21/09 Aula 11

EQUIVARIÂNCIA: UMA INTRODUÇÃO

Já vimos que se não restringirmos a classe de possíveis estimadores de $g(\theta)$, o estimador $\delta(X) \equiv g(\theta_0)$ terá risco nulo se $\theta = \theta_0$ e não haverá um estimador que minimize o risco uniformemente.

Restrição utilizada: não viciosidade

Agora usaremos outro tipo de restrição: simetria.

FAMÍLIA DE LOCAÇÃO

Seja $X = (X_1, \ldots, X_n)$ tendo distribuição conjunta com densidade de probabilidade

$$f(x - \xi) = f(x_1 - \xi, \ldots, x_n - \xi), \quad -\infty < \xi < +\infty,$$

$x = (x_1, \ldots, x_n)$, f é conhecida e ξ é um parâmetro de locação desconhecido.

$\delta(X)$: estimador "satisfatório" para estimar ξ com função de perda $L(\xi, d)$.

Considere as transformações

$$X_i' = X_i + a$$

e

$$\xi' = \xi + a.$$

Pasta 05

Nº cópias 75

23,

8, 18, 24, 30

A densidade conjunta de $X' = (X_1', \ldots, X_n')$ pode ser escrita como

$$f(x' - \xi') = f(x_1' - \xi', \ldots, x_n' - \xi'),$$

$x' = (x_1', \ldots, x_n')$. Tomando $x_i' = x_i + a$ temos

$x_i' - \xi' = x_i - \xi$, $i = 1, \ldots, n$, e

$$f(x' - \xi') = f(x - \xi), \text{ para todo } x \text{ e } \xi.$$

Vamos admitir que a estimativa de ξ' seja

$$d' = d + a.$$

precisamos uma f. de perda assim

* A perda resultante é $L(\xi', d')$. É razoável escolher L que satisfaça $L(\xi', d') = L(\xi, d)$, $\forall a$ ou seja,

$$L(\xi + a, d + a) = L(\xi, d), \text{ para todo } a.$$

Uma fç L satisfaz esse condição sse depende apenas das diferenças $d - \xi$, i.e, é da forma $L(\xi, d) = \rho(d - \xi)$.

De fato,

$$L(\xi, d) = \rho(d - \xi) \Rightarrow L(\xi, d) = \rho((d + a) - (\xi + a)) = L(\xi + a, d + a)$$

Por outro lado, definindo $\rho(d - \xi) = L(0, d - \xi)$ e tomando $a = -\xi$ temos

$$L(\xi + a, d + a) = L(\xi, d) \Rightarrow L(0, d - \xi) = L(\xi, d)$$

$$\Rightarrow L(\xi, d) = \rho(d - \xi).$$

Definição. Uma família de densidades $f(x|\xi)$, com parâmetro ξ, e uma função de perda $L(\xi, d)$ são invariantes por locação, se, respectivamente,

$$f(x'|\xi') = f(x|\xi) \quad e \quad L(\xi, d) = L(\xi', d')$$

sempre que

$$x' = x + a, \quad \xi' = \xi + a \quad e \quad d' = d + a.$$

Se as densidades e a função de perda são invariantes por locação, o problema de estimar ξ é dito ser invariante por locação sob as transformações

$$X_i' = X_i + a, \qquad \xi' = \xi + a \quad e \quad d' = d + a. \qquad (\ast)$$

Invariância é uma expressão de simetria: imutabilidade sob certas transformações.

No problema original (~~...~~ "sem linhas"), suponha que decidimos usar o estimador $\delta(X)$ para ξ. É desejável que nós ~~...~~ usemos

$$\delta(X') = \delta(X_1 + a, \ldots, X_n + a) \quad \text{a conhecido}$$

para estimar $\xi' = \xi + a$. Por outro lado, é natural estimar $\xi + a$ por $\delta(X) + a$.

É desejável então que

$$\delta(X_1 + a, \ldots, X_n + a) = \delta(X) + a. \qquad (\star\star)$$

equivariante por locação

Definição. Um estimador satisfazendo $(\star\star)$ é chamado de equivariante sob as transformações $(\ast)$, ou equivariante por locação.

Exemplos. média e mediana amostrais, média ponderada das estatísticas de ordem (com pesos que somam 1), estimador de máxima verossimilhança (se $\hat{\xi}$ maximiza $f(x-\xi)$, $\hat{\xi} + a$ maximiza $f(x - (\xi + a))$).

Teorema. Seja $X = (X_1, \ldots, X_n)$ com densidade $f(x - \xi) = f(x_1 - \xi, \ldots, x_n - \xi)$, $-\infty < \xi < \infty$, e seja δ um estimador equivariante por locação com perda $L(\xi, d) = \rho(d - \xi)$. Então, o vício, o risco e a variância de δ são constantes (i.e, não dependem de ξ).

Dem.: Se X tem densidade $f(x)$ (i.e, $\xi = 0$), então $X + \xi$ tem densidade $f(x - \xi)$. Então, o vício pode ser escrito como

$$b(\xi) = E_\xi[\delta(X)] - \xi \overset{\text{famí. loca}}{=} E_0[\delta(X + \xi)] - \xi$$

$$\overset{\text{inv. por loc.}}{=} E_0[\delta(X) + \xi] - \xi = E_0(\delta(X))$$

que não depende de ξ.

Ex: risco e variância.

Definição. Em um problema de estimação invariante por locação, se existe um estimador equivariante por locação que minimiza o risco (que é constante), ele é chamado de estimador equivariante de risco mínimo (ERM, MRE).

Lema. Seja δ_0 um estimador equivariante por locação, então uma condição necessária e suficiente para δ ser equivariante por locação é que

$$\delta(x) = \delta_0(x) + u(x),$$

onde $u(x)$ é qualquer função que satisfaz.

$$u(x+a) = u(x), \text{ para todo } x, a.$$

Dem. ($\Rightarrow$) Se δ e u são como dados acima

$$\begin{aligned} \delta(x+a) &= \delta_0(x+a) + u(x+a) \\ &= \delta_0(x) + a + u(x) \\ &= \delta(x) + a. \end{aligned}$$

($\Leftarrow$) Se δ é equivariante, seja $u(x) = \delta(x) - \delta_0(x)$. Então,

$$\begin{aligned} u(x+a) &= \delta(x+a) - \delta_0(x+a) = \\ &= \delta(x) + a - \delta_0(x) - a = u(x). \end{aligned}$$

Logo, δ e u são como dados no lema.

→ [Totalidade de estimadores equivariantes por locação]

Lema (caracterização das funções u)
Uma função u satisfaz

$$u(x+a) = u(x), \text{ p/ todo } x, a.$$

sse é uma função das diferenças $y_i = x_i - x_n$, para $i = 1, \dots, n-1$, $n \geq 2$; para $n = 1$, sse é uma constante.

Dem. exercício.

OBS. $u(X)$, que é invariante por locação, é uma estatística ancilar (lista 1).

Teorema (caracterização dos estimadores equivariantes por locação)
Se δ_0 é qualquer estimador equivariante por locação, então uma condição necessária e suficiente para δ ser equivariante por locação é que exista uma função v de $n-1$ argumentos para a qual

$$\delta(x) = \delta_0(x) - v(y) \text{ para todo } x.$$

OBS. $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_{n-1})$, $y_i = x_i - x_n$, $i = 1, \dots$, se $n \geq 2$ e $y = c$, se $n = 1$ (c: constante).

Ex 1. $n = 1$.
Os únicos estimadores equivariantes por locação são $X + c$, para alguma constante c.

Teorema. (Estimador equivariante por locação de risco mínimo).

Seja $X = (X_1, \dots, X_n)$ tendo distribuição com densidade

$$f(x-\xi) = f(x_1-\xi, \dots, x_n-\xi), \quad -\infty < \xi < \infty,$$

seja

$$Y_i = X_i - X_n, \quad i = 1, \dots, n-1$$

e

$$Y = (Y_1, \dots, Y_{n-1}).$$

Suponha que a perda é dada por

$$L(\xi, d) = \rho(d-\xi)$$

e que exista um estimador equivariante δ_0 de ξ de risco finito.

Assuma que, para cada y exista um número $v(y) = v^*(y)$ que minimiza

$$E_0\{\rho[\delta_0(X) - v(y)] \mid y\}.$$

Então, um estimador equivariante por locação δ^* de ξ com risco mínimo existe e é dado por

$$\delta^*(X) = \delta_0(X) - v^*(Y).$$

Dem. O estimador ERM é encontrado determinando v que minimiza

$$R_\xi(\delta) = E_\xi[\rho(\delta_0(X) - v(Y) - \xi)].$$

Como o risco é independente de ξ, basta minimizar

$$R_0(\delta) = E_0[\rho(\delta_0(X) - v(Y))]$$

$$= \int E_0[\rho(\delta_0(X) - v(y)) \mid y] \, dP_0(y).$$

Esta integral é minimizada, minimizando o integrando para cada y, i.e, minimizando

$$E_0[\rho(\delta_0(X) - v(y)) \mid y] \quad \text{para cada } y.$$

Como $\delta_0(X)$ tem risco finito, tem-se

$$E_0[\rho(\delta_0(X) \overset{v(y)}{)} \mid y] < \infty \quad (q.c.\ P_0),$$

e o problema de minimização acima faz sentido. ∎

Corolário.

Sob as suposições do teorema, suponha que ρ seja convexa e não monótona. Então, um estimador ERM de ξ existe; ele é único se ρ é estritamente convexa.

Dem. Teorema + propriedades de fξs convexas (ver Teo 7.15, Cap 1, TPE, p. 50).

Teorema (Teo 7.15, Cap 1, TPE, p. 50)

Seja ρ uma função convexa definida sobre $(-\infty, +\infty)$ e X uma v.a. tal que $\phi(a) = E[\rho(X-a)]$ é finita para algum a.

Se ρ não é monótona, $\phi(a)$ atinge um valor mínimo e o conjunto de valores para os quais ϕ é mínimo é um intervalo fechado.

Se ρ é estritamente convexa, o valor mínimo de ϕ é único.

Ex: perda quadrática

Seja $\rho(t) = t^2$ e suponha que $E(X^2) < \infty$. Como ρ é estritamente convexa, segue que $\phi(a)$ tem um único valor mínimo. Se $E(X) = \mu$ $(<\infty)$, temos

$$\phi(a) = E[\rho(X-a)] = E[(X-a)^2]$$
$$= E[(X-\mu)^2] + (\mu - a)^2$$

que é mínimo sse $a = \mu$.

Ex. perda do valor absoluto

Seja $\rho(t) = |t|$ e suponha que $E(|X|) < \infty$. Aqui, ρ é convexa mas não estritamente convexa.

O conjunto S de valores de "a" que minimizam $\phi(a)$ é um intervalo fechado. Este é o conjunto das medianas de X (Problemas 1.7 e 1.8, TPE, Cap 1, p. 62).

Corolário. Sob as suposições do Teorema acima, suponha que ρ é uma função par e X é simétrica em torno de μ; $\rho(t) = \rho(-t)$.

Então $\phi(a)$ atinge o valor mínimo em $a = \mu$.

Ex 1. (cont) $n=1$.

Se X tem risco finito, o estimador ERM de ξ é $X - v^*$, onde v^* minimiza

$$E_0[\rho(X - v)]. \qquad (*)$$

Perda quadrática: $X - E_0(X)$ é ERM.
Perda do erro absoluto: $X - med_0(X)$ é ERM.

Alunos: ver Ex 7.17 e 7.18, Cap 1, TPE, p. 50.

Ainda, se X tem distribuição simétrica em torno de ξ e ρ é convexa e par $(\rho(-t) = \rho(t))$, então $(*)$ é minimizado tomando $v = 0$ e X é ERM.

Sob as mesmas suposições, se $n=2$, o estimador ERM é $(X_1 + X_2)/2$ (exercício).

Corolário (perda não convexa) (do Teo. da p. 120)

Sob as suposições do exemplo acima, suponha que $0 \le \rho(t) \le M$ para todo t, que $\rho(t) \to M$, qdo $t \to \pm\infty$ e que a densidade f de X é contínua (q.c.). Então, um estimador ERM de ξ existe.

Dem. Problema 1.8, p. 207, TPE, Cap 3.

Ex. Perda 0-1.

$$\rho(d-\xi) = \begin{cases} 1, & \text{se } |d-\xi| > k \\ 0, & \text{c.c.} \end{cases}$$

$$E_0[\rho(X-v)] = P_0[|X-v| > k] = 1 - P_0[|X-v| \le k]$$

Então, v que minimiza $E_0[\rho(X-v)]$, maximiza $P_0[|X-v| \le k]$.

Suponha que a dens. f é simétrica em torno de zero. Se f é unimodal, então $v=0$ e X é ERM.
Se f tem forma de U, digamos $f(x)=0$, para $|x| > c > k$ e estritamente crescente para $0 < x < c$.
Então há 2 valores de v que maximizam $P_0[|X-v| \le k]$: $v = c-k$ e $v = -c+k$ e, $\therefore$, $X-c+k$ e $X+c-k$ são ERM.

Ex. Normal.
$X_1, \ldots, X_n$: iid, $N(\xi, \sigma^2)$, $\sigma^2 > 0$ conhecido.
Seja $\delta_0 = \bar{X}$. Pelo Teo. de Basu, é indep. de $Y = (Y_1, \ldots, Y_{n-1})$, $Y_i = X_i - X_n$.
Então, v^* que minimiza

$$E_0[\rho(\delta_0(X) - v(y)) | y]$$

é o que minimiza

$$E_0[\rho(\bar{X} - v)].$$

Logo, $\bar{X}$ é ERM para toda função de perda $L(\xi, d) = \rho(d-\xi)$ tq ρ é convexa e par.
Também é MRE p/ a fç de perda do Ex acima.

Ex. Exponencial
$X_1, \ldots, X_n$: iid, $X_i \sim E(\xi, b)$, b conhecido
Tomamos
$$\delta_0 = X_{(1)},$$
que é independente de Y (T. Basu).
Então $v = v(y)$ é determinado minimizando
$$E_0[\rho(X_{(1)} - v)].$$

Perda quadrática: $v = E_0(X_{(1)}) = b/n$.
$X_{(1)} - b/n$ é ERM.

Outras perdas: TPE, p. 153, Ex 1.18.

Nota: Sob perda quadrática, o estimador ERM é
$$\delta^*(x) = \delta_0(X) - E_0[\delta_0(X) | Y] \qquad [\blacktriangle]$$
sendo $\delta_0(X)$ um estimador equivariante.
Este pode ser obtido mais explicitamente usando o teorema a seguir.

Teorema: Sob as suposições do Teorema da p. 120, com $L(\xi, d) = (d-\xi)^2$, o estimador [Δ] é dado por

$$\delta^*(x) = \frac{\int_{-\infty}^{+\infty} u\, f(x_1-u, \dots, x_n-u)\, du}{\int_{-\infty}^{+\infty} f(x_1-u, \dots, x_n-u)\, du}$$

ñ dep de ξ

e, nesta forma, é chamado de estimador de Pitman de ξ.

Dem.: Seja $\delta_0(X) = X_n$.
Para calcular $E_0(X_n | y)$, vamos fazer a seguinte mudança de variáveis:

$$y_i = x_i - x_n, \quad i = 1, \dots, n-1, \quad \text{e} \quad y_n = x_n.$$

O jacobiano da transformação é 1. A densidade conjunta dos Y's é

$$p_Y(y_1, \dots, y_n) = f(y_1 + y_n, \dots, y_{n-1} + y_n, y_n).$$

A densidade condicional de Y_n dado $Y = (y_1, \dots, y_{n-1})$ é

$$\frac{f(y_1 + y_n, \dots, y_{n-1} + y_n, y_n)}{\int_{-\infty}^{+\infty} f(y_1 + t, \dots, y_{n-1} + t, t)\, dt}.$$

Então,

$$E_0[X_n | y] = E_0[Y_n | y] = \frac{\int_{-\infty}^{+\infty} t\, f(y_1 + t, \dots, y_{n-1} + t, t)\, dt}{\int_{-\infty}^{+\infty} f(y_1 + t, \dots, y_{n-1} + t, t)\, dt}$$

$$= \frac{\int_{-\infty}^{+\infty} t\, f(x_1 - x_n + t, \dots, x_{n-1} - x_n + t, t)\, dt}{\int_{-\infty}^{+\infty} f(x_1 - x_n + t, \dots, x_{n-1} - x_n + t, t)\, dt}$$

δ^* é obtido fazendo a mudança de var. $u = x_n - t$.

Ex. Normal ver p. 124)

$X_1, \dots, X_n$: iid, $X_i \sim N(\xi, \sigma^2)$, σ^2 conhecido

$$f(x_1 - u, \dots, x_n - u) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{\sum (x_i - u)^2}{2\sigma^2} \right\}$$

$\sum (x_i + \bar{x} - \bar{x} - u)^2$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{(u-\bar{x})^2}{2\sigma^2/n} - \frac{\sum (x_i - \bar{x})^2}{2\sigma^2} \right\}$$

$$= \frac{1}{(2\pi\delta^2)^{1/2}} \exp\left\{ -\frac{(u-\bar{x})^2}{2\delta^2} \right\} h(\sigma, x, n),$$

com $h(\sigma, x, n) = \dfrac{n^{-1/2}}{(2\pi\sigma^2)^{(n-1)/2}} \exp\left\{ -\dfrac{\sum (x_i - \bar{x})^2}{2\sigma^2} \right\}$ e $\delta = \dfrac{\sigma}{\sqrt{n}}$.

Então,

$$f(x_1 - u, \dots, x_n - u) = p(u; \bar{x}, \delta)\, h(\sigma, x, n)$$

com $p(u; \bar{x}, \delta)$ sendo a densidade de uma v.a. com distr. $N(\bar{x}, \delta^2)$.

O estimador de Pitman $\delta^*(x)$ é dado por

$$\delta^*(x) = \frac{\int_{-\infty}^{+\infty} u\, p(u; \bar{x}, \delta)\, du}{\int_{-\infty}^{+\infty} p(u; \bar{x}, \delta)\, du} = \bar{x}.$$

$\delta^*(X) = \bar{X}$ (coincide c/ o resultado obtido na p.124)

Ex. Uniforme

$X_1, \ldots, X_n$: i.i.d, $U(\xi - b/2; \xi + b/2)$; b conhecido, $b>0$.

$$f(x_1-\xi, \ldots, x_n-\xi) = \begin{cases} \frac{1}{b^n}, & \text{se } \xi - \frac{b}{2} \le x_{(1)} \le x_{(n)} \le \xi + \frac{b}{2} \\ 0, & \text{c.c.} \end{cases}$$

O estimador de Pitman $\delta^*(x)$ é dado por

$$\delta^*(x) = \int_{x_{(n)}-b/2}^{x_{(1)}+b/2} u \, du \left[\int_{x_{(n)}-b/2}^{x_{(1)}+b/2} du \right]^{-1} = \frac{1}{2} [x_{(1)}+ \ldots$$

$$\delta^*(x) = \frac{X_{(1)} + X_{(n)}}{2}.$$

Obs: Em muitos casos as integrais são difíceis de calcular.

$$f_X(x/\xi) = f(x-\xi) = \prod_{i=1}^{n} \frac{1}{b} \mathbb{I}_{(\xi-\frac{b}{2}, \xi+\frac{b}{2})}(x_i)$$

$$= \frac{1}{b^n} \mathbb{I}_{(\xi-b/2, \infty)}(x_{(1)}) \, \mathbb{I}_{(-\infty, \xi+b/2)}(x_{(n)})$$

$$x_{(n)} \qquad x_{(1)}$$

$$f_X(x-w) = \frac{1}{b^n} \mathbb{I}_{(w-b/2, x_{(n)})}(x_{(1)}) \, \mathbb{I}_{(x_{(1)}, w+b/2)}(x_{(n)})$$

Notas:

1. Caso existe ENVVUM, ele tipicamente minimiza o risco p/ toda perda convexa, mas não para perdas limitadas (pode nem existir ENVVLM).

2. Tipicamente existe ERM mesmo p/ certas funções de perda não convexas.

3. O estimador ERM tipicamente varia dependendo da ff de perda.

4. ENVVUM são relevantes no contexto das famílias exponenciais enquanto que os ERM (num sentido mais amplo) são importantes em famílias de grupo de transformações.

Fam. l. expon. → posto completo
$\Downarrow$ $\qquad\qquad\qquad\qquad$ $\Downarrow$
T(.) é suf. → completa

Lema. Considere a função de perda quadrática. Seja $\delta(x)$ um estimador equivariante com vício constante b e risco finito.

(a) $\delta(x) - b$ é equivariante e tem risco menor do que $\delta(x)$.

(b) O único estimador ERM é não viciado.

(c) Se um ENVVUM existe e é equivariante, ele é ERM.

Dem.: (a) $\delta^*(x) = \delta(x) - b$ é equivariante pois

$$\delta^*(x+a) = \delta(x+a) - b = \delta(x) + a - b \quad (\text{pois } \delta \text{ é equiv.})$$
$$= \delta^*(x) + a.$$

δ^* é não viciado pois, p/ $\forall \xi$,

$$E_\xi(\delta^*(x)) = E_\xi(\delta(x)) - b = \xi \quad (\text{pois o vício de } \delta \text{ é } b$$

O risco de δ é dado por

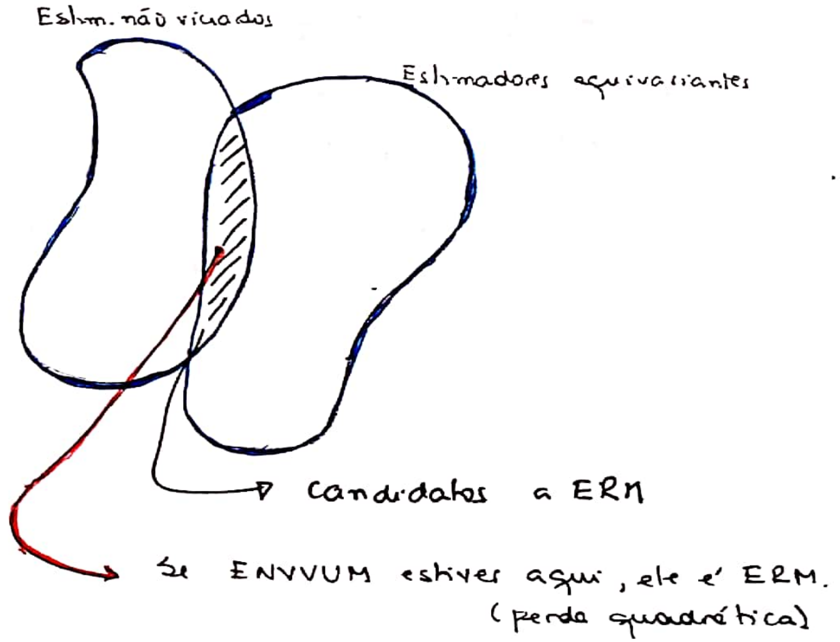
$$R(\xi, \delta) = E_\xi\left[(\delta(x) - \xi)^2\right] = E_\xi\left[((\delta^*(x) - \xi) + b)^2\right]$$
$$= E_\xi\left[(\delta^*(x) - \xi)^2\right] + 2bE_\xi\left[\delta^*(x) - \xi\right] + b^2$$
$$= R(\xi, \delta^*) + b^2, \quad \forall \xi$$

Como $R(\xi, \delta) < \infty$ e, admitindo que $b \neq 0$, temos

$$R(\xi, \delta^*) < R(\xi, \delta), \quad \forall \xi.$$

(b) Unicidade do estimador ERM: corolário (p.121) do Teorema da p.120.
Não viciosidade segue de (a).

(c) Consequência de (b). ∎



Nota: O conceito de equivariância pode ser estendido para além da família de locação: ver TPE, Secão 2, Cap. 3; em particular para a família de escala, locação - escala e modelos de regressão normais lineares.

Ver:

Azevedo, C.L.N.; Nobre, J.S. (2005). O princípio da equivariância: conceitos e aplicações. RT. MAE, 2005-19.

[publicado em 2006 na Revista Colombiana de Estadística]

OTIMALIDADE DE RISCO MÉDIO

Contexto: Estimação de $g(\theta)$ a valores reais
$\theta \in \Omega$
X: va com distribuição $P_\theta \in \mathcal{P}$
$\mathcal{P} = \{ P_\theta : \theta \in \Omega \}$

δ: estimador de $g(\theta)$; $\delta = \delta(X)$
$L(\theta, d)$: função de perda

$L(\theta, d) \geq 0$, para todo θ, d
$L(\theta, g(\theta)) = 0$, para todo θ

$L(\theta, d)$ mede as consequências (perda, custo) de se estimar $g(\theta)$ pelo valor $d = \delta(x)$: estimativa de $g(\theta)$ obtida da amostra x; $x \in \mathcal{X}$.

A perda média, segundo a distribuição de X e dado o valor de θ,

$$R(\theta, \delta) = E_\theta \{ L(\theta, \delta(X)) \}, \quad \theta \in \Omega,$$

é chamado de risco do estimador δ.

$$R(\theta, \delta) = \int L(\theta, \delta(x)) \, dP_\theta(x).$$

Já vimos que, em geral, não é possível encontrar um estimador δ que minimize o risco uniformemente em θ (i.e, p/ todo $\theta \in \Omega$).

Algumas possibilidades:

1. Restringir a classe dos possíveis estimadores de $g(\theta)$ à classe dos estimadores não viciados e procurar aquele que minimiza o risco para todo θ (ENVVUM; estimador não viciado de risco uniformemente mínimo).

2. Considerar aspectos de invariância das distribuição dos dados e da função de perda e restringir a classe dos possíveis estimadores a classe dos estimadores equivariantes. Sob condições adequadas, o risco dos estimadores equivariantes não depende de θ. Procurar o estimador equivariante que tem risco mínimo (ERM).

3. Procurar o estimador que tem risco médio mínimo. O risco médio é obtido usando uma função de peso (em θ). (Estimador de Bayes - otimalidade de risco médio).

OBS: função de peso não negativa.

Risco médio

$$r(\Lambda, \delta) = \int R(\theta, \delta)\, d\Lambda(\theta). \qquad [*]$$

Vamos assumir que Λ é uma distribuição de probabilidade e que, portanto,

$$\int d\Lambda(\theta) = 1.$$

$r(\Lambda, \delta)$ é chamado de risco de Bayes; depende do estimador δ e da distribuição de probabilidade Λ; NÃO depende de θ.

Um estimador δ^* que minimiza o risco de Bayes é chamado de estimador (solução) de Bayes com respeito a Λ.

Nota: Em [*] vamos considerar que o estimador δ tem risco $R(\theta, \delta)$ finito para todo θ.

O parâmetro θ pode ser visto como a realização de uma v.a. Θ com distribuição Δ (em lugar de ser visto como uma constante desconhecida).

Δ pode refletir a informação a priori sobre θ quer seja obtida por experimentos anteriores ou por conhecimento (crença) prévio do pesquisador.

Vamos denotar a densidade de X por $f(x|\theta)$ (densidade condicional de X dado $\Theta = \theta$) e a densidade a priori por $\pi(\theta)$, ou por $\pi(\theta|\lambda)$ caso esta dependa de um outro parâmetro (hiperparâmetro) λ.

Denotemos a densidade marginal de X por $g(x)$ (ou $g(x|\lambda)$).

$$g(x) = \int f(x|\theta)\, \overbrace{d\Delta(\theta)}^{\pi(\theta)\,d\theta}.$$

Então, a densidade condicional de θ, dado que $X = x$, denotada por $\pi(\theta|x)$ (ou $\pi(\theta|x,\lambda)$) é

$$\pi(\theta|x) = \frac{f(x|\theta)\,\pi(\theta)}{g(x)}$$

$$\propto f(x|\theta)\,\pi(\theta)$$

ou, caso a distribuição a priori dependa de um hiperparâmetro λ,

$$\pi(\theta|x,\lambda) = \frac{f(x|\theta)\,\pi(\theta|\lambda)}{g(x,\lambda)}$$

$$\propto f(x|\theta)\,\pi(\theta|\lambda).$$

Esta é chamada de densidade a posteriori de θ.

IDE'IA

Experiência prévia

- A priori (antes de observar $X=x$)
  Θ tem distribuição (a priori) Λ com densidade $\pi(\theta)$ que pode depender de outros parâmetros (hiperparâmetros).

Dados

- $X=x$ é observado

  $f(x|\theta)$: função de verossimilhança

Experiência atualizada

- A posteriori (após observar $X=x$)
  Dado $X=x$, Θ tem distribuição (a posteriori) condicional com densidade

$$\pi(\theta|x) = \frac{f(x|\theta)\,\pi(\theta)}{g(x)} \propto f(x|\theta)\,\pi(\theta)$$

$$= \text{verossimilhança} \times \text{priori}$$

Tudo que se conhece sobre θ, após $X=x$ ter sido observado, está na sua distribuição a posteriori.

Faz sentido buscar o estimador δ tal que o risco a posteriori

$$E\left\{ L(\Theta, \delta(x)) \mid X=x \right\}$$

seja mínimo.

Mas a idéia inicial era de procurar um estimador que minimizasse o risco de Bayes $r(\Lambda, \delta)$:

$$r(\Lambda, \delta) = \int R(\theta, \delta)\, d\Lambda(\theta) .$$

Pergunta: um estimador que minimiza o risco a posteriori é um estimador de Bayes (ou seja, minimiza $r(\Lambda, \delta)$)?

Vejamos (caso contínuo):

$d\Lambda = \pi(\theta)d\theta$

$\Lambda = \int \pi(\theta)$

$$
\begin{aligned}
\underline{r(\Lambda, \delta)} &= \int_{\Omega} R(\theta, \delta)\, \pi(\theta)\, d\theta \\
&= \int_{\Omega} \int_{\mathcal{X}} L(\theta, \delta(x))\, f(x|\theta)\, dx\, \pi(\theta)\, d\theta \\
&= \int_{\mathcal{X}} \int_{\Omega} L(\theta, \delta(x))\, \underline{\frac{f(x|\theta)\, \pi(\theta)}{q(x)}}\, d\theta\, \underline{q(x)}\, dx \\
&= \int_{\mathcal{X}} \int_{\Omega} L(\theta, \delta(x))\, \pi(\theta|x)\, d\theta\, q(x)\, dx \\
&= \int_{\mathcal{X}} E\{L(\theta, \delta(x)) \,|\, X = x\}\, q(x)\, dx
\end{aligned}
$$

$r(\Lambda, \delta)$ é minimizado minimizando o integrando $E\{L(\theta, \delta(x)) \,|\, X = x\}$ para cada x. Ou seja, o estimador que minimiza o risco a posteriori é um estimador de Bayes.

Temos, portanto, o seguinte teorema (p. 142).

Formalmente, o argumento utilizado na página anterior (troca da ordem das integrais) é justificado pelo **Teorema de Fubini** (TPE, p. 13, Teo 2.8) que afirma essencialmente que em uma integral repetida (dupla, por exemplo) de uma função não negativa (aqui, a função de perda) a ordem de integração é irrelevante.

Essa forma do teorema de Fubini (que exige que a função seja não negativa) é também conhecida como Teorema de Tonelli.

Com esse argumento, prova-se o seguinte teorema.

Teorema.
Suponha que Θ tem uma distribuição Λ e, dado $\Theta = \theta$, suponha que X tem distribuição P_θ. Suponha ainda que as suposições abaixo sejam válidas para o problema de estimar $g(\Theta)$ com função de perda $L(\theta, d)$, não negativa:

(a) existe um estimador δ_0 com risco finito;

(b) para quase todo x (i.e., exceto para todo x em um conjunto A tq $Q(A)=0$, sendo Q a distribuição marginal de X), existe um $\delta_\Lambda(x)$ que minimiza

$$E\{L(\Theta, \delta(x)) \mid X = x\}.$$

Então, $\delta_\Lambda(x)$ é um estimador de Bayes.

Aula 12

Corolário. Sob as condições do Teorema, se

$L(\theta, d) = w(\theta)[d - g(\theta)]^2$, $w(\theta) > 0$, p/ todo $\theta \in$

então

$$\delta_\Lambda(x) = \frac{E[w(\Theta)g(\Theta) \mid X = x]}{E[w(\Theta) \mid X = x]}.$$

Em particular, se $w(\theta) = 1$, para todo $\theta \in \Omega$, (perda quadrática)

$$\delta_\Lambda(x) = E[g(\Theta) \mid X = x].$$ (média a posteriori de $g(\Theta)$).

Dem. Para provar desenvolva

$$E[w(\Theta)[\delta(x) - g(\Theta)]^2 \mid X = x]$$

e obtenha um polinômio de 2º grau em $\delta(x)$.

Nota. Para outras funções de perda, ver TPE, Cor. 1.2, p. 228.

Perda do valor absoluto $\to$ estimador: mediana a posteriori.

Corolário. Se a função de perda é estritamente convexa em d (em particular quadrática), então uma solução de Bayes δ_Λ é única q.c.-$\mathcal{P}$, onde $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ desde que δ_Λ tenha risco de Bayes finito e q.c.-Q implique em q.c.-$\mathcal{P}$.

Dem. Segue do fato de que

$$E[L(\theta, d) \mid X = x]$$

é estritamente convexa em d para x fixo.

Ex1. Poisson

$X_1, \dots, X_n$: iid, $X_i \mid \theta \sim \text{Poisson}(\theta)$, $\theta > 0$.

[dado $\Theta = \theta$, $X_1, \dots, X_n$ são iid e têm distrib. Poisson(θ)].

$\Theta \sim \text{Gama}(a, b)$, $a, b > 0$

$$\pi(\theta) = \frac{1}{\Gamma(a)\, b^a} \theta^{a-1} e^{-\theta/b}, \quad \theta > 0.$$

$$L_k(\theta, \delta) = \frac{(\theta - \delta)^2}{\theta^k} \qquad \text{(função de perda)}$$

$$f(x \mid \theta) = \prod_{i=1}^{n} \frac{e^{-\theta} \theta^{x_i}}{x_i!}, \quad x_i \in \{0, 1, 2, \dots\} \text{ p/ } i = 1, \dots, n,$$

$$x = (x_1, \dots, x_n).$$

$$f(x \mid \theta) = \frac{e^{-n\theta}\, \theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

$$\pi(\theta \mid x) \propto f(x \mid \theta)\, \pi(\theta) \propto \theta^{a + n\bar{x} - 1} e^{-\theta\left(\frac{1 + nb}{b}\right)}$$

$$\theta \mid X = x \sim \text{Gama}\left(a + n\bar{x}, \frac{b}{1 + nb}\right)$$

$\Theta / X = x$

Estimador de Bayes de θ sob a perda L_k:

$$\delta^{(k)}(\bar{x}) = \frac{E[\Theta^{-k}\, \Theta \mid \bar{x}]}{E[\Theta^{-k} \mid \bar{x}]} = \frac{E[\Theta^{1-k} \mid \bar{x}]}{E[\Theta^{-k} \mid \bar{x}]}$$

Para $j+a>0$

$$E[\Theta^j \mid \bar{x}] = \int_0^{+\infty} \frac{1}{\Gamma(a+n\bar{x})} \left[\frac{1+nb}{b}\right]^{a+n\bar{x}} \theta^{j+a+n\bar{x}-1}$$

$$\times \exp\left\{-\theta \frac{1+nb}{b}\right\} d\theta$$

$$= \frac{1}{\Gamma(a+n\bar{x})} \Gamma(j+a+n\bar{x}) \left(\frac{1+nb}{b}\right)^{-j} \times$$

$$\times \int_0^{+\infty} \frac{1}{\Gamma(j+a+n\bar{x})} \left(\frac{1+nb}{b}\right)^{j+a+n\bar{x}} \times$$

$$\times \theta^{j+a+n\bar{x}-1} \exp\left\{-\theta \frac{1+nb}{b}\right\} d\theta$$

$$E[\Theta^j \mid \bar{x}] = \frac{\Gamma(j+a+n\bar{x})}{\Gamma(a+n\bar{x})} \left(\frac{b}{1+nb}\right)^j, \text{ p/ } j+a>0.$$

$$\frac{E[\theta^{1-k} \mid \bar{x}]}{E[\theta^{-k} \mid \bar{x}]} = \frac{\Gamma(1-k+a+n\bar{x})}{\Gamma(\cancel{a+n\bar{x}})} \left(\frac{b}{1+nb}\right)^{1-k} \times$$

$$\times \left\{ \frac{\Gamma(-k+a+n\bar{x})}{\Gamma(\cancel{a+n\bar{x}})} \left(\frac{b}{1+nb}\right)^{-k} \right\}^{-1}$$

$$= (n\bar{x}+a-k) \frac{b}{1+nb}, \text{ para } -k+a>0$$

▶ Estimador de Bayes sob perda L_k, $k<a$:

$$\delta^{(k)}(\bar{X}) = b \, \frac{n\bar{X}+a-k}{1+nb}.$$

Em particular, se $k=0$ (perda quadrática), o estimador de Bayes é

$$\delta^{(0)}(x) = \frac{b}{1+nb}(n\bar{X}+a) = \frac{1}{1+nb} \underbrace{ab}_{\substack{\text{média de } \theta \\ \text{(a priori)}}} + \frac{nb}{1+nb} \underbrace{\bar{X}}_{\text{EMV de } \theta}$$

pesos

▶ Média ponderada da média de θ a priori e o EMV de θ.

▶ n grande: $\delta^{(0)}(\bar{X}) \approx \bar{X}$ (EMV).

▶ Ex2. Binomial

$X \mid p \sim b(p, n)$

$p \sim \text{Beta}(a, b)$, $\quad a, b > 0$

$$\pi(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1.$$

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

$$\pi(p \mid x) \propto p^{x+a-1}(1-p)^{n-x+b-1}$$

▶ $p \mid x \sim \text{Beta}(a', b')$, $\quad a' = a + x$
$\quad b' = b + n - x$

Estimador de Bayes de p sob perda quadrática:

▶ $\delta(x) = E(p \mid X) = \dfrac{a+X}{a+b+n}$,

pois se $Y \sim \text{Beta}(a', b')$, $E(Y) = a'/(a'+b')$.

Note que

$$\delta(x) = \frac{a+b}{a+b+n} \underbrace{\frac{a}{a+b}}_{\substack{\text{média de } p \\ \text{a priori}}} + \frac{n}{a+b+n} \underbrace{\frac{X}{n}}_{\text{EMV}}$$

pesos.

▶ Ex3 Normal

▶ Dado θ, $X_1, \dots, X_n$ são iid, $X_i \sim N(\theta, \sigma^2)$, $\sigma^2 > 0$ conhecido

$\Theta \sim N(\mu, b^2)$; $\quad x = (x_1, \dots, x_n)$

$$\pi(\theta \mid x) \propto f(x \mid \theta)\, \pi(\theta)$$

$$\propto \exp\left[-\frac{1}{2\sigma^2} \sum (x_i - \theta)^2\right] \exp\left[-\frac{1}{2b^2}(\theta - \mu)^2\right]$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{b^2}\right)\left(\theta - \frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}\right)^2\right\}$$

▶ $\Theta \mid X = x \sim N\left(\dfrac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}, \dfrac{1}{n/\sigma^2 + 1/b^2}\right)$.

Estimador de Bayes sob perda quadrática:

▶ $\delta(\bar{X}) = \dfrac{n/\sigma^2}{n/\sigma^2 + 1/b^2} \underset{\text{EMV}}{\bar{X}} + \dfrac{1/b^2}{n/\sigma^2 + 1/b^2} \underset{\substack{\text{média de } \theta \\ \text{a priori}}}{\mu}$

pesos

▶ Nota: Voltemos aos exemplos

- Ex 1: $\Theta \sim Gama(a,b)$ $\Rightarrow$ $\theta | x \sim Gama(a', b')$
  $X_1, \dots, X_n | \theta \sim Poisson(\theta)$, iid

- Ex 2: $\Theta \sim Beta(a,b)$ $\Rightarrow$ $\Theta | x \sim Beta(a', b')$
  $X | \theta \sim b(\theta, n)$

- Ex 3: $\Theta \sim N(\mu, b^2)$ $\Rightarrow$ $\Theta | x \sim N(\mu', b'^2)$
  $X_1, \dots, X_n | \theta \sim N(\theta, \sigma^2)$, σ^2 conh, iid

▶ DISTRIBUIÇÕES A PRIORI CONJUGADAS:
"priori e posteriori na mesma família de distribuições".

▶ Teorema. Qualquer estimador de Bayes que é único é admissível. [se quaisquer 2 estimadores de Bayes são iguais qc - P].

▶ Dem. Se δ é o único estimador de Bayes com respeito a uma distribuição a priori Λ e é dominado por δ', então

$$R(\theta, \delta') \leq R(\theta, \delta) \quad \text{para todo } \theta \in \Omega$$

e

$$\int R(\theta, \delta') \, d\Lambda(\theta) \leq \int R(\theta, \delta) \, d\Lambda(\theta).$$

Logo, δ' é um estimador de Bayes. Isso contraria a hipótese de unicidade.

▶ Exemplos: Poisson (p. 147), Binomial (p. 148).

▶ Ex. Normal (p. 149)

▶ $$\delta(\bar{X}) = \frac{n/\sigma^2}{n/\sigma^2 + 1/b^2} \bar{X} + \frac{1/b^2}{n/\sigma^2 + 1/b^2} \mu.$$

Como μ e b^2 podem ser quaisquer ($\mu \in \mathbb{R}$, $b^2 > 0$), temos que todos os estimadores da forma $a\bar{X} + c$, com $0 < a < 1$ e $c \in \mathbb{R}$, são admissíveis.

▶ Nota: Admissibilidade de $a\bar{X} + c$ com "a" fora do intervalo $(0,1)$ é discutida em TPE, Ex 2.5, 2.7, 2.8 e Teo 2.6, p. 323-326.
$\bar{X}$ é admissível.

▶ ESTIMADOR DE BAYES E SUFICIÊNCIA

$X = (X_1, \dots, X_n)$ com densidade $f(x \mid \theta)$; $x = (x_1, \dots, x_n)$

Verossimilhança: $L(\theta \mid x) = f(x \mid \theta)$

Estatística suficiente: $T = T(X)$

Se é observado $T = t$, então

$$f(x \mid \theta) = L(\theta \mid x) = g(t \mid \theta)\, h(x)$$

onde $h(.)$ não depende de θ. Para uma densidade $\pi(\theta)$ qualquer, a densidade a posteriori é

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\, \pi(\theta)}{\int f(x \mid \theta')\, \pi(\theta')\, d\theta'} = \frac{L(\theta \mid x)\, \pi(\theta)}{\int L(\theta' \mid x)\, \pi(\theta')\, d\theta'} =$$

$$= \frac{g(t \mid \theta)\, \pi(\theta)}{\int g(t \mid \theta')\, \pi(\theta')\, d\theta'}.$$

▶ A distribuição a posteriori só depende de x através de t (tanto faz obter a dist. a posteriori a partir de x ou de t). Depende também da função de verossimilhança e da priori. Nada mais. ◀

▶ Todas as "quantidades" Bayesianas são funções de uma estatística suficiente minimal. ◀

▶ Ex 3. Normal (cont)

$$\bar{X} \mid \theta \sim N(\theta, \sigma^2/n)$$

$$\pi(\theta \mid \bar{x}) \propto \exp\left\{ -\frac{n}{2\sigma^2} (\bar{x} - \theta)^2 \right\} \exp\left\{ -\frac{1}{2b^2} (\theta - \mu)^2 \right\}$$

[verificar que é a mesma distribuição a posteriori obtida anteriormente].

▶ PRIORIS IMPRÓPRIAS

Λ: medida tal que $\int d\Lambda(\theta) = +\infty$.

▶ Ex 2. Binomial (cont) - estimação de p com perda quadrática

Se $a=b=0$, a densidade a priori de θ é

$\pi(\theta) \propto p^{-1}(1-p)^{-1}$ ▶ (Beta (0,0) !) ◀

e

$$\int_0^1 p^{-1}(1-p)^{-1}\,dp = \int_0^1 \left(\frac{1}{p} + \frac{1}{1-p}\right) dp =$$

$$= \log \frac{p}{1-p} \Big|_0^1 = +\infty$$

A densidade a posteriori é

$\pi(\theta|x) \propto p^{x-1}(1-p)^{n-x-1}$.

Se $x = 1, \ldots, n-1$, essa é uma densidade a posteriori própria ▶(Beta $(x, n-x)$) ◀

E se $x=0$ ou $x=n$?

Se $x=0$, a densidade a posteriori é proporcional a

$$\frac{(1-p)^{n-1}}{p} = \frac{1}{p}\sum_{k=0}^{n-1}\binom{n-1}{k} 1^k (-p)^{n-1-k}$$

$$= \frac{1}{p}\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{n-1-k}\,p^{n-1-k}$$

$$= \sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{n-1-k}\,p^{n-2-k}$$

$$= \sum_{k=0}^{n-2}\binom{n-1}{k}(-1)^{n-1-k}\,p^{n-2-k} + \frac{1}{p}.$$

A integral de cada parcela do somatório em $(0,1)$ é finita e a da parcela $1/p$ é infinita.

$$\int_0^1 \frac{(1-p)^{n-1}}{p}\,dp = K + \int_0^1 \frac{1}{p}\,dp = K + \log p\Big|_0^1 = +\infty,$$

onde K é um nº real (finito).

❗ Alunos: fazer o mesmo para $x=n$ (escrever $q = 1-p$ e $p^{n-1}/(1-p) = (1-q)^{n-1}/q$).

Tomando um estimador qualquer tal que $\delta(0)=0$ e $\delta(n)=1$, temos:

$$E\{ L(p, \delta(x)) \mid X=x\} =$$

$$= E\{ (p-\delta(x))^2 \mid X=x\}$$

$$= K(x) \int_0^1 (p-\delta(x))^2 \, p^{x-1}(1-p)^{n-x-1} dp$$

$< \infty$ (Lista 4)

e esta esperança é minimizada em $\boxed{\delta(x)=\frac{x}{n}}$

(note que $\delta(0)=0$ e $\delta(n)=1$).

Faz sentido tomar X/n como um estimador de Bayes.

Nota: se $\delta(0)\neq 0$ ou $\delta(n)\neq 1$, a esperança explode em $x=0$ ou $x=n$.

Definição.

Um estimador δ_Λ é um estimador de Bayes generalizado com respeito a uma medida Λ (mesmo se não for uma medida de probabilidade) se a perda esperada a posteriori

$$E\{ L(\Theta, \delta(x)) \mid X=x\}$$

é minimizada em $\delta(x)=\delta_\Lambda(x)$ para todo x.

Ex. Normal

$X_1,\dots,X_n \mid \theta$ são iid, com distr $N(\theta,\sigma^2)$

$\Theta \sim$ uniforme na reta (imprópria, "não informativa").

$\Theta \mid x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$ (própria)

$\bar{X}$: Estimador de Bayes Generalizado (perda quadrática)

Definição: Um estimador δ é um limite de estimadores de Bayes se existe uma seqüência de distribuições a priori Λ_ν e estimadores de Bayes δ_{Λ_ν} tais que

$$\delta_{\Lambda_\nu}(x) \longrightarrow \delta(x) \quad \text{para quase todo } x$$

(com respeito à densidade $f(x|\theta)$) quando $\nu \to +\infty$.

▶ Ex Normal (cont.) ; $\theta \sim N(\mu, b^2)$

$$\lim_{b \to +\infty} \frac{n/\sigma^2}{n/\sigma^2 + 1/b^2}\,\bar{x} + \frac{1/b^2}{n/\sigma^2 + 1/b^2}\,\mu = \bar{x}.$$

▶ Ex. Binomial (cont.)

$$\lim_{\substack{a \to 0 \\ b \to 0}} \frac{a+b}{a+b+n}\,\frac{a}{a+b} + \frac{n}{a+b+n}\,\frac{x}{n} = \frac{x}{n}.$$

▶ Nota: Priori não informativa.

Intuitivamente: $\pi(\theta) \propto$ constante

Problemas: • não é invariante sob reparametrização

$\phi = \phi(\theta)$: tranf 1-a-1.

$$\pi^*(\phi) = \pi(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right| \text{ (não constante)}$$

• pode ser imprópria.

Priori de Jeffreys: $\pi(\theta) \propto \sqrt{I(\theta)}$, $\theta \in \Omega$ (θ escalar)

E' invariante sob transformações 1-a-1.

$$\phi^* = \phi(\theta) \Rightarrow \pi^*(\phi) \propto \sqrt{I^*(\phi)}$$

28/09/2015 - Aula 13

▶ Priori de Jeffreys

Idéia: fornecer o mínimo possível de informação a priori sobre θ relativamente à informação da amostra

Justificativa: envolve informação de Kullback-Leibler e de Shannon.

▶ Informação de Kullback-Leibler para discriminação entre duas densidades f e g:

$$\bullet\; K[f,g] = \int \log\left[\frac{f(t)}{g(t)}\right] f(t)\,dt. \bullet$$

▶ Interpretação: qto maior $K[f,g]$, mais fácil é discriminar entre f e g.

Prioris que têm pouco efeito sobre $\pi(\theta|x)$ produzem $K[\pi(\theta|x), \pi(\theta)]$ grande; os dados têm muita influência sobre $\pi(\theta|x)$.

▶ Problema: $K[\pi(\theta|x), \pi(\theta)]$ depende de x.

▶ Informação de Shannon

$$\bullet\; S(\pi) = \int K[\pi(\theta|x), \pi(\theta)]\, q_\pi(x)\,dx; \bullet$$

• $q_\pi(x) = \int f(x|\theta)\,\pi(\theta)\,d\theta$ é a densidade marginal de x.

▶ Teo (TPE, p. 260)

Se $(X_1, \ldots, X_n)$ é uma a.a. de $f(x|\theta)$ e $S_n(\pi)$ é a informação de Shannon para a amostra. Então, qdo $n \to \infty$

$$S_n(\pi) = \frac{1}{2} \log \frac{n}{2\pi e} + \int \pi(\theta) \log \frac{I(\theta)^{1/2}}{\pi(\theta)} d\theta + o(1);$$

$I(\theta)$: informação de Fisher.

A segunda parcela é a única que envolve $\pi(\theta)$.
A idéia é maximizar esta parcela, ou seja,

▶ $E\left[\log \dfrac{I(\theta)^{1/2}}{\pi(\theta)} \right]$

em que a esperança é calculada sob a distr. a priori.

Mas, para qualquer v.a. Y não negativa e tq $E[Y] < \infty$, temos $E[\log Y] \leq \log E[Y]$, com igualdade apenas se $Y = c$ com probab. 1.
[desigualdade de Jensen, Cor. 1.7.6, p. 47, TPE).

Isto sugere tomarmos $\boxed{\pi(\theta) \propto I(\theta)^{1/2}}$, desde que $\int I(\theta)^{1/2} d\theta$ seja finita.

Priori de Jeffreys

▶ Ex: Família de locação

Dado $\Theta = \theta$, X tem densidade $f(x-\theta)$

Já vimos que a informação de Fisher não depende de θ (é constante)

Priori de Jeffreys: $\pi(\theta) \propto$ constante

• (se o espaço paramétrico é ilimitado, a priori é imprópria).

• Nota: Se θ é vetorial, a priori de Jeffreys é

$$\pi(\theta) \propto \sqrt{|I(\theta)|}.$$

BAYES HIERÁRQUICO

$X \mid \theta \sim f(x \mid \theta)$
$\theta \mid \gamma \sim \pi(\theta \mid \gamma)$
$\gamma \sim \psi(\gamma)$

Ex. Normal

$X_1, \dots, X_n \mid \theta$, independentes, com distr $N(\theta, \sigma^2)$,
σ^2 conhecido

$\Theta \mid \tau \sim N(0, \tau^2)$

$\frac{1}{\tau^2} \sim$ Gama (a,b), a,b conhecidos.

O estimador de Bayes sob perda quadrática é

$$E\{\Theta \mid x\} = \iint \theta \, \pi(\theta \mid x, \tau^2) \, d\theta \, \pi(\tau^2 \mid x) \, d\tau^2$$

$$= \int \frac{n \tau^2 \bar{x}}{n\tau^2 + \sigma^2} \, \pi(\tau^2 \mid x) \, d\tau^2$$

$$= E\left[E(\Theta \mid x, \tau^2) \right]$$

Ver problema 5.3, TPE, p 289.

Não há forma explícita para $E(\Theta \mid x)$.

Estimação Minimax

O estimador de Bayes minimiza o risco médio

$$r(\Lambda, \delta) = \int R(\theta, \delta) \, d\Lambda(\theta).$$
↳ priori

O estimador minimax minimiza o risco máximo.

Estimador Minimax

Definição. Um estimador δ^M de θ, que minimiza o risco máximo, ou seja,

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\theta} R(\theta, \delta^M),$$

é chamado de estimador minimax.

Problema: difícil obter estimador minimax em generalidade (caso a caso).

• Ex. Binomial

$X \sim b(p, n)$; perda quadrática
$p \sim Beta(a, b)$.

Estimador de Bayes: $\delta_1(x) = \dfrac{a+x}{a+b+n}$.

Considere um conjunto de 3 estimadores:

δ^{π_1}: est. de Bayes com priori Beta(1,3) $[E(p) = 1/4]$

δ^{π_2}: " " " " " Beta(2,2) $[E(p) = 1/2]$

δ^{π_3}: " " " " " Beta(3,1) $[E(p) = 3/4]$

Aqui

$$R(p, \delta_1) = E_p\left[\left(\frac{a+X}{a+b+n} - p\right)^2\right]$$

$$= Var_p\left[\frac{a+X}{a+b+n}\right] + \left\{E_p\left[\frac{a+X}{a+b+n} - p\right]\right\}^2$$

$$= \frac{np(1-p)}{(a+b+n)^2} + \left(\frac{a+np}{a+b+n} - p\right)^2$$

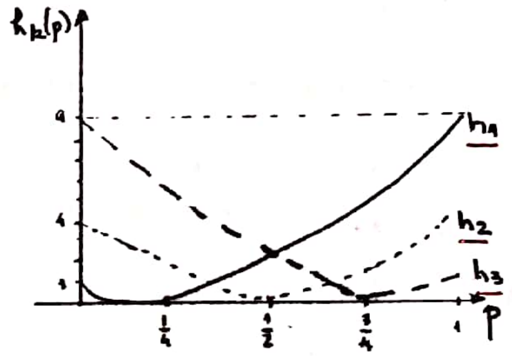
$$= \frac{1}{(a+b+n)^2}\left[np(1-p) + (a-(a+b)p)^2\right]$$

• $R(p, \delta^{\pi_1}) = \dfrac{1}{(4+n)^2}\left[np(1-p) + h_1(p)\right]$, $h_1(p) = (1-4p)^2$

• $R(p, \delta^{\pi_2}) = \dfrac{1}{(4+n)^2}\left[np(1-p) + h_2(p)\right]$, $h_2(p) = (2-4p)^2$

• $R(p, \delta^{\pi_3}) = \dfrac{1}{(4+n)^2}\left[np(1-p) + h_3(p)\right]$, $h_3(p) = (3-4p)^2$

$h_k(p) = (k-4p)^2$, $k = 1, 2, 3$.



Funções de risco não comparáveis.

• As funções de risco não são comparáveis pois se cruzam.

δ^{π_1} é preferível se (se acredita que) p está próximo de 1/4 (média da priori);
δ^{π_2}, se p está próximo de 1/2 e δ^{π_3}, se p está próximo de 3/4.

δ^{π_2} tem menor risco máximo: minimax entre os 3 estimadores, mas não entre todos [VERIFICAR]

O estimador minimax é o estimador de Bayes para a "pior" distribuição a priori possível (priori menos favorável).

Priori menos favorável

Definição. Uma distribuição a priori Λ é **menos favorável** se

- $r_\Lambda \geq r_{\Lambda'}$,

[Λ que produz o estimador Bayes com maior risco de Bayes]

para todas as distribuições a priori Λ'.

Aqui,

Est. Bayes seg. priori Λ.

- $r_\Lambda = r(\Lambda, \delta_\Lambda) = \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta)$

(risco médio, risco de Bayes do estimador de Bayes).

Relação entre estimador Bayes e Minimax

Teorema. Suponha que Λ é uma distribuição sobre Ω tal que

$$r(\Lambda, \delta_\Lambda) = \sup_\theta R(\theta, \delta_\Lambda), \quad (*)$$

onde δ_Λ é um estimador de Bayes com respeito a Λ. Então

(i) δ_Λ é minimax;

(ii) se δ_Λ é a única solução de Bayes com respeito a Λ, ele é o único estimador (procedimento) minimax;

(iii) Λ é menos favorável.

- **Dem.** Seja δ um outro estimador qualquer. Então

(i) $\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta)\, d\Lambda(\theta) \geq$

$$\geq \int R(\theta, \delta_\Lambda)\, d\Lambda(\theta) = \sup_\theta R(\theta, \delta_\Lambda).$$

(ii) idem, substituindo $\geq$ por $>$ em ( ).

(iii) Seja Λ' qualquer outra distribuição sobre Ω. Então,

mudei o estimador (δ_Λ ñ nec. o do Bayes)

$$r_{\Lambda'} = \int R(\theta, \delta_{\Lambda'})\, d\Lambda'(\theta) \leq \int R(\theta, \delta_\Lambda)\, d\Lambda'(\theta)$$

$$\leq \sup_\theta R(\theta, \delta_\Lambda) = r_\Lambda.$$

Risco médio é menor que sup. do risco

Nota: <*> significa que o risco médio iguala-se a seu máximo.
Quando a função de risco é constante, isso ocorre.

Se δ_Λ tem risco cte, é minimax.

▶ Corolário. Se a solução de Bayes δ_Λ tem risco constante, ela é minimax.

Dem. Se δ_Λ tem risco constante, vale <*>.

▶ Corolário. Seja

▶ $$w_\Lambda = \left\{ \theta \in \Omega : R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda) \right\}.$$

Então δ_Λ é minimax se $\Lambda(w_\Lambda) = 1$.

Em outras palavras, uma condição suficiente para δ_Λ ser minimax é que exista um conjunto w tq $\Lambda(w)=1$ e $R(\theta, \delta_\Lambda)$ atinja seu máximo em todo ponto de w.

Ex. Binomial (continuação) Procura da priori Λ tq o est. m. Bayes tenha risco cte

Tentemos uma priori Λ Beta(a,b). Vimos que

$$R(p, \delta_\Lambda) = \frac{1}{(a+b+n)^2} \left\{ np(1-p) + \left(a - (a+b)p\right)^2 \right\}.$$

Procuremos (a,b) tal que $R(p, \delta_\Lambda)$ seja constante em p.

$$np(1-p) + (a-(a+b)p)^2 = np - np^2 + a^2 - 2a(a+b)p + (a+b)^2 p^2$$
$$= p\left[n - 2a(a+b)\right] + p^2\left[(a+b)^2 - n\right] + a^2$$

é constante em p se

$$n = (a+b)^2 \quad \text{e} \quad n = 2a(a+b)$$

Como a e b são positivos temos $a+b = \sqrt{n}$ e $\therefore$ $a = b = \sqrt{n}/2$.

Então, o estimador

$$\delta = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} = \frac{X}{n} \frac{\sqrt{n}}{1+\sqrt{n}} + \frac{1}{2} \frac{1}{1+\sqrt{n}}$$

é de Bayes (com respeito à priori $B(\sqrt{n}/2, \sqrt{n}/2)$) e tem risco constante. Logo, é minimax. Como este estimador de Bayes é único, ele é o único estimador minimax para p.

Este estimador é viciado (X/n é o único não viciado que é função de X).

▸ Viés de δ:

$$E_p(\delta) - p = p\,\frac{\sqrt{n}}{1+\sqrt{n}} + \frac{1}{2}\,\frac{1}{1+\sqrt{n}} - p \xrightarrow[n\to\infty]{} 0 .$$

▸ Risco de δ:

$$r_n = E_p\left[(\delta - p)^2\right] = \frac{1}{4}\,\frac{1}{(1+\sqrt{n})^2} \qquad \text{(mostrar)}$$

Risco de X/n: $\quad R_n(p) = \dfrac{p(1-p)}{n}$.

$$r_n < R_n(p) \quad \text{para} \quad p \in I_n = \left(\frac{1}{2} - c_n;\ \frac{1}{2} + c_n\right),$$

onde $c_n \to 0$ qdo $n \to +\infty$ (obter c_n, prob. 1.11, p 399, TPE)

$$r_n \not< R_n(p) \quad p/ \quad p \notin I_n.$$

▸ Alunos: refazer usando a perda $L(p,d) = \dfrac{(d-p)^2}{p(1-p)}$.

Exercício.

Aqui, X/n tem risco constante e é um estimador de Bayes com respeito à distribuição a priori $U(0,1)$. É ∴ minimax.

Nota: Uma priori menos favorável não é necessariamente única.

Linearidade do Risco

▸ Lema. Seja δ um estimador de Bayes (respectivamente, NVVUM, minimax, admissível) de $g(\theta)$ com perda quadrática. Então, $a\delta + b$ é estimador de Bayes (respectivamente, NVVUM, minimax, admissível) de $ag(\theta)+b$.

Dem. Vem do fato de que

$$R\,(ag(\theta)+b,\ a\delta+b) = E\left[(a\delta+b - (ag(\theta)+b))^2\right]$$

$$= E\{[\,a(\delta - g(\theta))]^2\} = a^2\, R\,(g(\theta), \delta).$$

$R(\theta,\delta)$

Pode ñ existir priori menos favorável (Talvez exista, mas imprópria)

Nota: Para usar o Teorema (p.164) é necessário que exista uma priori menos favorável. Mas pode não existir; ela pode não ser própria.

▸ Definição: Uma seqüência de distribuições a priori $\{\Lambda_n\}$ é menos favorável se para cada distribuição a priori Λ temos

$$r_\Lambda \le r = \lim_{n\to+\infty} r_{\Lambda_n},$$

onde $\quad r_{\Lambda_n} = \int R(\theta, \delta_n)\, d\Lambda_n(\theta)$

é o risco de Bayes do estimador de Bayes δ_n sob Λ_n.

Teorema Suponha que $\{\Lambda_n\}$ seja uma seqüência de distribuições a priori com riscos de Bayes r_{Λ_n} satisfazendo

$$\lim_{n\to\infty} r_{\Lambda_n} = r$$

e que δ seja um estimador para o qual

$$\sup_\theta R(\theta,\delta) = r.$$

Então,

(i) δ é minimax;

(ii) a seqüência $\{\Lambda_n\}$ é menos favorável.

Dem.:

(i) Suponha que δ' seja qualquer outro estimador. Então

$$\sup_\theta R(\theta,\delta') \geq \int R(\theta,\delta')\, d\Lambda_n(\theta) \geq r_{\Lambda_n}$$

e isso vale para todo n. Então

$$\sup_\theta R(\theta,\delta') \geq \lim_{n\to\infty} r_{\Lambda_n} = r = \sup_\theta R(\theta,\delta)$$

ou seja, $\sup_\theta R(\theta,\delta') \geq \sup_\theta R(\theta,\delta)$ e δ é minimax.

(ii) Se Λ é qualquer distribuição sobre Ω, então

$$r_\Lambda = \int R(\theta,\delta_\Lambda)\, d\Lambda(\theta) \leq \int R(\theta,\delta)\, d\Lambda(\theta) \leq$$

$$\leq \sup_\theta R(\theta,\delta) = r.$$

Lema. Se δ_Λ é o estimador de Bayes de $g(\theta)$ com respeito a Λ e se

$$r_\Lambda = E\{[\delta_\Lambda(X) - g(\Theta)]^2\}$$

é o risco de Bayes, então.

$$r_\Lambda = \int var[g(\Theta) \mid x]\, dQ(x).$$

Em particular, se a variância a posteriori de $g(\Theta)$ (dado x) é independente de x, então

$$r_\Lambda = var[g(\Theta) \mid x].$$

obs: Q é a distr. marginal de X

Dem.

$$r_\Lambda = E\{[\delta_\Lambda(X) - g(\Theta)]^2\}$$

$$= \int E\{[\delta_\Lambda(x) - g(\Theta)]^2 \mid x\}\, dQ(x).$$

Como

$$\delta_\Lambda(x) = E[g(\Theta) \mid x]$$

então

$$r_\Lambda = \int var[g(\Theta) \mid x]\, dQ(x).$$

▶ Ex. Média da distr. normal; perda quadrática

$X = (X_1, \ldots, X_n)$, $x = (x_1, \ldots, x_n)$

Dado θ, $X_1, \ldots, X_n$ são iid, $X_i \sim N(\theta, \sigma^2)$, σ^2 conhecido.

θ: estimando.

• Mostremos que $\bar{X}$ é minimax.

Prioris: $N(\mu, b_k^2)$ (Λ_{b_k}); $b_k \xrightarrow[k \to +\infty]{} +\infty$

$$\delta_{\Lambda_{b_k}}(x) = \frac{n\bar{x}/\sigma^2 + \mu/b_k^2}{n/\sigma^2 + 1/b_k^2}$$

(est. de Bayes sob Λ_{b_k} e perda quadrática)

Risco de Bayes de $\delta_{\Lambda_{b_k}}$:

• $r_{\Lambda_{b_k}} = \dfrac{1}{n/\sigma^2 + 1/b_k^2}$ (variância a posteriori de θ; indep. de x)

▶ $r_{\Lambda_{b_k}} \nearrow \dfrac{\sigma^2}{n} = r$, qdo $k \to \infty$

e

▶ $r = \sup_\theta R(\theta, \bar{X})$

pois $R(\theta, \bar{X}) = \text{var}_\theta(\bar{X}) = \dfrac{\sigma^2}{n}$ (é independente de θ).

Então $\bar{X}$ é minimax.

Considere agora σ^2 desconhecido; $\sigma^2 > 0$.

O risco máximo de qualquer estimador δ é maior ou igual ao risco de $\bar{X}$ ($\forall \sigma^2$) pois $\bar{X}$ é minimax quando σ^2 é conhecido, e tem risco constante em θ.

Então

$$\sup_\theta R((\theta, \sigma^2), \delta) \geq R((\theta, \sigma^2), \bar{X}) = \frac{\sigma^2}{n}, \ \forall \sigma^2$$

e

$$\sup_{(\theta, \sigma^2)} R((\theta, \sigma^2), \delta) = +\infty,$$

a menos que σ^2 seja limitada.

Assuma que $\sigma^2 \leq M$ $(M > 0)$. Então

$$\sup_{(\theta, \sigma^2)} R((\theta, \sigma^2), \bar{X}) = \frac{M}{n}.$$

Para mostrar que $\bar{X}$ é minimax, usamos o lema abaixo.

Lema. Seja X v.a. com distribuição $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Omega\}$ e seja $g(\theta)$ o estimando. Suponha que δ é um estimador minimax de $g(\theta)$ quando θ está restrito a um subconjunto Ω_0 de Ω. Então, se

$$\sup_{\theta \in \Omega_0} R(\theta, \delta) = \sup_{\theta \in \Omega} R(\theta, \delta),$$

δ é minimax também quando se permite que θ varie em Ω.

Dem. Suponha que existe δ' com

$$\sup_{\theta \in \Omega} R(\theta, \delta') < \sup_{\theta \in \Omega} R(\theta, \delta),$$

ou seja, que δ não é minimax em Ω. Então,

$$\sup_{\theta \in \Omega_0} R(\theta, \delta') \leq \sup_{\theta \in \Omega} R(\theta, \delta') < \sup_{\theta \in \Omega} R(\theta, \delta) =$$

$$= \sup_{\theta \in \Omega_0} R(\theta, \delta)$$

e, portanto, δ não é minimax em Ω_0 (contradição).

Ex. normal (cont)

$$\Omega = \{(\theta, \sigma^2) : \theta \in \mathbb{R},\ 0 < \sigma^2 \leq M\}$$

$$\Omega_0 = \{(\theta, \sigma^2) : \theta \in \mathbb{R},\ \sigma^2 = M\}$$

$$\sup_{(\theta,\sigma^2) \in \Omega_0} R(\theta, \bar{X}) = \frac{M}{n} = \sup_{(\theta,\sigma^2) \in \Omega} R(\theta, \bar{X})$$

e $\bar{X}$ é minimax em Ω_0

Pelo Lema, $\bar{X}$ é estimador minimax qdo θ e σ^2 são desconhecidos e $\sigma^2 \leq M$.

▶ Ex: Normal, 2 amostras

$X_1, \dots, X_n$: aa de $X \sim N(\mu, \sigma^2)$

$Y_1, \dots, Y_n$: aa de $Y \sim N(\mu, \rho\sigma^2)$

amostras independentes

$\mu \in \mathbb{R}$, $\sigma^2 > 0$, $\rho > 0$ desconhecidos

▪ Problema: estimação de μ sob perda $\dfrac{(\hat{\mu}-\mu)^2}{\sigma^2 \max(1,\rho)}$.

▶ Mostremos que $\hat{\mu} = \dfrac{\bar{X}+\bar{Y}}{2}$ é minimax.

Seja $\theta = (\mu, \sigma^2, \rho)$. O espaço param. é:

$\Omega = \{\theta = (\mu, \sigma^2, \rho) : \mu \in \mathbb{R}, \sigma^2 > 0, \rho > 0\}$.

Fixemos $\sigma^2 = 1$, $\rho = 1$. O espaço param. reduzido é

$\Omega_0 = \{\theta = (\mu, \sigma^2, \rho) : \mu \in \mathbb{R}, \sigma^2 = 1, \rho = 1\}$.

Pelo ex. da p. 171, $\hat{\mu}$ é minimax em Ω_0.

$$R(\theta, \hat{\mu}) = \frac{Var(\hat{\mu})}{\sigma^2 \max(1,\rho)}$$

$$= \frac{\sigma^2 (1+\rho)}{4n\sigma^2 \max(1,\rho)} = \begin{cases} (1+\rho)/(4n\rho), & \rho \geq 1 \\ (1+\rho)/(4n), & \rho \in (0,1) \end{cases}$$

$$\sup_{\theta \in \Omega_0} R(\theta, \hat{\mu}) = \frac{1}{2n}$$

$$\sup_{\theta \in \Omega} R(\theta, \hat{\mu}) = \frac{1}{2n}.$$

Pelo Lema da p. 173, $\hat{\mu}$ é minimax em Ω.

Método dos Momentos

X: v.a. com distribuição na família
$\mathcal{P} = \{ P_\theta,\ \theta \in \Omega \subset \mathbb{R}^k \}$,
$E[|X|^k] < \infty$.

$X_1, \dots, X_n$: a.a. da distribuição de X.

j-ésimo momento amostral: $\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j$.

j-ésimo momento populacional: $\mu_j = E[X^j]$.

Note que $\hat{\mu}_j$ é um estimador não viciado de μ_j, $j = 1, \dots, k$.

Tipicamente,

$$\mu_j = h_j(\theta), \quad j = 1, \dots, k,$$

para certas funções h_j definidas sobre $\mathbb{R}^k$.

O estimador do métodos dos momentos $(\hat{\theta})$ de θ é obtido igualando μ_j a $\hat{\mu}_j$, isto é,

$$\hat{\mu}_j = h_j(\hat{\theta}), \quad j = 1, \dots, k.$$

Sejam $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_k)$ e $h = (h_1, \dots, h_k)$. Então

$$\hat{\mu} = h(\hat{\theta}).$$

Se a função inversa h^{-1} existe, então

$$\hat{\theta} = h^{-1}(\hat{\mu}).$$

Quando h não é um-a-um, qualquer solução de $\hat{\mu} = h(\hat{\theta})$ é um EMM (estimador do método dos momentos) de θ.

Se possível, escolhe-se uma solução $\hat{\theta}$ no espaço paramétrico Ω.

Em alguns casos EMM não existe.

Ex. $X_1, \dots, X_n$: v.a's iid com distr. P_θ,
$\theta = (\mu, \sigma^2)$,
$E[X_1] = \mu \in \mathbb{R}$, $\quad Var[X_1] = \sigma^2 \in (0, +\infty)$.

(a) Normal

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

(b) Exponencial Dupla

$$f(x;\theta) = \frac{1}{\sqrt{2}\,\sigma} \exp\left\{ -\frac{\sqrt{2}\,|x-\mu|}{\sigma} \right\}$$

(c) Logística

$$f(x;\theta) = \frac{1}{\gamma} \exp\left\{ -\frac{x-\mu}{\gamma} \right\} \left[1 + \exp\left\{ -\frac{x-\mu}{\gamma} \right\} \right]^{-2}$$

$$\gamma^2 = \frac{3\sigma^2}{\pi^2}$$

Como $\mu_1 = E[X_1] = \mu$ e $\mu_2 = E[X_1^2] = Var[X_1] + (E[X_1])^2 = \sigma^2 + \mu^2$,

o EMM de θ é solução de

$$\bar{X} = \mu \quad \text{e} \quad \frac{\sum X_i^2}{n} = \sigma^2 + \mu^2 ,$$

ou seja,

$$\hat{\theta} = \left(\bar{X}, \ \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right) = (\hat{\mu}, \hat{\sigma}^2) .$$

$\hat{\mu} = \bar{X}$: é estimador não viciado de μ. [redacted]

$\hat{\sigma}^2$: não é est. não viciado de σ^2.

Normal: $\hat{\theta}$ é estatística suficiente

Exponencial Dupla e Logística: $\hat{\theta}$ não é suficiente.

Suponha que μ é conhecido e $\mu = 0$. A 1ª equação ($\mu = \bar{X}$) não envolve parâmetro desconhecido.

Usando a 2ª equação, $\hat{\mu}_2 = \mu_2 = \sigma^2$, obtém-se

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 .$$

Normal: $\hat{\sigma}^2$ é estatística suficiente (ENVVUM)

Exponencial Dupla: $\hat{\sigma}^2$ não é estatística suficiente

Consideremos a transformação: $Y_i = |X_i|$

O EMM de σ^2 baseado em $Y_1, \dots, Y_n$ é

$$\tilde{\sigma}^2 = \left(\frac{1}{n} \sum_{i=1}^{n} |X_i| \right)^2 \qquad \text{(exercício)}$$

e é o ENVVUM de σ^2.

Ex. $X_1, \dots, X_n$: v.a's iid com distribuição $U(\theta_1, \theta_2)$, $-\infty < \theta_1 < \theta_2 < \infty$.

$$\mu_1 = E[X_1] = (\theta_1 + \theta_2)/2$$

$$\mu_2 = E[X_1^2] = \frac{\theta_1^2 + \theta_2^2 + \theta_1 \theta_2}{3}$$

EMM de $\theta = (\theta_1, \theta_2)$ é (exercício)

$$\hat{\theta}_1 = \bar{X} - \sqrt{\frac{3(n-1)\ S^2}{n}}$$

$$\hat{\theta}_2 = \bar{X} + \sqrt{\frac{3(n-1)\ S^2}{n}} ; \qquad S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} .$$

$\hat{\theta}$ não é função da est. suficiente completa $(X_{(1)}, X_{(n)})$.

Ex. Gama

$X_1, \dots, X_n$: v.a's iid com fdp

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

$$a, b > 0.$$

$\mu_1 = E[X_1] = ab$; $Var[X_1] = ab^2$.

EMM:

$$\begin{cases} \bar{X} = \mu_1 \\ \frac{\Sigma X_i^2}{n} = \mu_2 \end{cases} \Leftrightarrow \begin{cases} \bar{X} = \mu_1 \\ \hat{\sigma}^2 = \mu_2 - \mu_1^2 \end{cases}; \quad \hat{\sigma}^2 = \frac{\Sigma X_i^2}{n} - \bar{X}^2 = \frac{\Sigma (X_i - \bar{X})^2}{n}$$

$$\Leftrightarrow \begin{cases} \bar{X} = \mu_1 \\ \hat{\sigma}^2 = Var[X_1] \end{cases} \Leftrightarrow \begin{cases} \bar{X} = ab \\ \hat{\sigma}^2 = ab^2 \end{cases} \Leftrightarrow a = \frac{\bar{X}^2}{\hat{\sigma}^2}, \; b = \frac{\hat{\sigma}^2}{\bar{X}}.$$

Nota: Os EMM são comumente utilizadas como "chute inicial" para métodos numéricos de obtenção de EMV's.

Ex. Família não-paramétrica.

Estimação dos momentos centrais

$$c_j = E\{[X_1 - \mu_1]^j\}, \quad j = 2, \dots, k.$$

Os EMM's de $c_2, \dots, c_k$ são obtidos igualando μ_j a $\hat{\mu}_j$, $j = 1, \dots, k$.

Note que:

$$c_j = E\{(X_1 - \mu_1)^j\} = \sum_{t=0}^{j} \binom{j}{t} (-\mu_1)^t \mu_{j-t}$$

e

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t (X_i)^{j-t}$$

$$= \frac{1}{n} \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \sum_{i=1}^{n} X_i^{j-t}$$

$$= \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \hat{\mu}_{j-t}.$$

O EMM de c_j é

$$\hat{c}_j = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j, \quad j = 2, \dots, k.$$

Comentário: admissibilidade de $\bar{X}$ como estimador de $E[X]$

$X_1, \ldots, X_n$: a.a. de X

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

ii) $X \sim$ Bernoulli (θ)

iii) $X \sim$ Poisson (θ)

Nos 3 casos, $\bar{X}$ é EMV, ENVVUM, EMM
(normal: sob perda quadrática - ERM, minimax)

Pergunta: $\bar{X}$ é admissível?

Consideremos estimadores da forma $a\bar{X}+b$

Já vimos que qualquer estimador de Bayes que é único é admissível (p. 151).

Isso possibita mostrar, por exemplo, que, sob perda quadrática, se $0 < a < 1$:

i) Normal - $a\bar{X}+b$ é admissível, $\forall b \in \mathbb{R}$;

ii) Bernoulli - $a\bar{X}+b$ é admissível, $\forall\ 0<b<1$;

iii) Poisson - $a\bar{X}+b$ é admissível, $\forall\ b>0$.

Teorema. Seja Y uma v.a. com média θ e variância σ^2. Então $aY+b$ é um estimador inadmissível de θ sob perda quadrática quando

(i) $a>1$ ou

(ii) $a<0$ ou

(iii) $a=1$ e $b \neq 0$.

Prova: O risco de $aY+b$ é

$$\rho(a,b) = E[(aY+b-\theta)^2]$$

$$= E[\{a(Y-\theta) + [(a-1)\theta + b]\}^2]$$

$$= a^2\sigma^2 + [(a-1)\theta + b]^2$$

(i) Se $a>1$, então
$$\rho(a,b) \geq a^2\sigma^2 > \sigma^2 = \rho(1,0) \quad (\text{risco de } Y)$$

(ii) Se $a<0$, então $(a-1)^2 > 1$ e
$$\rho(a,b) \geq (a-1)^2\left[\theta + \frac{b}{a-1}\right]^2 > \left[\theta + \frac{b}{a-1}\right]^2$$
$$= \rho\left(0, -\frac{b}{a-1}\right)$$

(risco de $\delta(X) \equiv -b/(a-1)$.)

(iii) $a=1$ e $b \neq 0$

$$\rho(1,b) = \sigma^2 + b^2 > \sigma^2 \quad (\text{risco de } Y).$$

Para mostrar que $\bar{X}$ e' admissível nos 3 casos, pode-se usar o Corolário 2.18 de TPE (conseqüência do Teorema de Karlin (1958)) que fornece condição suficiente para admissibilidade de T como estimador de E[T] com perda quadrática na família exponencial unidimensional com fdp

$$f(x;\eta) = \exp\{\eta T(x) - A(\eta)\}\, h(x) \qquad (*)$$

Fato: Se o espaço paramétrico natural de (*) e' a reta real ($\mathbb{R}$) então T e' um estimador admissível de E[T] sob perda quadrática.

Normal : $\eta = \frac{\theta}{\sigma^2} \in \mathbb{R}$

Bernoulli : $\eta = \log \frac{\theta}{1-\theta} \in \mathbb{R}$

Poisson : $\eta = \log \theta \in \mathbb{R}$.

Prova direta da admissibilidade de $\bar{X}$ no caso normal : TPE, p. 324-326, Ex 2.8.