MAE 5834 - ESTATÍSTICA AVANÇADA I

PROFa. SILVIA FERRARI

INTRODUÇÃO (CAP1, TPE)

O PROBLEMA

Dados (observações): assume-se que são valores observados de variáveis aleatórias (v.a.'s) que seguem uma distribuição de probabilidades conjunta P pertencente a uma classe $\mathcal{P}$.

Frequentemente, P é indexada por um parâmetro $\theta \in \Omega$.

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

Exemplos.

1. Tempos de vida de lâmpadas
   Admite-se que $T \sim Exp(\theta)$ (exponencial de média θ).
   Observam-se $T_1, \ldots, T_n$ (réplicas de T)

   $T_1, \ldots, T_n$: v.a.'s i.i.d. ; $T_i \sim Exp(\theta)$, $\theta > 0$.

Pasta 05

Nº cópias 79

2. $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$

   Observam-se $x_1, \ldots, x_n$ (réplicas de x)
   $X_1, \ldots, X_n$: v.a.'s i.i.d., $X_i \sim N(\theta, 1)$.

3. Proporção de óleo bruto que se transforma em gasolina após processamento (X)

   $X \sim Beta(p, q)$, $\underset{\sim}{\theta} \in \{(p,q) \in \mathbb{R}^2 : p > 0 \text{ e } q > 0\} = \Omega$

   Observam-se $x_1, \ldots, x_n$: réplicas de X.
   $X_1, \ldots, X_n$: v.a.'s i.i.d., $X_i \sim Beta(p,q)$

4. $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ (modelo de regressão normal linear)

   $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)'$
   $\underset{\sim}{X}$: matriz $n \times p$ (de posto p) fixa
   $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$

   $\underset{\sim}{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$: vetor de v.a.'s ind.,
   $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma^2 > 0$

   $\underset{\sim}{\theta} = (\beta_1, \ldots, \beta_p, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^+$.

5. Regressão Logística

$\underset{\sim}{Y} = (Y_1, \ldots, Y_n)'$ tq $Y_i \sim \text{Bernoulli}(\mu_i)$, ind.

$$\log \frac{\mu_i}{1-\mu_i} = \underset{\sim}{x}_i' \underset{\sim}{\beta}, \quad i = 1, \ldots, n.$$

$\underset{\sim}{x}_i = (x_{i1}, \ldots, x_{ip})'$ fixados

$\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p = \Omega.$

OBJETIVO: fazer inferências sobre θ ou $g(\theta)$ com base nos dados observados.

- Estimação pontual e intervalar
- Testes de hipóteses - problema de decisão
- Previsão de novas observações.

ANÁLISE BAYESIANA: assume-se que θ é uma v.a. (não observável) com uma distribuição conhecida (distribuição a priori). Esta é modificada após a observação dos dados e é determinada então a distribuição a posteriori (distribuição condicional de θ dadas as observações).

Estimação Pontual

Ingredientes:

1. Uma função g definida sobre Ω (espaço paramétrico), a valores reais, cujo valor em θ, $g(\theta)$, será estimado.

$g(\theta)$: estimando

2. Uma v.a. (ou vetor aleatório) X, tomando valores em um espaço amostral $\mathcal{X}$ de acordo com a distribuição $P_\theta \in \mathcal{P}$.

O valor observado, x, de X constitui o conjunto de dados.

Objetivo: obter um bom estimador para $g(\theta)$

Um estimador é uma função δ, a valores reais, definida sobre $\mathcal{X}$. É usado para estimar um estimando $g(\theta)$.

$\delta(X)$: estimador de $g(\theta)$
$\delta(x)$: estimativa de $g(\theta)$

Obs: pode-se definir estimador como acima mas impondo que δ tome valores apenas no conjunto dos possíveis valores de $g(\theta)$.

E' desejável que δ "esteja perto" de $g(\theta)$.

Medidas de proximidade ou distância entre δ e $g(\theta)$:

- $P(|\delta(x) - g(\theta)| < c)$, para algum $c>0$;
- $E\{|\delta(x) - g(\theta)|^p\}$, para algum $p>0$;
- $E\left\{\left|\frac{\delta(x)}{g(\theta)} - 1\right|^p\right\}$ (se δ e g só tomam valores positivos), para algum $p>0$;
- $\lambda(\theta)\, E\{|\delta(x) - g(\theta)|^p\}$.

Suponha que as conseqüências (perda) de se estimar $g(\theta)$ por um valor d de δ são medidas por $L(\theta, d)$ (função de perda). Assumimos que

$$L(\theta, d) \geq 0, \quad \text{para todo } \theta, d$$

e

$$L(\theta, g(\theta)) = 0, \quad \text{para todo } \theta.$$

Função de risco (perda média)

$$R(\theta, \delta) = E_\theta\{L(\theta, \delta(x))\}$$

Um critério para escolha do estimador: tomar δ que minimize o risco para todo $\theta \in \Omega$.

Problema: não existe tal estimador.

Seja $\delta(x) = g(\theta_0)$, $\forall x \in \mathcal{X}$. Este tem risco nulo para $\theta = \theta_0$. Não existe um estimador com risco uniformemente menor ou igual aos demais.

Solução: restringir a classe dos estimadores.

Classe dos estimadores não viciados: requer que

$$\underbrace{E_\theta[\delta(X)] - g(\theta)}_{\text{vício (viés) do estimador}} = 0, \quad \text{para todo } \theta \in \Omega.$$

Classe dos estimadores não viciados com respeito à **mediana**: requer que

$$\text{Mediana}_\theta[\delta(X)] - g(\theta) = 0, \quad \text{para todo } \theta \in \Omega.$$

Outros critérios para escolha de estimadores

1. Escolher δ que minimiza

$$\int_\Omega R(\theta, \delta)\, w(\theta)\, d\theta$$

para alguma função de peso w. Se w é a fdp de θ, este é o estimador de Bayes.

2. Escolher δ que minimiza

$$\sup_\Omega R(\theta, \delta)$$

(estimador minimax).

MODELO ESTATÍSTICO

Para a descrição de um modelo estatístico seria natural tomar como ponto de partida um espaço mensurável $(\mathcal{E}, \mathcal{B})$ e uma família de distribuições de probabilidade definida sobre ele. Aqui $\mathcal{E}$ é o conjunto dos possíveis resultados de um experimento e $\mathcal{B}$ é uma σ-álgebra de sub-conjuntos de $\mathcal{E}$.

$(\mathcal{E}, \mathcal{B})$, sendo $\mathcal{B}$ uma σ-álgebra de sub-conjuntos de $\mathcal{E}$, é chamado de "espaço mensurável" e os conjuntos $B \in \mathcal{B}$ são "mensuráveis".

No entanto, toma-se como ponto de partida uma v.a. X (ou vetor, $X = (X_1, \dots, X_n)$) e todos os outros aspectos do experimento são ignorados.

A especificação do modelo começa com X (os dados), um espaço mensurável $(\mathcal{X}, \mathcal{Q})$, sendo $\mathcal{X}$ o conjunto dos possíveis valores de X e $\mathcal{Q}$ uma σ-álgebra de sub-conjuntos de $\mathcal{X}$, e uma família $\mathcal{P}$ de distribuições de probabilidades à qual a distribuição de X (desconhecida) pertence.

MODELO ESTATÍSTICO

$$(\mathcal{X}, \mathcal{Q}, \mathcal{P})$$

onde

$\mathcal{X}$: conjunto

$\mathcal{Q}$: σ-álgebra de sub-conjuntos de $\mathcal{X}$

$\mathcal{P}$: família de medidas (distribuições) de probabilidade.

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

Ω : espaço paramétrico

Suposição usual (identificabilidade): para quaisquer $\theta_1 \in \Omega$ e $\theta_2 \in \Omega$,

$$\theta_1 \neq \theta_2 \implies P_{\theta_1} \neq P_{\theta_2}.$$

Dizemos que $(\mathcal{X}, \mathcal{Q})$ é o espaço amostral. Muitas vezes, $\mathcal{Q}$ é omitido (sub-entendido) e dizemos que $\mathcal{X}$ é o espaço amostral.

Alunos: ler Seções 3 e 4, Cap 1 (TPE) sobre Teoria da Medida e Integração e Teoria da Pobabilidade.

Vamos trabalhar no contexto em que $\mathcal{P}$ é dominada por uma medida μ, σ-finita; em geral, a medida de contagem ou a medida de Lebesgue

Nota: μ é σ-finita se $\exists A_1, A_2, \ldots \in \mathcal{C}$, tais que

$$\cup A_i = \mathcal{X} \text{ e } \mu(A_i) < \infty.$$

$\mathcal{P}$ é dominada por μ se todos os elementos de $\mathcal{P}$ são absolutamente contínuos com respeito a μ, i.e., para cada $A \in \mathcal{C}$,

$$\mu(A) = 0 \Rightarrow P_\theta(A) = 0, \quad \forall \theta \in \Omega.$$

Também,

$$P_\theta(A) = \int_A p_\theta \, d\mu, \text{ para todo } \theta \in \Omega$$

(integral de Lebesgue),

sendo p_θ a derivada de Radon-Nikodym de P_θ com respeito a μ.

p_θ é chamada de densidade de probabilidade de P_θ com respeito a μ. Escrevemos

$$p_\theta = \frac{dP_\theta}{d\mu}.$$

Abordaremos os seguintes casos:

i) μ é a medida de contagem

$\mathcal{X}$ é contável; $\mathcal{C}$ é a classe de todos os subconjuntos de $\mathcal{X}$ e para cada $A \in \mathcal{C}$, $\mu(A)$ é o nº de pontos de A se A é finito e $\mu(A) = +\infty$, caso contrário.

Aqui,

$$P_\theta(A) = \int_A p_\theta \, d\mu = \sum p_\theta(x_i),$$

onde a soma varre todos os x_i's pertencentes a A.

Caso discreto: $\mathcal{X}$ e $\mathcal{C}$ como acima e $\mathcal{P}$ dominada pela medida de contagem.

ii) μ é a medida de Lebesgue

$\mathcal{X}$ é o espaço Euclidiano n-dimensional $\mathbb{R}^n$; $\mathcal{C}$ é a menor σ-álgebra contendo todos os retângulos abertos

$$A = \{(x_1, \ldots, x_n) : a_i < x_i < b_i\}, \quad -\infty < a_i < b_i < \infty.$$

Os conjuntos de $\mathcal{C}$ são chamados Borelianos.

Nota: Dizemos que "$\mathcal{X}$ (ou $(\mathcal{X}, \mathcal{C})$) é Euclidiano" para indicar que $\mathcal{X}$ é um Boreliano e assumimos que $\mathcal{C}$ é a classe dos sub-conjuntos Borelianos de $\mathcal{X}$.

Nota: $\mathcal{C}$ como definido acima contém, entre outras, todos os subconjuntos abertos e fechados de $\mathcal{X}$. É a menor σ-álgebra com essa propriedade.

Aqui,

$\mu(A) = (b_1 - a_1) \cdots (b_n - a_n)$ (volume) - Medida de Lebesgue

e

$\int p_\theta \, d\mu$ existe sempre que a integral de

Riemann existe e as duas coincidem (mas pode existir sem que a de Riemann exista).

Caso absolutamente contínuo: $(\mathcal{X}, \mathcal{C})$ é Euclidiano e $\mathcal{P}$ é dominada pela medida de Lebesgue.

Nota: Em ambos os casos (discreto e contínuo) p_θ é chamada de "densidade de probabilidade".

Notação para densidades:

$p_\theta(x)$, $p(x|\theta)$, $f_\theta(x)$, $f(x)$, $f(x|\theta)$, $p(x;\theta)$, $f(x;\theta)$.

Nota: $(\mathcal{X}, \mathcal{C}, \mu)$ é um espaço de medida.

Para definir um modelo estatístico, substituímos μ por $\mathcal{P}$, uma família de medidas de probabilidade. Admitimos que $\mathcal{P}$ é dominada por uma medida μ σ-finita.

Suporte. O suporte de uma distribuição P sobre $(\mathcal{X}, \mathcal{C})$ é o conjunto de todos os pontos x para os quais $P(A) > 0$ para todo retângulo aberto A que contém x.

Inferências sobre θ são feitas com base em estatísticas, que são transformações dos dados.

Nos exemplos 1 e 2, é razoável utilizar a média amostral para fazer inferências sobre θ.

Mais especificamente, variáveis (ou vetores) aleatórias, T digamos, que são transformações mensuráveis de $(\mathcal{X}, \mathcal{C})$ em algum $(\mathcal{T}, \mathcal{B})$ são chamadas estatísticas.

$\mathcal{B}$ é uma σ-álgebra de sub-conjuntos de $\mathcal{T}$ tq a imagem inversa

$$T^{-1}(B) = \{x \in \mathcal{X} : T(x) \in B\}$$

está em $\mathcal{C}$.

Em particular, estimadores são estatísticas.

Notas.

1. O foco do curso está em questões estatísticas. Alguns aspectos técnicos (ex.: verificação de mensurabilidade de transformações) não serão abordados.

2. Sempre que for mencionada uma medida, μ digamos, estará sub-entendido que é σ-finita, mesmo se não mencionado.

FAMÍLIAS DE MODELOS

FAMÍLIAS DE GRUPO

Uma "família de grupo" de distribuições é uma família obtida submetendo uma v.a. com uma dada distribuição a um grupo de transformações.

Uma classe de transformações 1-a-1 ($\mathcal{J}$) é um "grupo de transformações" se é fechada sob composição e inversão.

Fechada sob composição

$$g_1 \in \mathcal{J},\ g_2 \in \mathcal{J} \Rightarrow g_2 \cdot g_1 \in \mathcal{J}$$

Fechada sob inversão

$$g \in \mathcal{J} \Rightarrow g^{-1} \in \mathcal{J}.$$

(ler Seção 1.4, TPE)

(a) Família de locação

Seja U uma v.a. com distribuição F.
Sejam $a \in (-\infty, +\infty)$ e $X = U + a$.

$$P(X \leq x) = F(x-a).$$

O conjunto das distribuições $F(x-a)$, F fixo, com a variando de $-\infty$ a $+\infty$ constitui uma família de locação. (Um caso de família de grupo)

(b) Família de escala (outro caso)

É gerada pela transformação $X = bU$, $b>0$.

$$P(X \leq x) = F\left(\frac{x}{b}\right).$$

Exemplo: $U \sim N(0,1)$

(a) $X = U + a$

$P(X \leq x) = P(U \leq x-a) = \Phi(x-a)$ Locação

: Normais com média a e var. 1.

(b) $X = bU$

$P(X \leq x) = P(U \leq x/b) = \Phi\left(\frac{x}{b}\right)$ Escala

: Normais com $\mu = 0$ e $\sigma^2 = b^2$.

c) Família de locação-escala: é gerada pela transformação

$$X = a + bU \,, \quad b > 0 \,, \quad a \in (-\infty, +\infty)$$

$$P(X \leq x) = F\left(\frac{x-a}{b}\right) . \text{ Locação-Scala}$$

Em geral, F tem densidade f com respeito à medida de Lebesgue. No caso de locação-escala, a densidade é

$$p(x|a,b) = \frac{1}{b} f\left(\frac{x-a}{b}\right) .$$

Exemplos:

1. Normal: $N(a,b^2)$; $U \sim N(0,1)$, $X = bU + a$

$$p(x|a,b) = \frac{1}{\sqrt{2\pi}\, b} \exp\left\{ -\frac{(x-a)^2}{2b^2} \right\} , \quad -\infty < x < \infty$$

Normais $\mu = a$ e $\sigma^2 = b^2$.

2. Exponencial Dupla: $DE(a,b)$; $U \sim DE(0,1)$

$$p(x|a,b) = \frac{1}{2b} \exp\left\{ -\frac{|x-a|}{b} \right\} , \quad -\infty < x < \infty$$

3. Cauchy: $C(a,b)$; $U \sim C(0,1)$

$$p(x|a,b) = \frac{b}{\pi} \frac{1}{b^2 + (x-a)^2} , \quad -\infty < x < \infty$$

4. Logística: $L(a,b)$; $U \sim L(0,1)$

$$p(x|a,b) = \frac{1}{b} \frac{\exp\{-(x-a)/b\}}{[1 + \exp\{-(x-a)/b\}]^2} , \quad -\infty < x < +\infty$$

Outros exemplos: uniforme, exponencial, etc. (ver Tabela 4.1, TPE, p18).

Generalizações: $U = (U_1, \ldots, U_n)$

- $X = (U_1 + a, \ldots, U_n + a)$ - de locação
- $X = (bU_1, \ldots, bU_n)$ - de escala
- $X = (U_1 + a_1, \ldots, U_n + a_n)$ - de locação com comp. dif.
- $X = (b_1 U_1, \ldots, b_n U_n)$, ... - de escala com comp. dif.

Ex. Distribuição Normal Multivariada

$U = (U_1, \ldots, U_p)'$ tq $U_i \sim N(0,1)$, independentes.

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} + B \begin{pmatrix} U_1 \\ \vdots \\ U_p \end{pmatrix} ,$$

onde B é matriz $p \times p$ não singular, ou seja,

$$X = a + BU , \qquad (*)$$

com $X = (X_1, \ldots, X_p)'$, $a = (a_1, \ldots, a_p)'$. Aqui, $E(X) = a$ e $Cov(X) = \Sigma = E[(X-a)(X-a)'] = BB'$.

A densidade de U é (fazer os cálculos)

$$\frac{1}{(\sqrt{2\pi})^p} \exp\left\{ -\frac{1}{2} u'u \right\} , \quad u = (u_1, \ldots, u_n)',$$

e $U = B^{-1}(x-a)$. O ~~determinante~~ Jacobiano da transformação (*) é o determinante $|B|$ de B. Então, a densidade de X é

$$p(x|a,B) = \frac{|B|^{-1}}{(\sqrt{2\pi})^p} \exp\left\{ -\frac{1}{2} (x-a)' \Sigma^{-1} (x-a) \right\}$$

e escrevemos $X \sim N_p(a, \Sigma)$.

Ex. Modelo Linear

$U = (U_1, \dots, U_n)$ com uma dada distribuição conjunta

$$X_i = a_i + b\,U_i, \quad i = 1, \dots, n,$$

tais que
$$a_i = \sum_{j=1}^{p} d_{ij}\, \beta_j, \quad i = 1, \dots, n,$$

e os d_{ij} são fixos e os β_j são arbitrários. Sem perda de generalidade, assume-se que $D = (d_{ij})$ é de posto p.

Se $U_i \sim N(0,1)$ e são iid, a densidade de X é

$$\frac{1}{(\sqrt{2\pi}\, b)^n} \exp\left[-\frac{1}{2b^2} \sum (x_i - a_i)^2 \right]$$

[Compare com ex. 4, p. 1A - mod. de regressão normal linear].

Ex. Família não-paramétrica

U: v.a. com uma distribuição contínua (ex.: $N(0,1)$), cujo suporte é a reta toda.

G: classe de transformações tais que
$$X = g(U),$$
onde g é qualquer função contínua, estritamente crescente com
$$\lim_{u \to -\infty} g(u) = -\infty \quad \text{e} \quad \lim_{u \to +\infty} g(u) = +\infty.$$

Dá pra mostrar que:

- G é um grupo de transformações
- A classe $\{F_g : g \in G\}$ das distribuições de $g(U)$, fazendo g variar em G, é a classe de todas as distribuições contínuas cujo suporte é $(-\infty, +\infty)$, i.e., a classe de todas as distribuições com fda contínua e estritamente crescente em $(-\infty, +\infty)$.

Ex. Família não paramétrica simétrica

Igual ao exemplo anterior, mas com U tendo distribuição simétrica em torno de zero e G restrito às funções ímpares, i.e,

$g(-u) = -g(u)$, para todo u.

A classe resultante é a de todas as distribuições simétricas com relação à origem, com suporte em $(-\infty, +\infty)$.

Simetria:

F_X: fda de X ; F_U: fda de U ; $X = g(U)$. ímpar

P/ $\forall u \geq 0$, $F_U(u) = 1 - F_U(-u)$ (pois a distr de U é simétrica em torno ao 0).

quero provar que

qpq $F_X(x) = 1 - F_X(-x)$, $\forall x \geq 0$

Para $\forall x \geq 0$, seja $u = g^{-1}(x)$ e

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(g(U) \leq x) = P(U \leq u) \\ &= F_U(u) = 1 - F_U(-u) = 1 - P(U \leq -u) \\ &= 1 - P(g(U) \leq g(-u)) = \\ &= 1 - P(X \leq -g(u)) = 1 - P(X \leq -x) \\ &= 1 - F_X(-x). \quad \blacksquare \end{aligned}$$

Generalização : $X = g(U) + a$, $-\infty < a < \infty$.
(simetria em relação ao ponto a).

(Como mostrar que são todas?)

FAMÍLIAS EXPONENCIAIS

Uma família $\{P_\theta\}$ de distribuições forma uma família exponencial s-dimensional se as distribuições P_θ têm densidade da forma:

$$p_\theta(x) = \exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - B(\theta)\right] h(x) \qquad (*)$$

com respeito a alguma medida (comum) μ. Os η_i's e B são funções dos parâmetros a valores reais, T_i's são funções (estatísticas) a valores reais definidas sobre $\mathcal{X}$, o suporte da densidade, e x é um ponto de $\mathcal{X}$ e h é uma função definida sobre $\mathcal{X}$ a valores reais não negativos.

Ex 1. Poisson ; $X \sim$ Poisson (λ), $\lambda > 0$.

$$p_\lambda(x) = \frac{1}{x!} \lambda^x e^{-\lambda}, \quad x = 0, 1, 2, \ldots \qquad s = 1$$

unidimensional

(densidade de X com respeito à medida de contagem

$$p_\lambda(x) = \exp\left[(\log \lambda) x - \lambda\right] \frac{1}{x!}, \quad x = 0, 1, 2 \ldots$$

Tem a forma (*) com $T(x) = x$, $\eta(\lambda) = \log \lambda$, $B(\lambda) = \lambda$, $h(x) = \frac{1}{x!}$, $\mathcal{X} = \{0, 1, 2, \ldots\}$.

(família exponencial unidimensional)

Ex 2. Distribuição de Rayleigh (Ver em casa)

$$p_\theta(x) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\}, \quad x > 0$$

(densidade com respeito à medida de Lebesgue)

$$p_\theta(x) = \exp\left[-\frac{x^2}{2\theta^2} - \log\theta^2\right] x, \quad x > 0$$

Tem a forma (*) com $T(x) = x^2$, $\eta(\theta) = -\frac{1}{2\theta^2}$,

$B(\theta) = \log\theta^2$, $h(x) = x$, $\mathcal{X} = \{x \in \mathbb{R} : x > 0\}$

Ex 3. Suponha que $X = (Z, Y)$ onde $Y = Z + \theta W$, $\theta > 0$, Z e W são indep, $Z \sim N(0,1)$, $W \sim N(0,1)$.

(Ver em casa)

$$p_\theta(x) = p_\theta(z, y) = p(z)\, p_\theta(y \mid z) =$$

$$= \phi(z) \frac{1}{\theta} \phi\left((y - z)/\theta\right)$$

onde ϕ é a fdp $N(0,1)$.

$$p_\theta(x) = \frac{1}{(\sqrt{2\pi})^2 \theta} \exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\}$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{1}{2} z^2\right\} \exp\left\{-\frac{1}{2\theta^2}(y-z)^2 - \log\theta\right\}$$

Tem a forma (*) com $T(x) = (y - z)^2$, $\eta = -\frac{1}{2\theta^2}$,

$B(\theta) = \log\theta$, $h(x) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2} z^2\right\}$, $\mathcal{X} = \mathbb{R}^2$.

Ex 4. $X \sim N(\xi, \sigma^2)$, $\xi \in \mathbb{R}$, $\sigma^2 > 0$, $\theta = (\xi, \sigma^2)$

$$p_\theta(x) = \exp\left[\frac{\xi}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \frac{\xi^2}{2\sigma^2}\right] \frac{1}{\sqrt{2\pi}\,\sigma} \qquad \frac{1}{\sigma} = e^{-\log\sigma}$$

$s = 2$

bidimens.

Tem a forma (*) com $T_1(x) = x$, $T_2(x) = x^2$,

$\eta_1(\theta) = \frac{\xi}{\sigma^2}$, $\eta_2(\theta) = -\frac{1}{2\sigma^2}$, $B(\theta) = \frac{\xi^2}{2\sigma^2} + \log\sigma$, $h(x) = \frac{1}{\sqrt{2\pi}}$,

$\mathcal{X} = \mathbb{R}$.

(família exponencial bi-dimensional)

Outros exemplos importantes: Gama(a, b), Qui-quadrado χ^2_f, Beta(a, b), Bernoulli(p), Binomial(p, n), n fixo, Binomial negativa BN(p, m), m fixo (ver TPE Tabela 5.1, p. 25).

Forma canônica (forma natural): os η_i's são vistos como os parâmetros.

$$p(x \mid \eta) = \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x) \qquad (**)$$

$$\eta = (\eta_1, \dots, \eta_s)$$

$$\downarrow$$

$\eta_i = \eta_i(\theta)$, $B(\theta) = A(\eta)$

escrito em função de η

"reparametrização!"

Para que (★★) seja uma densidade de probabilidade, a integral com respeito a μ deve ser 1, i.e., (forma canônica)

$$\int_{\mathcal{X}} \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x)\, d\mu(x) = 1,$$

ou seja,

$$\exp[-A(\eta)] \int_{\mathcal{X}} \exp\left[\sum_{i=1}^{s} \eta_i T_i(x)\right] h(x)\, d\mu(x) = 1.$$

$A(\eta)$ tal que (★★) tenha integral 1 vai existir sse

$$\int_{\mathcal{X}} \exp\left[\sum_{i=1}^{s} \eta_i T_i(x)\right] h(x)\, d\mu(x) < \infty.$$

O conjunto Δ de pontos $\eta = (\eta_1, \ldots, \eta_s)$ tq isso ocorre é chamado de espaço paramétrico natural (canônico) da família e η é chamado de parâmetro natural (canônico).

Nota: Δ é convexo (TSH, Lema 7, p. 57).

Ex 1. Poisson (cont.)
$\eta = \log \lambda$ é o parâmetro natural e o espaço paramétrico natural é $\mathbb{R}$.

$$p(x|\eta) = \exp\{\eta x - \exp(\eta)\} \frac{1}{x!}$$

$$A(\eta) = \exp(\eta), \quad B(\lambda) = \lambda$$

Ex 4. $X \sim N(\xi, \sigma^2)$ (cont)

$$p(x|\eta) = \exp\left[\eta_1 x + \eta_2 x^2 - A(\eta)\right] \frac{1}{\sqrt{2\pi}}$$

$\eta = (\eta_1, \eta_2) = \left(\frac{\xi}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ é o parâmetro natural,

$$A(\eta) = \frac{\xi^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} + \log\sqrt{-\frac{1}{2\eta_2}},$$

e o espaço paramétrico natural é $\{(\eta_1, \eta_2) \in \mathbb{R}^2 : \eta_2 < 0\}$

Exercício: Colocar as densidades dos exemplos 2 e 3 na forma natural. Idem p/ as distribuições da Tabela 5.1, TPE.

Se nem os T's nem os η's satisfazem restrições lineares e o espaço paramétrico para os η's contém retângulos s-dimensionais, dizemos que a família é de posto completo.

Ex 4. $X \sim N(\xi, \sigma^2)$ (cont)

Família exponencial bi-dimensional de posto completo.

• **Nota.** Considere $(\eta = (\eta_1, \eta_2))$

$$p(x|\eta) = \exp\left\{ \eta_1 T_1(x) + \eta_2 T_2(x) - A(\eta_1, \eta_2) \right\} h(x)$$

(fam. exp. bi-dimensional)

i) Suponha que $T_2(x) = a\, T_1(x) + b$, a, b fixos, $a \neq 0$

$$p(x|\eta) = \exp\left\{ \eta_1 T_1(x) + \eta_2 (a\, T_1(x) + b) - A(\eta) \right\} h(x)$$

$$= \exp\left\{ (\eta_1 + a\eta_2)\, T_1(x) - A^*(\eta) \right\} h(x),$$

onde $A^*(\eta) = -\eta_2 b + A(\eta)$

Nesse caso, não é de posto completo. Pode ser reduzida em dim!

(fam. exp. uni-dimensional) (T_1, T_2 rest. lineares)

ii) Suponha que $\eta_2 = a\eta_1 + b$, a, b fixos, $a \neq 0$

$$p(x|\eta) = \exp\left\{ \eta_1 T_1(x) + (a\eta_1 + b)\, T_2(x) - A(\eta) \right\} h(x)$$

$$= \exp\left\{ \eta_1 T_1^*(x) - A(\eta) \right\} h^*(x),$$

onde $T_1^*(x) = T_1(x) + a\, T_2(x)$ e $h^*(x) = h(x) \exp\{ b\, T_2(x) \}$,

(fam. exp. uni-dimensional). rest. lineares

Ex 5. Família normal curvada

$X \sim N(\xi, \xi^2)$, $\xi > 0$

$$p_\xi(x) = \exp\left[\frac{1}{\xi} x - \frac{1}{2\xi^2} x^2 - \frac{1}{2} \right] \frac{1}{\sqrt{2\pi}\,\xi},$$

Tem a forma (*) com $T_1(x) = x$, $T_2(x) = x^2$, $B(\xi) = +\log \xi$,

$\eta_1(\xi) = \dfrac{1}{\xi}$, $\eta_2(\xi) = -\dfrac{1}{2\xi^2}$,

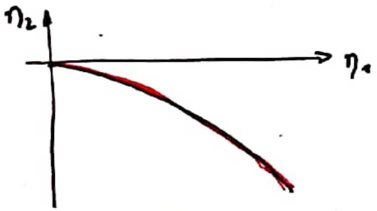
$h(x) = \exp\left\{ -\frac{1}{2} \right\} \frac{1}{\sqrt{2\pi}}$.

Família exponencial bi-dimensional com parâmetro natural $\eta = (\eta_1, \eta_2) = \left(\frac{1}{\xi}, -\frac{1}{2\xi^2} \right)$.

O espaço paramétrico ~~natural~~ é

$$\left\{ (\eta_1, \eta_2) \in \mathbb{R}^2 : \eta_1 > 0 \text{ e } \eta_2 = -\frac{1}{2} \eta_1^2 \right\}.$$

(não é de posto completo).



O espaço paramétrico não contém retângulos bi-dimensionais (aqui $s=2$).

Quando os parâmetros naturais da distribuição estão relacionados de uma forma não linear, dizemos que (*) ou (**) forma uma família exponencial curvada.

Nota. A estrutura exponencial é preservada sob amostragem aleatória. Se $X = (X_1, \dots, X_n)$ e $X_1, \dots, X_n$ são iid com distr. na família exponencial s-dimensional sua densidade conjunta é

$$\exp\left[\sum_{i=1}^{s} \eta_i T_i^*(x) - n A(\eta)\right] h(x_1) \cdots h(x_n)$$

onde

$$T_i^*(x) = \sum_{j=1}^{n} T_i(x_j), \quad x = (x_1, \dots, x_n).$$

Ex 6. $X_i \sim N(\xi, \sigma^2)$, $i = 1, \dots, n$, iid.

A densidade conjunta de $X_1, \dots, X_n$ (com respeito à medida de Lebesgue) é

$$\exp\left\{\frac{\xi}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 - \frac{n}{2\sigma^2} \xi^2\right\} \frac{1}{(\sqrt{2\pi}\,\sigma)^n}.$$

Aqui $T_1^*(x) = \sum_{i=1}^{n} x_i$, $T_2^*(x) = \sum_{i=1}^{n} x_i^2$.

Alunos: ver Ex 5.7, TPE, p. 27 (normal bivariada).

Teorema. Para qualquer função integrável f e qualquer η no interior de Δ, a integral

$$\int f(x) \exp\left[\sum_{i=1}^{s} \eta_i T_i(x)\right] h(x)\, d\mu(x)$$

é contínua e têm derivadas de todas as ordens com respeito aos η's e essas podem ser obtidas sob o sinal de integração.

Prova: TSH (Cap 2, Teo 9, p 59).

Tomemos η, ponto interior de Δ. Como

$$\frac{\partial}{\partial \eta_j} \int_x \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x)\, d\mu(x) = 1 \frac{\partial}{\partial \eta_j}$$

temos

$$\int_x \frac{\partial}{\partial \eta_j} \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x)\, d\mu(x) = 0$$

ie,

$$\int_x \left[T_j(x) - \frac{\partial}{\partial \eta_j} A(\eta)\right] \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x)\, d\mu(x) =$$

ou seja,

X maiúsculo

$$\boxed{E_\eta\left(T_j(X)\right) = \frac{\partial}{\partial \eta_j} A(\eta).}$$

Média das est. obtidas pelas derivadas em relação aos η_j's.

Mostrar que diferenciando ambos os lados da equação com respeito a η_k tem-se

$$\boxed{\operatorname{cov}_\eta\left(T_j(x), T_k(x)\right) = \frac{\partial^2}{\partial \eta_j \partial \eta_k} A(\eta).}$$

Função geradora de momentos (f.g.m.)

$T = (T_1, \dots, T_s)$

$$M_T(u_1, \dots, u_s) = E\left\{ \exp\,(u_1 T_1 + \dots + u_s T_s) \right\}, \quad [*]$$

$$u = (u_1, \dots, u_s) \in \mathbb{R}^s .$$

Suponha que a f.g.m. de T existe ([*] é finita) numa vizinhança da origem de $\mathbb{R}^s$, ie, $\|u\| < \delta$, $\delta > 0$, ou seja, $\left[\sum_{i=1}^{s} u_i^2\right]^{1/2} < \delta$.

A f.g.m. pode então ser expandida em uma série de potências multivariada da forma

$$M_T(u_1, \dots, u_s) = \sum \alpha_{r_1, \dots, r_s} \frac{u_1^{r_1} \dots u_s^{r_s}}{r_1! \dots r_s!},$$

onde

$$\alpha_{r_1, \dots, r_s} = E\left(T_1^{r_1} \dots T_s^{r_s} \right)$$

$$= \left. \frac{\partial^{r_1 + \dots + r_s} M_T(u_1, \dots, u_s)}{\partial u_1^{r_1} \dots \partial u_s^{r_s}} \right|_{u=(0,\dots,0)}$$

Note que o 1º termo da expansão é obtido qdo $r_1 = \dots = r_s = 0$ e é igual a 1.

χ

Os cumulantes $\chi_{r_1, \dots, r_s}$ são definidos como os coeficientes na expansão da função geradora de cumulantes

$$K_T(u_1, \dots, u_s) = \log M_T(u_1, \dots, u_s)$$

$$= \sum_{r_1, \dots, r_s} \chi_{r_1, \dots, r_s} \frac{u_1^{r_1} \dots u_s^{r_s}}{r_1! \dots r_s!}.$$

Note que

$$\chi_{r_1, \dots, r_s} = \left. \frac{\partial^{r_1 + \dots + r_s} K_T(u_1, \dots, u_s)}{\partial u_1^{r_1} \dots \partial u_s^{r_s}} \right|_{u=(0,\dots,0)}$$

Relações entre momentos e cumulantes podem ser obtidas comparando as duas séries de potências.

Ex. $s=1$, $T=T_1$

$$M_T(u) = \sum_{j=0}^{+\infty} \frac{\alpha_j u^j}{j!}, \quad \alpha_j = E(T^j), \quad j=1,2,\ldots, \quad \alpha_0 = 1. \qquad (1)$$

$$K_T(u) = \log M_T(u) \Rightarrow K_T(0) = 0.$$

$$\exp(K_T(u)) = M_T(u)$$

Seja $K_T(u) = x$. Expandindo e^x em torno de $x=0$ (em série de Taylor) vem

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots$$

Avaliando e^x em $x = K_T(u)$ temos:

$$\exp(K_T(u)) = 1 + K_T(u) + \frac{(K_T(u))^2}{2!} + \cdots$$

$$= 1 + \sum_{j=0}^{+\infty} \frac{k_j u^j}{j!} + \frac{1}{2!}\left(\sum_{j=0}^{+\infty} \frac{k_j u^j}{j!}\right)^2 + \cdots \qquad (2)$$

Igualando (1) e (2) vem

$\kappa_0 = 0$

$\kappa_1 = \alpha_1 = E(T)$

$$\frac{\kappa_2}{2!} + \frac{\kappa_1^2}{2!} = \frac{\alpha_2}{2!}, \text{ i.e., } \kappa_2 = \alpha_2 - \alpha_1^2 = Var(T)$$

κ_2 é a var.

Exercício: Mostrar que

1. $\kappa_3 = E[(T-\alpha_1)^3]$ 3º mom central de T

$\kappa_4 = E[(T-\alpha_1)^4] - 3[Var(T)]^2$ 4º mom central $- 3\sigma^4$

$\kappa_5 = E[(T-\alpha_1)^5] - 10\, Var(T)\, E[(T-\alpha_1)^3]$

$\kappa_6 = E[(T-\alpha_1)^6] - 15\, Var(T)\, E[(T-\alpha_1)^4] - 10\{E[(T-\alpha_1)^3]\}^2 + 30\{Var(T)\}^3$

2. $\kappa_j(T+a) = \kappa_j(T)$, $j = 2,3,\ldots$

3. $\mu_j(T+a) = \mu_j(T)$, $j = 2,3,\ldots$; $\mu_j = E[(T-\alpha_1)^j]$

4. $\kappa_j(a.T) = a^j \kappa_j(T)$, $j=1,2,\ldots$

5. Se T_1 e T_2 são independentes, então

$$\kappa_j(T_1+T_2) = \kappa_j(T_1) + \kappa_j(T_2), \quad j=1,2,\ldots$$

6. Se $T \sim N(\mu, \sigma^2)$ então

$$\kappa_1(T) = \mu, \quad \kappa_2(T) = \sigma^2 \text{ e } \kappa_j(T) = 0, \ j=3,4,\ldots$$

Ref.: Bening, V. E. (2000). Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss and Deficiency. VSP: Utrecht.

Teorema: TPE, T. 5.10, p28.

Se X tem densidade na forma

$$p(x|\eta) = \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x),$$

então para qualquer η no interior de Δ, as funções geradoras de momentos e de cumulantes dos T_i's, $M_T(u)$ e $K_T(u)$, existem (em alguma vizinhança da origem) e são dadas por

$$M_T(u) = \frac{\exp(A(\eta+u))}{\exp(A(\eta))}$$

e $K_T(u) = A(\eta+u) - A(\eta)$.

Dem. $M_T(u) = E[\exp(u_1 T_1 + \cdots + u_s T_s)]$

$$= \int_{\mathcal{X}} \exp\left(\sum_{i=1}^{s} (\eta_i+u_i) T_i(x) - A(\eta)\right) h(x)\, d\mu(x)$$

$$= \frac{\exp(-A(\eta))}{\exp(-A(\eta+u))} \int_{\mathcal{X}} \exp\left(\sum_{i=1}^{s} (\eta_i+u_i) T_i(x) - A(\eta+u)\right) h(x)\, d\mu(x);$$

$\eta = (\eta_1, \ldots, \eta_s)$; $u = (u_1, \ldots, u_s)$.

Como η é pto interior de Δ, se u ~~pode~~ está numa vizinhança adequada da origem de $\mathbb{R}^s$, $\eta+u$ tb é ponto interior de η. A integral acima vale 1 e segue o resultado.

$$K_T(u) = \log M_T(u) = A(\eta+u) - A(\eta). \quad \blacksquare$$

Ex. Binomial

Para $x = 0, 1, 2, \ldots, n$

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad 0<p<1$$

$$= \exp\{\underset{T(x)}{x} \log \frac{p}{1-p} + n \log(1-p)\} \binom{n}{x}$$

família exponencial com μ sendo a medida de contagem sobre $x = 0, 1, \ldots, n$ e com

$$\eta = \log \frac{p}{1-p} \quad ; \quad A(\eta) = n \log(1+e^{\eta}).$$

$$M_X(u) = \frac{\exp(A(\eta+u))}{\exp(A(\eta))} = \left[\frac{1+e^{\eta+u}}{1+e^{\eta}}\right]^n = (1-p+pe^u)^n.$$

$$K_X(u) = n \log(1-p+pe^u).$$

$$E(X) = \alpha_1 = M_X'(u)\Big|_{u=0} = n(1-p+pe^u)^{n-1} pe^u \Big|_{u=0} = np.$$

$$Var(X) = \chi_2 = K_X''(u)\Big|_{u=0} = np(1-p).$$

Alunos: Estudar exemplos 5.11 - 5.14 (TPE, p. 29-31).

MODELOS LINEARES GENERALIZADOS

Seja Y v.a. com f.d.p.

$$f(y;\theta) = \exp\{\phi\,(y\theta - b(\theta)) + c(y,\phi)\}, \qquad [*]$$

$\phi > 0$ (fixado). par. de precisão (Maior ϕ, maior precisão)

A f.d.p. de Y está na forma de família exponencial (*) (p. 17) com

$$\eta(\theta) = \phi\theta \quad ; \quad T(y) = y \quad ;$$

$$B(\theta) = \phi b(\theta) \quad ; \quad h(y) = \exp\{c(y,\phi)\}.$$

Na forma natural da fam. exponencial, temos que a fdp de Y pode ser re-escrita como

$$f^*(y;\eta) = \exp\left\{\eta y - \phi\, b\left(\frac{\eta}{\phi}\right)\right\} \exp\{c(y,\phi)\}.$$

Aqui, $\eta = \phi\theta$ e $A(\eta) = \phi\, b\left(\frac{\eta}{\phi}\right)$.

$$E(Y) = A'(\eta) = \frac{\partial}{\partial\eta}\left[\phi\, b\left(\frac{\eta}{\phi}\right)\right] = b'(\theta) = \mu.$$

$$Var(Y) = A''(\eta) = \frac{\partial^2}{\partial\eta^2}\left[\phi\, b\left(\frac{\eta}{\phi}\right)\right] = \frac{1}{\phi}\, b''(\theta)$$

$$= \frac{1}{\phi}\,\frac{d\mu}{d\theta}.$$

$$\boxed{\begin{array}{l} E(Y) = \mu = b'(\theta) \\ Var(Y) = \dfrac{1}{\phi}\, V \quad ; \quad V = \dfrac{d\mu}{d\theta} \end{array}}$$

f. de variância

ϕ : parâmetro de ~~pre~~ precisão
μ : média
V : função de variância

$\forall i \in \{1, \dots, n\}$ Y_i satisfaz $[*]$.

$\forall i$, Y_i está na família exponencial, Mas ñ quer uma $\Rightarrow$ $T(y)$ deve ser uma função linear da observação

Contra-exemplo: Beta ($T(x)$ ñ-lin (depende de μ) em x)

Nos MLG's admite-se que $Y_1, \dots, Y_n$ são v.a.'s com fdp $f(y;\theta_i)$ (ver $[*]$) e que

$$g(\mu_i) = \eta_i = x_i^T \beta \qquad \mu_i = E(Y_i) = b'(\theta_i)$$

em que $\mu_i = b'(\theta_i)$, $x_i = (x_{i1}, \dots, x_{ip})^T$, $p < n$, é um vetor de constantes (variáveis explicativas), $\beta = (\beta_1, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação.

- Casos particulares de $[*]$:

- Normal (μ, σ^2) ; $\phi = \sigma^{-2}$ (Fix. ϕ, deve estar na fam. exp linear)
- Poisson (μ)
- Binomial (p, n) ; n fixo
- Gama, Normal inversa

Modelos Lin. Gen. considera v.a. ~~na~~ ~~exp~~ família expon. linear.

Ligação canônica: $\theta_i = \eta_i = \sum_{j=1}^{p} x_{ij} \beta_j$

Nesse caso, a fdp conjunta de $\underset{\sim}{Y} = (Y_1, \dots, Y_n)$ é

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}; \beta) = \exp\left\{ \phi\left(\sum_{i=1}^{n} y_i \theta_i - \sum_{i=1}^{n} b(\theta_i) \right) + \sum_{i=1}^{n} c(y_i, \phi) \right\}$$

$\underset{\sim}{y} = (y_1, \dots, y_n)$; $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_n)$

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}; \beta) = \exp\left\{ \phi \sum_{j=1}^{p} \left(\sum_{i=1}^{n} y_i x_{ij} \right) \beta_j - \phi \sum_{i=1}^{n} b\left(\sum_{j=1}^{p} x_{ij} \beta_j \right) + \sum_{i=1}^{n} c(y_i, \phi) \right\}$$

Nesse caso, tem uma

- **FAMÍLIA EXPONENCIAL p-DIMENSIONAL DE POSTO COMPLETO** (estamos admitindo aqui que ϕ é conhecido e que o espaço paramétrico é $\mathbb{R}^p$).

No caso, da ligação canônica, vale (2)

OBS: (1) Nos MLG's pode-se admitir que ϕ é desconhecido

(2) A estatística $S = (S_1, \dots, S_p)$, com $S_j = \sum_{i=1}^{n} y_i x_{ij}$, $j = 1, \dots, p$, é uma estatística suficiente completa para β (conceito que será visto a seguir)

ESTATÍSTICAS SUFICIENTES

Uma estatística $T = T(X)$ é suficiente para θ [ou para x, ou para a família $\mathcal{P} = \{ P_\theta, \ \theta \in \Omega \}$] se a distribuição condicional de X dado $T = t$ é independente de θ para todo t.

Idéia: Redução dos dados sem perda de informação sobre θ.

"Depois que observou $T = t$, $X / T = t$ se torna irrelevante para θ"

Ex 1. Poisson

$X \sim \text{Poisson}(\lambda)$, $\lambda > 0$

Foram feitas n observações de X

$X_1, \dots, X_n$: v.a's iid, $X_i \sim \text{Poisson}(\lambda)$

a) Perdeu-se a última observação de X. Houve "perda de informação"?

Sim → Basta calc. a dist. cond. de X da $X_1, \dots, X_{n-1}$, que dep. de λ

b) Perderam-se todas as observações, mas ficou o registro da média $\bar{X}$. Houve "perda de informação"?

Não → Dist. cond. de $X / \bar{X} = \bar{x}$ não dep. de λ.

b) Para $t = 0, 1, 2, \ldots$, $x_1, x_2, \ldots, x_n \in \{0, 1, \ldots\}$, $\sum_{i=1}^{n} x_i = t$:

$$P(X_1 = x_1, \ldots, X_n = x_n \mid \sum_{i=1}^{n} X_i = t) =$$

$$\frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(\sum_{i=1}^{n} X_i = t)} = \frac{\prod_{i=1}^{n} P(X_i = x_i)}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}}$$

$$= \frac{\prod_{i=1}^{n} e^{-\lambda}\lambda^{x_i}/x_i!}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}} = \frac{1}{n^t}\,\frac{t!}{x_1!\cdots x_n!}$$

(não depende de λ)

"Se $\sum X_i \neq t$, $P(X_1 = x_1, \ldots, X_n = x_n \mid \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i) = 0$, que indep. de λ"

Se $\sum_{i=1}^{n} x_i \neq t$, tal probabilidade condicional e' zero. Logo, $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ e' uma estatística suficiente para λ.

E' fácil ver que, como $\bar{X}$ e' função 1-a-1 de T, $\bar{X}$ e' estatística suficiente para λ.

Resposta: não houve "perda de informação".

OBS. Suponha que $n=2$. (caso (b))

A distribuição de X_1 dado $T = X_1 + X_2 = t$ e' binomial $(1/2, t)$ ([se $t=0$ essa distr. e' degenerada em zero, i.e, $P(X_1 = 0 \mid T=0) = 1$].) → Exercício

Para "recuperar" (X_1, X_2) a partir de $T=t$, pode-se gerar uma v.a. X_1' com distr. $b(1/2, t)$ e definir $X_2' = t - X_1'$. A distribuição condicional de (X_1', X_2') dado $T=t$ e' a mesma de (X_1, X_2) dado $T=t$. Logo, as distribuições marginais de (X_1, X_2) e (X_1', X_2') coincidem. (Sempre é possível obter a condicional?) PP.

a) Para $x_1, \dots, x_n \in \{0, 1, 2, \dots\}$:

$$P(X_1 = x_1, \dots, X_n = x_n \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}) =$$

$$\frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(X_1 = x_1, \dots, X_{n-1} = x_{n-1})} =$$

$$\frac{\prod_{i=1}^{n} P(X_i = x_i)}{\prod_{i=1}^{n-1} P(X_i = x_i)} = P(X_n = x_n) = \frac{e^{-\lambda} \lambda^{x_n}}{x_n!}$$

(depende de λ)

Logo, $H(X_1, \dots, X_n) = (X_1, \dots, X_{n-1})$ não é uma estatística suficiente para λ.

Resposta: sim, houve "perda de informação".

Exercício: $X_1, \dots, X_n$ iid, $X_i \sim$ Binomial (p, n), n fixo $0 < p < 1$. (Fazer!)

Mostrar que $\bar{X}$ é suficiente para p, mas $(X_1, \dots, X_{n-1})$ não é.

Ex 2. $X \sim N(0, \sigma^2)$, $\sigma^2 > 0$; $T = |X|$. ↑ dist em rel. à origem (Intuitivo)

A distribuição condicional de $X \mid T = t$, $t > 0$, é tq

redação: "tirar o sinal"

$$P(X = t \mid T = t) = P(X = -t \mid T = t) = 1/2.$$

Simetria da normal.

Como se fosse: "(Bernoulli (1/2)). em $\{-t, t\}$."

Verificação:

- Para $t > 0$:

$$F_T(t) = P(|X| \leq t) = P(-t \leq X \leq t)$$

$$\overset{\text{Padron.}}{=} \underbrace{\Phi}_{\text{ACUMULADA DA NORMAL PADRÃO}}\left(\frac{t}{\sigma}\right) - \Phi\left(-\frac{t}{\sigma}\right).$$

$$f_T(t) = \frac{1}{\sigma} \underbrace{\phi}_{\text{DENSIDADE DA NORMAL PADRÃO}}\left(\frac{t}{\sigma}\right) + \frac{1}{\sigma}\phi\left(\frac{t}{\sigma}\right) = \frac{2}{\sigma}\phi\left(\frac{t}{\sigma}\right).$$

- Vamos construir a distr. de X combinando a distr. de $X \mid T = t$ (supondo Bernoulli (1/2)) e a distr. marginal de T (acima) para verificar se, de fato, chegamos a $N(0, \sigma^2)$.

Para $t>0$:

$$f_T(t) \,.\, P(X=t \mid T=t) = \frac{1}{\sigma}\,\phi\left(\frac{t}{\sigma}\right)$$

$$f_T(t) \,.\, P(X=-t \mid T=t) = \frac{1}{\sigma}\,\phi\left(\frac{t}{\sigma}\right)$$

$$\therefore\ f_X(t) = \frac{1}{\sigma}\,\phi\left(\frac{t}{\sigma}\right),\quad t\in\mathbb{R},\quad \text{i.e.},$$

$$X \sim N(0,\sigma^2).$$

Devido a prob. de arbitrariedade da def. de prob. cond., uma vez que o ev. condição tem prob. nula,

A definição de estatística suficiente envolve o conceito de probabilidade condicional.

A definição de probabilidade condicional pode ser alterada para qualquer conjunto de valores de t que tenha probabilidade zero, sem alterar a distribuição de X, esta obtida combinando-se a distribuição de T com a distribuição condicional de X dado T.

Note que, no ex. 2, $P_\theta(T=t)=0$.

Uma estatística T será dita suficiente se existir uma determinação das funções distribuições condicionais de X dado T que seja independente de θ.

TPE, p. 34-35; TSH, Seção 2.6.

Teorema. CRITÉRIO DA FATORAÇÃO

Uma condição necessária e suficiente para que uma estatística T seja suficiente para a família $\mathcal{P}=\{P_\theta,\ \theta\in\Omega\}$ de distribuições de X dominada por uma medida μ σ-finita é que existam funções g_θ e h tais que as densidades p_θ de P_θ satisfaçam

tem uma dens.

pode ser fatorada.

$$p_\theta(x) = g_\theta[T(x)]\,h(x) \quad (\text{q.c.}\,\mu) \qquad \boxed{*}$$

posso mudar dens. nos pts com medida nula

Dem. (caso discreto)

μ: medida de contagem sobre $\mathcal{X}$, o conjunto de possíveis valores de X.

($\Rightarrow$) Vale $\boxed{*}$ (ida)

Para todo $x\in\mathcal{X}$ tq $p_\theta(x)\neq 0$, defina $t=T(x)$. Temos

$$P_\theta\{T=t\} = \sum_{\{x\in\mathcal{X}:\,T(x)=t\}} p_\theta(x) = \sum_{\{\dots\}} g_\theta[T(x)]\,h(x)$$

$$= g_\theta(t) \sum_{\{\dots\}} h(x).$$

$$P_\theta\{X=x \mid T=t\} = \begin{cases} \dfrac{p_\theta(x)}{g_\theta(t)\sum_{\{\dots\}} h(x)} = \dfrac{\cancel{g_\theta(t)}\,h(x)}{\cancel{g_\theta(t)}\sum_{\{\dots\}} h(x)}, & T(x)=t \\ 0\ ; & T(x)\neq t \end{cases}$$

é independente de θ.

($\Leftarrow$) T é suficiente (volta)

A distribuição condicional de X dado T é independente de θ

$$P_\theta(X=x) = P_\theta(X=x,\ T=t) = \underbrace{P(X=x \mid T=t)}_{\text{ñ dep } \theta \text{ da suf.}}\ \underbrace{P_\theta(T=t)}_{g_\theta(t)}$$

$$= h(x)\ g_\theta(t).$$

Dem. no caso geral: TSH, secção 2.6, Teo 8, Cor. 1.

Ex 1. Poisson (cont)

$$P_\lambda(X_1=x_1, \dots, X_n=x_n) = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}, \quad x_i \in \{0,1,\dots\},\ i=1,\dots,n.$$

Tem a forma $\boxed{*}$ com $T = \sum X_i$ (est. suficiente). $g_\theta(t) = e^{-n\theta}\,\theta^t$

Ex 2. $N(0, \sigma^2)$ (cont.)

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

$h(x) = \frac{1}{\sqrt{2\pi}}$ $g_\sigma(t) = \frac{1}{\sigma} e^{-\frac{t^2}{2\sigma^2}}$

$|X|$ é est. suficiente.

Ex 3. $X_1, \dots, X_n$ iid, $X_i \sim U(0,\theta)$, $\theta > 0$.

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

$$p_\theta(x) = \frac{1}{\theta^n}\prod_{i=1}^{n} I(x_i \geq 0)\, I(x_i < \theta)$$

$$= \frac{1}{\theta^n} I(x_{(n)} \leq \theta) \prod_{i=1}^{n} I(x_i \geq 0), \quad x_{(n)} = \max(x_1, \dots, x_n)$$

Tem a forma $\boxed{*}$ com $T = X_{(n)} = \max(X_1, \dots, X_n)$.

Ex 4. $X_1, \dots, X_n$ iid, $X_i \sim N(\xi, \sigma^2)$ (Tarefa)

$T_1 = (\sum X_i, \sum X_i^2)$ é estat. suficiente

f 1:1 ($T_2 = (\sum X_i, \sum (X_i - \bar{X})^2)$ " "

g 1:1 ($T_3 = \left(\bar{X}, \frac{\sum (X_i - \bar{X})^2}{n}\right)$ " "

Podem existir diferentes estatísticas suficientes.

Nota: As 3 estatísticas são equivalentes

(para cada par de estatísticas, uma é função da outra).

Elas identificam os mesmos pontos do espaço amostral, i.e., se $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$, tem-se

$$T_1(x) = T_1(y) \iff T_2(x) = T_2(y).$$

(idem para os pares (T_1, T_3) e (T_2, T_3)).

Ex 5. Estatísticas de ordem

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

$$p_\theta(x) = \prod_{i=1}^{n} p(x_i | \theta), \quad x = (x_1, \dots, x_n)$$

sendo $p(x_i|\theta)$ a densidade de X_i, avaliada em x_i.

$$p_\theta(x) = \prod_{i=1}^{n} p(x_{(i)} | \theta)$$

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$X_{(i)}$: i-ésima estatística de ordem

$(X_{(1)}, \dots, X_{(n)})$ é uma estatística suficiente.

↓redução: só informar ordem em que for. obtidos

Compara c/ Ex 4 (aqui, pouca redução).

Nota. Na família exponencial s-dimensional a densidade de uma a.a. $X_1, \dots, X_n$ é

$$\exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i^*(x) - n B(\theta) \right] \prod_{i=1}^{n} h(x_i)$$

onde $T_i^*(x) = \sum_{j=1}^{n} T_i(x_j)$, $x = (x_1, \dots, x_n)$.

Pelo critério da fatoração,

$$\left(\sum_{j=1}^{n} T_1(X_j), \dots, \sum_{j=1}^{n} T_s(X_j) \right)$$

é uma estatística suficiente, qualquer que seja o tamanho da amostra.

Redução de dimensionalidade: de n p/ s.

Redução máxima dos dados sem perda de informação sobre θ: estatística suficiente minimal.

Aula 04, 19/08 (Até 15)

SUFICIÊNCIA MINIMAL

Uma estatística T é dita ser suficiente minimal se é suficiente e, entre todas as estatísticas suficientes, é aquela que produz maior redução nos dados, ou seja, para qualquer estatística suficiente U, existe uma função H tal que $T = H(U)$ (q.c. $\mathcal{P}$).

Teorema. Seja $\mathcal{P}$ uma família finita com densidades p_i, $i = 0, 1, \dots, k$, todas tendo o mesmo suporte. Então, a estatística

$$T(X) = \left(\frac{p_1(X)}{p_0(X)}, \frac{p_2(X)}{p_0(X)}, \dots, \frac{p_k(X)}{p_0(X)} \right)$$

é suficiente minimal.

Para provar o teorema vamos usar o seguinte corolário do Critério da Fatoração (CF).

Corolário do CF. Sob as condições do C.F., uma condição necessária e suficiente para uma estatística U ser suficiente é que para quaisquer θ e θ_0 fixados, a razão $p_\theta(x) / p_{\theta_0}(x)$ seja uma função apenas de $U(x)$.

Quarta: 8:30 → 10:00

Terça: 14 → 15:30

Dem. ($\Rightarrow$) U é suficiente

Pelo CF, $\exists$ g_θ e h, f$\underline{\text{ç}}$s não negativas, tq

$$p_\theta(x) = g_\theta[U(x)]\, h(x)$$

e

$$p_{\theta_0}(x) = g_{\theta_0}[U(x)]\, h(x) \qquad (\text{q.c. } \mu)$$

Então, p/ $\forall\, \theta, \theta_0$ fixados

$$\frac{p_\theta(x)}{p_{\theta_0}(x)} = \frac{g_\theta[U(x)]}{g_{\theta_0}[U(x)]} \qquad (\text{q.c. } \mu)$$

e é $\therefore$ uma f$\underline{\text{ç}}$ apenas de $U(x)$.

($\Leftarrow$) Para $\forall\, \theta, \theta_0$, $\exists$ f$\underline{\text{ç}}$ H_{θ,θ_0} tq

$$\frac{p_\theta(x)}{p_{\theta_0}(x)} = H_{\theta,\theta_0}(U(x))$$

Tomando um θ_0 fixado

$$p_\theta(x) = p_{\theta_0}(x)\, H_{\theta,\theta_0}(U(x))$$

$$= h(x)\, g_\theta[U(x)],$$

onde $h(x) = p_{\theta_0}(x)$ e $g_\theta[U(x)] = H_{\theta,\theta_0}(U(x))$.

Dem. do Teorema.

$$T = T(X) = (T_1, \dots, T_k), \quad T_i = T_i(X) = \frac{p_i(X)}{p_0(X)}, \; i=1,\dots,k.$$

Sejam $g_0(T) = 1$ e $g_i(T) = T_i$, $i = 1, \dots, k$.

Então $p_i(x) = g_i(T(x))\, p_0(x)$ e $\therefore$,

pelo C.F., $T(X)$ é suficiente.

Pelo corolário, U é uma estatística suficiente sse T é uma f$_c$ de U. Isso prova que T é minimal. ■

Extensão do Teorema para $\mathcal{P}$ contável é imediata (exercício).

Para $\mathcal{P}$ não contável, é útil o Lema a seguir.

Lema. Se $\mathcal{P}$ é uma família de distribuições com suporte comum (o mesmo para todos os membros da família) e $\mathcal{P}_0 \subset \mathcal{P}$, e se T é suficiente minimal para $\mathcal{P}_0$ e suficiente para $\mathcal{P}$, então é suficiente minimal para $\mathcal{P}$.

Dem. esboço | Se U é suficiente para $\mathcal{P}$, também é para $\mathcal{P}_0$. Então, T é f$_c$ de U, pois T é suficiente minimal para $\mathcal{P}_0$. ■

[conj. de medida nula em $\mathcal{P}_0$ também será med. nula em $\mathcal{P}$, dado o sup. comum

Ex1. $N(\theta, 1)$, $\theta \in \mathbb{R}$

$X_1, \dots, X_n$: iid, $X_i \sim N(\theta, 1)$

$\mathcal{P}_0$: consiste de 2 distribuições $N(\theta_0, 1)$ e $N(\theta_1, 1)$.

$$T(X) = \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} \text{ é suficiente minimal;} \quad X = (X_1, \dots, X_n).$$

$$T(X) = \exp\left\{ -\sum \frac{(X_i - \theta_1)^2}{2} + \sum \frac{(X_i - \theta_0)^2}{2} \right\} \text{ expandir,}$$

$$= \exp\left\{ n\bar{X}(\theta_1 - \theta_0) - \frac{n}{2}(\theta_1^2 - \theta_0^2) \right\}, \text{ rearranjar}$$

que é equivalente a $\bar{X}$ (f$_c$ 1 a 1 de $\bar{X}$).

Como $\bar{X}$ é suficiente para $\mathcal{P} = \{N(\theta, 1) : \theta \in \mathbb{R}\}$,

então é suficiente minimal para $\mathcal{P}$.

O Teorema abaixo (Shervish, p. 92) dá uma maneira bem geral de se obter estatísticas suficientes minimais.

Teorema. Sob as condições do C.F., assuma que

$$T(x) = T(y) \iff y \in D(x)$$

onde

$$D(x) = \{y \in \mathcal{X} : p_\theta(y) = p_\theta(x)\, h(x,y),\ \forall\, \theta \in \Omega\}$$

para alguma função h não negativa. Então $T(X)$ é suficiente minimal.

Dem. Devemos mostrar que para ∀ estatística suficiente S, existe uma função v tq $T = v(S)$ e que T é suficiente.

Sendo x e y tq $S(x) = S(y)$ temos do CF que, para $\forall\, \theta \in \Omega$,

$$p_\theta(x) = g(S(x),\theta)\, h(x) = g(S(y),\theta)\, h(x)$$

$$= g(S(y),\theta)\, h(y)\, \frac{h(x)}{h(y)} = p_\theta(y)\, h(x,y).$$

Então,

$$S(x) = S(y) \Rightarrow p_\theta(x) = p_\theta(y)\, h(x,y), \forall\, \theta \in \Omega \Rightarrow$$

$$T(x) = T(y).$$

Logo, $T = v(S)$ para alguma fç v.

Falta mostrar que T é suficiente. Faremos isso no caso discreto ($\mathcal{X}$ é contável).

Para um $x_0 \in \mathcal{X}$ arbitrário, denotemos por c_0 $T(x_0)$.

$$P_\theta\left[X = x_0 \mid T(X) = c_0\right] = \frac{P_\theta[X = x_0,\ T(X) = c_0]}{P_\theta[T(X) = c_0]}$$

$$= \frac{P_\theta[X = x_0]}{P_\theta[T(X) = c_0]}$$

$$= \frac{P_\theta[X = x_0]}{\sum_{d \in \{x \in \mathcal{X}:\, T(x) = c_0\}} P_\theta[X = d]} = \frac{1}{\sum_{d \in \{\dots\}} \frac{p_\theta(d)}{p_\theta(x_0)}}$$

$$= \frac{1}{\sum_{d \in \{\dots\}} h(d, x_0)} \quad \text{é indep. de } \theta.$$

$$P_\theta[X = x \mid T(X) = c] = 0, \quad \text{para } c \neq c_0 \text{ (ind. de } \theta).$$

∎

Ver: - Dudewicz & Mishra (1988). Modern Mathematical Statistics. Wiley (p. 401. Teo 8.1.23).

- Shervish, Teo 2.29, p. 92 (demonstração mais geral).

Ex 1. (cont) $X_1, \dots, X_n$: iid, $X_i \sim N(\theta, 1)$, $\theta \in \mathbb{R}$

$T(X) = \sum X_i$ é suficiente para θ ; $X = (X_1, \dots, X_n)$

$$p_\theta(x) = \frac{1}{(\sqrt{2\pi})^n} \exp\left\{ - \sum \frac{(x_i - \theta)^2}{2} \right\}$$

$$= \frac{1}{(\sqrt{2\pi})^n} \exp\left\{ - \frac{\sum x_i^2}{2} + \theta \sum x_i - \frac{n\theta^2}{2} \right\}.$$

$$p_\theta(y) = \frac{1}{(\sqrt{2\pi})^n} \exp\left\{ - \frac{\sum x_i^2}{2} + \theta \sum x_i - \frac{n\theta^2}{2} \right\} \times \exp\left\{ - \frac{\sum y_i^2 - \sum x_i^2}{2} + \theta\left(\sum y_i - \sum x_i\right) \right\}$$

$$= p_\theta(x) \exp\left\{ - \frac{\sum y_i^2 - \sum x_i^2}{2} + \theta\left(\sum y_i - \sum x_i\right) \right\}. \quad (*)$$

Voltamos à $(*)$

Note que $T(x) = T(y) \Leftrightarrow \sum x_i = \sum y_i \Leftrightarrow$

$y \in \{ y = (y_1, \dots, y_n) \in \mathbb{R}^n : \sum y_i = \sum x_i \} \Leftrightarrow$

$y \in D(x) = \{ y \in \mathbb{R}^n : p_\theta(x) = p_\theta(y) \cdot h(x,y), \forall \theta \in \mathbb{R} \}$,

onde $h(x,y) = \exp\left\{ \frac{\sum y_i^2 - \sum x_i^2}{2} \right\}$.

$\therefore$ $T(X) = \sum X_i$ (e $\bar{X}$) é suficiente minimal.

Exercício.

Ex 2. $X_1, \dots, X_n$: iid, $X_i \sim$ Bernoulli(θ), $0 < \theta < 1$.

$\sum X_i$ e $\bar{X}$ são estatísticas suficientes mínimas (exercício).

Ex 3. $X_1, \dots, X_n$: iid, $X_i \sim U(0, \theta)$, $\theta > 0$.

$$p_\theta(x) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}), \quad x_i > 0, \ i = 1, \dots, n$$

Seja $T(X) = X_{(n)}$ (é uma estat. suficiente).

Defina $D(x) = \{ y = (y_1, \dots, y_n) \in \mathbb{R}^n : y_i > 0, \ p/ \ i = 1, \dots, n$ e

$$p_\theta(y) = p_\theta(x) \cdot 1, \ \forall \theta > 0 \}.$$

Assim $T(x) = T(y) \Leftrightarrow y \in D(x)$

e $\therefore$ $T(X) = X_{(n)}$ é suficiente minimal.

$(\Rightarrow)$ $\max(X_1, \dots, X_n) = \max(Y_1, \dots, Y_n)$ $\Rightarrow I_{(0,\theta)}(x_{(n)}) = I_\theta(0, y_{(n)})$, $\forall \theta$

$(\Leftarrow)$ $p_\theta(y) = p_\theta(y)$ $\Leftrightarrow I_{(0,\theta)}(x_{(n)}) = I(y_{(n)})$ $\Rightarrow x_{(n)} = y_{(n)}$. $T(X) = T(Y)$

Ex 4. Exponencial, $X_i \sim E(\theta, 1)$, $\theta \in \mathbb{R}$, iid, $i = 1, \dots, n$.

Se $X \sim E(a, b)$, $a \in \mathbb{R}$, $b > 0$,

$$p_{a,b}(x) = \frac{1}{b} \exp\left\{ - \frac{(x-a)}{b} \right\} I_{[a, +\infty)}(x).$$

Aqui, $X_{(1)}$ é estatística suficiente minimal (exercício).

Ex 5. Uniforme, $X_i \sim U(\theta - 1/2, \theta + 1/2)$, iid, $i = 1, \dots, n$.

$(X_{(1)}, X_{(n)})$ é suf. minimal (exercício)

Ex 6. Logístico, $X_i \sim L(\theta,1)$, $i=1,\dots,n$, iid.

Se $X \sim L(a,b)$, $a \in \mathbb{R}$, $b>0$,

$$p_{a,b}(x) = \frac{1}{b}\,\frac{\exp\{-(x-a)/b\}}{[1+\exp\{-(x-a)/b\}]^2}$$

$$p_\theta(x) = \exp\left[-\sum x_i + n\theta\right] \frac{1}{\prod\{1+\exp\{-(x_i-\theta)\}\}^2}$$

Sejam

$T(x) = (x_{(1)},\dots,x_{(n)})$, $u_i = e^{-y_i}$, $v_i = e^{-x_i}$,

$\xi = e^{\theta}$, $\eta = \xi^{-1}$.

Note que

$$\frac{p_\theta(y)}{p_\theta(x)} = \exp\left[+\sum x_i - \sum y_i\right]\prod_{i=1}^n \frac{[1+\exp\{-(y_i-\theta)\}]^{-2}}{[1+\exp\{-(x_i-\theta)\}]^{-2}}$$

$$= \exp\left[+\sum x_i - \sum y_i\right]\prod_{i=1}^n \left[\frac{1+u_i\xi}{1+v_i\xi}\right]^{-2}$$

$$= \exp\left[+\sum x_i - \sum y_i\right]\prod_{i=1}^n \left[\frac{\eta+v_i}{\eta+u_i}\right]^{2}.$$

$T(x) = T(y) \Rightarrow \dfrac{p_\theta(y)}{p_\theta(x)} = 1$, $\forall\, \theta \in \mathbb{R}$.

Por outro lado,

$\dfrac{p_\theta(y)}{p_\theta(x)} = h(x,y)$, $\forall\, \theta \in \mathbb{R} \Rightarrow \prod(\eta+u_i) = \prod(\eta+v_i)$, $\forall\, \eta \geq 0$

(igualdade de 2 polinômios de grau n em η)

$\Rightarrow$ $(u_1,\dots,u_n)$ deve ser uma permutação de $(v_1,\dots,v_n)$
(os 2 polinômios têm as mesmas raízes).

$\Rightarrow$ $(x_{(1)},\dots,x_{(n)}) = (y_{(1)},\dots,y_{(n)}) \Rightarrow T(x) = T(y)$.

$\therefore$ $T(x) = T(y) \Leftrightarrow y \in D(x) = \{y \in \mathbb{R}^n : p_\theta(y) = p_\theta(x)\, h(x,y)$
$\forall\, \theta \in \mathbb{R}\}$,

onde $h(x,y) = \exp\{\sum x_i - \sum y_i\}$.

Conclui-se que o vetor de estatísticas de ordem é suficiente minimal.

Alunos: usar esta ideia para Cauchy e Exponencial Dupla. (família de locação).

Porquê $(X_1, X_2, \dots, X_n)$ não o é?

Ex 7. Família exponencial s-dimensional de posto completo

$$p_\theta(x,\eta) = \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right\} h(x)$$

$\eta \in \Delta$ (não necessariam. o espaço paramétrico natural).

$$p(y,\eta) = p(x,\eta) \exp\left\{\sum_{i=1}^{s} \eta_i \left(T_i(y) - T_i(x)\right)\right\} h(x,y),$$

$$\forall \eta \in \Delta, \quad h(x,y) = \frac{h(y)}{h(x)}.$$

$D(x) = \{ y \in \mathcal{X} : \ p(y,\eta) = p(x,\eta)\, h(x,y), \ \forall \eta \in \Delta \}$

$T(x) = (T_1(x), \dots, T_s(x))$

- $T(x) = T(y) \Rightarrow y \in D(x)$
- $y \in D(x) \Rightarrow \sum_{i=1}^{s} \eta_i \left(T_i(y) - T_i(x)\right) = 0, \ \forall \eta \in \Delta$

$\xRightarrow{\text{posto completo}}$ $T_i(x) = T_i(y), \ i=1,\dots,s \Rightarrow T(x) = T(y).$

Logo, $T = (T_1(x), \dots, T_s(x))$ é uma estatística suficiente minimal.

Ex 8. Família exponencial curvada (Exercício)

Igual ao ex anterior, mas aqui a família não é de posto completo necessariamente.

Assumimos que o espaço paramétrico contém $s+1$ pontos $\eta^{(j)}$, $j = 0, 1, \dots, s$, que geram $\mathbb{R}^s$.

Mostrar que $T = (T_1(X), \dots, T_s(X))$ é uma estatística suficiente minimal.

Ex 9. Família normal curvada

$X_1, \dots, X_n$ iid, $X_i \sim N(\xi, \xi^2)$, $\xi > 0$.

$(\Sigma X_i, \Sigma X_i^2)$ é est. suficiente.

É também suficiente minimal pelo Ex 8.
De fato, como $\eta = \left(\frac{1}{\xi}, -\frac{1}{2\xi^2}\right)$ (ver p. 22),

podemos escolher

$$\eta^{(0)} = (1, -1/2), \quad \eta^{(1)} = (1/2, -1/8) \ \text{e} \ \eta^{(2)} = (1/3, -1/18)$$

e temos que os vetores $\eta^{(1)} - \eta^{(0)}$ e $\eta^{(2)} - \eta^{(0)}$ são L.I. Note que a matriz

$$\begin{pmatrix} 1/2 - 1 & 1/3 - 1 \\ -\frac{1}{8} + \frac{1}{2} & -\frac{1}{18} + \frac{1}{2} \end{pmatrix}$$

é inversível (tem posto igual a 2).

ANCILARIDADE

Uma estatística $V(X)$ é dita ser ancilar para a família $\mathcal{P} = \{ P_\theta,\ \theta \in \Omega \}$ de possíveis distribuições para X se sua distribuição não depende de θ. É dita ser ancilar de primeira ordem se $E_\theta[V(X)]$ não depende de θ.

Idéia: Uma estatística ancilar não contém informação sobre θ.

Nota: Estatísticas suficientes mínimas podem conter algum "material ancilar". Por exemplo, no Ex 6 (Logística $L(\theta,1)$), mostramos que $(X_{(1)}, \dots, X_{(n)})$ é suficiente minimal, mas as diferenças $X_{(n)} - X_{(i)}$, $i = 1, \dots, n-1$ são ancilares (mostrar).

Ex 1. (cont). $X_1, \dots, X_n$ iid, $X_i \sim N(\theta, 1)$.

$$\Sigma (X_i - \bar{X})^2 \sim \chi^2_{n-1} \qquad \text{(é estatística ancilar)}$$

Idem c/ $X_i \sim N(\theta, \sigma^2)$, $\sigma^2 > 0$ fixado

$$\frac{\Sigma (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \text{(é est. ancilar)}$$

$$\Sigma \frac{(X_i - \bar{X})^2}{n-1} \text{ é est. ancilar.}$$

(variância amostral)

OBSERVAÇÕES SOBRE ANCILARIDADE

1. O que chamamos aqui de "estatística ancilar" é chamado por alguns autores de "estatística de distribuição constante" (distribution constant statistic).

Ex:

Pace, L. & Salvan, A. (1997) Principles of Statistical Inference from a Neo-Fisherian Perspective. World Scientific. Singapore.

Pace & Salvan (Definition 2.6, p. 52):

"Under a parametric model $\mathcal{F}$ with minimal sufficient statistic s, a statistic a is called auxiliary if $(\hat{\theta}, a)$ is a one-to-one funct of s. If a is auxiliary and distribution constant, a is said to be ancillary.

$\hat{\theta}$: estimador de máxima verossimilhança de θ (veremos depois).

Comentário: Relevância da estatística ancilar

(Casella & Berger, p. 285)

X_1, X_2: 2 v.a.'s ind com a mesma distr. de X:

$P_\theta(X=\theta) = P_\theta(X=\theta+1) = P_\theta(X=\theta+2) = \frac{1}{3}$;

θ inteiro, desconhecido.

$X_{(1)} \leq X_{(2)}$: estatísticas de ordem

$$\left.\begin{array}{l} R = X_{(2)} - X_{(1)} \\ M = \dfrac{X_{(1)} + X_{(2)}}{2} \end{array}\right\}$$ Note que (R, M) é equivalente a $(X_{(1)}, X_{(2)})$! (Camarelo)

- (R, M) é estatística suficiente minimal
- R é estatística ancilar
- R sozinho não fornece informação sobre θ, pois sua distribuição não depende de θ.

Suponha que foi observado $(R, M) = (r, m)$ (m inteiro).
Mas considere apenas a observação de M.
Para esse ponto ter probabilidade positiva, θ deve ser igual a m, $m-1$ ou $m-2$.

Note que

* $P(M=m \mid \theta = m-2) = 1/9$;

$P(M=m \mid \theta = m-1) = 2/9$;

$P(M=m \mid \theta = m) = 1/9$.

Suponha agora que usamos a informação adicional de que $R = r = 2$.

O único valor possível para θ é $m-1$.

[Note que $(R, M) = (2, m) \Rightarrow X_{(1)} = m-1$ e $X_{(2)} = m+1$]

O conhecimento (adicional) de que $R=2$ aumentou a informação sobre θ.

$(\theta, \theta+1, \theta+2)$

(

Aula 05, 24/08

Seria interessante procurar estatísticas suficientes T tais que nenhuma função não constante de T fosse ancilar, nem mesmo ancilar de primeira ordem, ou seja, t.q.

se $E_\theta[f(T)] = c$, p/ todo $\theta \in \Omega$, então

$f(T) = c$ (q.c. $\mathcal{P}$).

Subtraindo c, esta condição é equivalente à definição de estatística completa.

COMPLETIVIDADE

Uma estatística T é completa se, para qualquer função f de T a valores reais

$E_\theta[f(T)] = 0$, p/ todo $\theta \in \Omega \Rightarrow f(T) = 0$ (q.c. $\mathcal{P}$),

onde $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$.

Nota. Se existe uma estatística suficiente minimal, então uma condição necessária para que uma estatística suficiente seja completa é que ela seja suficiente minimal (ver Problema 6.29, p.72; Scherwish, Teo. 2.36, p. 94 - Teorema de Bahadur).

Teorema de Basu

Se T é uma estatística suficiente completa para a família $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$, então qualquer estatística ancilar V é independente de T.

Dem. Se V é ancilar,

$p_A = P(V \in A)$

é independente de θ para todo A.

Seja $\eta_A(t) = P(V \in A \mid T = t)$

(note que $\eta_A(T)$ não depende de θ pois T é suficiente)

Então,

- $E_\theta[\eta_A(T)] = p_A$ e, como T é completa,

$\eta_A(T) = p_A$ (q.c. $\mathcal{P}$).

Isso estabelece a independência de T e V.

- $$E_\theta[\eta_A(T)] = E_\theta[P(V \in A \mid T = t)]$$
$$= E_\theta[E[\mathbb{I}_A(V) \mid T]]$$
$$= E_\theta[\mathbb{I}_A(V)] = P(V \in A) = p_A$$

Teorema. Se X tem distribuição numa família exponencial s-dimensional de posto completo, então $T = (T_1(X), \dots, T_s(X))$ é suficiente completa.

Dem. TSH, Teo 1, Seção 4.3.

Ex 1. $X_i \sim$ Bernoulli (p), $0 < p < 1$

$\mathcal{P} = \{$ Bernoulli (p), $0 < p < 1\}$

$T(X) = \sum X_i$: estatística suficiente (minimal) completa

Ex 2. $X_i \sim$ Poisson (θ), $\theta > 0$

$\mathcal{P} = \{$ Poisson (θ), $\theta > 0\}$

$T(X) = \sum X_i$: est. suf. (minimal) completa

Ex 3. $X_i \sim E(a, b)$, $a \in \mathbb{R}$ fixado, $b > 0$ (Exp. de dois par.)

$\mathcal{P} = \{ E(a,b);\ b > 0\}$, $a \in \mathbb{R}$.

$$p_{a,b}(x) = \frac{1}{b} \exp\left\{ -\frac{(x-a)}{b} \right\}, \quad x > a; \ b > 0.$$

$T(X) = \sum X_i$: est. suf (minimal) completa.

"Mostrar que é de posto completo
$\Rightarrow T(x)$ é suf. completa"

Ex 4. $X_1, \dots, X_n$ iid, $X_i \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

$\mathcal{P} = \{ N(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}$

$(\bar{X}, S^2)$ é estatística suficiente completa.

$$\bar{X} = \frac{\sum X_i}{n}, \qquad S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}.$$

Ex 5. $X_1, \dots, X_n$: iid, X_i com distr. uniforme discreta em $1, 2, \dots, N$; $N \in \{1, 2, \dots\}$.

Para $x = (x_1, \dots, x_n)$, $x_i \in \{1, 2, \dots\}$, $i = 1, \dots, n$:

$$p_N(x) = \frac{1}{N^n} I_{\{1, \dots, N\}}(x_{(n)}) .$$

Suficiência de $X_{(n)}$ segue do C.F.

Para $y = 1, 2, \dots, N$:

$$P_N(X_{(n)} \le y) = \prod_{i=1}^{n} P_N(X_i \le y) = \left(\frac{y}{N}\right)^n$$

e $$P_N(X_{(n)} = y) = \left(\frac{y}{N}\right)^n - \left(\frac{y-1}{N}\right)^n = \frac{1}{N^n}\left[y^n - (y-1)^n \right].$$

$E_N[f(X_{(n)})] = 0$, para todo $N \in \{1, 2, \dots\} \Rightarrow$

$\sum_{y=1}^{N} f(y) \left[y^n - (y-1)^n \right] = 0$, " " " $\Rightarrow$

$N = 1 \Rightarrow f(1) = 0$
$N = 2 \Rightarrow f(1) + f(2)(2^n - 1) = 0$
$N = 3 \Rightarrow f(1) + f(2)(2^n - 1) + f(3)(3^n - 2^n) = 0 \quad \Rightarrow$
$\vdots$

$\Rightarrow f(y) = 0$, p/ $y = 1, 2, \dots$.

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

Ex 6. Suf. minimal, mas não completa

$X_1, \dots, X_n$ iid, $X_i \sim U(\theta - 1/2, \theta + 1/2)$, $\theta \in \mathbb{R}$.

$T = (X_{(1)}, X_{(n)})$ é suficiente minimal (mostrar)

T não é completa pois

$$E_\theta\left[X_{(n)} - X_{(1)} - \frac{(n-1)}{(n+1)}\right] = 0 \quad \text{para todo } \theta \in \mathbb{R}.$$

[redacted]

Ex 7. Suf. minimal, mas não completa

$X_1, \dots, X_n$ iid, $X_i \sim N(\mu, \mu^2)$, $\mu > 0$.

Mostrar que $T = (\sum X_i, \sum X_i^2)$ é suficiente minimal, mas não é completa.

Ex 8. $X_1, \dots, X_n$: iid, $X_i \sim U(0, \theta)$, $\theta > 0$

$\mathcal{P} = \{U(0,\theta), \theta > 0\}$

$T =$ [redacted] $X_{(n)}$ é estat. suficiente (pelo CF).

T é completa?

$$P_\theta(T \leq t) = \frac{t^n}{\theta^n}, \quad 0 < t < \theta;$$

$$p_\theta(t) = \frac{n t^{n-1}}{\theta^n}, \quad 0 < t < \theta$$

$$E_\theta[f(T)] = 0, \ \forall \theta > 0 \Rightarrow \int_0^\theta f(t) t^{n-1} dt = 0, \ \forall \theta > 0$$

$\Rightarrow \int_A f(t)\, t^{n-1} dt = 0$, [redacted], $\forall A$ Boreliano de $\mathbb{R}_+$

$\Rightarrow f(t)\, t^{n-1} = 0$ qc (segundo a medida de Lebesgue)

$\Rightarrow f(t) = 0$ qc $\Rightarrow X_{(n)}$ é completa.

Ex 9. $Y_1, \dots, Y_n$ iid; $Y_i \sim E(\eta, 1)$

$\mathcal{P} = \{E(\eta, 1); \eta \in \mathbb{R}\}$

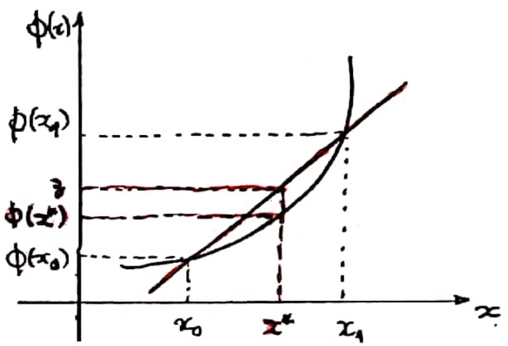
Se $X_i = e^{-Y_i}$ e $\theta = e^{-\eta}$, então $X_1, \dots, X_n$ são iid, $X_i \sim U(0,\theta)$. (mostrar). Pelo ex acima $X_{(n)}$ é suficiente completa. Mas $X_{(n)}$ é equivalente a $Y_{(1)}$. Então, $Y_{(1)}$ é suficiente completa.

FUNÇÕES DE PERDA CONVEXAS

Def. Uma função ϕ a valores reais definida sobre um intervalo aberto $I = (a,b)$ com $-\infty \leq a < b \leq +\infty$ é convexa se para quaisquer $a < x < y < b$ e qualquer $0 < \gamma < 1$

$$\phi(\gamma x + (1-\gamma) y) \leq \gamma \phi(x) + (1-\gamma)\phi(y).$$

A função ϕ é dita ser estritamente convexa se a desigualdade acima é estrita. A função ϕ é côncava sobre (a,b) se $-\phi$ é convexa.



$x^* = \gamma x_0 + (1-\gamma) x_1$

$z = \gamma \phi(x_0) + (1-\gamma)\phi(x_1)$

OBS: se ϕ é convexa sobre (a,b), então é contínua em (a,b).

Teorema.

(i) Se ϕ é definida e diferenciável sobre (a,b) então uma condição necessária e suficiente para ϕ ser convexa é que

$(\star)$ $\phi'(x) \leq \phi'(y)$, para todo $a < x < y < b$

(estritamente convexa: $<$ no lugar de $\leq$).

(ii) Se, adicional/, ϕ é duplamente diferenciável, então, a condição $(\star)$ é equivalente a

$\phi''(x) \geq 0$, para todo $a < x < b$,

com desigualdade estrita sendo condição suficiente (mas não necessária) para convexidade estrita.

Obs. Em geral, uma função convexa é estritamente convexa a menos que seja linear num subconjunto de (a,b).

Exemplos: As seguintes funções são convexas (mostrar). Quais são estritamente convexas?

$\phi(x) = |x|$, $x \in \mathbb{R}$ → " " " " do valor absoluto

→ $\phi(x) = x^2$, $x \in \mathbb{R}$ → vai me levar a perda quadrática

$\phi(x) = x^p$, $x > 0$ $(p \geq 1)$

$\phi(x) = \dfrac{1}{x^p}$, $x > 0$ $(p > 0)$

$\phi(x) = e^x$, $x \in \mathbb{R}$

$\phi(x) = -\log x$, $x > 0$.

Nota: Propriedades interessantes de certos estimadores são obtidas desde que a função de perda considerada seja (estritamente) convexa.

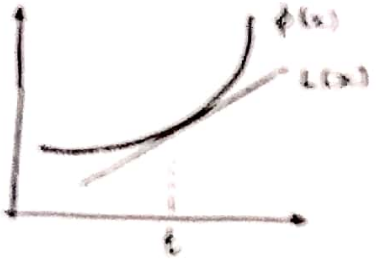
Teorema Seja ϕ uma função convexa definida sobre $I = (a,b)$ e seja t um ponto em I. Então, existe uma linha reta

$$y = L(x) = c(x-t) + \phi(t)$$

que passa no ponto $[t, \phi(t)]$ tal que

$L(x) \leq \phi(x)$, para todo x em I.

Se ϕ é estritamente convexa $L(x) < \phi(x)$, exceto em $x = t$.



Teorema. Desigualdade de Jensen

Se ϕ é uma função convexa definida sobre um intervalo aberto I e X é uma v.a. com $P(X \in I) = 1$ e tem esperança finita, então

$$\phi[E(X)] \leq E[\phi(X)].$$

Se ϕ é estritamente convexa a desigualdade é estrita a menos que X seja uma constante com probabilidade 1.

Dem. Usaremos a proposição: Y e Z v.a's

(i) $Y \geq Z$ q.c. $\Rightarrow E(Y) \geq E(Z)$ desde que as esperanças existam.

(ii) $X \geq 0$ q.c. e $E(X) = 0 \Rightarrow X = 0$ q.c.

Seja $y = L(x) = c(x-t) + \phi(t)$, sendo $t = E(X)$ que assumimos ser finita. Note que $E[L(X)]$ é finita.

Por (i), se $E[\phi(X)]$ é finita:

$$E[\phi(X)] \geq E[L(X)] = L[E(X)] = L(t)$$
$$= \phi(t) = \phi(E(X)). \quad \blacksquare$$

Note que a desigualdade também vale se $E[\phi(X)] = +\infty$.

Suponha agora que ϕ é estritamente convexa.
Por (*) temos

$$E[\phi(X) - L(X)] \geq 0 .$$

Por (ii),

$$E[\phi(X) - L(X)] = 0 \Rightarrow \phi(X) = L(X) \text{ q.c.} \quad (**)$$

Como ϕ é estritamente convexa, temos que $\phi(x) > L(x)$ exceto em $x = t$ e $\phi(t) = L(t)$.

Assim, (**) implica que $X = t = E(X)$ com probabilidade 1.

Conclui-se que se ϕ é estritamente convexa

$$E[\phi(X)] > E[L(X)] = \phi[E(X)] \qquad \text{(ver (*))}.$$

a menos que $X = t$ com probabilidade 1

- $E(\eta(T)) = E[E(\delta(X)|T)] = E[\delta(X)]$

Teorema de Rao-Blackwell

Seja X uma v.a. observável com distribuição $P_\theta \in \mathcal{P} = \{P_{\theta'},\ \theta' \in \Omega\}$ e seja T uma estatística suficiente para $\mathcal{P}$. Seja δ um estimador de um estimando $g(\theta)$ e seja $L(\theta, d)$ uma função de perda estritamente convexa de d. Então, se δ tem média finita e risco

$$R(\theta, \delta) = E[L(\theta, \delta(X))] < \infty$$

e se

$$\eta(t) = E[\delta(X) \mid T = t],$$

então o risco do estimador $\eta(T)$ satisfaz

$$R(\theta, \eta) < R(\theta, \delta)$$

a menos que $\delta(X) = \eta(T)$ com probab. 1

Dem.: Seja $\phi(d) = L(\theta, d)$ estritamente convexa; sejam $\delta = \delta(X)$ e X uma v.a.
Pela desigualdade de Jensen

$$L[\theta, \eta(t)] < E\{L[\theta, \delta(X)] \mid T = t\}$$

a menos que $\delta(X) = \eta(T)$ com probabilidade 1
Tomando esperança em ambos os lados vem

$$R(\theta, \eta) < R(\theta, \delta)$$

a menos que $\delta(X) = \eta(T)$ com probabilidade 1. ▪

Nota: Suficiência de T foi usada para assegurar que $\eta(T)$ não depende de θ (é um estimador de $g(\theta)$).

Nota: L convexa mas não estritamente convexa
$\Rightarrow R(\theta,\eta) \leq R(\theta,\delta)$.
(o estimador η é pelo menos "tão bom quanto" δ).

Nota: O teorema sugere que estimadores que não sejam funções de estatísticas suficientes não devem ser usados.

PERDA QUADRÁTICA. $L(\theta,d) = (d-g(\theta))^2$
(estritamente convexa) $\Rightarrow g''(d) = (2(d-\theta))' = 2 > 0$

Seja δ um estimador não viciado de $g(\theta)$, isto é:
$E_\theta[\delta(X)] = g(\theta)$, $\forall \theta \in \Omega$. Seja

$$\eta(t) = E[\delta(X) \mid T=t],$$

com T sendo uma estatística suficiente.
Aqui,
$$R(\theta,\delta) = E_\theta[[\delta(X) - g(\theta)]^2]$$
$$= Var_\theta[\delta(X)] + \{\underbrace{E_\theta[\delta(X)] - g(\theta)}_{\text{viés}}\}^2 = Var_\theta[\delta(X)]$$

Se $Var_\theta[\delta(X)] < \infty$,

$$Var_\theta(\eta(T)) < Var_\theta(\delta(X)),$$

a menos que $\delta(X) = \eta(T)$ com prob. 1.

ADMISSIBILIDADE

Um estimador é dito ser inadmissível se existe outro estimador δ' que o domina, i.e., tal que $R(\theta,\delta') \leq R(\theta,\delta)$ para todo θ, com desigualdade estrita para algum θ e é admissível se tal estimador não existe.

Ex. Estimador inadmissível
$X_1,\dots,X_n$: iid, $X_i \sim N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$
$X = (X_1,\dots,X_n)$
Estimando: σ^2
$L(\theta,\delta) = (\delta-\sigma^2)^2$.

$$\delta_1(X) = \frac{1}{n-1}\sum(X_i-\bar{X})^2 \quad ; \quad \delta_2(X) = \frac{1}{n+1}\sum(X_i-\bar{X})^2.$$

Seja $S^2 = \sum(X_i-\bar{X})^2/\sigma^2 \sim \chi^2_{n-1}$.

$$R(\theta,\delta_1) = E[(\delta_1(X)-\sigma^2)^2] = E\left[\left[\frac{\sigma^2 S^2}{n-1} - \sigma^2\right]^2\right]$$
$$= \frac{\sigma^4}{(n-1)^2} E[[S^2 - E(S^2)]^2] = \frac{\sigma^4}{(n-1)^2} Var(S^2)$$
$$= \frac{\sigma^4}{(n-1)^2}\, 2(n-1) = \frac{2\sigma^4}{n-1}.$$

$$\underbrace{R(\theta,\delta_2)}_{\text{viesado}} = \text{exercício} = \frac{2\sigma^4}{n+1} < R(\theta,\delta_1),\ \forall\theta$$

$\therefore \delta_1$ é inadmissível.

Nota: Os conceitos e resultados acima podem ser generalizadas para o caso em que

$$g(\theta) = [g_1(\theta), \dots, g_k(\theta)]$$

e

$$\delta(x) = [\delta_1(x), \dots, \delta_k(x)]$$

Ver Def. 7.11, Teo. 7.13, TPE, p.49.