le \text{var}_{\theta_0}(\delta'(X))$$

para qualquer outro estimador não viciado $\delta'(X)$.

OBS. O ENVVUM é único (prova usa teorema a seguir).

<u>Teorema</u>. Seja X v.a. com distribuição P_θ, $\theta \in \Omega$; seja δ um estimador em Δ, a classe dos estimadores δ com $E_\theta(\delta^2) < \infty$, para todo $\theta \in \Omega$; seja $\mathcal{U}$ o conjunto de todos os estimadores não viciados do zero que estão em Δ.

Então, uma condição necessária e suficiente para que δ seja um <u>ENVVUM</u> de sua esperança $g(\theta)$ é que

$$\underline{E_\theta(\delta U) = 0}, \quad \underline{\text{para todo } U \in \mathcal{U} \text{ e todo } \theta \in \Omega}.$$

<u>Obs:</u> Como $E_\theta(U) = 0$, a condição acima equivale a que δ seja não correlacionado com todo $U \in \mathcal{U}$.

Dem. (i) Necessidade

suponha que $\delta(X) \in \Delta$ é ENVVUM de $g(\theta) = E_\theta(\delta(X))$

Fixe $U \in \mathcal{U}$ e, para $\forall \lambda$ real arbitrário, seja $\delta' = \delta + \lambda U$. Então, δ' é um ENV de $g(\theta)$ e, para todo $\theta \in \Omega$,

(1) $$\mathrm{var}_\theta(\delta') = \mathrm{var}_\theta(\delta + \lambda U) \geq \mathrm{var}_\theta(\delta), \text{ para todo } \lambda,$$

pois δ é ENVVUM de $g(\theta)$. Mas

(2) $$\mathrm{var}_\theta(\delta + \lambda U) = \mathrm{var}_\theta(\delta) + \lambda^2 \mathrm{var}_\theta(U) + 2\lambda\, \mathrm{cov}_\theta(\delta, U).$$

De (1) e (2) e como $\mathrm{var}_\theta(\delta) < \infty$, temos

$$h(\lambda) = \lambda^2 \mathrm{var}_\theta(U) + 2\lambda\, \mathrm{cov}_\theta(\delta, U) \geq 0, \text{ p/ todo } \lambda \text{ e todo } \theta \in \Omega.$$

$h(\lambda)$ tem duas raízes: $\lambda = 0$ e $\lambda = -\dfrac{2\,\mathrm{cov}_\theta(\delta, U)}{\mathrm{var}_\theta(U)}$.

Então $h(\lambda)$ terá valores negativos a menos que $\mathrm{cov}_\theta(\delta, U) = 0$, para todo $\theta \in \Omega$.

(ii) Suficiência

Suponha que

(3) $E_\theta(\delta U) = 0$, para todo $U \in \mathcal{U}$ e todo $\theta \in \Omega$.

Seja δ' um ENV de $E_\theta(\delta(X)) = g(\theta)$.

Se $\mathrm{var}_\theta(\delta') = +\infty$, então $\mathrm{var}_\theta(\delta') \geq \mathrm{var}_\theta(\delta)$.

Se $\mathrm{var}_\theta(\delta') < +\infty$, temos $\delta' - \delta \in \mathcal{U}$ e

$$E_\theta[\delta(\delta - \delta')] = 0 \quad (\text{por } (3)).$$

Logo, $E_\theta[\delta^2] = E_\theta[\delta\delta']$. Como $E_\theta[\delta] = E_\theta[\delta']$, temos

$$\mathrm{var}_\theta[\delta] = E_\theta[\delta^2] - \{E_\theta[\delta]\}^2 = E_\theta[\delta\delta'] - E_\theta[\delta]E_\theta[\delta']$$
$$= \mathrm{cov}_\theta[\delta, \delta'].$$

Mas $\{\mathrm{cov}_\theta[\delta, \delta']\}^2 \leq \mathrm{var}_\theta[\delta]\, \mathrm{var}_\theta[\delta']$

e, então, $\mathrm{var}_\theta[\delta]^2 \leq \mathrm{var}_\theta[\delta]\, \mathrm{var}_\theta[\delta']$

e $\therefore$ $\mathrm{var}_\theta[\delta] \leq \mathrm{var}_\theta[\delta']$,

já que $\mathrm{var}_\theta[\delta] < \infty$, pois $\delta \in \Delta$.

UNICIDADE DO ENVVUM

δ : ENVVUM de $g(\theta)$, $\delta \in \Delta$.

$\delta^* = \delta - U$, outro ENVVUM de $g(\theta)$, para algum $U \in \mathcal{U}$.

Para todo $\theta \in \Omega$,

$$\mathrm{var}_\theta(\delta^*) = \mathrm{var}_\theta(\delta) + \mathrm{var}_\theta(U) - 2\,\mathrm{Cov}_\theta(\delta, U).$$

Como $\mathrm{cov}_\theta(\delta, U) = 0$ (teorema), temos

$$\mathrm{var}_\theta(\delta^*) = \mathrm{var}_\theta(\delta) + \mathrm{var}_\theta(U) > \mathrm{var}_\theta(\delta)$$

a menos que $U = 0$ q.c.-P_θ.

Portanto, $\delta^* = \delta$ q.c.-$\mathcal{P}$.

Ex. (cont). Que funções $g(p)$ têm ENVVUM?

Pelo teorema uma condição necessária e suficiente para que um estimador δ seja ENVVUM de $E_\theta[\delta]$ é que

$E_p[\delta U] = 0$, para todo p e todo $U \in \mathcal{U} = \{U(x):$

$U(k) = ak,\ k = -1, 0, 1, \dots,\ a \in \mathbb{R}\}$.

ou seja,

$E_p[\delta X] = 0$, para todo p.

Mas,

$$E_p[\delta X] = \sum_{k=-1}^{+\infty} \delta(k)\, k\, P_p(X = k) =$$

$$= -\delta(-1)\,p + q^2 \sum_{k=0}^{+\infty} \delta(k)\, k\, p^k.$$

Pode-se mostrar (exercício) que

$$-\delta(-1)\,p + q^2 \sum_{k=0}^{+\infty} \delta(k)\, k\, p^k = 0, \overset{\forall p}{\iff} k\,\delta(k) = k\,\delta(-1),$$

p/ $k = 0, 1, 2, \dots$

o qual é satisfeito desde que

$\delta(k) = \delta(-1)$, p/ $k = 1, 2, \dots$

($\delta(0)$ e $\delta(-1)$ quaisquer).

Tomando $\delta(-1) = a$ e $\delta(0) = b$, temos

$$g(p) = E_p[\delta] = ap + bq^2 + \sum_{k=1}^{+\infty} a q^2 p^k = \text{exercício} =$$

$$= a + bq^2 - aq^2 = a + cq^2, \quad c = b - a.$$

Note que a e c são arbitrários.

$g(p)$ tem ENVVUM sse $g(p) = a + c(1-p)^2$, para algum a e algum c.

Lema: Seja X uma v.a. com distribuição na família $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ e seja T uma estatística suficiente completa para $\mathcal{P}$. Então, toda função estimável $g(\theta)$ tem um e somente um estimador não viciado que é função de T.

(Aqui unicidade significa que quaisquer funções de T, estimadores não viciados de $g(\theta)$, são iguais $q.c.\ \mathcal{P}$).

Dem.: Seja δ um estimador não viciado de $g(\theta)$.
Então,
$E(\delta \mid T)$ é um estimador não viciado de $g(\theta)$
(suficiência de T garante que $E(\delta|T)$ não depende de θ), e é fç de T.

Se $\delta_1(T)$ e $\delta_2(T)$ são dois est. não viciados de $g(\theta)$, então:

$$f(T) = \delta_1(T) - \delta_2(T)$$

satisfaz

$$E_\theta[f(T)] = 0, \quad \text{para todo } \theta \in \Omega.$$

Como T é completa, $\delta_1(T) = \delta_2(T)$ $q.c.\ \mathcal{P}$.

Teorema.
Seja X uma v.a. distribuída de acordo com uma distribuição em $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ e suponha que T é uma estatística suficiente completa para $\mathcal{P}$.

(a) Para toda função estimável $g(\theta)$, existe um estimador não viciado que minimiza o risco uniformemente para qualquer função de perda $L(\theta, d)$, convexa no segundo argumento. Este estimador é, em particular, ENVVUM.

(b) O ENVVUM em (a) é o único estimador não viciado que é função de T; ele é o único estimador não viciado com risco mínimo desde que seu risco seja finito e L seja estritamente convexa em d.

Dem. Lista 2.

Corolário. Se $\mathcal{P}$ é uma família exponencial de posto completo, s-dimensional, então as conclusões do teorema valem com $\theta=(\theta_1,\dots,\theta_s)$ e $T=(T_1,\dots,T_s)$.

Dem. T é uma estatística suficiente completa p/ $\mathcal{P}$.

Ex1. $X_1,\dots,X_n$ iid
$X_i \sim N(\xi,\sigma^2)$;

$(\bar{X}, S^2)$: Estatística suficiente completa
$\bar{X}=\frac{\sum X_i}{n}$, $\quad S^2=\sum(X_i-\bar{X})^2$

$E_\theta(\bar{X})=\xi \quad ; \quad E_\theta(S^2/(n-1))=\sigma^2.$

$\bar{X}$ é ENVVUM de ξ
$\frac{S^2}{n-1}$ é ENVVUM de σ^2.

* (são também estimadores não viciados de risco uniformemente mínimo para qualquer função de perda convexa).

Métodos para obtenção de ENVVUM

Método 1. Resolvendo equações em δ

Se T é estatística suficiente completa, o ENVVUM de $\forall$ função estimável $g(\theta)$ é unicamente determinado pela equação

$$E_\theta[\delta(T)]=g(\theta) \quad \text{para todo } \theta\in\Omega.$$

Ex 2. Binomial ; $T\sim b(p,n)$; T é est. suf. completa
$g(p)=pq$; $q=1-p$.

$E_p[\delta(T)]=pq$, para todo $0<p<1 \iff$

(*) $\displaystyle\sum_{t=0}^{n}\binom{n}{t}\delta(t)\,p^t q^{n-t}=pq$, p/ todo $0<p<1 \iff$

Se $\rho=\frac{p}{q}$, de tal forma que $p=\frac{\rho}{1+\rho}$ e $q=\frac{1}{1+\rho}$,
(*) pode ser escrita como

$$\sum_{t=0}^{n}\binom{n}{t}\delta(t)\rho^t=\rho(1+\rho)^{n-2}, \text{ p/ todo } 0<\rho<\infty$$

Mas $\displaystyle\rho(1+\rho)^{n-2}=\rho\sum_{k=0}^{n-2}\binom{n-2}{k}\rho^k\,1^{n-2-k}=\sum_{k=0}^{n-2}\binom{n-2}{k}\rho^{k+1}$
$\displaystyle\underset{(t=k+1)}{=}\sum_{t=1}^{n-1}\binom{n-2}{t-1}\rho^t.$

Então, (*) equivale a

(pula p. 75)

$$\sum_{t=0}^{n} \binom{n}{t} \delta(t) \rho^t = \sum_{t=1}^{n-1} \binom{n-2}{t-1} \rho^t, \text{ p/ todo } 0<\rho<\infty$$

Comparando os coeficientes de ρ^t, para $t=0,1,\dots,n$ vem

$$\delta(t) = \frac{t(n-t)}{n(n-1)}, \quad t=0,1,\dots,n.$$

Outra solução: $\frac{T}{n}$ é ENV de p; $1-\frac{T}{n}$ é ENV de q

Considere o estimador $\frac{T}{n}\left(1-\frac{T}{n}\right) = \frac{T(n-T)}{n^2}$ de pq.

$$E_p\left[\frac{T(n-T)}{n^2}\right] = \frac{1}{n^2}\left\{n E_p(T) - E_p(T^2)\right\}$$
$$= \frac{1}{n^2}\left\{n^2 p - var_p(T) - [E_p(T)]^2\right\}$$
$$= \frac{1}{n^2}\left\{n^2 p - np(1-p) - n^2p^2\right\} = \frac{p(1-p)(n-1)}{n}$$

Então

$$\frac{n}{n-1}\,\frac{T(n-T)}{n^2} = \frac{T(n-T)}{n(n-1)} \text{ é ENVVUM de } pq.$$

Ex 3. $T \sim$ Poisson (θ), $\theta>0$

$g(\theta) = e^{-3\theta}$

T: est. suficiente completa

$$E_\theta[\delta(T)] = \sum_{k=0}^{+\infty} \delta(k) \frac{e^{-\theta}\theta^k}{k!} = e^{-3\theta}, \text{ p/ todo } \theta>0$$

$$\Leftrightarrow \sum_{k=0}^{+\infty} \delta(k) \frac{\theta^k}{k!} = e^{-2\theta}, \text{ para todo } \theta>0$$

$$\Leftrightarrow \sum_{k=0}^{+\infty} \delta(k) \frac{\theta^k}{k!} = \sum_{k=0}^{+\infty} \frac{(-2\theta)^k}{k!}, \quad " \quad "$$

$$\Leftrightarrow \delta(k) = (-2)^k, \text{ para } k=0,1,2,\dots$$

$\delta(T) = (-2)^T$ é ENVVUM de $e^{-3\theta}$.

Não é um bom estimador pois assume valores negativos enquanto que $e^{-3\theta}>0$, $\forall\theta>0$.

Aula 07, 31/08

Método 2. Condicionamento

* Toma-se um estimador não viciado $\delta(X)$ para $g(\theta)$ e o ENVVUM é obtido como a esperança condicional de $\delta(X)$ dado T, uma estatística suficiente completa.

Ex 4. Uniforme

$X_1, \dots, X_n$: iid, $X_i \sim U(0,\theta)$

$g(\theta) = E_\theta(X_1) = \theta/2.$

$T = X_{(n)}$ é est. suficiente completa.

$\delta(X) = X_1$ é ENV de $g(\theta)$.

* $E_\theta[X_1 \mid X_{(n)}]$ é ENVVUM de $g(\theta)$.

Se $X_{(n)} = t$, então $X_1 = t$ com prob. $1/n$ e é distribuído uniformemente em $(0,t)$ com prob. $\frac{n-1}{n}$.

$$E_\theta\left[X_1 \mid X_{(n)} = t\right] = \frac{1}{n}\,t + \frac{n-1}{n}\,\frac{t}{2} = \frac{n+1}{n}\,\frac{t}{2}.$$

$\frac{n+1}{2n}T$ é ENVVUM de $\frac{\theta}{2}$.

→ Outra solução: calcular $E_\theta[T]$; mostrar que $E_\theta[T] = \frac{n}{n+1}\theta$ e corrigir o vício de T.

* T: est. suf. completa

$E_\theta[\delta(X)] = g(\theta),\ \forall \theta \in \Omega$; $E[\delta(X)/T]$: est. de $g(\theta)$; $E_\theta[E(\delta(X)/T)] = g(\theta)$,

Ex 5. $X_1, \dots, X_n$ iid; $X_i: N(\xi, \sigma^2)$, σ^2 conhecido

$\bar{X}$: estatística suficiente completa (fam. exponencial)

$E_\xi(\bar{X}) = \xi$

$\bar{X}$ é ENVVUM de ξ

Considere $g(\xi)$ uma fç estimável de ξ. Existe um único ENVVUM de $g(\xi)$ que é função apenas de $\bar{X}$.

Ex. Se $g(\xi)$ é um polinômio de grau r em ξ, o ENVVUM $\delta(\bar{X})$ de $g(\xi)$ também é um polinômio de grau r em $\bar{X}$ (Lista 2).

* Obs: a prova é feita por indução em r $(r = 2,3,\dots)$

Ex 6. $X_1, \dots, X_n$ iid; $X_i \sim N(\xi, \sigma^2)$, ξ conhecido.

Seja $S^2 = \sum_{i=1}^{n}(X_i - \xi)^2$.

S^2 é estatística suficiente completa (mostrar).

$\frac{S^2}{\sigma^2} \sim \chi^2_n,\ n \geq 1$ (esta distr. não depende de σ^2)

Seja $K_{n,r} = \left\{E\left(\frac{S^r}{\sigma^r}\right)\right\}^{-1}$ (não depende de σ^2)

Então $K_{n,r}S^r$ é ENVVUM de σ^r.

Mas $K_{n,r}^{-1} = E\left(\frac{S^r}{\sigma^r}\right) = E\left[\left(\chi^2_n\right)^{r/2}\right] = \frac{\Gamma[(n+r)/2]}{\Gamma(n/2)}\,2^{r/2}$,

↓ result. sobre χ^2_n.

p/ $r > -n$

Nota. Ver Ex 5.14, p. 30-31 de TPE, que dá a f.g.m. de uma v.a. Gama (α, b) com densidade

$$\frac{1}{\Gamma(\alpha)\, b^{\alpha}}\, x^{\alpha-1} e^{-x/b}, \quad x>0;\ \alpha>0,\ b>0.$$

A densidade χ^2_f (f inteiro) se iguala à da Gama (α, b) com $\alpha = f/2$ e $b=2$.

No Ex 5.14, a fórmula de $K_{n,r}$ é obtida para $r>0$, mas vale p/ $n>-r$.

Ex 7. $X_1, \dots, X_n$ iid; $X_i \sim N(\xi, \sigma^2)$, ξ e σ^2 desconhecidos.

$(\bar X, S^2)$: est. suficiente completa (fam. exp.)
$S^2 = \sum (X_i - \bar X)^2$

$\bar X$: ENVVUM de ξ (ver Ex 1).

$\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$, $n \geq 2$

$K_{n-1,r}\, S^r$: ENVVUM de σ^r, p/ $n > -r+1$.

Em particular, $S^2/(n-1)$ é ENVVUM de σ^2 (ver Ex 1).

Considere $g(\xi, \sigma) = \frac{\xi}{\sigma}$ (p/ medir ξ em unidades de desvio padrão).

Como $\bar X$ é ENV de ξ e $K_{n-1,-1}/S$ é ENV de $1/\sigma$, pela independência de $\bar X$ e S^2 vem
$K_{n-1,-1}\, \frac{\bar X}{S}$ é ENV de $\frac{\xi}{\sigma}$, $n-1>1$ ($n>2$)
É tb ENVVUM pela suf e completividade de $(\bar X, S^2)$.

Teorem Basu: T: suf. completa, V: ancilar ⇒ T e V são independentes

Ex 8. Igual ao Ex 7; Estimação de quantis de $X_1 \sim N(\xi, \sigma)$

$$p = P(X_1 \leq u) = \Phi\left(\frac{u-\xi}{\sigma}\right)$$

$u = g(\xi, \sigma) = \xi + \sigma\, \Phi^{-1}(p)$ (estimando)

$\bar X + K_{n-1,1}\, S\, \Phi^{-1}(p)$: ENVVUM de u; $n \geq 2$.

$$K_{n-1,1} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{2^{1/2}\, \Gamma\left(\frac{n}{2}\right)}.$$

Ex 9. $X_1, \dots, X_n$, iid; $X_i \sim N(\xi, \sigma^2)$, σ conhecido, $\sigma = 1$.
Estimação de p.

$$\delta(X) = \begin{cases} 1, & \text{se } X_1 \leq u \\ 0, & \text{c.c.} \end{cases}$$ é ENV de p.

Como $\bar X$ é est. suficiente completa, o ENVVUM de p é
$$E[\delta \mid \bar X] = P[X_1 \leq u \mid \bar X].$$

Mas $X_1 - \bar X$ e $\bar X$ são independentes (Basu) e $X_1 - \bar X \sim N\left(0, \frac{n-1}{n}\right)$.

Então,
$$P[X_1 \leq u \mid \bar X = \bar x] = P[X_1 - \bar X \leq u - \bar x \mid \bar X = \bar x] =$$
$$= P[X_1 - \bar X \leq u - \bar x] = \Phi\left[\sqrt{\frac{n}{n-1}}\,(u - \bar x)\right]$$: ENVVUM de p.

$X_1, \dots, X_n$ iid., $X_i \sim N(\xi, \sigma^2)$, $\xi \in \mathbb{R}$, $\sigma^2 > 0$. $\to \sigma^2 = \sigma_0^2$ fixo
F.E. de dimensão 1 posto completo, com suf. completa $\bar X$. Como σ_0^2 é arbitrário, vale.
$\bar X$ suf. compl. S^2 ancilar (dist. ñ dep ξ) $\overset{BASU}{\Rightarrow}$ $\bar X$ e S^2 ind. $\forall \xi \in \mathbb{R}$ e $\sigma^2 = \sigma_0^2$.

Ex 10. Igual ao Ex 9; ξ e σ^2 desconhecidos.

$(\bar{X}, S^2)$: est. suficiente completa
$S^2 = \Sigma (X_i - \bar{X})^2$

ENVVUM de p: $E[\delta(X) \mid \bar{X}, S] = P[X_1 \leq u \mid \bar{X}, S]$

Seja $Z = \dfrac{X_1 - \bar{X}}{S}$. A densidade de Z é

$$f(z) = \begin{cases} \dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \sqrt{\dfrac{n}{n-1}} \left(1 - \dfrac{n z^2}{n-1}\right)^{\frac{n}{2}-2}, & \text{se } 0 < |z| < \sqrt{\dfrac{n-1}{n}} \\ 0, & \text{c.c.} \end{cases}$$

$\dfrac{X_1 - \bar{X}}{S}$ é ancilar e, pelo Teorema de Basu, é independente de $(\bar{X}, S)$.

$$P[X_1 \leq u \mid \bar{X} = \xi, S = s] = P\left[\frac{X_1 - \bar{X}}{S} \leq \frac{u - \bar{x}}{s}\right]$$

Então, o ENVVUM de p é

$$\int_{-\infty}^{\frac{u-\bar{x}}{s}} f(z)\, dz.$$ avaliado em $(\bar{x}, s)$

Ex 11. $X_1, \dots, X_m$, $Y_1, \dots, Y_n$: v.a's ind
$X_i \sim N(\xi, \sigma^2)$
$Y_i \sim N(\eta, \tau^2)$

ξ, η, σ, τ desconhecidos (fam. exponencial de dimensão 4, posto completo)

$\bar{X}, \bar{Y}, S_x^2 = \Sigma (X_i - \bar{X})^2, S_y^2 = \Sigma (Y_i - \bar{Y})^2$: est. suf. completa.

ENVVUM de
ξ : $\bar{X}$
σ^r : $K_{m-1,r} S_x^r$, $m > -r$
η : $\bar{Y}$
τ^r : $K_{n-1,r} S_y^r$, $n > -r$
$\eta - \xi$: $\bar{Y} - \bar{X}$
$\dfrac{\tau^r}{\sigma^r}$: ENVVUM de τ^r × ENVVUM de σ^{-r}
independentes

Ex 12. Igual ao Ex 11, $\sigma = \tau$.

$\bar{X}, \bar{Y}, S^2 = \Sigma (X_i - \bar{X})^2 + \Sigma (Y_j - \bar{Y})^2$: est. suf. completa.
χ_{m+n-2}

ENVVUM de
ξ : $\bar{X}$
η : $\bar{Y}$
σ^r : encontrar (Ex 2.11, p 132, TPE) $K_{m+n-2,r} S^r$
$\eta - \xi$: $\bar{Y} - \bar{X}$
$\dfrac{\eta - \xi}{\sigma}$: encontrar (Ex 2.11, p 132).

Aqui, usar T. Basu

Ex 13. Igual ao Ex 11, com $\eta = \xi$.

Note que

$(\bar{X}, \bar{Y}, S_x^2, S_y^2)$: é est. suficiente minimal mas não é completa pois $E(\bar{Y} - \bar{X}) = 0$.

Suponha que $\frac{\sigma^2}{\tau^2} = \gamma$ é conhecido. Então

$\rightarrow \delta_\gamma = \alpha \bar{X} + (1-\alpha)\bar{Y}$ é um ENV de ξ. $\leftarrow$

$$g(\alpha) = Var(\delta_\gamma) = \alpha^2 Var(\bar{X}) + (1-\alpha)^2 Var(\bar{Y})$$

$$= \alpha^2 \frac{\gamma \tau^2}{m} + (1-\alpha)^2 \frac{\tau^2}{m}.$$

$$g'(\alpha) = 2\alpha \frac{\gamma \tau^2}{m} - 2(1-\alpha)\frac{\tau^2}{n} = 0 \iff$$

$$\alpha\left[\frac{\gamma \tau^2}{m} + \frac{\tau^2}{n}\right] - \frac{\tau^2}{n} = 0 \iff$$

$$\alpha \tau^2 \left[\frac{\gamma}{m} + \frac{1}{n}\right] - \frac{\tau^2}{n} = 0 \iff$$

$$\alpha = \frac{1}{n}\left(\frac{\gamma}{m} + \frac{1}{n}\right)^{-1} = \frac{\tau^2/n}{\sigma^2/m + \tau^2/n}$$

$g''(\alpha) > 0$

$var(\delta_\gamma)$ é minimizado tomando α como acima.

Este estimador é ENVVUM?

Ex. Note que a densidade conjunta de $\underbrace{(X_1, \dots, X_m)}_{X}, \underbrace{(Y_1, \dots, Y_n)}_{Y}$ é da forma

$$\exp\{\eta_1(\theta) T_1(x,y) + \eta_2(\theta) T_2(x,y) - B(\theta)\} h(x,y)$$

onde $\theta = (\eta, \tau^2)$, $x = (x_1, \dots, x_m)$, $y = (y_1, \dots, y_n)$,

$$\eta_1(\theta) = -\frac{1}{2\tau^2\gamma}, \quad \eta_2(\theta) = \frac{\xi}{\gamma \tau^2}$$

$$T_1(x,y) = \sum x_i^2 + \gamma \sum y_i^2, \quad T_2(x,y) = \sum x_i + \gamma \sum y_i$$

$B(\theta) =$ completar, $h(x,y) = 1$.

(família exponencial de dimensão 2, posto completo)

Note que o espaço paramétrico p/ (η_1, η_2) é $\{(\eta_1, \eta_2) \in \mathbb{R}^2 : \eta_1 < 0 \text{ e } \eta_2 \in \mathbb{R}\}$ e contém retângulos bi-dimensionais

Note ainda que

$$\delta_\gamma = \alpha \bar{X} + (1-\alpha)\bar{Y}$$

com α dado acima pode ser escrito como

$$\delta_\gamma = \frac{1}{nm}\left[\frac{\gamma}{m} + \frac{1}{n}\right]^{-1}\left[\sum X_i + \gamma \sum Y_i\right]$$

e é, $\therefore$, função da est. suficiente completa (T_1, T_2).

$\therefore$ É ENVVUM de ξ.

Quando $\sigma^2/\tilde{\sigma}^2$ é desconhecido, $\nexists$ ENVVUM de ξ (problema 2.13, p.132, TPE).

Pode-se estimar α (por $\hat{\alpha}$) e usar o estimador

$$\hat{\xi} = \hat{\alpha}\bar{X} + (1-\hat{\alpha})\bar{Y}.$$

Se $\hat{\alpha}$ é função apenas de S_x^2 e S_y^2 temos

$$E(\hat{\xi}) = E(\hat{\alpha})E(\bar{X}) + (1 - E(\hat{\alpha}))E(\bar{Y}) = \xi,$$

pois $\hat{\alpha}$ e $(\bar{X}, \bar{Y})$ são independentes.

Por exemplo, $\tilde{\sigma}^2$ e σ^2 em α podem ser substituídos por $\frac{S_x^2}{m-1}$ e $\frac{S_y^2}{n-1}$ respectivamente.

Alunos: ver ex 24, p. 96, TPE (Normal multivariada)

Ex 14. • Exponencial

• $X_1, \dots, X_n$ iid, $X_i \sim E(a,b)$ com densidade

• $\frac{1}{b}\exp\left\{-\frac{(x-a)}{b}\right\}$, $x > a$; b conhecido

$X_{(1)}$: estatística suficiente completa (ex 1.6.24, TPE, p. 43)

ta: $X_{(1)}$ e $\Sigma(X_i - X_{(1)})$ são independentes

$$\frac{n[X_{(1)} - a]}{b} \sim E(0,1)$$

$$\frac{2\,\Sigma[X_i - X_{(1)}]}{b} \sim \chi^2_{2(n-1)} \qquad \text{(Lista 1)}$$

$E[X_{(1)}] = \frac{b}{n} + a$ e, $\therefore$, $X_{(1)} - \frac{b}{n}$ é ENVVUM de a.

Ex 15. • Igual a 14; a conhecido •

$\Sigma(X_i - a)$ é est. suficiente completa para b.

$$2\,\Sigma\,\frac{X_i - a}{b} \sim \chi^2_{2n}$$

$\Sigma\,\frac{X_i - a}{n}$ é o ENVVUM de b.

Ex 16. • Igual a 14; a, b desconhecidos •

$(X_{(1)}, \Sigma(X_i - X_{(1)}))$ é estatística suficiente completa (Exemplo 1.6.24, p 43, TPE)

$\hat{b} = \frac{\Sigma[X_i - X_{(1)}]}{n-1}$ é o ENVVUM de b; $X_{(1)} - \frac{\hat{b}}{n}$ é o ENVVUM de a.

$X \sim P_\theta \in \mathcal{P} = \{P_\theta, \theta \in \Omega\}$

Vimos que uma condição necessária e suficiente para que um estimador $\delta(X)$ de variância finita e $E_\theta(\delta(X)) = g(\theta)$ seja ENVVUM de $g(\theta)$ é que

$E_\theta(\delta(X)\, U(X)) = 0$ para todo $U \in \mathcal{U}$ e todo $\theta \in \Omega$ em que

$$\mathcal{U} = \{U(X) : E_\theta(U(X)) = 0 \text{ e } Var_\theta(U(X)) < \infty, \text{ para todo } \theta \in \Omega\}.$$

[Fato. (Shao, p. 166, Teo 3.2)]

Suponha que $T = T(X) = h(S)$, em que $S = S(X)$ é uma estatística suficiente para $\mathcal{P}$.

Seja $\mathcal{U}_S$ o subconjunto de $\mathcal{U}$ consistindo de funções de S.

Então, uma condição necessária e suficiente para que T seja ENVVUM de sua esperança é que

$E_\theta(T(X)\, U(X)) = 0$, para todo $U \in \mathcal{U}_S$ e todo $\theta \in \Omega$.

Dem. Basta mostrar que

$E_\theta(TU) = 0$, para todo $U \in \mathcal{U}_S$ e $\theta \in \Omega$

$\Rightarrow E_\theta(TU) = 0$, para todo $U \in \mathcal{U}$ e $\theta \in \Omega$.

Para todo $U \in \mathcal{U}$ e todo $\theta \in \Omega$:

$$\begin{aligned} E_\theta[TU] &= E_\theta[E_\theta(TU \mid S)] \\ &= E_\theta[E_\theta(h(S)U \mid S)] \\ &= E_\theta[h(S)\, E_\theta(U \mid S)] \\ &= E_\theta[h(S)\, U_S] \\ &= E_\theta[T U_S], \end{aligned}$$

em que

$U_S = E_\theta[U \mid S] \in \mathcal{U}_S$.

Redução por suficiência.

Estimadores de Máxima Verossimilhança

Suponha que X tem distribuição P_θ, $\theta \in \Omega \subset \mathbb{R}^k$. Assuma que P_θ tem fdp p_θ com respeito a uma medida (σ-finita).

Para cada $x \in \mathcal{X}$ (espaço amostral), $p_\theta(x)$, vista como função de θ, é chamada de função de verossimilhança e é denotada por $L(\theta; x)$.

Um $\hat{\theta} \in \Omega$ satisfazendo $L(\hat{\theta}; x) = \sup_{\theta \in \Omega} L(\theta; x)$ é chamado de estimativa de máxima verossimilhança de θ (EMV). $\hat{\theta}$ visto como função de X é chamado de estimador de máxima verossimilhança (EMV) de θ.

Seja g uma função de θ em $\mathbb{R}^p$, $p \leq k$. Se $\hat{\theta}$ é um EMV de θ, então $\hat{g} = g(\hat{\theta})$ é, por definição, um EMV de $g(\theta)$.

OBS. Usualmente, $p_\theta(x)$ é denotada por $f(x;\theta)$ ou $f(x|\theta)$.

OBS. Se $X = (X_1, \ldots, X_n)$ e os X_i's são iid com fdp $f(x_i;\theta)$, a função de verossimilhança para a amostra $x = (x_1, \ldots, x_n)$ é

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Omega.$$

$p_\theta(x) = P_\theta[X = x]$, x a amostra observada

$L(\theta; x) = p_\theta(x)$, $\theta \in \Omega$.

Procurar $\theta \in \Omega$ que maximiza $L(\theta; x)$, para a amostra observada x, equivale a procurar $\theta \in \Omega$ que maximiza

$$\ell(\theta; x) = \log L(\theta; x)$$

pois log é uma função estritamente crescente. Aqui log é o logaritmo na base e.

Em muitas situações, $\hat{\theta}$ pode ser obtido como solução de

$$\frac{\partial \ell(\theta; x)}{\partial \theta} = 0. \quad \text{(Equações de verossimilhança)}$$

Nota: O vetor coluna $\partial \ell / \partial \theta$ é chamado de vetor escore.

$$\frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \partial \ell / \partial \theta_1 \\ \vdots \\ \partial \ell / \partial \theta_k \end{pmatrix} \; ; \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}.$$

89

Ex 1. $X_1, \dots, X_n$ iid; $X_i \sim \text{Bernoulli}(p)$, $p \in [0,1]$

$$L(p;x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

para $x = (x_1, \dots, x_n)$, $x_i \in \{0,1\}$, $i=1,\dots,n$.

a) Se $\bar{x} \neq 0$ e $\bar{x} \neq 1$, para $p \in (0,1)$

$$\ell(p;x) = n\bar{x}\log p + n(1-\bar{x})\log(1-p)$$

$$\frac{\partial \ell}{\partial p} = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} = 0 \iff p = \bar{x}$$

$$\frac{\partial^2 \ell}{\partial p^2} = -\frac{n\bar{x}}{p^2} - \frac{n(1-\bar{x})}{(1-p)^2} \quad e \quad \left.\frac{\partial^2 \ell}{\partial p^2}\right|_{p=\bar{x}} < 0.$$

Para $p=0$ ou $p=1$, $L(p;x) = 0$.

b) Se $\bar{x} = 0$

$L(p;x) = (1-p)^n$ é função estrita/decrescente em p e é maximizada em $p = 0 = \bar{x}$.

c) Se $\bar{x} = 1$

$L(p;x) = p^n$ é função estritamente ~~de~~crescente em p e é maximizada em $p = 1 = \bar{x}$

Logo, $\hat{p} = \bar{X}$ é EMV de p. É também ENVVUM.

Ex 2. Idem, com $p \in (0,1) = \Omega$

Se $\bar{x} \neq 0$ e $\bar{x} \neq 1$, então $\bar{x} \in (0,1) = \Omega$ e $\bar{x}$ é EMV de p.

Se $\bar{x} = 0$, $L(p;x)$ é maximizada em $p=0$ (fora do Ω)

Aqui, $\sup_{p \in (0,1)} L(p;x) = 1$.

Para mostrar que $\nexists$ estimativa MV qdo $\bar{x}=0$, admita, por absurdo, que esta estimativa é $\hat{p} = p_0 \in (0,1)$.

Tome $p_1 \in (0, p_0)$. Então

$$L(p_1;x) > L(p_0;x)$$

pois L é estritamente decrescente em p.

Logo $\hat{p} = p_0$ não é estimativa MV de p.

[Fazer demonstração análoga p/ o caso em que $\bar{x}=1$]

Note que, se $p \in (0,1)$

$$P_p(\bar{X}=0) = (1-p)^n \quad e \quad P_p(\bar{X}=1) = p^n$$

que convergem rapidamente p/ zero qdo $n \to \infty$.

O exemplo mostra que, para n fixo, um EMV pode não existir.

Nota: Em geral, para n grande, os EMVs têm boas propriedades:

- consistência
- normalidade assintótica
- eficiência assintótica

Cap 6, TPE, Seções 6.1 a 6.6.

Lema. Seja $\hat{\theta} = \hat{\theta}(X)$ um EMV de θ e admite que $\hat{\theta}$ seja o <u>único</u> EMV. Então, se $\hat{\theta}$ é uma estatística suficiente, ele é uma estatística suficiente minimal.

Dem. Conseqüência do Teorema 8.1.30 de Dudewicz & Mishra, 1988, p-403.

Ex 3. Uniforme

$X_1, \dots, X_n$ iid, $X_i \sim U[0,\theta]$.

Para $x_1, \dots, x_n > 0$,

$$L(\theta; x) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[x_{(n)}, +\infty)}(\theta)$$

$\hat{\theta} = X_{(n)}$ é EMV (único) de θ e é suficiente

Então, $X_{(n)}$ é estatística suficiente minimal.

Note que $\hat{\theta}$ é um estimador viciado de θ.

Ex 4. Uniforme

$X_1, \dots, X_n$, iid ; $X_i \sim U[\theta - 1/2, \theta + 1/2]$, $\theta \in \mathbb{R}$

$$L(\theta; x) = \prod_{i=1}^{n} I_{[\theta-1/2, \theta+1/2]}(x_i) =$$

$$= I_{[\theta-1/2, +\infty)}(x_{(1)}) \, I_{(-\infty, \theta+1/2]}(x_{(n)}) = I_{(-\infty, x_{(1)}+1/2]}(\theta) \, I_{[x_{(n)}-1/2, +\infty)}(\theta)$$

Qualquer estatística $T(X)$ tq

$$X_{(n)} - \frac{1}{2} \leq T(X) \leq X_{(1)} + \frac{1}{2}$$

é um EMV de θ.

EMV não é necessariamente único.

Ex 5. Gama

$X_1, \ldots, X_n$ iid; $X_i \sim$ Gama (α, γ) com fdp

$$f(x;\theta) = \frac{1}{\Gamma(\alpha)} \gamma^{-\alpha} x^{\alpha-1} \exp\left\{-\frac{x}{\gamma}\right\}, \quad x>0$$

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

Equações de verossimilhança

$$-n\log\gamma - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum\log x_i = 0$$

e

$$-\frac{n\alpha}{\gamma} + \frac{1}{\gamma^2}\sum x_i = 0.$$

A 2ª equação leva a $\gamma = \bar{x}/\alpha$. Substituindo na 1ª, temos

$$\log\alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \frac{1}{n}\sum\log x_i - \log\bar{x} = 0$$

que não tem solução explícita.

Pode-se mostrar que $\exists$ solução e é única (Ex. 3.20, p502, TPG)

Método numérico deve ser usado para obter estimativa MV de α e γ para uma dada amostra (Newton Raphson, scoring de Fisher, etc.).

Ex 6. Normal

$X_1, \ldots, X_n$ iid; $X_i \sim N(\mu, \sigma^2)$,

$\theta = (\mu, \sigma^2) \in \Omega = \mathbb{R} \times (0, +\infty)$

$x = (x_1, \ldots, x_n) \in \mathbb{R}^n$

Log-verossimilhança

$$\ell(\theta) = \ell(\theta; x) = -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{n}{2}\log\sigma^2 - \frac{n}{2}\log(2\pi).$$

Equações de verossimilhança

$$\frac{1}{\sigma^2}\sum(x_i-\mu) = 0 \quad \text{e} \quad \frac{1}{\sigma^4}\sum(x_i-\mu)^2 - \frac{n}{2\sigma^2} = 0.$$

A 1ª equação tem como única solução $\hat{\mu} = \bar{x}$. Substituindo μ por $\hat{\mu}$ na 2ª equação vem $\hat{\sigma}^2 = \sum(x_i-\bar{x})^2/n$.

Note que Ω é um conjunto aberto; $\ell(\theta)$ é diferenciável para todo $\theta \in \Omega$; qdo $\mu \to -\infty$, ou $\mu \to +\infty$ ou $\sigma^2 \to \infty$, $L(\theta; x)$ tende a zero.

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = -\begin{pmatrix} \frac{n}{\sigma^2} & \frac{\Sigma(x_i-\mu)}{\sigma^4} \\ \frac{\Sigma(x_i-\mu)}{\sigma^4} & \frac{\Sigma(x_i-\mu)^2}{\sigma^6} - \frac{n}{2\sigma^4} \end{pmatrix}$$

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Bigg|_{\theta=(\hat\mu,\hat\sigma^2)} = -\begin{pmatrix} n/\hat\sigma^2 & 0 \\ 0 & n/2\hat\sigma^4 \end{pmatrix}$$

Como o determinante dessa matriz é positivo e $\left.\frac{\partial^2 \ell(\theta)}{\partial \mu^2}\right|_{\theta=(\hat\mu,\hat\sigma^2)} < 0$, temos que

$$\hat\mu = \frac{\Sigma x_i}{n} \quad e \quad \hat\sigma^2 = \frac{\Sigma(x_i-\bar X)^2}{n}$$

são EMV's de μ e σ^2 respectivamente.

Ex 6. Normal - COMPARAÇÃO entre ENVVUM e EMV.

$X_1,\dots,X_n$, iid; $X_i \sim N(\mu,\sigma^2)$

EMV de μ e σ^2: $\hat\mu = \bar X$ e $\hat\sigma^2 = \frac{1}{n}\Sigma(X_i-\bar X)^2$

$\bar X$ é EMV e ENVVUM de μ

$\frac{S^2}{n-1} = \frac{\Sigma(X_i-\bar X)^2}{n-1}$ é ENVVUM e $\frac{S^2}{n}$ é EMV (viciado).

Considere a classe dos estimadores de σ^2 da forma cS^2. O EQM de cS^2 é

$$E[(cS^2-\sigma^2)^2] = \sigma^4[(n^2-1)c^2 - 2(n-1)c+1] \quad \text{(exercício)}$$

que é minimizado p/ $c = 1/(n+1)$. Então

$\frac{S^2}{n+1}$ é o estimador de σ^2 que tem risco uniformemente mínimo sob perda quadrática.

Além disso, S^2/n tem risco uniformemente menor que $S^2/(n-1)$.

Considere σ^2 conhecido e o estimando $g(\mu,\sigma^2)=\mu^2$.

EMV de μ^2 : $\bar X^2$ (viciado)

ENVVUM de μ^2 : $\bar X^2 - \sigma^2/n$.

O risco (EQM) do EMV é uniformemente maior que o do ENVVUM (ver Lema abaixo)

Lema. Considere o risco sob perda quadrática (EQM). Se δ é um estimador não viciado de $g(\theta)$ (função de θ a valores reais) e se $\delta^* = \delta + b$, onde o viés b $(b \neq 0)$ é independente de θ, então δ^* tem risco uniformemente maior que δ. De fato;

$$R_{\delta^*}(\theta) = R_\delta(\theta) + b^2.$$

Ex 6. Normal (cont)

Se σ^2 é conhecido: $\bar{X}^2 - \frac{\sigma^2}{n}$ é ENVVUM de μ^2

Se σ^2 é desconhecido: $\bar{X}^2 - \frac{S^2}{n(n-1)}$ " " " μ^2.

Em ambos os casos, o ENVVUM pode tomar valores negativos com prob. positiva, embora $\mu^2 > 0$.

Amostras grandes: isso ocorre com prob. pequena
ENVVUM $\approx$ EMV.

Considere σ^2 conhecido, $\sigma^2 = 1$, e o estimando $p = \Phi(u - \mu)$

ENVVUM de p: $\Phi\left[\sqrt{\frac{n}{n-1}}\,(u - \bar{X})\right]$

EMV de p: $\Phi(u - \bar{X})$.

EMV é viesado: pela completividade, pode existir apenas 1 estimador não viciado de p e este é função de $\bar{X}$.

Alunos: Ler Seção 2.3, TPE: ENVVUM em modelos discretos

ENVVUM em Famílias Não-Paramétricas

$X_1, \ldots, X_n$: iid com distribuição $F \in \mathcal{F}$.

Faremos apenas suposições bem gerais sobre $\mathcal{F}$ (por ex.: $\mathcal{F}$ é a família de distribuições F que têm densidade com respeito à medida de Lebesgue; ou que têm certos momentos, etc.).

Estimando: $g(F)$ (p.ex. $E(X_i) = \int x\, dF(x)$; $Var(X_i)$, $P(X_i \leq a) = F(a)$).

Pode-se mostrar que para a família $\mathcal{F}_0$ de todas as distribuições que têm densidade com respeito à medida de Lebesgue, as estatísticas de ordem $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ são uma estatística suficiente completa e isso também vale se exigirmos que as distribuições em $\mathcal{F}_0$ tenham alguns momentos.

Um estimador $\delta(X_1, \ldots, X_n)$ é uma função das estatísticas de ordem sse é simétrico em seus n argumentos.

Para as famílias para as quais as estatísticas de ordem são suficientes e completas, pode existir apenas um estimador não viciado, função de $X_{(1)}, \ldots, X_{(n)}$ de qualquer estimando, e ele é ENVVUM.

Basta procurar entre os estimadores "simétricos" (funçã de $X_{(1)}, \dots, X_{(n)}$) de uma dada funçã estimável $g(F)$, para encontrar o ENVVUM de $g(F)$.

Ex 1. Estimando: $g(F) = P(X_1 \leq a) = F(a)$, a fixo.
Estimador natural: nº de X_i's $\leq a$ dividido por n.

$T(X)$ = nº de X_i's menores ou iguais a "a".
$T(X) \sim$ Binomial $(F(a), n)$.

$E(T(X)/n) = F(a)$ e $\therefore$ $\frac{T(X)}{n}$ é ENVVUM de $F(a)$.

Ex 2. Idem; $\mathcal{F}_0$ restrita à classe de distribuições (com densidade em relaçã à med. de Lebesgue), com $E|X| < \infty$.

Estimando: $g(F) = \int x f(x)\, dx = E(X_1)$

$\bar{X}$: estimador não viciado e simétrico
(Note que $\bar{X} = \Sigma X_{(i)}/n$).
$\bar{X}$: ENVVUM de $g(F)$.

Outra soluçã: X_1 é ENV de $g(F)$

$$E[X_1 \mid X_{(1)}, \dots, X_{(n)}] = \frac{1}{n} \Sigma X_{(i)} = \bar{X}$$

já que, dado $(X_{(1)}, \dots, X_{(n)})$, X_1 assume cada um desses valores com prob. $1/n$.

Alunos: Ler secçã 3.4, TPE (outros exemplos em famílias não paramétricas).

DESIGUALDADE DA INFORMAÇÃO

Suponha que X tem distribuiçã com densidade $p_\theta = p(x, \theta) = p_\theta(x)$ com respeito a uma medida μ e que p_θ é diferenciável com respeito a θ, $\theta \in \Omega \subset \mathbb{R}$.

Seja
$$\psi(x, \theta) = \frac{\partial p(x,\theta)}{\partial \theta} \frac{1}{p(x,\theta)} = \frac{\partial}{\partial \theta} \log p(x, \theta),$$
para todo x e θ tais que $p(x,\theta) \neq 0$.

Seja
$$I(\theta) = E_\theta\left[\psi(X,\theta)^2\right] = E_\theta\left[\left(\frac{\partial}{\partial\theta} \log p(X,\theta)\right)^2\right]$$
$$= \int \left(\frac{p'_\theta}{p_\theta}\right)^2 p_\theta \, d\mu,$$
em que $p'_\theta = \frac{\partial}{\partial \theta} p(x, \theta)$

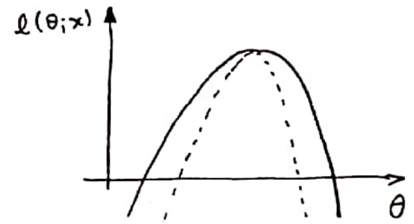
$I(\theta)$: informação de Fisher (informação esperada) que X contém sobre θ.

$\psi(x, \theta)$ é a taxa de variação relativa da densidade p_θ (no ponto x) em relaçã a θ.

OBS. Veremos a seguir que, sob condições adequadas

$$I(\theta) = E_\theta\left[-\frac{\partial^2 \log p(X,\theta)}{\partial\theta^2}\right].$$

$I(\theta)$ é interpretada como o valor médio da curvatura da log-verossimilhança. Quanto maior a curvatura, mais "informação" os dados têm sobre θ.



boca p/ baixo; 2ª derivada negativa; daí sinal negativo em $E_\theta[-\cdots]$.

Veremos também que a variância de estimadores não viciados de θ é sempre $\geq I(\theta)^{-1}$. $I(\theta)$ é a maior precisão possível para estimadores não viciados de θ.

De maneira mais geral, se δ é um estimador não viciado de $g(\theta)$, então, sob condições adequadas,

$$var_\theta(\delta) \geq \frac{[g'(\theta)]^2}{I(\theta)}.$$

(Esta é a chamada desigualdade da informação; ver p. 109)

Limites inferiores para $var_\theta(\delta)$ podem ser obtidos da "desigualdade da covariância".

Se X e Y são duas v.a's com segundo momento finito, então

$$cov(X,Y)^2 \leq var(X)\, var(Y).$$

(Problema 1.5, TPE, p. 130: Desigualdade de Cauchy-Schwarz).

Se uma função $\psi(x,\theta)$ tem segundo momento finito

$$var_\theta(\delta) \geq \frac{(cov_\theta[\delta,\psi])^2}{var_\theta(\psi)}.$$

O problema é que o lado direito envolve δ. Mas se $cov_\theta[\delta,\psi]$ só depende de δ através de sua média $E_\theta[\delta] = g(\theta)$, o lado direito da inequação fornece um limite inferior para as variâncias de todos os estimadores não viciados de $g(\theta)$.

Mas quando isso ocorre?

Teorema: Suponha que δ tem 2º momento finito para todo $\theta \in \Omega$.

Uma condição necessária e suficiente para que $cov_\theta(\delta, \psi)$ dependa de δ somente através de $g(\theta) = E_\theta(\delta)$ para todo θ é que

$$cov_\theta(U, \psi) = 0 \text{ para todo } U \in \mathcal{U},$$

onde

$$\mathcal{U} = \{U : E_\theta(U) = 0 \text{ e } E_\theta(U^2) < \infty, \text{ para todo } \theta \in \Omega\}.$$

Dem. Afirmar que $cov_\theta(\delta, \psi)$ depende de δ somente através de $g(\theta)$ é equivalente a dizer que para quaisquer 2 estimadores δ_1 e δ_2 com $E_\theta(\delta_1) = E_\theta(\delta_2)$ para todo θ temos $cov_\theta(\delta_1, \psi) = cov_\theta(\delta_2, \psi)$.

A prova segue do fato de que

$$cov_\theta(\delta_1, \psi) - cov_\theta(\delta_2, \psi) = cov_\theta(\delta_1 - \delta_2, \psi) = cov_\theta(U, \psi), \quad U = \delta_1 - \delta_2$$

e que, portanto,

$$cov_\theta(\delta_1, \psi) = cov_\theta(\delta_2, \psi), \ \forall \delta_1, \delta_2, \ \forall \theta \iff$$

$$cov_\theta(U, \psi) = 0, \quad \forall U, \forall \theta.$$

Ex. Suponha que X tem densidade $P_\theta = p(x, \theta)$ e $p(x, \theta) > 0$ para todo x, $E_\theta(\delta) = g(\theta)$, $E_\theta(\delta^2) < \infty$

Suponha que θ e $\theta + \Delta$ sejam dois valores p/ os quais $g(\theta) \neq g(\theta + \Delta)$.

Seja

$$\psi(x, \theta) = \frac{p(x, \theta + \Delta)}{p(x, \theta)} - 1.$$

$$E_\theta(\psi) = \int \left[\frac{p(x, \theta + \Delta)}{p(x, \theta)} - 1\right] p(x, \theta)\, d\mu(x) = 0$$

$$cov_\theta(U, \psi) = E_\theta(\psi U) = E_{\theta+\Delta}(U) - E_\theta(U) = 0$$

e

$$cov_\theta(\delta, \psi) = E_\theta(\delta\psi) = g(\theta + \Delta) - g(\theta).$$

Então:

$$var_\theta(\delta) \geq \frac{(g(\theta + \Delta) - g(\theta))^2}{E_\theta\left\{\left[\frac{p(x, \theta + \Delta)}{p(x, \theta)} - 1\right]^2\right\}}, \quad \forall \Delta \qquad (\bullet)$$

O lado direito pode ser substituído por seu supremo em Δ. (desigualdade de Hammersley - Chapman - Robbins).

A desigualdade (·) também é obtida tomando

$$\psi(x,\theta) = \frac{p(x,\theta+\Delta) - p(x,\theta)}{\Delta \quad p(x,\theta)}$$

que tende a $\left[\frac{\partial}{\partial\theta} p(x,\theta)\right] \frac{1}{p(x,\theta)}$, desde

que $p(x,\theta)$ seja diferenciável em θ. Isso sugere considerar a seguinte função ψ:

$$\psi(x,\theta) = \left[\frac{\partial}{\partial\theta} p(x,\theta)\right] \frac{1}{p(x,\theta)}$$

$$= \frac{\partial}{\partial\theta} \log p(x,\theta) \ .$$

Esta especificação para ψ, com as suposições adequadas, leva à desigualdade da informação.

Suposições:

(a) Ω é um intervalo aberto (finito, infinito ou semi-finito).

(b) As distribuições P_θ têm suporte comum, de tal forma que o conjunto
$$A = \{x,\ p_\theta(x) > 0\}$$
é independente de θ.

(c) Para qualquer $x \in A$ e $\theta \in \Omega$, a derivada
$$p'_\theta(x) = \frac{\partial p_\theta(x)}{\partial\theta} \qquad \langle * \rangle$$
existe e é finita.

Lema: (i) Se (a)-(c) estão satisfeitas e a derivada com respeito a θ do lado esquerdo de
$$\int p_\theta(x)\, d\mu(x) = 1 \qquad \langle \bullet \rangle$$
pode ser obtida diferenciando sob o sinal de integração, então
$$E_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(X)\right] = 0$$
e
$$I(\theta) = \operatorname{var}_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(X)\right] \qquad \langle ** \rangle$$

(ii) Se, adicional/, a 2ª derivada com respeito a θ de $\log p_\theta(x)$ existe para todo $x \in A$ e $\theta \in \Omega$ e a 2ª derivada c/ respeito a θ do lado esquerdo de $\langle * \rangle$ pode ser obtida diferenciando 2 vezes sob o sinal de integração, então
$$I(\theta) = -\,E_\theta\left[\frac{\partial^2}{\partial\theta^2} \log p_\theta(X)\right].$$

Dem.: (i) $\int p_\theta(x)\, d\mu(x) = 1 \Rightarrow \frac{\partial}{\partial\theta}\int p_\theta(x)\, d\mu(x) = 0$

$\Rightarrow \int \frac{\partial}{\partial\theta} p_\theta(x)\, d\mu(x) = 0 \Rightarrow \int \left[\frac{\partial}{\partial\theta} p_\theta(x)\right] \frac{p_\theta(x)}{p_\theta(x)}\, d\mu(x) = 0$

$\Rightarrow \int \frac{\partial}{\partial\theta} \log p_\theta(x)\; p_\theta(x)\, d\mu(x) = 0 \Rightarrow E_\theta\left[\frac{\partial \log p_\theta(x)}{\partial\theta}\right] = 0.$

<**>: imediato usando igualdade acima.

(ii) $\frac{\partial^2}{\partial\theta^2} \log p_\theta(x) = \frac{\partial}{\partial\theta}\left[\frac{p'_\theta(x)}{p_\theta(x)}\right] =$

$\left\{\left[\frac{\partial^2}{\partial\theta^2} p_\theta(x)\right] p_\theta(x) - \left[\frac{\partial}{\partial\theta} p_\theta(x)\right]^2\right\} \frac{1}{(p_\theta(x))^2} =$

$\left[\frac{\partial^2}{\partial\theta^2} p_\theta(x)\right] \frac{1}{p_\theta(x)} - \left\{\left[\frac{\partial}{\partial\theta} p_\theta(x)\right] \frac{1}{p_\theta(x)}\right\}^2$

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p_\theta(x)\right] = \int \frac{\partial^2}{\partial\theta^2} p_\theta(x)\, d\mu(x) - \int\left[\frac{\partial}{\partial\theta} p_\theta(x)\right]^2 \frac{1}{(p_\theta(x))^2}\, p_\theta(x)\, d\mu(x)$$

$$= \underbrace{\frac{\partial^2}{\partial\theta^2}\int p_\theta(x)\, d\mu(x)}_{=0} - \int\left(\frac{\partial \log p_\theta(x)}{\partial\theta}\right)^2 p_\theta(x)\, d\mu(x)$$

$$= -E_\theta\left[\left[\frac{\partial \log p_\theta(x)}{\partial\theta}\right]^2\right] = -I(\theta).$$

Nota. Na família exponencial unidimensional

$$p_\eta(x) = \exp(\eta\, T(x) - A(\eta))\, h(x).$$

Se $A(\eta)$ é diferenciável, a troca de integral e derivada c/ respeito a η em $\int p_\eta(x)\, d\mu(x)$ é válida.

Se $A(\eta)$ é duplamente diferenciável, a troca da 2ª derivada c/ respeito a η em $\int p_\eta(x)\, d\mu(x)$ é válida.

Alunos: verificar (conseqüência do Teorema 5.8, p. 27, TPE).

Nota: $I(\theta)$ depende da parametrização utilizada.

Se $\theta = h(\xi)$ e h é diferenciável, então a informação de Fisher na parametrização ξ (informação que X contém sobre ξ) é

$$I^*(\xi) = E_\xi\left\{\left[\frac{\partial \log p(X, h(\xi))}{\partial \xi}\right]^2\right\}$$

$$= E_\xi\left\{\left[\frac{\partial \log p(X, h(\xi))}{\partial h(\xi)} \frac{\partial h(\xi)}{\partial \xi}\right]^2\right\}$$

$$= E_\xi\left\{\left[\frac{\partial \log p(X, h(\xi))}{\partial h(\xi)}\right]^2\right\} h'(\xi)^2$$

$$= \underbrace{I(h(\xi))}_{\text{informação que } X \text{ contém sobre } h(\xi) := \theta} h'(\xi)^2$$

$$I(h(\xi)) = \frac{I^*(\xi)}{h'(\xi)^2}.$$

$\theta = e^{\xi}$

$I(\theta) = \frac{1}{\theta^2}$

Informação sob reparametrização

Aula 09, 14/09

Teorema. Suponha que X tem distribuição na família exponencial unidimensional (ver p. 17). Seja
$$\tau(\theta) = E_\theta(T).$$
Então, a informação que X contém sobre $\tau(\theta)$, denotada por $I[\tau(\theta)]$, é
$$I[\tau(\theta)] = \frac{1}{var_\theta(T)}.$$

T na famil. exponencial é ñ viciado e tem v. unif. mínima.

Dem.: $I(\theta) = var_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(X)\right] = var_\theta\left[\eta'(\theta) T(X) - B'(\theta)\right]$

$$= (\eta'(\theta))^2 \, var_\theta(T).$$

Mas $$I[\tau(\theta)] = \frac{I(\theta)}{(\tau'(\theta))^2} = \left[\frac{\eta'(\theta)}{\tau'(\theta)}\right]^2 var_\theta(T). \quad (\cdot)$$

Além disso (lista 1),
$$\tau(\theta) = E_\theta(T) = \frac{B'(\theta)}{\eta'(\theta)} \quad (\cdot\cdot)$$
e
$$var_\theta(T) = \frac{B''(\theta)}{[\eta'(\theta)]^2} - \frac{\eta''(\theta) B'(\theta)}{[\eta'(\theta)]^3}. \quad (\cdot\cdot\cdot)$$

Substituindo $(\cdot\cdot)$ e $(\cdot\cdot\cdot)$ em $(\cdot)$, vem o resultado (exercício).

Resultado mais geral.
$h(\theta)$ diferenciável

$$I[h(\theta)] = \left[\frac{\eta'(\theta)}{h'(\theta)}\right]^2 var_\theta(T).$$

Ex1. Gama

$X \sim \text{Gama}(\alpha, \beta)$, α conhecido. Para $x > 0$

$$p_\beta(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}.$$

$$= \exp\left\{-\frac{x}{\beta} - \alpha \log \beta\right\} h(x),$$

em que $h(x) = x^{\alpha-1}/\Gamma(\alpha)$.

Aqui: $\eta(\beta) = -\frac{1}{\beta}$, $T(x) = x$, $B(\beta) = \alpha \log \beta$.

$\tau(\beta) = E_\beta(T) = \alpha\beta$ e $\text{var}_\beta(T) = \alpha\beta^2$ (lista 1)

A informação de fisher sobre $\alpha\beta$ é

$$I[\alpha\beta] = \left[\frac{\eta'(\beta)}{\tau'(\beta)}\right]^2 \text{var}_\beta(T) = \frac{1}{\alpha\beta^2}.$$

Para obter a informação sobre β, tomamos $\eta(\beta) = -\frac{\alpha}{\beta}$ e $T(x) = \frac{x}{\alpha}$.

Aqui, $E_\beta(T) = \beta$ e $\text{var}_\beta(T) = \beta^2/\alpha$.

$$I[\beta] = \left[\frac{\alpha}{\beta^2}\right]^2 \text{var}_\beta(T) = \frac{1}{\text{var}_\beta(T)}.$$

Ver outros exemplos na Tab 5.1, p. 118, TPE.

Ex 2. Normal; estimação de ξ^2

$X \sim N(\xi, \sigma^2)$, σ^2 conhecido

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\xi)^2\right\}, \quad x \in \mathbb{R}$$

$$= \exp\{\eta(\xi)\,T(x) - B(\xi)\}\, h(x)$$

em que

$\eta(\xi) = \xi$, $T(x) = \frac{x}{\sigma^2}$, $B(\xi) = \frac{1}{2}\frac{\xi^2}{\sigma^2}$ e

$h(x) = \exp\{-x^2/2\sigma^2\} \frac{1}{\sqrt{2\pi}}$.

A informação sobre ξ^2 que X contém é

$$I[\xi^2] = \left[\frac{\eta'(\xi)}{g'(\xi)}\right]^2 \text{var}_\xi(T),$$

em que $g(\xi) = \xi^2$.

$$I[\xi^2] = \left(\frac{1}{2\xi}\right)^2 \frac{1}{\sigma^2} = \frac{1}{4\xi^2\sigma^2}.$$

Ex 3. Poisson

$$P_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \exp\{x \log \lambda - \lambda\}\frac{1}{x!}, \quad x = 0,1,2,\ldots$$

$$P_\lambda(x) = \exp\{\eta(\lambda)T(x) - B(\lambda)\}\, h(x)$$

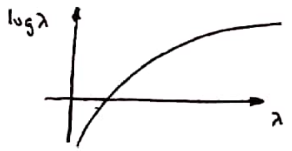
com $\eta(\lambda) = \log \lambda$, $T(x) = x$, $B(\lambda) = \lambda$, $h(x) = \frac{1}{x!}$.

$$\underline{I[\eta(\lambda)]} = I[\log\lambda] = \left[\frac{\eta'(\lambda)}{\eta'(\lambda)}\right]^2 \mathrm{var}_\lambda(T(X)) = \lambda.$$

Kar) $I[\lambda] = 1/\lambda$: decrescente em λ. Estima-se λ de forma mais acurada qdo λ é pequeno.

$I[\log\lambda] = \lambda$: crescente em λ. Estima-se $\log\lambda$ de forma mais acurada qdo λ é grande.

Motivo: qdo λ é grande $\log\lambda$ varia muito lentamente.



Tomemos $h(\lambda) = \sqrt{\lambda}$

$$I[\sqrt{\lambda}] = \left[\frac{\eta'(\lambda)}{h'(\lambda)}\right]^2 \mathrm{Var}_\lambda(T(X)) = \left[\frac{1/\lambda}{1/[2\sqrt{\lambda}]}\right]^2 \cdot \lambda = \text{constante}.$$

(não depende de λ).

FAMÍLIA DE LOCAÇÃO

X tem densidade $f(x-\theta)$; $f(x) > 0$, p/ todo $x \in \mathbb{R}$; $\theta \in \Omega = \mathbb{R}$.

As condições da p. 100 estão satisfeitas desde que $f'(x)$ exista para todo x.

$$I(\theta) = I_f = \int_{-\infty}^{+\infty}\left[\frac{f'(x-\theta)}{f(x-\theta)}\right]^2 f(x-\theta)\,dx$$

$$= \int_{-\infty}^{+\infty}\frac{(f'(x-\theta))^2}{f(x-\theta)}\,dx = \int_{-\infty}^{+\infty}\frac{(f'(x))^2}{f(x)}\,dx$$

(não depende de θ) Ver Tab 5.2, p. 119, TPE.

FAMÍLIA DE LOCAÇÃO-ESCALA (parâmetro b fixado)

X tem densidade $b^{-1} f(b^{-1}(x-\theta))$

$$I(\theta) = \frac{I_f}{b^2} \quad \text{(mostrar)}$$

$$I(\theta) = E\left[\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2\right]$$

Teorema. Sejam X e Y variáveis aleatórias independentes com densidades p_θ e q_θ (com respeito às medidas μ e ν) respectivamente, satisfazendo as condições (a) - (c) (p.100) e

$$E_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta(x)\right]=0 \quad e \quad E_\theta\left[\frac{\partial}{\partial\theta}\log q_\theta(y)\right]=0. \quad [*]$$

Se $I_1(\theta)$, $I_2(\theta)$ e $I(\theta)$ são as informações sobre θ contidas em X, Y e (X,Y) respectivamente, então

$$I(\theta)=I_1(\theta)+I_2(\theta).$$

Dem. Alunos $\quad p_{X,Y}(x,y;\theta)=p_\theta(x)p_\theta(y) \rightarrow$

Corolário: Se $X_1,\dots,X_n$ são iid satisfazendo (a)-(c) (p.100) e condições como [*] acima e se cada uma tem informação $I(\theta)$ sobre θ, então a informação contida em $(X_1,\dots,X_n)$ é $nI(\theta)$.

Dem. Alunos.

Teorema. DESIGUALDADE DA INFORMAÇÃO

Suponha que $\{p_\theta,\ \theta\in\Omega\}$, é uma família de densidades dominada pela medida μ para a qual as condições (a) - (c) valem e

$$E_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta\right]=0$$

e tal que $I(\theta)>0$. Seja δ uma estatística t.q. $E_\theta(\delta^2)<\infty$, para a qual $E_\theta(\delta)$ existe e pode ser diferenciada sob o sinal de integração, i.e.,

$$\frac{\partial}{\partial\theta}E_\theta(\delta)=\int\frac{\partial}{\partial\theta}\delta\, p_\theta\, d\mu.$$

Então
$$\operatorname{var}_\theta(\delta)\geq\frac{\left[\frac{\partial}{\partial\theta}E_\theta(\delta)\right]^2}{I(\theta)} \qquad (✪)$$

Dem. Alunos. Usar $\operatorname{var}_\theta(\delta)\geq\dfrac{[\operatorname{cov}_\theta(\delta,\psi)]^2}{\operatorname{var}_\theta(\psi)}$, se $\operatorname{var}_\theta(\delta)<\infty$ e $0<\operatorname{var}_\theta(\psi)<\infty$, com $\psi=\partial\log p_\theta(x)/\partial\theta$.

Nota: Alguns livros colocam $I(\theta)\in(0,+\infty)$. Mas se $I(\theta)=+\infty$, o lado direito de (✪) é tomado como zero e a desigualdade é trivial.

Nota: Se $E_\theta(\delta)=\theta$, então $\operatorname{var}_\theta(\delta)\geq\dfrac{1}{I(\theta)}$.

(✪): Limite inferior para a variância dos estimadores não viciados de uma função dada de θ.

Limite inferior de Cramér-Rao.

Se δ é um estimador de $g(\theta)$ com

$$\underline{E_\theta(\delta)} = \underline{g(\theta)} + \underbrace{b(\theta)}_{\text{vié's}}$$

então

$$\underline{\operatorname{var}_\theta(\delta)} \geq \frac{[b'(\theta) + g'(\theta)]^2}{I(\theta)}$$

(Limite inferior para a variância de qualquer estimador em termos de seu vié's e de $I(\theta)$).

Caso iid: $X_1, .., X_n$ iid ; $\delta = \delta(X_1, ..., X_n)$

$$\operatorname{var}_\theta(\delta) \geq \frac{[b'(\theta) + g'(\theta)]^2}{n I_1(\theta)}$$

$I_1(\theta)$: informação sobre θ contida em X_1

Nota: o LICR é invariante sob reparametrização (idem para os 2 limites acima).

De fato, como, se $\theta = h(\xi)$ com h diferenciável,

$$\frac{\partial}{\partial \xi} E_{h(\xi)}(\delta) = \frac{\partial}{\partial \theta} E_\theta(\delta)\, h'(\xi),$$

temos

$$\frac{\left[\frac{\partial}{\partial \theta} E_\theta(\delta)\right]^2}{I(\theta)} = \frac{\left[\frac{\partial}{\partial \theta} E_\theta(\delta)\, h'(\xi)\right]^2}{I(\theta)\, h'(\xi)^2} = \frac{\left[\frac{\partial}{\partial \xi} E_{h(\xi)}(\delta)\right]^2}{I^*(\xi)}$$

(ver p. 102)

Nota: O LICR é atingido, ou seja,

$$\operatorname{var}_\theta(\delta) = \frac{[\partial E_\theta(\delta)/\partial\theta]^2}{I(\theta)}$$

sse

$$\frac{\partial}{\partial \theta} \log p_\theta(x) = a(\theta)\, \delta(x) + b(\theta) \quad \text{q.c.-}P_\theta.$$

A prova vem do fato de que, para 2 v.a's X e Y tq $\operatorname{var}(X) < \infty$ e $\operatorname{var}(Y) < \infty$,

$$\operatorname{cov}(X, Y)^2 \leq \operatorname{var}(X)\operatorname{var}(Y)$$

e a igualdade ocorre sse $Y = aX + b$.
(Tomar $X = \delta$ e $Y = \frac{\partial}{\partial\theta} \log p_\theta(X)$).

Alunos: mostrar.

Ex. Binomial
$X \sim b(p, n)$, $\theta = p$, $0 < \theta < 1$.

Seja $\delta = a(\theta) \frac{\partial \log p_\theta(x)}{\partial \theta} + b(\theta)$, p/ algum a, b

$$\delta = a(\theta)\left[\frac{\partial}{\partial\theta}\{x \log\theta + (n-x)\log(1-\theta)\}\right] + b(\theta)$$

$$= a(\theta)\left[\frac{x}{\theta} - \frac{n-x}{1-\theta}\right] + b(\theta) =$$

$$= a(\theta)\frac{x}{\theta(1-\theta)} + b(\theta) - a(\theta)\, n \frac{\theta}{\theta(1-\theta)}$$

não depende de θ (é estimador) $\Leftrightarrow$
$a(\theta) = c\,\theta(1-\theta)$ e $b(\theta) - cn\theta = d$,
com c e d não dependendo de θ.

achievement

$\Leftrightarrow \quad \delta = cx + e$, onde c e e não dependem de θ.

Aqui, $E_\theta(\delta(X)) = cn\theta + e$ $\quad g(\theta)$

e

$$var_\theta(\delta(X)) = c^2 n\theta(1-\theta).$$

Apenas funções lineares de θ (essencial/ θ) têm estimadores não viciados cuja variância coincide com o LICR.

Aqui $\quad LICR = \dfrac{[\partial E_\theta(\delta)/\partial\theta]^2}{I(\theta)} = \dfrac{(cn)^2}{n/[\theta(1-\theta)]}$

$$= c^2 n\, \theta(1-\theta).$$

(Alunos: obter $I(\theta)$).

Estimação de $\theta(1-\theta)$ [lista 2]

O ENVVUM de $\theta(1-\theta)$ é $X(n-X)/[n(n-1)]$;

sua variância é $\dfrac{\theta(1-\theta)}{n}\left[(1-2\theta)^2 + \dfrac{2\theta(1-\theta)}{n-1}\right]$.

Por outro lado,

$$LICR = \frac{[\partial\theta(1-\theta)/\partial\theta]^2}{I(\theta)} = \frac{(1-2\theta)^2}{n/[\theta(1-\theta)]} = \frac{(1-2\theta)^2\theta(1-\theta)}{n}$$

$$= var_\theta\left(\frac{X(n-X)}{n(n-1)}\right) - \frac{2\theta^2(1-\theta)^2}{n(n-1)} < var_\theta\left(\frac{X(n-X)}{n(n-1)}\right).$$

Não viciado função de est. suf. completa $\to$ ENVVUM

Nota: Integrando ambos os lados de

$$\longrightarrow \quad \frac{\partial}{\partial\theta}\log p_\theta(x) = a(\theta)\,\delta(x) + b(\theta)$$

pode-se mostrar que, sob condições adequadas,

$$p_\theta(x) = C(\theta)\exp\{\varphi(\theta)\,\delta(x)\}\,h(x)$$

(família exponencial unidimensional com $\delta(x)$ estatística suficiente)

Ver Teorema 5.12, p. 121, TPE.

CONDIÇÕES DE REGULARIDADE

Para a validade da desigualdade de informação foi exigido que $E_\theta(\delta^2) < \infty$ (condição não restritiva).

Foi exigido também que

$$\frac{\partial}{\partial\theta} E_\theta[\delta(X)] = \int \frac{\partial}{\partial\theta} \delta(x)\, p_\theta(x)\, d\mu(x) = g'(\theta). \quad (*)$$

(troca)

Na demonstração da D.I. é necessário assegurar que

$$\mathrm{cov}_\theta(\delta, \psi) = \frac{\partial}{\partial\theta} E_\theta(\delta); \qquad \psi(x,\theta) = \frac{\partial}{\partial\theta} \log p_\theta(x)$$

Vejamos:

$$\mathrm{cov}_\theta(\delta, \psi) = \int \delta(x) \frac{\partial}{\partial\theta}\left[\log p_\theta(x)\right] p_\theta(x)\, d\mu(x)$$

$$= \int \delta(x) \left[\lim_{\Delta\to 0} \frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta\, p_\theta(x)}\right] p_\theta(x)\, d\mu(x)$$

(?) $$\overset{?}{=} \lim_{\Delta\to 0} \int \delta(x) \left[\frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta\, p_\theta(x)}\right] p_\theta(x)\, d\mu(x)$$

$$= \lim_{\Delta\to 0} \left[E_{\theta+\Delta}(\delta(X)) - E_\theta(\delta(X))\right] \frac{1}{\Delta}$$

$$= \frac{\partial}{\partial\theta} E_\theta(\delta(X)).$$

Lema. Suponha que as condições (a) e (b) (p. 100) sejam válidas e seja δ um estimador para o qual $E_\theta(\delta^2) < \infty$. Seja $q_\theta(x) = \frac{\partial}{\partial\theta} \log p_\theta(x)$ e, para cada $\varepsilon > 0$, seja b_θ uma função que satisfaz

(**) $$E_\theta\left[(b_\theta(X))^2\right] < \infty \quad \text{e} \quad \left|\frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta\, p_\theta(x)}\right| \leq b_\theta(x),$$

para todo $|\Delta| < \varepsilon$.

Então,

$$E_\theta[q_\theta(X)] = 0$$

e

$$\frac{\partial}{\partial\theta} E_\theta[\delta(X)] = E_\theta[\delta(X)\, q_\theta(X)] = \mathrm{cov}_\theta(\delta, q_\theta).$$

e, portanto, vale (*).

Dem. Como $\left|\delta(x) \frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta\, p_\theta(x)}\right| \leq |\delta(x)|\,|b_\theta(x)|$

e $E_\theta[|\delta(X)|\,|b_\theta(X)|] \leq E_\theta[(\delta(X))^2]^{1/2} E_\theta[(b_\theta(X))^2]^{1/2} < \infty$

segue do Teorema da Convergência Dominada (Teo 2.5, Cap 1, TPE) que a troca de limite e integral em (?) é válida.

<u>Teorema</u>. Suponha que $p_\theta(x)$ é uma família de densidades com uma medida dominante $\mu(x)$ que satisfaz as condições (a)-(c) (p. 100), $I(\theta)>0$ e que existam uma função b_θ e um $\varepsilon>0$ para os quais $(**)$ seja válido. Se δ é uma estatística para a qual $E_\theta(\delta^2)<\infty$, a desigualdade da informação é válida.

Nota: A condição $(**)$ é válida na família exponencial unidimensional, ou seja, com

$$p_\eta(x) = \exp\{\eta\, T(x) - A(\eta)\}\, h(x).$$

Ex. Cauchy

$$p_\theta(x) = \frac{1}{\pi}\,\frac{1}{1+(x-\theta)^2} \quad \text{satisfaz } (**).$$

$$\left|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta\, p_\theta(x)}\right| = \left|\frac{1}{\Delta}\left(\frac{1+(x-\theta)^2}{1+(x-\Delta-\theta)^2}-1\right)\right|$$

$$= \left|\frac{1}{\Delta}\,\frac{1+(x-\theta)^2-1-(x-\Delta-\theta)^2}{1+(x-\Delta-\theta)^2}\right|$$

$$= \left|\frac{1}{\cancel{\Delta}}\,\frac{2\cancel{\Delta}(x-\theta)-\Delta^{\cancel{2}}}{1+(x-\Delta-\theta)^2}\right|$$

$$\leq 2\,\frac{|x-\Delta-\theta|}{1+(x-\Delta-\theta)^2} + \frac{|\Delta|}{1+(x-\Delta-\theta)^2}$$

$$\leq 2+\varepsilon$$

pois (i) $|\Delta|<\varepsilon$, o que garante que a 2ª parcela é $\leq$

(ii) $\dfrac{|y|}{1+y^2}\leq 1$, p/ $\forall y$ [faça $y = x-\Delta-\theta$].

Basta tomar $b_\theta(x) = 2+\varepsilon$.

O CASO MULTIPARAMÉTRICO

$\theta \in \Omega \subset \mathbb{R}^s$

Suposições:

(a) Ω é um retângulo aberto.

(b) As distribuições P_θ têm suporte comum, de tal forma que
$$A = \{x, \ p_\theta(x) > 0\}$$
é independente de θ.

(c) Para qualquer $x \in A$, $\theta \in \Omega$ e $i = 1, \dots, s$, as derivadas $\partial p_\theta(x)/\partial\theta_i$ existem e são finitas.

Matriz de Informação

$I(\theta) = \{I_{ij}(\theta)\}$ (matriz $s \times s$)

com $I_{ij}(\theta) = E_\theta\left[\frac{\partial}{\partial\theta_i} \log p_\theta(x) \frac{\partial}{\partial\theta_j} \log p_\theta(x)\right]$.

Se valem (a)-(c) (p. 113 E) e as derivadas com respeito a θ_i, $i = 1, \dots, s$ do lado esquerdo de (∘)
$$\int p_\theta(x)\, d\mu(x) = 1$$
podem ser obtidas diferenciando sob o sinal de integração, então

$$E_\theta\left[\frac{\partial}{\partial\theta_i} \log p_\theta(x)\right] = 0 \qquad (\bullet)$$

e

$$I_{ij}(\theta) = \operatorname{cov}_\theta\left[\frac{\partial}{\partial\theta_i} \log p_\theta(x), \frac{\partial}{\partial\theta_j} \log p_\theta(x)\right].$$

Em notação matricial: seja $\theta = (\theta_1, \dots, \theta_s)'$.

$$I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(x) \frac{\partial}{\partial\theta'} \log p_\theta(x)\right]$$
$$= \operatorname{Cov}_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(x)\right].$$

$\frac{\partial}{\partial\theta} \log p_\theta(x)$ é chamado de vetor escore.

Se valem (a) - (c) (p. 113E) e (*), e as segundas derivadas com respeito a θ de $\log p_\theta(x)$ existem para todo x e θ e as 2as derivadas com respeito a θ do lado esquerdo de (o) podem ser obtidas diferenciando sob o sinal de integração, então

$$\underline{I_{ij}(\theta)} = - E_\theta\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p_\theta(x)\right],$$

isto é,

$$\underline{I(\theta)} = - E_\theta\left[\frac{\partial^2}{\partial\theta\, \partial\theta'} \log p_\theta(x)\right].$$

O Teorema da p. 108 (aditividade da matriz de informação sob amostragem aleatória, i.e., caso i.i.d.) continua válido.

Alunos: enunciar o Teorema e o corolário da p. 108 e demonstrá-los.

REPARAMETRIZAÇÃO

$\theta_i = h_i(\xi_1, \dots, \xi_s)$, $i = 1, \dots, s$

Matriz de informação para $(\xi_1, \dots, \xi_s)'$: $I^*(\xi) = \{I^*_{ij}(\xi)\}$.

$I^*(\xi) = J I J'$, com $J = \dfrac{\partial \theta'}{\partial \xi}$

(ver TPE, p. 125).

Teorema. Suponha que X tem distribuição na família exponencial de dimensão s e seja $\tau = (\tau_1, \dots, \tau_s)$ com $\tau_i = E(T_i(X))$, para $i = 1, \dots, s$. Então:

$$I(\tau) = C^{-1}$$

onde C é a matriz de covariâncias de $T_1, \dots, T_n$.

Dem. Alunos.

Ex.: $N(\xi, \sigma^2)$

$$I(\xi, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix} \qquad \text{(alunos)}$$

Ver Tabela 6.1, p. 127, TPE (Gama e Beta: alunos).

Ex. Família de locação-escala (alunos: p. 126, TPE).

Teorema. DESIGUALDADE DA INFORMAÇÃO

Suponha que sejam válidas as condições (a)-(c) (p. 113E) e $I(\theta)$ seja positiva definida. Seja δ uma estatística (unidimensional) tq $E_\theta(\delta^2) < \infty$. Suponha ainda que, para $i=1,\dots,s$, $(\partial/\partial\theta_i)\, E_\theta(\delta)$ exista e possa ser obtida diferenciando sob o sinal de integração.

Então, $E_\theta\left[\frac{\partial}{\partial\theta_i} \log p_\theta(x)\right] = 0$ e

$$\mathrm{var}_\theta(\delta) \geq \alpha' I^{-1}(\theta)\, \alpha,$$

onde α' é ~~o~~ ~~[illegible]~~ vetor linha com i-ésimo elemento

$$\alpha_i = \frac{\partial}{\partial\theta_i} E_\theta[\delta(x)].$$

Dem. TPE, p 127.

Nota. Condições sobre $p_\theta(x)$ para a validade da desigualdade da informação: ver TPE, p. 127.

Nota: Dizemos que $\theta_1,\dots,\theta_s$ são ortogonais se $I(\theta)$ é uma matriz diagonal, i.e.

$$I(\theta) = \{I_{ij}(\theta)\}, \quad \text{com} \quad I_{ij}(\theta) = 0, \; i \neq j, \; \forall \theta \in \Omega.$$

Nesse caso, $I^{-1}(\theta) = \{I^{ij}(\theta)\}$ com

$$I^{ii}(\theta) = I_{ii}(\theta)^{-1} \quad \text{e} \quad I^{ij}(\theta) = 0, \; i \neq j.$$

$$\alpha' I^{-1}(\theta)\, \alpha = \sum_{i=1}^{s} \alpha_i^2 I_{ii}(\theta)^{-1}$$

Em particular, para um estimador δ_i, não viciado de θ_i, temos

$$\mathrm{var}_\theta(\delta_i) \geq I_{ii}(\theta)^{-1}. \qquad \alpha_j = \frac{\partial}{\partial\theta_j} E_\theta(\delta_i), \; j=1,\dots,s$$

Nota: Pode-se mostrar que, sob condições adequadas, o limite inferior dado na desigualdade da informação é atingido apenas na família exponencial (análogo ao Teorema 5.12, TPE).

Nota: Importância da Informação de Fisher no estudo das propriedades assintóticas dos EMV's: Cap 6, TPE (ver Teo 5.1, p. 463).