MAE 5834 - Estatística Avançada I
2º semestre de 2015
Prof. Silvia L.P. Ferrari

1 Programa Resumido

1. Introdução
2. Modelos estatísticos; famílias de modelos
3. Suficiência, suficiência minimal, completividade e ancilaridade
4. Estimação pontual, admissibilidade, não viciosidade, estimação ótima, estimação minimax, equivariância
5. Verossimimilhança e estimação por máxima verossimilhança
6. Informação de Fisher, desigualdade da informação
7. Estimação pontual Bayesiana
8. Intervalos de confiança e de credibilidade
9. Formulação geral do problema de testes de hipóteses
10. Lema de Neyman Pearson e testes uniformemente mais poderosos
11. Testes da razão de verossimilhanças, escore, Wald e gradiente
12. Outros tópicos selecionados

2 Bibliografia

1. **Referências Principais**

Lehmann, E.L. e Casella, G. (1998). *Theory of Point Estimation*. Segunda Edição. Springer-Verlag: New York.

Lehmann, E.L., Romano, J.P. (2005). *Testing Statistical Hypotheses*. Terceira Edição. Springer-Verlag: New York.

ou

Lehmann, E.L. (1997). *Testing Statistical Hypotheses*. Segunda Edição. Springer-Verlag: New York.

2. **Outras referências**

Bickel, P. J. e Doksum, K.A. (2000). *Mathematical Statistics: Basic Ideas and Selected Topics*. Segunda Edição. Prentice Hall: Upper Sadlle River.

Schervish, M.J. (1997). *Theory of Statistics*. Springer-Verlag: New York.

Shao, J. (2003). *Mathematical Statistics*. Segunda Edição. Springer-Verlag: New York.

3 Avaliação

Provas nº 5 ; 5 listas longas ao longo do semestre

A avaliação será baseada em duas provas (70%), listas de exercícios (10%) e um trabalho (20%). A primeira prova terá peso 1 e a segunda, peso 2. As listas de exercícios e o trabalho poderão ser feitos em grupos de até 3 alunos. Os trabalhos deverão ser escritos em formato de artigo científico e serão apresentados oralmente. Para cada trabalho, um grupo de alunos será designado para discutí-lo após a apresentação oral. Os trabalhos escritos devem ser entregues pelo menos 7 dias antes da respectiva apresentação. As datas de entrega das listas de exercícios e do trabalho e das apresentações serão fixadas posteriormente.

4 Datas de Provas

As provas serão realizadas nos dias **30 de setembro** e **18 de novembro**.

Estatística Avançada I

Aula 01

10/08/2015

Primeira parte - livro do Lehmann, Cap. 1.

$\mathcal{P} = \{P_\theta, \theta \in \Omega\}$: classe de distribuições conjuntas indexadas por um parâmetro θ. (pelo vetor paramétrico θ).

Obs: Fazer inferência sobre θ ou $g(\theta)$ com base nos dados.

- Estimação pontual e intervalar
- Teste de hipóteses - problema de decisão
- Previsão de novas observações (Não será tratado).

Para inf., os dados e a suposição das possíveis dist. (No clássica)
Do ponto de vista bayesiano, o parâmetro é assumido como uma v.a. (não observável). A distribuição inicial (à priori) é atualizada a partir das observações para se obter a dist. a posteriori.

Estimação Pontual

Ingredientes:

1. $g: \Omega \to \mathbb{R}$, cujo valor em θ, $g(\theta)$, será estimado.
   estimando

2. Dados

2. Estimador: uma transf. dos dados δ usada para estimar $g(\theta)$.

Ex. média da Beta (p,q) $\mu = \frac{p}{p+q}$

Seria interessante que $\delta(x) \in [0,1]$. Não será impostas tais restrições, i.e., $\delta(x) \in \operatorname{Im} g(\theta)$ não é obrigatório.

É desejável que δ "est. perto" de $g(\theta)$.

Medidas de proximidade ou dist. entre δ e $g(\theta)$

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

Consequências (perdas) ao estimar $g(\theta)$ por um valor d de δ. são medidas por $L(\theta, d)$ (função de perda).

Assumimos que

$$L(\theta, d) \geq 0, \quad \forall \theta, d$$

e

$$L(\theta, g(\theta)) = 0, \quad \forall \theta$$

Função de Risco

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

. Indexação em θ para indicar o valor

Buscar δ que minimiza perda média.

Problema: Estimador ideal não existe.

Seja $\delta(x) = g(\theta_0)$, $\forall x \in \mathcal{X}$.
$\forall \theta \in \Omega$, encontro $\delta(X)$ que tem risco mínimo.

Solução: restringir a classe dos estimadores. Destes, encontrar o ideal (risco unif. mínimo dentre os não viesados)

Restrição: Classe dos estimadores não viciados.
Em média, acerta em $g(\theta)$.

- Classe dos estimadores não viciados com resp. à mediana.
$$\text{Mediana}_\theta[\delta(X)] - g(\theta) = 0, \quad \forall \theta \in \Omega$$

Outros critérios p/ escolha de estimadores
1. Est. de Bayes
2. δ que minimiza $\sup\limits_{\Omega} R(\theta, \delta)$: minimax

Modelo Estatístico

- Dados a partir de um experimento.

Como ponto de partida, um espaço mensurável $(\mathcal{E}, \mathcal{B})$ e uma família de dist. de prob. definidas sobre este espaço.

Vamos começar com os dados mesmo!

$(\mathfrak{X}, \mathcal{A}, \mathcal{P})$

onde

$\mathfrak{X}$: cjto

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

$\mathcal{P}$: família de medidas (dist.) de prob.

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

Ω: espaço paramétrico

p/ inf., quanto mais vasta $\mathcal{P}$, mais difícil o proc. de inf.

Suposição usual (identificabilidade): $\forall \theta_1, \theta_2 \in \Omega$,

$\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$. (rel. biunívoca entre θ e dst. correspondente)

Tarefa: Ler seções 3 e 4, Cap 1, Lehmann.

T é um estatística se é uma transf. mensurável de $(\mathfrak{X}, \mathcal{A})$ em $(\mathcal{T}, \mathcal{B})$.

O foco do curso será nos conceitos entendidos.
Sempre que mencionar uma medida, ela será σ-finita.

Na próxima aulas começar a falar sobre os mod. estatísticos

Família de Grupos e Famílias Exponenciais.

Tarefa Aula 01: ler seções 3 e 4 Lehmann, Theory of Point Estimation

3. Probability Theory

Para o trab. em estatística, a aplicação mais import. da teoria da medida é a teoria da probabilidade. Uma medida P def. num espaço mensurável $(\mathcal{X}, \mathcal{A})$ satisfazendo

$$P(\mathcal{X}) = 1$$

é uma medida de probabilidade. e $P(A)$ indica a prob. de A. Se P é absolutamente contínua com respeito à medida μ com a deriva de Radon Nikodyn p, segue

$$P(A) = \int_A p \, d\mu, \quad \text{em geral, a med. de Lebesgue}$$

p é função densidade de P com respeito à μ.
(ou dist. de prob.)

Estaremos preoc. apenas com sit. em que $\mathcal{X}$ é um espaço Euclideano, e tipicamente as distrib. serão ou discretas (no caso em que μ é tomado como uma med. de contagem) ou abs. cont. com resp. à med. de Lebesgue.

. Problemas estatísticos preocupam-se não apenas com uma única distribuição de prob. mas com famílias de tais distribuições

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

def sobre um comum espaço de prob $(\mathcal{X}, \mathcal{A})$.

Quando todos as dist. de $\mathcal{P}$ são abs. cont. com resp. à uma medida comum μ, como será usualmente, a família $\mathcal{P}$ é dita dominada por μ.

Ex.

med de contagem

i) Caso Discreto. $\mathcal{X}$ é enum., $\mathcal{A}$ é a classe de subc. de $\mathcal{X}$, e $P \in \mathcal{P}$ é domi. por μ.

ii) Caso Abs. Cont. $\mathcal{X}$ é um boreliano de um esp. Euclideano, $\mathcal{A} = \mathcal{B}(\mathcal{X})$ e $P \in \mathcal{P}$ é dominada por λ em $(\mathcal{X}, \mathcal{A})$

Lebesgue measure.

Conjunto N é dito $\mathcal{P}$-nulo se

$$P(N) = 0 \quad \forall P \in \mathcal{P}.$$

Se algo é válido, exceto em N, dizemos que tal "algo" vale q.c.

Se $\mathcal{P}$ é domin. por μ, então $\mu(N) = 0$ implica N ser $\mathcal{P}$-nulo. Quando vale o reverso, dizemos que $\mathcal{P}$ e μ são equivalentes

Variáveis Aleatórias (Vetores Aleatórios)

$$(\mathcal{E}, \mathcal{B}) \xrightarrow{X} (\mathcal{X}, \mathcal{A})$$

possív. resu. do exp.

uma func à valores reais mensurável

Para X, podemos def.

$$F_X(a) = P(X \le a).$$
$$P\{e : X(e) \le a\}.$$

f.d. acumulada de X.

Outro conc. de cont. absoluta: f real em $(-\infty, \infty)$ é dita abs. cont. se $\forall \varepsilon > 0, \exists \delta > 0$ t.q. $\forall$ coleção finita de (a_i, b_i) disjuntos

$$\sum (b_i - a_i) < \delta \Rightarrow \sum |f(b_i) - f(a_i)| < \varepsilon$$

f.d.

Conexão: F na reta é abs. contínua $\Leftrightarrow$ a medida gerada por F (digamos $\mathbb{P}$) é absol. contínua com respeito à med. de Lebesgue.

abs. cont: f abs. contínua $\Rightarrow$ f contínua ; f cont. $\not\Rightarrow$ f abs. cont.

$$(\mathcal{E}, \mathcal{B}) \xrightarrow{(X_1, \dots, X_n)} (\mathcal{X}, \mathcal{A})$$

$$F_X(a_1, \dots, a_n) = \mathbb{P}(X_1 \leq a_1, \dots, X_n \leq a_n) =$$
$$= \mathbb{P}\{e \in \mathcal{E} : X_1(e) \leq a_1, \dots, X_n(e) \leq a_n\}.$$

F_X determina as prob. de $(X_1, \dots, X_n) \in A$ e coincide com a prob. dos ev.

$$\{e : [X_1(e), \dots, X_n(e)] \in A\}.$$

Desta desc. mat., pode-se esperar começar a modela. a partir de $(\mathcal{E}, \mathcal{B})$ e uma família de prob. $\mathcal{P}$ def. sobre isto. No entanto, a análise est. de um experimento não é tip. baseada numa desc. completa dos resultados do experimento (que ~~[illegible]~~ incluiria, por exemplo, os menores detalhes acerca de todos os subsecl. exp.) repres. por ptos $e \in \mathcal{E}$.

De modo mais freq., o pto de part. é o cjto de obs., repres. por $X = (X_1, \dots, X_n)$ com todos os asp. do exp. ignorados. A especificação do modelo irá, portanto, começar a partir de X - os dados; o espaço mens. $(\mathcal{X}, \mathcal{A})$ em que X toma seus valores e uma família

espaço amostral

de prob. $\mathcal{P}$ à qual a dist. de X pertence.

. Real-valoradas ou vetor-valoradas funções de $T = (T_1, \dots, T_k)$ de X são chamadas estatísticas. Em part., estimadores são estatísticas.
$T(X) \overset{(*)}{\in} \mathcal{A} \Rightarrow T(X) \in \mathcal{B}$. No entanto, iremos seguir com $(*)$ uma vez que

$$\int T(X(e))\, d\mathbb{P}(e) = \int T(x)\, d\mathbb{P}_X(x), \quad \mathbb{P}_X \text{ denota dist. prob. de } X.$$

~~[illegible]~~ i) Est. mensuráveis ~~[illegible]~~ ñ verificaremos
ii) Est. integráveis

4. Famílias de Grupos

Foco em: Família Exponencial } cobrem maior parte dos modelos est. comuns
Família de Grupo.

+ detalhes Johnson e Kolz

Princ. razões: Exp: efetuar grande simplif. dos dados - ∃ nº fixos de estatísticas ,usualm. menor
~~grupo~~ à quais os dados podem ser reduzidos sem perda de informação, a menos do tamanho amostral;

<u>Grupo</u>: $P \in \mathcal{P}$ possuem uma regra alt. simétrica. Esta sim. outra vez leva a redução da dimensão dos dados uma vez ser nat. impor a requisição de sim. ao estimador

Família de Grupo: de distribuições é um família obtida submetendo uma v.a. com uma dada dist. a um grupo de transf.

Exemplo 4.1. Famílias de locação-escala

U uma v.a. com $U \sim F$. Seja

$$X = U + a \text{. É claro que } \mathbb{P}(X \leq x) = F(x-a) \quad (*)$$

A totalidade das distrib. da forma 4.2, p/ F fixada e $a \in \mathbb{R}$, é dita Família de locação.

Dei modo análogo, uma Família de escala é gerada pela transf.

$$X = bU, \quad b > 0 \Rightarrow P(X \leq x) = P(U \leq x/b)$$

Comb. as duas últimas transf., i.e.,

$$X = bU + a, \quad b > 0 \Rightarrow P(X \leq x) = \mathbb{P}\left(U \leq \frac{x-a}{b}\right)$$

obtemos a Família de locação-escala.

Em geral, F tem f.d. f com resp. a λ. Daí

$$f_X(x) = \frac{1}{b} f\left(\frac{x-a}{b}\right)$$

Exemplos: Normal, Exp dupla, Cauchy, Logística, Exponencial, Uniforme, ...

i.) Fechada sob composição $\overset{\text{transf 1-1}}{g_1}: \mathcal{X} \to \mathcal{X}$; $\overset{\text{transf 1-1}}{g_2}: \mathcal{X} \to \mathcal{X}$ $g_2 \cdot g_1: \mathcal{X} \to \mathcal{X}$

Uma classe $\mathcal{G}$ de transf. é dita fechada se $g_1 \in \mathcal{G}$, $g_2 \in \mathcal{G} \Rightarrow g_2 \cdot g_1 \in \mathcal{G}$ $x \mapsto g_2(g_1(x))$

ii.) Fechada sob inversão

$\mathcal{G}$ é dita ↑ se $g \in \mathcal{G} \Rightarrow g^{-1} \in \mathcal{G}$.

A estrutura de classes de transf. que possuem estas prop. é um caso especial de um obj matemát mais geral, simplesmente chamado de grupo

Def. 4.2. Um cjto G de elementos é chamado um grupo se satisfaz as seguintes 4 condições

i. Está def. uma operação, group multiplication, em que opera dois elementos $a, b \in G$ associa $c \in G$. $c = ab$

ii. $(ab)c = a(bc)$ (associativa)

iii. $\exists\, e \in G$ (ident.), t.q $ae = ea = a$ $\forall a \in G$.

iv. $\forall a \in G$, $\exists\, a^{-1}$ (inverso) t.q $aa^{-1} = a^{-1}a = e$.

a^{-1} e e são únicos!

Os grupos de int primário em est. são os grupos de transformações

Def. 4.3. Uma classe G de transf. é cham. um grupo de transf. se é fechado sob compos. e invers.

Obs: Se G satisfaz $g_2 \cdot g_1 = g_1 \cdot g_2$, é dito comutativo.

- Extende a ideia ao caso multivariado
- Exemplo de Normal Multivariada
- ~~Exemplo do modelo de Reg. Linear~~

  (4.1) e (4.2) e (4.5) levam a dist. diferente

  (4.18) não nec. caso a matriz de $\underset{\sim}{a}_1 = \underset{\sim}{a}_2$ e $B_1 B_1' = B_2 B_2'$

- Ex. Modelo Linear

- Nonpar. i.i.d. family.

  $\{U_k, k \in \{1 \dots n\}\}$ i.i.d, $U_i \sim N(0,1)$. G a classe das transf.

  $X_i = g(U_i)$, g conti. e estrit. cresc.

  com $\lim_{u \to -\infty} g(u) = -\infty$ e $\lim_{u \to \infty} g(u) = \infty$.

G é um grupo. Dai $\{X_i\}$ é iid com com dis. F_g. $\{F_g : g \in G\}$ é a classe das dist. cont. em $(-\infty, \infty)$

Estatística Avançada I

Aula 02 12/08/2015

X: dados (v.a)

x: dados obs.

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

Se $X \sim$ Poisson (λ), então

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

$\eta = \log \lambda$ e $A(\eta) = e^{\eta}$.

Logo,

$$M_x(\omega) = \frac{e^{A(\eta+\omega)}}{e^{A(\eta)}} = \frac{e^{e^{(\eta+\omega)}}}{e^{e^{\log\lambda}}} = \frac{e^{e^{(\log\lambda+\omega)}}}{e^{\lambda}} =$$

$$= \frac{e^{\lambda . e^{\omega}}}{e^{\lambda}} = e^{-\lambda} e^{\lambda e^{\omega}} = e^{\lambda(e^{\omega}-1)}.$$

$$K_x(\omega) = A(\eta+\omega) - A(\eta) = e^{\eta+\omega} - e^{\eta} = e^{\eta}(e^{\omega}-1)$$
$$= \lambda(e^{\omega}-1)$$

Daí, $M_x(\omega) = e^{\lambda(e^{\omega}-1)}$

$E_\lambda(X) = K'_x(\omega)|_{\omega=0} = \lambda e^{\omega}|_{\omega=0} = \lambda$

$Var_\lambda(X) = K''_x(\omega)|_{\omega=0} = \lambda e^{\omega}|_{\omega=0} = \lambda$

$K_j(X) = \frac{d^j K_x(\omega)}{d\omega^j}\Big|_{\omega=0} = \lambda, \quad j = 1, 2, \ldots$

Estatística Avançada I

Aula 03 17/08/15

Ex. $X \sim N(\xi, \xi)$, $\xi > 0$

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sqrt{\xi}} e^{-\frac{1}{2}\left(\frac{x-\xi}{\xi}\right)^2} \mathbb{1}_{\mathbb{R}}(x)$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{x^2 - 2x\xi + \xi^2}{\xi}\right) - \log \xi\right\} \frac{1}{\sqrt{2\pi}}$$

$$= \exp\left\{-\frac{1}{2}\frac{x^2}{\xi} + x - \frac{\xi}{2} - \frac{1}{2}\log \xi\right\} \frac{1}{\sqrt{2\pi}}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\xi} + x - \frac{\xi}{2} - \frac{1}{2}\log \xi\right\}$$

que está na forma (*) com $\eta(\xi) = -1/\xi$, $B(\xi) = \xi/2 + \frac{1}{2}\log \xi$, $h(x) = \frac{1}{\sqrt{2\pi}}$.

- Está na ~~forma~~ família de exponencial de posto completo

$\eta \in \{\eta \in \mathbb{R}, \eta < 0\}$

⋮

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

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

$$\prod I_{(0,\theta)}(x_i) = \begin{cases} 1, & \text{se } 0 < x_i < \theta \ \forall i \iff x_{(n)} < \theta \iff \\ 0, & \end{cases}$$

$$\prod_{i=1}^{n} \mathbb{I}_{(0,\theta)}(x_i) = \mathbb{I}_{(0,\theta)}(x_{(n)}) \prod_{i=1}^{n} \mathbb{I}_{(0,\infty)}(x_i)$$

Estatística Avançada I

Aula 04 19/08/15

Recapitulando Est. Suf.

X: dados

T(X): estatística

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

Est. suf: dist. X | T(X) não dep. de θ

T(X): est. suf. minimal

suf. e função de Δ est. suf. todas.

- Critério para Encontra Estat. Suf.

Ex. 8.

$$p_\theta(y) = p_\theta(x), \forall \theta > 0 \overset{?}{\Rightarrow} x_{(n)} = y_{(n)}$$

$$p_\theta(y) = \frac{1}{\theta^n} I_{(0,\theta)}(y_{(n)}) = p_\theta(x) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}), \forall \theta > 0$$

$$\Rightarrow I_{(0,\theta)}(y_{(n)}) = I_{(0,\theta)}(x_{(n)}) \ \forall \theta > 0 \overset{?}{\Rightarrow} x_{(n)} = y_{(n)}$$

- Suponha $x_{(n)} \neq y_{(n)}$ e $\theta = \frac{x_{(n)} + y_{(n)}}{2}$. Então $I_{(0,\theta)}(x_{(n)}) \neq I_{(0,\theta)}(y_{(n)})$.

- Até aqui, resume de dados sem perda de inf. do parâmetro.

- Estatística Ancilar (No sent. oposto, estat. que não traz inf. sobre θ)
- Próx. Aula. Estat. Suf. Completa (Ex até 89 X {Est. Completa}) $\ast$ 82

Aula 5, Est. Avançada I

24/08

X: dados

- $X \sim P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Omega\}$
- $T(X)$: est.
- $T(X)$ é est. suf. se a dist. de $X|T$ ind. de θ
- $T(X)$ é suf. minimal se $T(X)$ é suf. e, $\forall$ est. suf $U(X)$, $\exists H : T(X) = H(U(X))$, q.c. $\mathcal{P}$
- $V(X)$ é uma est. ancilar se a dist. de $V(X)$ não dep. de θ
- Uma est. suf. pode conter "material ancilar"

Estatística suf. completa: "tem a ver com a ideia de ñ ter mat. ancilar"

T. Basu. Se T suf. completa e V: ancilar $\Rightarrow$ T e V são ind.
Dist. de T e T/V coincidem

Teorema:

Se $T(x) = T(y) \Leftrightarrow \{y \in D(x) = \{y \in \mathcal{X} : p_\theta(y)/p_\theta(x) = h(x,y), \forall \theta\}\}$
então $T(X)$ é suf. minimal.

Estimação Pontual.

$\theta \in \Omega$, $g(\theta)$: estimando ; $d(X)$: estimador de $g(\theta)$

Função de Perda:

$$L(\theta; d) \geq 0, \ \forall \theta, d$$
$$= 0, \text{ se } d = g(\theta).$$

Risco: $E[L(\theta, d(X))]$, $\theta \in \Omega$.

δ estim. de $g(\theta)$. δ é inadmissível se $\exists\ \delta'$ t.q.

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

$$R(\theta, \delta') < R(\theta, \delta) \text{ , para algum } \theta.$$

Aula 09, Avanç. I.

14/09

- Na aula passada, informação de Fisher.
- Útil para desigualdade interessante acerca da variância do estimador.
- Desig. de Hammersley - Chapman

- Resultado sobre Inf. de Fisher. Desig. da Inf.
- Ex. Poisson
- Família de Locação - não depende do a
- Família " " - escala - depende de b
- $E_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta(x)\right]=0$ e $E_\theta\left[\frac{\partial}{\partial\theta}\log q_\theta(y)\right]=0$.
- $X_1,\dots X_n$ i.i.d. $J(\theta)$, então a inf. cont. em $(X_1,\dots,X_n)$ é $nJ(\theta)$.

- Desigualdade da Informação: prova. desigualdade surge da desig. da covariância.
- Limite Inferior de Cramér - Rao

$Cov(X,Y) \leq Var X\, Var Y$
igualdade ocorre sse $Y = aX + b$

- Limite Inf. atingido qdo $\frac{\partial}{\partial\theta} p_\theta(x) = a(\theta)\delta(x)+b(\theta)$ ⇒ Só atingida quando tiver na família exponencial e $\delta(x)=T(x)$.

- Apenas funções lineares de θ têm est. não viciado cuja var. ating. LICR.
- ENVUM não necessariamente atinge LICR: Exemplo $\theta(1-\theta)$ binomial(n,θ)
- Retomar a Des. da Inf. evitando a exig. da troca de derivada e integral. (113) → Troca de integral e limite; Teorema da Conv. Dominada
- Caso Multiparamétrico. → Matriz de Inf. de Fisher
$Var_\theta(\delta) \geq \alpha' J^{-1}(\theta)\alpha$, $\alpha_i = \frac{\partial}{\partial\theta_i}E_\theta(\delta(x))$

Exercícios para Lista 2. 8, 18, 23, 24, 30.

1. $\delta(x) = g(x) + xg'$

Estatística Avançada

Aula 11 — 21/09/2015

Seminário

Data	Apresenta	Discute
26/10	Jaime e Marcon	William e Fábio
26/10	Jeniffer e Lina	Sandra
04/11	A Marcos e Mara	Hugo, Estefania e Sérgio.
09/11	Hugo, Es; Sérgio	A. Marcos e Mara
11/11	Sandra	Jeniffer e Lina
16/11	William e Fábio	Jaime e Maicon

- Equivariância: uma introdução
- Nova restrição: simetria.
- Invariante por locação.

- Equivariante por locação
- Estimador equivariante de Risco mínimo (ERM, MRE)
- Como encontrar todos os equivariates: $\delta(x) = \delta_0(x) + u(x)$, $u(x+a) = u(x)$.
- Como encontra o equivariante de Risco mínimo
- Se ρ for convexa, $\exists$ EMR (é único se ρ for estritamente convexa) e ~~não~~ monótona
  - Perda quadrática $\overset{n=1}{\Rightarrow}$ $X - E_0(X)$ é ERM
  - Perda absoluta $\Rightarrow$ $X - med_0(X)$ é ERM
- Forma de encontrar EMR se a perda for quadrática (teo.) (Estimador de Pitman) (Normal, Uniforme)

- Otimalidade do Risco médio.
- Estimador de Bayes - aquele que minimiza o risco médio com respeito a Λ

- θ como realização de uma v.a Θ
- dist. à priori e a posteriori
- Faz sentido tbm encontrar estimador que minimize risco à posteriori
- Decorre que ele tbm será o que minimiza o risco médio, def. antes
- levam ao estimador de Bayes

Estatística Avançada

Teoria da ~~Decisão~~, Aula 12

- Revisão distribuição a priori, a posteriori, Risco de Bayes
- Buscar δ' que minimize o Risco de Bayes
- Tudo que se conhece sobre Θ, após $X=x$ ter observado, está na sua dist. a posteriori
- Buscar δ'' t.q. o risco à posteriori $E[L(\Theta, \delta(x)) \mid X=x]$ seja mínimo
- Minimizar risco à post. $\Rightarrow$ minimizar risco de Bayes $\Rightarrow \delta'' = \delta'$
- Teorema: Sob perda quadrática, $\delta_\Lambda = E[g(\Theta) \mid X=x]$
- Teorema: Se perda estrit. convexa, então uma solução de δ_Λ é única.
- Exemplo. $X_i \mid \Theta \sim \text{Poisson}(\Theta)$; ~~priori~~ $\Theta \sim \text{Gama}(a,b) \Rightarrow \Theta \mid X=x \sim \text{Gama}\left(a+n\bar{x}, \frac{b}{1+nb}\right)$

  $\delta(X) = b(n\bar{X}+a)/(1+nb)$ sob perda quad.

  $X_i \mid p \sim \text{Binomial}$, $p \sim \text{Beta}(a,b) \Rightarrow p \mid X=x \sim \text{Beta}(a+x, b+n-x)$

  $\delta(X) = a+X/(a+b+n)$ sob. perda quad

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

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

- Teorema: Qqer estimador de Bayes, que é único é admissível.

  [δ admissível] se

  $\nexists\ \delta' : R(\Theta, \delta') \leq R(\Theta, \delta),\ \forall \Theta \in \Omega$ e

  $R(\Theta, \delta') < R(\Theta, \delta)$ para algum $\Theta \in \Omega$]

- Relação entre Estimador de Bayes e Suficiência
- Prioris impróprias ("Beta(0,0)",

Lista 3 - 4, 5, 10, 24, 25, 26

Capítulo 1 - Preparações.

Capítulo 2 - Não Viciosidade

Capítulo 3 - Equivariância

Aluno: Maicon Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

O Problema: Assume-se que os dados (observações) são valores observados de v.a. que seguem uma distribuição de prob. P pertencente a uma classe $\mathcal{P}$. Freq., P é indexada por um parâmetro $\Theta \in \Omega$, de modo que

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

Exemplo: Proporção de óleo bruto que se transf. em gasolina após proc.(X). $X \sim Beta(p,q)$, $\Theta \in \{(p,q) \in \mathbb{R}^2 : p>0, q>0\} = \Omega$
Obs. $X_1, \dots, X_n$ réplicas de X. $X_1 \dots X_n$ i.i.d, $X_i \sim Beta(p,q)$

Objetivo: Fazer inferência sobre Θ ou $g(\Theta)$ com base nos dados observados.

1. Estimação pontual e intervalar
2. Testes de hipótese - problema de decisão
3. Previsão de novas observações (Não será abordada)

1 (especificar um valor plausível para Θ);
2 (ou determinar um conjunto de Ω t.q. possamos declarar que de modo provável, este conterá ou não o valor de Θ).

Para inferir, os dados e a suposição das possíveis distribuições (Na clássica)
Do ponto de vista bayesiano, o parâmetro é assumido como uma v.a. (não observável). A dist. inicial (a priori) é atualizada a partir das observações para se obter a dist. a posteriori.

Estimação Pontual. (Ingredientes)

1. $g: \Omega \to \mathbb{R}$, cujo valor em Θ, $g(\Theta)$, será estimado (estimando)
2. Uma v.a (ou vetor a.) X, tomando valores em um espaço amostral $\mathcal{X}$ de acordo com a dist. $P_\Theta \in \mathcal{P}$. ($X = x$ é o conj. de dados)
3. Estimador: uma transf. dos dados δ usada p/ estimar $g(\Theta)$.

Objetivo: Obter δ a partir de 1 e 2.

Na vdd, foram 4 exemplos

1. Tempos de vida de lâmpadas ($T_1, \dots T_n$ réplicas de T, $T \sim exp(\theta)$)
2. $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$ ($X_1, \dots X_n$ réplicas de X)
4. $\underset{\sim}{Y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ (regr. normal linear)

   $\underset{\sim}{Y} = (Y_1, \dots Y_n)$

   $\underset{\sim}{X}$: matriz $n \times p$ (de posto p) fixo

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

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

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

5. Regressão logística.

   $\underset{\sim}{Y} = (Y_1, \dots, Y_n)'$ t.q. $Y_i \sim \text{Bernoulli}(\mu_i)$ ind.

   $$\log \frac{\mu_i}{1 - \mu_i} = \underset{\sim}{x_i} \underset{\sim}{\beta}, \quad i = 1, \dots, n$$

   $\underset{\sim}{x_i} = (x_{i1}, \dots, x_{ip})'$, fixados

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

Estatística Avançada I - Resumo 01-11, A1

Aluno: Maicon Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

Estimador: uma função δ, à valores reais, definida sobre $\mathcal{X}$. É usado para estimar um estimando $g(\theta)$. ($\delta(X)$ é o estimador de $g(\theta)$ e $\delta(x)$ a estimativa de $g(\theta)$). (No curso, será impostas restrições sobre $\delta(x)$, como por exemplo, $\delta(x) \in [0,1]$ - que faz sent. quando $\delta(x)$ é um estimador para $\theta \in [0,1]$, p. exemplo)

Desejável que δ "esteja perto" de $g(\theta)$; o que fomenta as medidas de proximidade ou distância entre δ e $g(\theta)$:

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

Função de Perda

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

$$\begin{cases} L(\theta, d) \geq 0, & \text{p/ todo } \theta, d \text{ e} \\ L(\theta, g(\theta)) = 0, & \text{p/ todo } \theta \end{cases}$$

Função de Risco: (perda média) $R(\theta, \delta) = E_\theta\{L(\theta, \delta(X))\}$.

. Não $\exists$ δ que minimiza $R(\theta, \delta)$ $\forall \theta \in \Omega$. (Seja $\delta(x) = g(\theta_0)$, $\forall x \in \mathcal{X}$). $R(\theta, \delta) = 0$ p/ $\theta = \theta_0$
$\forall \theta \in \Omega$, encontro $\delta(x)$ que tem risco mínimo.

Solução: Restringir a classe dos estimadores. Destes, encontrar o ideal (risco uniformemente mínimo). Restrição: (Não Viesados - $E_\theta(\delta(x)) = g(\theta)$, $\forall \theta \in \Omega$; Não Viciados com resp. à mediana $\text{Mediana}_\theta(\delta(x)) = g(\theta)$; Est. Bayes; minimax)

Es. Bayes: $\int_\Omega R(\theta, \delta) w(\theta) d\theta$, w é a fdp de θ: δ que miniza.

Minimax: δ que minimiza $\sup_\Omega R(\theta, \delta)$.

Modelo Estatístico:

Seria natural tomar como ponto de partida $(\mathcal{E}, \mathcal{B})$ e uma fam. de dist. de prob. def. sobre ele. Aqui $\mathcal{E}$ é o conjunto dos possíveis resultados de um experimento e $\mathcal{B}$ uma σ-álgebra de conjuntos de $\mathcal{E}$.

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

$$(\mathcal{E}, \mathcal{B}) \xrightarrow{X} (\mathcal{X}, \mathcal{G}, \mathcal{P})$$

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

$(\mathcal{X}, \mathcal{G}, \mathcal{P})$: $\mathcal{X}$: conjunto,

$\mathcal{G}$: σ-álg. de sub-cjtos de $\mathcal{X}$

esp. paramétrico

$\mathcal{P}$: família de medidas (dist.) de prob. $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$

Identificabilidade: para quaisquer $\theta_1 \in \Omega$ e $\theta_2 \in \Omega$,

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

Diremos que $(\mathcal{X}, \mathcal{G})$ é o espaço amostral. $\mathcal{G}$ será omitido (sub-entendido), e diremos $\mathcal{X}$ é o esp. amostral.

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

. μ é σ-finita se $\exists A_1, A_2 \dots \in \mathcal{G}$, t.q. $\cup A_i = \mathcal{X}$ e $\mu(A_i) < \infty$.

. $\mathcal{P}$ é dominada por μ se todos os el. de $\mathcal{P}$ são abs. cont. com resp. à μ, i.e., $\forall A \in \mathcal{G}$, $\mu(A) = 0 \Rightarrow P_\theta(A) = 0, \forall \theta \in \Omega$.

Tbm,

$P_\theta(A) = \int_A p_\theta \, d\mu$, $\forall \theta \in \Omega$ (int. de Lebesgue) sendo p_θ a der. R-N. com res à μ.

Estatística Avançada - Resumo

Aluno: Marcos Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

i. Caso Discreto: $p_\theta(A) = \int_A p_\theta \, d\mu = \sum_{x_i \in A} p_\theta(x_i)$, -

$\mathcal{X}$ e $\mathcal{G}$ (contável) e $\mathbb{P}$ dominada pela med. de contagem $\mu(A)$ é o nº de pontos de A, $A \in \mathcal{G}$

ii. Caso Contínuo: μ é a medida de Lebesgue.

$\mathcal{X} = \mathbb{R}^n$, $\mathcal{G} = \mathcal{B}(\mathbb{R}^n)$ = menor σ-álg. contendo todos os ret. abertos

$A = \{(x_1, \dots x_n) : a_i < x_i \leq b_i\}$, $-\infty < a_i < b_i < \infty$

$\mu(A) = (b_1 - a_1) \dots (b_n - a_n)$

Caso Absolutamente Cont. $\mathbb{P}$ é dominada pela medida de Lebesgue.

- Inferências sobre Θ são feitas com base em estatísticas.

. Estatística: variáveis (ou vetores) aleatórios, T digamos, que são transformações mensuráveis de $(\mathcal{X}, \mathcal{G})$ em algum $(\mathcal{J}, \mathcal{B})$ são chamadas estatísticas.

$\mathcal{B}$ é uma σ-álg. de sub-conj. de $\mathcal{J}$ t.q.

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

Em part., estimadores são estatísticas.

Estatística Avançada I - Resumo 12-20, A1

Aluno: Maicon Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

Famílias de Grupo: de distribuições: família obtida submetendo uma v.a. com uma dada distribuição a um grupo de transformações.

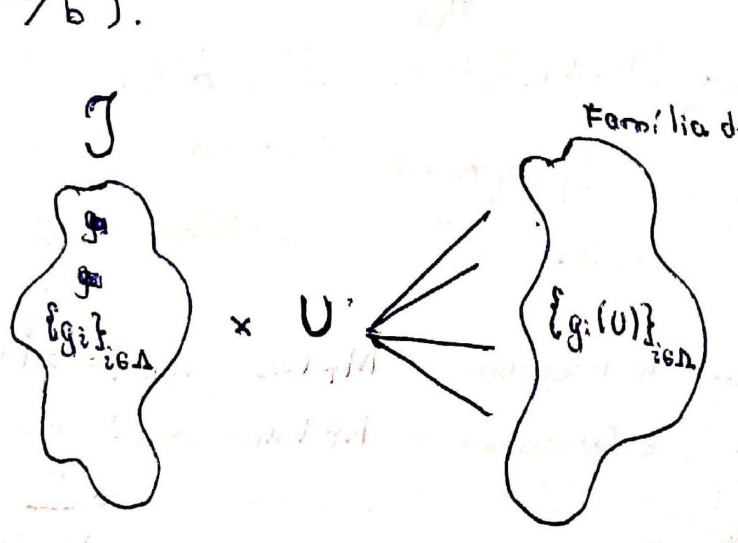
Grupo de Transformações: uma classe de transformações 1-a-1 ~~é um~~ "grupo" que é fechado sob composição e inversão, i.e.,

- $g_1 \in \mathcal{J},\ g_2 \in \mathcal{J} \Rightarrow g_2 \circ g_1 \in \mathcal{J}$
- $g \in \mathcal{J} \Rightarrow g^{-1} \in \mathcal{J}$.

Família de locação: É gerada pela transformação $X = U + a$, $a \in (-\infty, +\infty)$, U com dist. F. $\mathbb{P}(X \leq x) = F_U(x - a)$

Família de escala: É gerada pela transformação $X = bU$, $b > 0$.
$\mathbb{P}(X \leq x) = F_U(x/b)$,

Família de locação-escala: É gerada pela transformação $X = bU + a$, $a \in \mathbb{R}$, $b > 0$.
$\mathbb{P}(X \leq x) = F_U\left(\frac{x-a}{b}\right)$.



Família de grupo de dist.

Se $g_i(x) = x + a_i$, $a \in \mathbb{R}$
Fam. locação

Se $g_i(x) = b_i x$, $b_i > 0$,
Fam. escala

Se $g_i(x) = b_i x - a_i$,
Fam. loc.-escala.

Famílias Exponenciais: Uma fam. $\{P_\theta\}$ de dist. ~~da~~ forma uma família exponencial s-dimensional se as distribuições P_θ têm densidades da forma

$$p_\theta(x) = \exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - B(\theta)\right] h(x)$$ com resp. à alg. μ com...

$\forall \theta \in \Theta, \; p_\theta(x) = \exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - B(\theta)\right] h(x)$ ★

onde η_i's e B são funções dos parâmetros a val. reais, T_i's são funções (estatísticas) a val. reais definidas sobre $\mathcal{X}$, o sup. da densidade e x é um ponto de $\mathcal{X}$ e h é uma função def. sobre $\mathcal{X}$ a val. reais ñ negativos.

Forma Canônica (natural): os η_i's são vistos como parâmetros

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

$\eta = (\eta_1, \dots, \eta_s)$. $\eta_i = \eta_i(\theta)$, $B(\theta) = A(\eta)$ (reparametrização)

$B(\theta)$ reescrito em f. de η.

Para que ★★ seja dens., é prec. $\int_{\mathcal{X}} \exp\left[\sum^{s} \eta_i T_i(x)\right] h(x) d\mu(x) < \infty$ (○) $\Delta = \{\eta : \text{vale } ○\}$ esp. par. natural canônico.

Família de posto completo: 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.

Teorema ⇒ $E_\eta(T_j(X)) = \frac{\partial}{\partial \eta_j} A(\eta)$

$$Cov_\eta(T_j(X), T_k(X)) = \frac{\partial^2}{\partial \eta_j \partial \eta_k} A(\eta)$$

P/ qqer f integrável e qqer η no int. de Δ, a int. $\int f(x) ★★ \, d\mu(x)$ é cont. e tem der. de todas as ordens com resp. aos η_i's.

Função Geradora de Momentos: $M_T(w_1, \dots, w_n) = E(e^{w_1 T_1 + \dots + w_n T_n})$

Função Geradora de Cumulantes: $K_T(u_1, \dots, u_s) = \log M_T(u_1, \dots, u_n)$

$$= \sum_{r_1, \dots, r_s} K_{r_1, \dots, r_s} \frac{w_1^{r_1} \dots w_s^{r_s}}{r_1! \dots r_s!},$$

$$K_{r_1, \dots, r_s} = \frac{\partial^{r_1 + \dots + r_s}}{\partial u_1^{r_1} \dots \partial u_s^{r_s}} K_T(u_1, \dots, u_s) \Big|_{w = (0, \dots, 0)}.$$

Se X tem forma canônica

$$M_T(w) = \frac{\exp(A(\eta + w))}{\exp(A(\eta))} \; ; \; K_T(w) = A(\eta + w) - A(\eta).$$

Estatística Avançada - Resumo

Aluno: Maicon Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

Aula 03

Estatística Suficiente: 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.

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

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

$$p_\theta(x) = g_\theta(T(x))\, h(x) \quad (q.c.\ \mu)$$

Aula 04

Suficiência minimal: redução máxima dos dados sem perda de informação sobre θ.

Estatística Suficiente Minimal: 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 t.q. $T = H(U)$ (q.c. $\mathcal{P}$).

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

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

é suficiente minimal.

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

Extensão para $\mathcal{P}$ contável - Exercício!
Para $\mathcal{P}$ não contável, 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 ~~suf~~ suficiente minimal para $\mathcal{P}_0$ e suficiente para $\mathcal{P}$, então é suficiente minimal para $\mathcal{P}$.

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

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

$T(X) = \dfrac{p_{\theta_1}(X)}{p_{\theta_0}(X)}$ é suf. minimal $\Leftrightarrow T(X) = \bar{X}$

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

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

Teorema: Sob as condições do G.F, assuma que

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

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.

Ver Exemplo Logístico (pg 45)

Estatística Avançada - Resumo

Aluno: Maicon Aparecido Pinheiro Professora: Silvia Lopes de Paula Ferrari

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.

Ideia: Uma est. 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 suf. minimal, mas as dif. $X_{(n)} - X_{(i)}$, $i=1,\dots, n-1$ são ancilares (mostrar).

Ex: $\{X_i\}_{i=1}^n$ iid, $X_i \sim N(\theta, 1)$; $X_i \sim N(\theta, \sigma^2)$, $\sigma^2 > 0$ fixado

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

$\dfrac{\sum (X_i - \bar{X})^2}{n-1}$ é est. ancilar

Comentário: Exemplo X_1 e X_2 com mesma dist. com (R, M) uma est. suf. minimal, onde R é ancilar. O conhecimento (adicional) de que $R=2$ aumenta a inf. sobre θ.

Aula 05.

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

$$E_\theta[f(T)] = 0, \ \forall \theta \in \Omega \Leftrightarrow f(T) = 0 \ (q.c.\ \mathcal{P}),$$

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

Nota: Se existe uma suf. minimal, então uma condição necessária para que uma estatística seja completa é que ela seja suf. minimal.

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.

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.

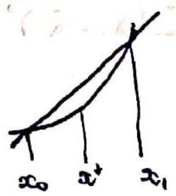
Rever ex. uniforme discreta!

Funções de Perda Convexas

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

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

$<$ (estritamente convexa)



Teorema.

i.

Estimador: uma função δ, à valores reais, definida sobre $\mathfrak{X}$. É usado para estimar um estimando $g(\theta)$. ($\delta(X)$ é o estimador de $g(\theta)$ e $\delta(x)$ a estimativa de $g(\theta)$.

Função de Perda: Suponha que as conseq. (perdas) de estimar $g(\theta)$ por um valor d de δ são medidas por $L(\theta,d)$ (f. de perda). Assumimos que $L(\theta,d) \geq 0$ $\forall \theta, d$ e $L(\theta, g(\theta)) = 0$ $\forall \theta$.

Função de Risco: (perda média) $R(\theta,\delta) = E_\theta\{L(\theta,\delta(X))\}$.

$(\mathcal{F},\mathcal{B}) \rightsquigarrow (\mathfrak{X}, \mathcal{C}, \mathcal{P})$, $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$, $\{\mathfrak{X},\mathcal{C}\}$ esp. am.

Identificabilidade: Para quaisquer $\theta_1 \in \Omega$ e $\theta_2 \in \Omega$, $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.

Estatística: variáveis (ou vetores) aleatórios, T digamos que são transformações mensuráveis de $(\mathfrak{X},\mathcal{C})$ em algum $(\mathcal{I},\mathcal{B})$.

Famílias de Grupo: de distribuições: família obtida submetendo uma v.a. com uma dada distribuição a um grupo de transformações.

Grupo de Transformações: uma classe de transformações 1-a-1 que é fechada sob composição e inversão, i.e., $g_i \in \mathcal{G}$, $i=1,2 \Rightarrow g_1 \circ g_2 \in \mathcal{G}$; e $g \in \mathcal{G} \Rightarrow g^{-1} \in \mathcal{G}$.

Família de locação: Seja U uma v.a. com dist. F. Sejam $a \in (-\infty,\infty)$ e $X = U + a$. $P(X \leq x) = F(x-a)$

O conjunto das dist. $F(x-a)$, F fixa, $a \in \mathbb{R}$, constitui uma família de locação.

Família de escala: É gerada pela transf. $X = bU$, $b > 0$

Família de locação-escala: Pela transf. $X = bU + a$, $b>0, a \in \mathbb{R}$

Famílias Exponenciais: Uma família $\{P_\theta\}$ de dist. forma uma família exponencial s-dimensional se as distribuições P_θ tem densidades da forma

$\star$ $p_\theta(x) = \exp\left[\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - B(\theta)\right] h(x)$ c.r.à μ.

Forma Canônica (natural) os η_i's são vistos como parâmetros

$p(x|\eta) = \exp\left[\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right] h(x)$, $\eta = (\eta_1, \dots, \eta_s)$,

$\eta_i = \eta_i(\theta)$, $B(\theta) = A(\eta)$. (reparametrização) $\star\star$

$\Delta = \{\eta : p(x|\eta)$ é dens.$\}$ esp. paramétrico natural (canônico)

Família de posto completo: 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.

Família Exponencial Curvada: Qdo os parâmetros naturais da dist. estão relacionados de uma forma ñ linear, dizemos $\star$ ou $\star\star$ forma uma família exponencial curvada.

Teo: f integ. e η no int. de Δ, $\int f(x) \exp\left[\sum \eta_i T_i(x)\right] h(x) d\mu(x)$ é contínua e tem der. todas as ordens c.r. aos η's. Decorre que $E_\eta\left[T_j(X)\right] = \frac{\partial}{\partial \eta_j} A(\eta)$; $\operatorname{cov}_\eta\left[T_j(X), T_k(X)\right] = \frac{\partial^2}{\partial \eta_j \partial \eta_k} A(\eta)$

Função Geradora de Momentos: $M_T(u_1, \dots, u_n) = E(e^{\sum u_i T_i})$

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

Função Geradora de Cumulantes: $K_T(u_1, \dots, u_n) = \log M_T(u_1, \dots, u_n)$

$= \sum_{r_1 \dots r_s} \kappa_{r_1 \dots r_s} \frac{u_1^{r_1} \dots u_s^{r_s}}{r_1! \dots r_s!}$; $\kappa_{r_1 \dots r_s} = \frac{\partial^{r_1+\dots+r_s}}{\partial u_1^{r_1} \dots \partial u_s^{r_s}} K_T(u_1 \dots u_s)\Big|_{u=(0,\dots,0)}$

$(s=1)$ $\kappa_1 = ET$, $\kappa_2 = \operatorname{Var} T$, $\kappa_3 = E(T-ET)^3$, $\kappa_4 = E(T-ET)^4 - 3[\operatorname{Var} T]^2$

Teorema: Se X tem densid. na forma $\star\star$, então p/ qq η no int. de Δ, as f.g. de mom. e cum. de T_i's existem (em alguma viz. da origem) e são dadas por $M_T(u) = \exp(A(\eta+u)) / \exp(A(\eta))$ e $K_T(u) = A(\eta+u) - A(\eta)$.

Estatísticas Suficientes: $T(X)$ é suf. p/ θ [ou p/ X, ou p/ a família $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$] se a dist. cond. de X dado $T=t$ é ind. de θ $\forall t$.

Teo (Critério da Fatoração) Uma cond. nec. e suf. para que T seja suf. p/ a fam. $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ de dist. de X dominada por uma μ σ-finita é que existam funções g_θ e h t.q. as dens. p_θ de P_θ satisfaçam $\left[p_\theta(x) = g_\theta(T(x)) h(x)\right]$ (q.c. μ)

Suficiência Minimal: T é suf. minimal se p/ qqer est. suf U, existe uma f. H t.q. $T = H(U)$ (q.c. $\mathcal{P}$) [é suf. e, entre todas as suf., é aquela que produz maior red. nos dados]

Teorema. Seja $\mathcal{P}$ uma fam. finita com dens. p_i, $i=0,\dots,k$ todas com mesmo suporte. Então, a est.

$T(X) = (p_1(X)/p_0(X), p_2(X)/p_0(X), \dots, p_k(X)/p_0(X))$ é suf. mín.

Lema: $\mathcal{P}$ fam. de dist. com sup. comum e $\mathcal{P}_0 \subset \mathcal{P}$, e se T é suf. minimal p/ $\mathcal{P}_0$ e suf. p/ $\mathcal{P}$, então é suf. min. para $\mathcal{P}$.

Teorema: Sob as condições do C.F. assuma que $T(x) = T(y) \Leftrightarrow y \in D(x)$, onde

$D(x) = \{y \in \mathfrak{X} : p_\theta(y) = p_\theta(x) h(x,y), \forall \theta \in \Omega\}$ p/ alguma função h não negativa. Então $T(X)$ é suf. minimal.

Ancilaridade: $V(X)$ é dita ser ancilar p/ a fam. $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ de possíveis dist. p/ X se sua dist. ñ dep. de θ. É dita ancilar de 1ª ordem se $E_\theta[V(X)]$ ñ depende de θ.

Completividade: Uma est. T é completa se, p/ qqer função f de T a val. reais

$E_\theta[f(T)] = 0$, $\forall \theta \in \Omega \Leftrightarrow f(T) = 0$ (q.c. $\mathcal{P}$)

Se existe uma est. suf. min. então uma cond. necessária p/ que uma est. seja comp. é q. ela seja suf. minimal

T. Basu. Se T é suficiente completa então qqer est. ancilar V é independente de T.

Teorema. Se $X \sim F \in$ s-dim de posto completo, então $T = (T_1(X), \dots, T_s(X))$ é suf. completa.

Função Convexa: ϕ é convexa se

$\phi(\gamma x + (1-\gamma) y) \leq \gamma \phi(x) + (1-\gamma)\phi(y)$, $a < x < y < b$, $0 < \gamma < 1$

ou se $\phi'(x) \leq \phi'(y)$. (ou $\phi''(x) \geq 0$).

Desigualdade de Jensen. se ϕ convexa em I e $P(X \in I) = 1$ e tem esp. finita $\Rightarrow \phi(EX) \leq E(\phi(X))$

Teo. Rao-Blackwell: $X \sim P_\theta \in \mathcal{P}$, T est. suf. δ estimador de $g(\theta)$, $L(\theta,d)$ função de perda estrit. convexa em d $\Rightarrow$ se δ tem média finita e $R(\theta,\delta) = E[L(\theta,\delta(X))] < \infty$ e se $\eta(t) = E[\delta(X) / T = t] \Rightarrow R(\theta,\eta) < R(\theta,\delta)$, a menos que $\delta = \eta$ com prob. 1.

Admissibilidade: um est. é dito ser inadmissível se existe outro est. δ' que o domina, i.e., t.q. $R(\theta,\delta') \leq R(\theta,\delta)$

p/ todo θ, com desi. estrit. p/ algum θ e é admissível se δ' ñ existe.

F. Perda Quadrática: $L(\theta,\delta)=(\delta-\theta)^2$, $R(\theta,\delta)=E_\theta(L(\theta,\delta))$

Def: (ñ viciado). Um est. $\delta(X)$ de $g(\theta)$ é dito ser ñ viciado (ñ viesado) se $E_\theta(\delta(X))=g(\theta)$, p/ todo $\theta\in\Omega$.

& Se $\exists$ est. ñ viciado de g, diz que g é estimável.

Lema: Se δ é qqer est. ñ viesado de $g(\theta)$, a tot. dos est. ñ vies. é dado por $\delta=\delta_0-U$, onde rep. qqer est. ñ viesado do zero ($E_\theta(U)=0, \forall\theta\in\Omega$)

Sob $L(\theta,d)=E_\theta(\delta-\theta)^2$, $R(\theta,\delta)=Var_\theta(\delta)=Var_\theta(\delta_0-U)=$ ~~...~~ $=E_\theta[(\delta_0-U)^2]-g(\theta)^2$ minimizado qdo min. $E_\theta[(\delta_0-U)^2]$.

ENNVUM: $V_\theta(\delta(X))\leq V_\theta(\delta^*(X))$, δ^*: qqer outro ñ viesado de $g(\theta)$. "O ENVVUM é único"

ENVVLM: δ não viciado de v. localmente mínima em $\theta=\theta_0$ se $Var_{\theta_0}(\delta(X))\leq Var_{\theta_0}(\delta'(X))$ p/ qqer outro $\delta'(X)$ ñ viciado.

Teorema. Seja $X\sim P_\theta$, $\theta\in\Omega$, S um est. em Δ, a classe dos est. com $E_\theta(\delta^2)<\infty$ $\forall\theta\in\Omega$; $\mathcal{U}$ o cjto de todos os est. ñ viciados do zero em Δ. Então, uma cond. nec. e suf. p/ que δ seja um ENVVUM de sua esp. $g(\theta)$ é que $E_\theta(\delta U)=0$, p/ todo $U\in\mathcal{U}$ e todo $\theta\in\Omega$.

Como $E_\theta(U)=0$, a cond. acima eq. à δ e U serem ñ correlacionad.

Lema: X - v.a $\sim P_\theta\in\mathcal{P}$, T est. suf completa p/ $\mathcal{P}$ $\Rightarrow$ toda função est. $g(\theta)$ tem um e som. um est. ñ viciado que é função de T.

Teo. v.a. $X\sim P_\theta\in\mathcal{P}$, T est. suf completa p/ $\mathcal{P}$:

(a) $\forall g(\theta)$ estimável $\exists$ um estimador ñ viesado que minimiza o risco unif. p/ qqer f. de perda $L(\theta,d)$, convexa em d. Este est. é o ENVVUM.

(b) o ENVVUM em (a) é o único estimador ñ viesado com risco min desde que seu risco seja finito e L estritamente convexa em d.

Corolário: Se $\mathcal{P}$ é uma F.E. de posto completo, s-dimensional, então o teo. ant. vale p/ $\theta=(\theta_1\ldots\theta_s)$ e $T=(T_1\ldots T_s)$.

Métodos p/ obter o ENVVUM

1. Se T é suf. completa $\forall$ função est. $g(\theta)\Rightarrow$ o ENVVUM é único determinado pela equação (Resolu. Eq. em δ)

$$E_\theta(\delta(T))=g(\theta) \text{ p/ todo } \theta\in\Omega.$$

2. (Condicionamento) Toma-se um est. ñ viciado p/ $g(\theta)$ e o ENNVUM é obtido como a esp cond. de $\delta(X)$ dado T, uma est. suf. completa.

. Se $g(\xi)$ é fc estimável de ξ e é pol. grau r em ξ, ENVVUM tbm é pol. grau r em $\bar{x}$ (fam. expon.)
prova por indução

EMV Para cada $x\in\mathcal{X}$, $p_\theta(x)$ vista como função de θ é chamada de func de veross. ($L(\theta,x)$).

$\hat{\theta}\in\Omega$ t.q. $L(\hat{\theta};x)=\sup_{\theta\in\Omega} L(\theta,x)$ é chamado de estimador de máxima verossimilhança de θ.

X tem dis $P_\theta, \theta\in\Omega\subset\mathbb{R}^k$.

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

Vetor escore: $(\partial/\partial\theta_1, \ldots, \partial/\partial\theta_k)$

Lema: $\hat{\theta}=\hat{\theta}(X)$ um EMV de θ, $\hat{\theta}$ único $\Rightarrow$ se $\hat{\theta}$ é uma est. suf. $\Rightarrow$ ela é suf. minimal

Lema: Se δ é um est. ñ viesado de $g(\theta)$, $\Theta\to\mathbb{R}$, e se $\delta^*=\delta+b$, onde $b\neq0$ é o vies e é ind. de $\theta\Rightarrow\delta^*$ tem risco sob perda quadrática uniform. maior: $R_{\delta^*}(\theta)=R_\delta(\theta)+b^2$.

Desigualdade da Informação.

$$\psi(x,\theta)=\frac{\partial}{\partial\theta}p(x,\theta)\frac{1}{p(x,\theta)}=\frac{\partial}{\partial\theta}\log p(x,\theta) \text{ v.a. dep. de } \theta$$

$$J(\theta)=E_\theta\left[\psi(X,\theta)^2\right]\overset{*}{=}E_\theta\left[-\frac{\partial^2\log p(X,\theta)}{\partial\theta^2}\right]$$

De modo geral, se δ é ENV de $g(\theta)$, então sob certas cond. adequadas (*), $Var_\theta(\delta)\geq\frac{[g'(\theta)]^2}{J(\theta)}$, (D.I)

Teo. δ tem 2° momento finito $\forall\theta\in\Omega$. Uma cond. nec. e suf. para que $Cov_\theta(\delta,\psi)$ dep. de δ somente através de $g(\theta)=E_\theta(\delta)$ $\forall\theta$ é que $Cov(U,\psi)=0$ $\forall U\in\mathcal{U}=\{U: E_\theta(U)=0 \text{ e } E_\theta(U^2)<\infty, \forall\theta\in\Omega\}$.

$\Rightarrow$ δ_1 e δ_2 com $E_\theta(\delta_1)=E_\theta(\delta_2)\forall\theta\Rightarrow Cov(\delta_1,\psi)=Cov(\delta_2,\psi)$

Suposições: (a) Ω é um int. aberto (finito, inf. ou semi-inf.) (b) $P_\theta, \theta\in\Omega$ tem suporte comum t.q. $A=\{x, p_\theta(x)>0\}$ é ind. de θ. (c) Para qqer $x\in A$ e $\theta\in\Omega$,

$$p'(\theta)=\frac{\partial p_\theta(x)}{\partial\theta} \text{ existe e é finita.} \quad *$$

Lema: Se (a)-(c) estão satisfeitas e a derivada com resp. à θ do lado esq. de $\int p_\theta(x)d\mu(x)=1$ pode ser obtida dif. sob o sinal de integração então

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

$$J_\theta=Var_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta(x)\right]$$

. Se, adicional, a 2ª derivada com resp à θ de $\log p_\theta(x)$ existe p/ todo $x\in A$ e $\theta\in\Omega$ e a 2ª derivada c. resp. a θ do lado esq. de * pode ser obt. dif. 2 vezes sob sinal de integração então

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

Nota: Na família exp. unidimensional, $p_\eta(x)=\exp\{\eta T(x)-A(\eta)\}h(x)$. Se $A(\eta)$ é dif. a troca de int. e deriv. c/ resp. à η em $\int p_\eta(x)d\mu(x)$ é válida.

Nota: $I(\theta)$ depende da parametrização utilizada. Se $\theta = h(\xi)$ então

$$I^*(h(\xi)) = I(h(\xi))\, h'(\xi)^2, \quad h \text{ diferenc.}$$

Teorema. $X \sim FE$ unid., $\tau(\theta) = E_\theta(T)$, então

$$I[\tau(\theta)] = 1/Var_\theta(T).$$

+ geral. $I(h(\theta)) = \left[\frac{h'(\theta)}{\tau'(\theta)}\right]^2 Var_\theta(T)$, $h(\theta)$ dif.

Família de locação: X tem dens. $f(x-\theta)$; $f(x) > 0$, $\forall x \in \mathbb{R}$, $\theta \in \Omega = \mathbb{R}$. As condições (a)-(c) estão satisf. desde que $f'(x)$ exista p/ todo x.

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

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

não dep. de θ.

Família Locação escala: X tem dens. $b^{-1} f(b^{-1}(x-\theta))$

$\Rightarrow I(\theta) = I_f / b^2$.

Teo. Sejam X e Y v.a. ind. com dens. p_θ e q_θ (com resp. à μ e ν), satisf. (a)-(c) e

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

se $I_1(\theta)$, $I_2(\theta)$ e $I(\theta)$ são as inf. de θ contidas em X, Y e (X,Y) resp. então

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

Corolário: válido p/ $X_1 \dots X_n$, $I(\theta) = n I_1(\theta)$.

Teorema: Desig. da Inf. Sup. que $\{p_\theta, \theta \in \Omega\}$ é uma fam. de dens. dominada por μ p/ qual (a)-(c) valem e $E_\theta[\partial/\partial\theta \log p_\theta] = 0$ e t.q. $I(\theta) > 0$. Seja δ um est. t.q. $E_\theta(\delta^2) < \infty$, p/ a qual $E_\theta(\delta)$ existe e pode ser diferenciada sob sinal de int., i.e.

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

$$Var_\theta(\delta) \geq \frac{\left[\partial/\partial\theta\, E_\theta(\delta)\right]^2}{I(\theta)}$$

(*) Se $E(\theta) = \theta$, então $Var_\theta(\delta) \geq 1/I(\theta)$ (Limite inferior de Cramer-Rao).

(*) Se δ estimador de $g(\theta)$ com $E_\theta(\delta) = g(\theta) + b(\theta)$

$$\Rightarrow Var(\delta) \geq \frac{[b'(\theta) + g'(\theta)]^2}{I(\theta)}$$

(*) $X_1, \dots, X_n$ iid, $\delta(x) : \delta(X_1, \dots, X_n)$

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

Nota: O LICR é atingido, ou seja

$$Var_\theta(\delta) = \frac{[\partial E_\theta(\delta)/\partial\theta]^2}{I(\theta)} \Leftrightarrow$$

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

$$\Rightarrow \delta = a(\theta)\frac{\partial}{\partial\theta}\log p_\theta(x) + b(\theta) \text{ p/ al. } a, b \text{ atinge LICR}$$

Nota: Integrando ambos os lados de

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

pode-se mostrar que, sob cond. adequadas,

$$p_\theta(x) = c(\theta)\exp\{\psi(\theta)\delta(x)\}h(x)$$

(fam. exp. unid. com $\delta(x)$ est. suficiente).

Condições de regularidade. Para val. D.I. foi exigido que $E_\theta(\delta^2) < \infty$ (cond. ñ restritiva) e que

$$\frac{\partial}{\partial\theta}E_\theta(\delta(x)) = \int \frac{\partial}{\partial\theta}\delta(x)\, p_\theta(x)\, d\mu(x) = g(\theta) \quad (\circledast)$$

Lema: (a)-(b) sejam vál. e seja δ um est. $E(\delta^2) < \infty$. Seja $q_\theta(x) = \partial/\partial\theta \log p_\theta(x)$ e, p/ cada $\varepsilon > 0$, b_θ uma função que satisfaz

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

$\forall |\Delta| \leq \varepsilon$. Então $E_\theta[q_\theta(X)] = 0$ e $\partial/\partial\theta\, E_\theta(\delta(X)) = E_\theta(\delta(X) q_\theta(X)) = cov_\theta(\delta, q_\theta)$ e, portanto, vale $(\circledast)$

Nota: A condição $(\ast\ast)$ é válida na fam. exp. unidim.

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

Supos:

(a) Ω é ret. aberto, (b) P_θ têm sup. comum, def. forma que $A = \{x, p_\theta(x) > 0\}$ é ind. de θ, (c) P/ ∀ $x \in A$, $\theta \in \Omega$ e $i = 1 \dots s$, $\partial p_\theta(x)/\partial\theta_i$ existem e são finitas

Matriz Informação

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

vet. escore

$= Cov_\theta(\partial/\partial\theta \log p_\theta(x))$

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

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

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

$\Rightarrow I^*(\xi) = J I' J$, $J = \partial\theta'/\partial\xi$

Teo. $\tau = (\tau_1, \dots \tau_s)$, $\tau_i = E(T_i(x))$, $i = 1 \dots s$. Então $I(\tau) = C^{-1}$, C a matriz cov. de $T_1 \dots T_s$

$N(\xi, \sigma^2)$

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

Teo. Des. Inf. $E_\theta\left[\frac{\partial}{\partial\theta_i}\log p_\theta(x)\right] = 0$ e Var. dif.

e $Var_\theta(\delta) \geq \alpha' I^{-1}(\theta)\alpha$, $\alpha_i = \partial E_\theta[\delta(x)]/\partial\theta_i$

(a-c), $I(\theta)$ pos. def., δ unidimens. $E(\delta^2) < \infty$.

Nota: O Limite inf. ating. ap. na família exponencial

Def. $f(x|\xi)$ e $L(\xi, d)$ são inv. por locação se respectivamente $f(x'|\xi') = f(x|\xi)$ e $L(\xi, d) = L(\xi', d')$ sempre que $x' = x + a$, $\xi' = \xi + a$ e $d' = d + a$.

Def. Um estimador satisfazendo $\delta(x_1+a,\dots,x_n+a) = \delta(x)+a$ sob as transf. $X_i' = X_i+a$, $\xi' = \xi+a$ e $d' = d+a$ é chamado equivariante por locação

Teo. $X = (X_1,\dots,X_n) \sim f(x-\xi)$, $\xi = (\xi_1 \dots \xi_n)$, δ equiv. por locação com perda $L(\xi,d) = \rho(d-\xi) \Rightarrow$ o viés, o risco e a variância de δ são cte ($\tilde{n}$ dep. de ξ).

Def. Estimador equiv. de risco mínimo (EMR)

Lema: δ_0 equiv. por locação $\Rightarrow$ uma condição necessária e suf. para δ ser equiv. por locação é que: $\delta(x) = \delta_0(x) + u(x)$, $u(x) = u(x+a)$. $\forall x, a$.

Lema. $u(x+a) = u(x) \Rightarrow$ é uma função dos $y_i = x_i - x_n$ ou um cte se $n=1$ ($u(x)$ inv. por locação é ancilar).

Teo. Se δ_0 equiv. por locação $\Rightarrow$ uma condi. necessária e suf. p/ δ ser equiv. por locação é que $\exists$ uma função V de $n-1$ argumentos t.q. $\delta(x) = \delta_0(x) - V(y)$, $\forall x$

(*) Teo. $X \sim f(x-\xi)$, $Y_i = X_i - X_n$, $L(\xi,d) = \rho(d-\xi)$, δ_0 est. equiv. de risco finito, p/ cada y $\exists$ $V(y) = V^*(y)$ que minimiza $E_0[\rho[\delta_0(x) - V(y)] | y]$ $\Rightarrow \exists$ um est. por locação de risco min $= \delta^*(x) = \delta_0(x) - V^*(y)$.

Corol. ρ convexa e $\tilde{n}$ monot. $\Rightarrow \exists$ um ERM de ξ e é único se ρ estrit. convexa.

Teorema. Sob sup. de ($\tilde{*}$) com $L(\xi,d) = (d-\xi)^2$, o estimador

(A) é

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

Nota: Sob perda quadrática, o estimador ERM é
$\delta^*(x) = \delta_0(x) - E_0[\delta_0(x)|Y]$, sendo A $\delta_0(X)$ um est. equiv.

Unif. $(\xi - b/2, \xi + b/2)$ - A $= Y_{(1)} + Y_{(n)}/2$.

ENVVUM - fam exponen.
EAM - contexto + amplo - grupo de transformações

Lema: f. perda quadr., δ est. equiv. com viés $\neq b$, cte, e risco $<\infty$.
(a) $\delta(x) - b$ é equiv. e tem risco $\leq \delta(x)$,
(b) o único est. ERM é não viesado
(c) se ENVVUM existe e equiv $\Rightarrow$ É o ERM

$\Rightarrow X_1 \dots X_n$ a.a. de uma família de locação esc. com f. de dist. $F((x-a)/b)$ $(a \in \mathbb{R}, b>0)$. Sejam $Y_j = \phi_j(x)$, $j = 1,\dots,t$. q. $\phi_j =$ ... $\phi = (x_1-\alpha/\beta,\dots,x_n-\alpha/\beta)$ $\forall \alpha,\beta$. $(Y_1 \dots Y_n)$ é ancilar.

$\Rightarrow X_1 \dots X_n$ de $F((x-a)/b) \Rightarrow$ se b conhec. $(X_1 - x_i)/b$, $i = 1 \dots n$ são ancilares. Se a é conh. $(X_1 - X_1)/(X_2 - X_1)$ $i = 3 \dots n$ são ancil.

$\Rightarrow \bar{X}$ e $\Sigma(X_i - \bar{X})^2$ são ind. se $X_i \sim N(\theta,\sigma^2)$
$X_{(1)}$ e $\Sigma(X_i - X_{(1)})$ ... $X_i \sim E(a,b)$

$\Rightarrow$ A subcj. de $\mathcal{X}$; P_θ^A é P_θ trunc. sob A
. se T é suf p/ $P \Rightarrow$ é suf por P^+
. se T é com. p/ $P \Rightarrow$ é com. para P^+

$\Rightarrow Var(\delta(X)) = Var[E(\delta(X)|T)] + E[Var(\delta(X)|T)]$

$\Rightarrow$ se $f_\theta(x) = P_\theta(X=x) = a(x)\theta^x / c(\theta)$, $x = 0,\dots$, $a(x) \geq 0$, $\theta > 0$, $M_X(u) = c(\theta e^u)/c(\theta)$.

$\Rightarrow X \sim F.E \Rightarrow E_\theta(T(X)) = B'(\eta)/\eta'(\theta)$

$\Rightarrow \Sigma(X_i - \bar{X})^2/\sigma^2 \sim \chi^2_{n-1}$

$\Rightarrow X_1 \dots X_n \sim E(a,b)$, $X_{(1)} \sim E(a, b/n)$

$\Rightarrow X_i \sim E(a,b) \Rightarrow Y_i = (n-i+1)[X_{(i)} - Y_{(i-1)}] \overset{ind}{\sim} E(0,1)$
$\Sigma(X_{(i)} - X_{(1)}) \sim Gama(n-1,1)$.

$\Rightarrow X \sim Pa(a,b) \Rightarrow \hat{c}_{mv} \sim P(na, c)$ e $2na/\hat{a}_{mv} \sim \chi^2_{2(n-1)}$ $V_i = a(\log X_i - \log c) \sim E(0,1)$

$\Rightarrow$ Se $E(\delta) = g(\theta)$ e $Var(\delta)$ at. lim da des. da Inf. $\Rightarrow \delta(x) = g(\theta) \pm \frac{g'(\theta)}{I(\theta)} \frac{\partial \log f(x)}{\partial\theta}$

$\Rightarrow$ famil escala:
$f_X(x;\tau) = \frac{1}{\tau^n} f(x/\tau)$, $\tau > 0$, $x \in \mathbb{R}^n$

$\delta(bx) = b^r \delta(x) \Rightarrow \delta$ é equi. por escala e $v(cx) = v(x) \geq 0$. $\forall c > 0$ $v(x) = w(z)$

$\Rightarrow$ ERM de τ^r
$$\delta^*(x) = \delta_0(x) \frac{E_1[\delta_0(x)|y)}{E_1(\delta_0^2(x)|y)}$$

-Est. Pitman
$$\delta^*(x) = \frac{\int_0^{+\infty} v^{n+r-1} f(vx_1 \dots vx_n)\,dv}{\int_0^{\infty} v^{n+2r-1} f(vx_1 \dots vx_n)\,dv}$$
é est. eq. por escal risco mínimo sob perda $\left(\frac{d-\tau^r}{\tau^r}\right)^2$

Marcon A. Pinheiro

Otimalidade do Risco Médio

Contexto: Est. de $g(\theta)$ a valores reais, $\theta \in \Omega$, X uma v.a. com dist. $P_\theta \in \mathcal{P}$, $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$.
δ um est. de $g(\theta)$; $\delta = \delta(X)$
$L(\theta,d)$: função de perda: $L(\theta,d) \geq 0 \ \forall \theta, d$, $L(\theta, g(\theta)) = 0 \ \forall \theta$. (mede as cons. de estimar $g(\theta)$ pelo valor $d = \delta(x)$).

$R(\theta,\delta) = E_\theta\{L(\theta,\delta(X))\}$, $\theta \in \Omega$, é a **perda média**, cg. a dist. de X e dado o valor de θ.
$R(\theta,\delta) = \int L(\theta,\delta(x))\, dP_\theta(x)$

Restrição: Procurar estimador com **risco médio mínimo**, obtido usando uma f. de peso (em θ).

Risco Médio: (Risco de Bayes): Seja $R(\theta,\delta) < \infty$, $\forall \theta \Rightarrow$ $r(\Lambda,\delta) = \int R(\theta,\delta)\, d\Lambda(\theta)$, em que $\int d\Lambda(\theta) = 1$ depende de δ e da distribuição de prob. Λ. Não depende de θ.

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

Densidade marginal de X, $g(x)$ ou $q(x/\lambda)$:
$$q(x) = \int f(x/\theta)\, d\Lambda(\theta)$$

Densidade à posteriori de θ.
$$\pi(\theta/x) = f(x/\theta)\pi(\theta)/q(x) \quad \text{ou}$$
$$\pi(\theta/x,\lambda) = f(x/\theta)\,\pi(\theta/\lambda)/q(x/\lambda)$$

$\Rightarrow$ Faz sentido buscar δ t.q. $E\{L(\Theta,\delta(x))/X=x\}$ seja mínimo, será o est. de Bayes.
Risco à posteriori: $E\{L(\Theta,\delta(x))/X=x\}$.

Teorema: Suponha que $\Theta \sim \Lambda$ e, dado $\Theta = \theta$, que $X \sim P_\theta$. Se quer estimar $g(\Theta)$ com f. de perda $L(\theta,d)$. Se
(a) existe um estimador δ_0 com risco finito
(b) p/ quase todo x (exceto p/ $x \in A$ t.q. $Q(A) = 0$, Q a dist. marginal de X), existe um $\delta_\Lambda(x)$ que minim. $E\{L(\Theta,\delta(x))/X=x\}$,

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

Corolário: Sob as condições do Teorema, se
$$L(\theta,d) = w(\theta)[d - g(\theta)]^2, \ w(\theta) > 0, \ \forall \theta \in \Omega$$
então
$$\delta_\Lambda(x) = \frac{E[w(\Theta)g(\Theta)/X=x]}{E[w(\Theta)/X=x]}$$
. Se $w(\theta) = 1$, $\forall \theta \in \Omega$ (perda quadrática),
$\delta_\Lambda(x) = E[g(\Theta)/X=x]$ (média a post. de $g(\theta)$)

Corolário: Se a função de perda é estritamente convexa em d (em particular quadrática), então uma est. 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. $\mathcal{P}$.

. $X_i/\Theta \sim Poi(\theta)$, $\theta > 0$.
$\Theta \sim Gama(a,b)$, $a,b > 0 \Rightarrow \Theta/X=x \sim Gama(a + n\bar{x}, b/(1+nb))$
$L_k(\theta,\delta) = (\theta-\delta)^2/\theta^k \Rightarrow \delta^{(k)} = \dfrac{E[\Theta^{-k+1}/X=x]}{E[\Theta^{-k}/X=x]} = b \cdot \dfrac{n\bar{X} + a - k}{1+nb}$.
$\pi(\theta) \propto \theta^{a-1} e^{-\theta/b}$.

$X/p \sim Bin(p,n)$; $p \sim Beta(a,b)$, $a,b > 0$. $p/x \sim Beta(a+x, b+n-x)$

• $X_i \sim N(\theta,\sigma^2)$, $\sigma^2 > 0$, $\Theta \sim N(\mu, b^2)$
$$\Theta/X=x \sim N\left(\frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}, \frac{1}{n/\sigma^2 + 1/b^2}\right)$$

Distribuições a priori conjugadas: priori e posteriori na msm família de distribuições.

Teorema: Qualquer estimador de Bayes, que é único é admissível [se quaisquer 2 estim. de Bayes são iguais q.c. $\mathcal{P}$]

Est. de Bayes e Suficiência: Seja $T = T(x)$ uma est. suf., então
$f(x/\theta) = L(\theta/x) = g(t/\theta)h(x) \Rightarrow$
$\pi(\theta/x) = g(t/\theta)\pi(\theta) / \int g(t/\theta')\pi(\theta')d\theta'$
↳ só depende de x através de t
Todas as "qtdes" bayesianas são funções de uma estatística suficiente minimal.

Prioris Impróprias: Λ: medida t.q. $\int d\Lambda(\theta) = +\infty$
$Beta(0,0) \Rightarrow x = 0$, $\pi(\theta/x) \propto (1-p)^{n-1}/p$.

Def: Estimador de Bayes Generalizado: δ_Λ é um est. de Bayes general. c.r. a uma med. Λ (msm se não for de prob.) se a perda esperada a posteriori
$$E[L(\Theta,\delta(x))/X=x]$$
é minimizada em $\delta(x) = \delta_\Lambda(x) \ \forall x$.
• $\Theta \sim U(\mathbb{R})$, $X_i/\Theta \sim N(\theta,\sigma^2)$, $\Theta/x \sim N(\bar{x}, \sigma^2/n)$ própria.
$\Rightarrow \bar{X}$ est. Bayes general (perda quadrática).

Limite de Estimadores de Bayes: δ é lim. Es. Bayes se existe uma seq. de dist. a priori Λ_ν e estimadores de Bayes δ_{Λ_ν} tais que $\delta_{\Lambda_\nu}(x) \to \delta(x)$ p/ q. todo x (c.r. $f(x/\theta)$) qdo $\nu \to \infty$.
• Ex: Norm.; $\Theta \sim N(\mu, b^2)$. $\lim_{b\to\infty} \dfrac{n/\sigma^2}{n/\sigma^2 + 1/b^2}\bar{x} + \dfrac{1/b^2}{n/\sigma^2 + 1/b^2}\mu = \bar{x}$

Informação de Kullback-Leibler: p/ discriminação entre densidades g e f
$$K[f,g] = \int \log\left[\frac{f(t)}{g(t)}\right] f(t)\, dt \geq 0.$$
qto maior K, mais fácil discriminar entre f e g.
Prioris com pouco efeito sobre $\pi(\theta/x) \Rightarrow K[\pi(\theta/x),\pi(\theta)]$ grande, os dados tem mta influência sobre $\pi(\theta/x)$.
Problema: $K[\pi(\theta/x),\pi(\theta)]$ depende de x.

Informação de Shannon:
$S(\pi) = \int K(\pi(\theta/x),\pi(\theta))\, q_\pi(x)dx$, onde
$q_\pi(x) = \int f(x/\theta)\pi(\theta)d\theta$

Priori de Jeffreys: $\pi(\theta) \propto [I(\theta)]^{1/2}$, maximiza
$$S_n(\pi) = \frac{1}{2}\log\frac{n}{2\pi e} + \int \pi(\theta)\log\frac{I(\theta)^{1/2}}{\pi(\theta)}d\theta + o(1)$$
Caso vetorial: $\pi(\theta) \propto \sqrt{|I(\theta)|}$

Marcos A. Pinheiro, Parte 2.

Estimação Minimax

Estimador Minimax: estimador δ^M de θ que minimiza o risco máximo, ou seja,

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

- Um est. minimax é o est. de Bayes p/ pior dist. a priori possível (prior menos favorável)

Prior menos favorável: Uma dist. a priori Λ é menos favorável se $r_\Lambda \geq r_{\Lambda'}$, $\forall$ dist. a priori Λ'.

Teo. (Rel. est. Bayes e minimax) Suponha Λ uma dist. sob Ω t.q. $r(\Lambda, \delta_\Lambda) = \sup_\theta R(\theta, \delta_\Lambda)$, onde δ_Λ é o est. de Bayes c.r. a Λ. Então

(i) δ_Λ é minimax

(ii) se δ_Λ é a única sol. de B. c.r. a Λ ele é o único est. minimax

(iii) Λ é menos favorável.

* = risco médio iguala-se o seu máximo = Função risco constante.

Corol. Se a sol. de Bayes δ_Λ tem risco constante, ela é minimax.

Corol. Seja $\omega_\Lambda = \{\theta \in \Omega : R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda)\}$, então δ_Λ é minimax se $\Lambda(\omega_\Lambda) = 1$.

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

Def: Uma seq. de dist. a priori $\{\Lambda_n\}$ é menos favorável se p/ cada dist. a priori Λ, $r_\Lambda \leq r = \lim_{n\to\infty} r_{\Lambda_n}$; $r_{\Lambda_n} = \int R(\theta, \delta_n) d\Lambda_n(\theta)$

Teorema: Suponha que $\{\Lambda_n\}$ seja uma seq. de dist. a priori com riscos de Bayes r_{Λ_n} satisfazendo $\lim_{n\to\infty} r_{\Lambda_n} = r$ e que δ seja um estimador p/ o qual $\sup_\theta R(\theta, \delta) = r$, então

(i) δ é minimax

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

Lema: Se δ_Λ é o est. de Bayes de $g(\theta)$ com resp. a Λ e se $r_\Lambda = E\{[\delta_\Lambda(x) - g(\theta)]^2\}$ é o risco de Bayes, então $r_\Lambda = \int var[g(\theta)/x] \, dQ(x)$

Em particular, se a var a posteriori de $g(\theta)$ é ind. de X, então $r_\Lambda = var[g(\theta)|x]$.

Lema: Se δ é minimax de $g(\theta)$ qdo $\theta \in \Omega_0 \subset \Omega_1$, então se $\sup_{\theta \in \Omega_0} R(\theta, \delta) = \sup_{\theta \in \Omega_1} R(\theta, \delta)$, δ é minimax em Ω_1.

Método dos momentos: $E[|X|^k] < \infty$, $\mu_j = E(X^j)$, $\hat{\mu}_j = 1/n \sum X_i^j$, $\mu_j = h_j(\theta)$, $\hat{\mu}_j = h_j(\hat{\theta})$, $\hat{\theta} = h^{-1}(\hat{\mu}_j)$.

Em geral

$\hat{\mu}_j = 1/n \sum X^j$ j-ésimo mom. amostral

$\mu_j = E[X^j]$ " " populacional

O estimador mt. mom. ($\hat{\theta}$) de θ é obt. igual. μ_j a $\hat{\mu}_j$, i.e. $\hat{\mu}_j = h_j(\hat{\theta})$, uma vez que geral. $\mu_j = h_j(\theta)$

Se h^{-1} existe, $\hat{\theta} = h^{-1}(\hat{\mu}_j)$.

(MM) usu. útil como chutes iniciais p/ mt. num. (iter.) p/ obt. EMV's

Teo. V.v.a com média θ e var. σ^2. Então $aY + b$ é um est. inadmissível de θ sob perda quadrática quando (i) $a > 1$ ou (ii) $a < 0$ ou (iii) $a = 1$ e $b \neq 0$

Testes de Hipóteses:

Tamanho do Teste: $\alpha = \sup_{\theta \in \Omega_H} P_\theta\{x \in S_1\}$, em que S_1 é a R. Crítica ou de rejeição

Poder do Teste: $P_\theta\{x \in S_1\}$ para um $\theta \in \Omega_K$ dado. Como função de θ $\forall \theta \in \Omega_K$, é a **função de poder** $\beta(\theta)$

Testes Aleatorizados: rejeita H com prob. ϕ se ocorre R_1, H é rej. com prob. ϕ

É totalmente especificado por ϕ (f. crítica), $0 \leq \phi(x) \leq 1, \forall x \in X$

Teste Não Aleatório: $\phi(x) = \begin{cases} 1, & x \in S_1 \\ 0, & x \in S_0 \end{cases}$

Prob de rejeição: $E_\theta\{\phi(x)\} = \int \phi(x) dP_\theta(x)$.

- O proc. usual é maximizar o poder $\beta_\phi(\theta) = E_\theta[\phi(x)]$ sujeito à $E_\theta\{\phi(x)\} \leq \alpha$ $\forall \theta \in \Omega_H$

Se o teste que max. o poder depende de θ ($\in \Omega_K$), não existe um teste mais poderoso $\forall \theta \in \Omega_K$.

Teste Unif. Mais Poderoso UMP: maximiza o poder $\forall \theta \in \Omega_K$. (ver + abaixo)

Lema Neyman-Pearson: P_0 e P_1 dist. de prob. com densidade p_0 e p_1 respect.

(i) Existência: Para testar $H: p_0$ contra $K: p_1$ existe um teste ϕ e uma cte k tais que

$E_0\{\phi(x)\} = \alpha$ [1] e $\phi(x) = \begin{cases} 1, & \text{qdo } p_1(x) > k p_0(x) \\ 0, & \text{"} < \text{"} \end{cases}$ [2]

(ii) Suficiência: se um teste satisfaz [1] e [2] p/ algum k ⇒ ele é mais poderoso de nível α.

(iii) Necessária: Se ϕ é mais poderoso de nível α então para algum k ele satisfaz [2] q.c. μ.

Também satisfaz [1], a menos que exista um teste de tamanho menor do que α e com poder 1.

Corolário: Seja β o poder do teste mais poderoso de nível α p/ testar P_0 contra P_1 ⇒ $\alpha < \beta$ a menos $P_0 = P_1$.

O teste mais poderoso é determinado de forma única por [1] e [2] sempre que $p_1(x) = k p_0(x)$ tem medida nula.

Teste Unifor. Mais Poderoso: $H: \theta \in \Omega_H$ e $K: \theta \in \Omega_K$ (compostas)

Um teste de H contra K é UMP de nível α se

$E_\theta\{\phi(x)\} \leq \alpha$, $\forall \theta \in \Omega_H$ e se $\forall \phi^*$ t.q. $E_\theta\{\phi^*(x)\} \leq \alpha$, $\forall \theta \in \Omega_H$,

$$E_\theta\{\phi(x)\} - E_\theta\{\phi^*(x)\} \geq 0, \forall \theta \in \Omega_K.$$

Dist. com razão de verossimilhança monótona: A família de densidades $p_\theta(x)$ que se existe uma função $T(x) \to \mathbb{R}$, tais que p/ quaisquer θ e θ', $\theta < \theta'$ as dist. P_θ e $P_{\theta'}$ são distintas e a razão $p_{\theta'}(x)/p_\theta(x)$ é não decrescente em $T(x)$

Teorema: $\theta \in \mathbb{R}$, X com $p_\theta(x)$ com razão de verossim. monótona em $T(x)$

(i) Para testar $H: \theta \leq \theta_0$ contra $K: \theta > \theta_0$, existe um teste UMP que é dado por

$$\phi(x) = \begin{cases} 1, & \text{qdo } T(x) > c \\ \gamma, & T(x) = c \\ 0, & T(x) < c \end{cases} \quad [3]$$

onde c e γ são determinados por

$$E_{\theta_0}\{\phi(x)\} = \alpha, \quad [4]$$

(ii) a função de poder $\beta(\theta) = E_\theta\{\phi(x)\}$ deste teste e estritamente crescente $\forall \theta$ para os quais $0 < \beta(\theta) < 1$.

(iii) $\forall \theta'$, o teste determinado por [3] e [4] é UMP p/ testar H contra K com nível $\alpha' = \beta(\theta')$

(iv) Para qualquer $\theta < \theta_0$ o teste minimiza $\beta(\theta)$ entre todos os testes que satisfazem [4]

Corolário: F.E. unid. Se $p_\theta(x) = \exp\{\eta(\theta)T(x) - B(\theta)\}h(x)$. Se $\eta(\theta)$ é estritamente monotona $\Rightarrow$ existe UMP e

se $\eta(\theta)$ é crescente . Se $\eta(\theta)$ é decrescente

$$\phi(x) = \begin{cases} 1, & T(x) > c \\ \gamma, & T(x) = c \\ 0, & T(x) < c \end{cases} \qquad \phi(x) = \begin{cases} 1, & T(x) < c \\ \gamma, & T(x) = c \\ 0, & T(x) > c \end{cases}$$

Nível descritivo (p-valor) $\hat{p} = \hat{p}(x)$: menor nível de significância para o qual H seria rejeitada para a amostra observada x.

Em testes não Aleatorizados: As regiões críticas S_α são aninhadas: $S_\alpha \subset S_{\alpha'}$ se $\alpha < \alpha'$ [5] então

$$\hat{p} = \inf\{\alpha : x \in S_\alpha\}$$

Em testes aleatorizados; $\phi_\alpha(x) \leq \phi_{\alpha'}(x)$ $\forall x$ e $\alpha < \alpha' \Rightarrow$

$$\hat{p} = \inf\{\alpha : \phi_\alpha(x) = 1\}.$$

Lema: $H: \theta \in \Omega_H$. Se as regiões críticas satisfazem [5]

(i) se $\sup_{\theta \in \Omega_H} P_\theta\{x \in S_\alpha\} \leq \alpha$ $\forall\, 0 < \alpha < 1$ [6] então

$$P_\theta(\hat{p} \leq u) \leq u, \ \forall\, 0 \leq u \leq 1$$

(ii) Se para $\theta \in \Omega_H$, $P_\theta(X \in S_\alpha) = \alpha$ $\forall\, 0 < \alpha < 1$ [7] então $P_\theta(\hat{p} \leq u) = u$, $\forall\, 0 \leq u \leq 1$, i.e. $\hat{p} \sim U(0,1)$

Limites de Confiança

Lim. inferior de conf. unif / mais acurado p/ θ: Uma função $\underline{\theta}(x)$ tal que, $P\{\underline{\theta}(x) \leq \theta'\} = \min$, $\forall \theta' < \theta$ sujeito a

$$P_\theta\{\underline{\theta}(x) \leq \theta\} \geq 1-\alpha, \ \forall \theta \quad \langle 1 \rangle$$

Estes minimizam $E_\theta\{L(\theta, \underline{\theta})\}$ $\langle 2 \rangle$ suj. a $\langle 1 \rangle$

Uma família de subconjuntos $S(x)$ do e.p. paramétrico Ω é dita constituir uma famil. de regiões de conf. com coef. de conf $1-\alpha$ se

$$P_\theta\{\theta \in S(x)\} \geq 1-\alpha, \forall \theta \in \Omega$$

$S(x) = \{\theta \in \Omega : \underline{\theta}(x) \leq \theta < \infty\}$ lim. inf. de conf

Teorema: (i) Para cada $\theta_0 \in \Omega$, seja $A(\theta_0)$ a região de aceitação de um teste de nível α de $H(\theta_0): \theta = \theta_0$ e para cada amostra x, seja $S(x) = \{\theta : x \in A(\theta), \theta \in \Omega\}$ $\Rightarrow$ $S(x)$ é uma família de regiões de confiança p/ θ com coef. de conf. $1-\alpha$.

(ii) Se $\forall \theta_0$, $A(\theta_0)$ é UMP para testar $H(\theta_0)$ de nível α contra as alternativas $K(\theta_0)$, então para cada $\theta_0 \in \Omega$, $S(x)$ minimiza $P_\theta\{\theta_0 \in S(x)\}$ $\forall \theta \in K(\theta_0)$ entre todas as regiões de conf. para θ com coef. de conf. $1-\alpha$

Corolário: $p_\theta(x)$, $\theta \in \Omega$ com razão de veros. monotona em $T(x)$. $F_\theta(t)$ de $T = T(x)$ contínua em cada var. t e θ quando a outra é fixada.

(i) Existe um L.I.C uniformemente mais acurado $\underline{\theta}$ para θ com coef de conf. $1-\alpha$.

(ii) Se x é o valor obs. de X e $t = T(x)$ e se a equação $F_\theta(t) = 1-\alpha$ tem uma solução $\theta = \hat{\theta}$ em Ω, então esta sol. é única e $\underline{\theta}(x) = \hat{\theta}$.

Para o lim superior unif. mais acurado com coef. de conf. $1-\alpha$, $P_\theta(T \geq t) = 1-\alpha$ ou $F_\theta(t) = \alpha$.

Intervalos de Confiança: Sejam $\underline{\theta}$ e $\bar{\theta}$ lim. inf. e sup. de conf. com coef de conf. $1-\alpha_1$ e $1-\alpha_2$, respect. Dado a amostra x, $F_{\underline{\theta}}(t) = 1-\alpha_1$ e $F_{\bar{\theta}}(t) = \alpha_2$. Se $\alpha_1 + \alpha_2 < 1$, $\underline{\theta} < \bar{\theta}$

Os int. $(\underline{\theta}, \bar{\theta})$ são ditos I.C p/ θ com conf. de conf. $1-\alpha_1-\alpha_2$

Se $\bar{\theta}$ e $\underline{\theta}$ são unif// mais acurados e minimizam $E_\theta[L_1(\theta, \underline{\theta})]$ e $E_\theta[L_2(\theta, \bar{\theta})]$. Seja $L(\theta; \underline{\theta}, \bar{\theta}) = L_1(\theta, \underline{\theta}) + L_2(\theta, \bar{\theta})$. Então os int. $(\underline{\theta}, \bar{\theta})$ minimizam $E\{L(\theta; \underline{\theta}, \bar{\theta})\}$ sujeito a $P_\theta\{\underline{\theta} > \theta\} \leq \alpha_1$ e $P(\bar{\theta} < \theta) \leq \alpha_2$.

Quantidade Pivotal: função $h(X, \theta) \to \mathbb{R}$ cuja dist. não depende de θ.

Seja $F_h(\cdot)$ a f.d.a de $h(X, \theta)$, estrit. crescente e contínua. Então

$$S(x) = \{\theta \in \Omega : h(x, \theta) \leq F_h^{-1}(1-\alpha)\}$$

é uma região de conf. para θ com coef. de conf. $1-\alpha$

Hipóteses Bilaterais:

Teorema: (i) para testar $H: \theta \leq \theta_1$ ou $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ contra $K: \theta_1 < \theta < \theta_2$ na F.E. unid. com $\eta(\theta)$ estrit. crescente, existe um teste UMP dado por

$$\phi(x) = \begin{cases} 1, & \text{qdo } c_1 < T(x) < c_2, \ (c_1 < c_2) \\ \gamma_i, & T(x) = c_i, \ i = 1, 2 \\ 0, & T(x) < c_1 \text{ ou } T(x) > c_2 \end{cases}$$

em que os c_i e γ_i são determinados por $E_{\theta_1}\{\phi(x)\} = E_{\theta_2}\{\phi(x)\} = \alpha$ $\langle * \rangle$

(ii) Este teste minimiza $E_\theta\{\phi(x)\}$ sujeito a $\langle * \rangle$ $\forall\, \theta < \theta_1$ e $\forall\, \theta > \theta_2$

(iii) Para $0 < \alpha < 1$, a função de poder deste teste tem um máximo em um ponto θ_0 entre

$\theta = \theta_0$ e decresce estritamente quando θ se afasta de θ_0 em qqr direção a menos que existam ... t.q. $P_\theta(T(x)=t_1) + P_\theta(T(x)=t_2) = 1 \ \forall \theta$.

Testes não viciados: teste ϕ tal que $\beta_\phi(\theta) \leq \alpha$ se $\theta \in \Omega_H$ e $\beta_\phi(\theta) \geq \alpha$ se $\theta \in \Omega_K$.
Sempre que um teste UMP existe, ele é não viciado.

Região de Credibilidade: Qualquer conjunto $S(x)$ t.q. $P(\theta \in S(x) \mid X = x) \geq 1-\alpha$. $\forall x$ é uma região de cred. p/ θ com prob. $1-\alpha$.

Regiões HPD: $S(x) = \{\theta \in \Omega \mid \pi(\theta|x) \geq k\}$ com k determinado por $P\{\theta \in S(x) \mid X = x\} = 1-\alpha$.

Teste da Razão de Verossimilhanças:

$$\lambda(x) = \sup_{\theta\in\Omega_H} L(\theta) \Big/ \sup_{\theta\in\Omega} L(\theta), \quad 0 \leq \lambda(x) \leq 1.$$

Um T.R.V é qqer teste que rejeita H se e somente se $\lambda(x) \leq c$, em que $c \in [0,1]$.

Para nível α: $\sup_{\theta\in\Omega_H} P_\theta(\lambda(x) \leq c_\alpha) = \alpha$.

Estatística da razão de verossimilhaças:
$$\omega(x) = 2\{\log L(\hat\theta) - \log L(\theta_0)\}$$

Estatística de Wald:
$$W(x) = (\hat\theta - \theta_0)^2 I(\hat\theta)$$

Estatística escore (Rao): $S_R(x) = U(\theta_0)^2 I(\theta_0)^{-1}$
$U(\theta) = \partial \log L(\theta)/\partial\theta$, f. escore

Estatística gradiente: $S_T(x) = U(\theta_0)(\hat\theta - \theta_0)$

. $Q \sim \text{Gama}(a,b) \Rightarrow 2bQ \sim \chi^2_{2a}$

$X_1, \dots, X_n$ a.a. $U(0,\theta)$.
a) Mostre que p/ testar $H: \theta \leq \theta_0$ contra $K: \theta > \theta_0$ qqer teste p/ o qual $E_{\theta_0}(\phi(X)) = \alpha$, $E_\theta(\phi(X)) \leq \alpha$, p/ $\theta \leq \theta_0$, e $\phi(x) = 1$ qdo $\max(x_1, \dots, x_n) > \theta_0$, é UMP de nível α.
Sug. Use LNP com $k = (\theta_0/\theta_1)^n$.
Pelo LNP temos que p/ testar $H: \theta = \theta_0$ vs. $K: \theta = \theta_1$, sendo $\theta_1 > \theta_0$, o teste UMP é dado por

$$\phi(x) = \begin{cases} 1, & p_{\theta_1}(x) > k p_{\theta_0}(x) \\ \gamma, & p_{\theta_1}(x) = k p_{\theta_0}(x) \\ 0, & p_{\theta_1}(x) < k p_{\theta_0}(x) \end{cases}$$

sendo

$$p_{\theta_j}(x) = 1/\theta_j^n \prod_{i=1}^n I_{(0,\theta_j)}(x_i), \quad j = 0, 1$$

$$= 1/\theta_j^n \, I_{(0,\theta_j)}(x_{(n)}), \quad j = 0, 1.$$

Pela sug., p/ $k = (\theta_0/\theta_1)^n$, temos que

$\phi(X) = 1 \Leftrightarrow 1/\theta_1^n I_{(0,\theta_1)}(x_{(n)}) > (\theta_0/\theta_1)^n \, 1/\theta_0^n I_{(0,\theta_0)}(x_{(n)})$
$\Leftrightarrow I_{(0,\theta_1)}(x_{(n)}) > I_{(0,\theta_0)}(x_{(n)})$
$\Leftrightarrow \theta_0 < x_{(n)} < \theta_1$

. $\phi(X) = \gamma \Leftrightarrow I_{(0,\theta_1)}(x_{(n)}) = I_{(0,\theta_0)}(x_{(n)}) \Leftrightarrow$
$x_{(n)} < \theta_0 < \theta_1$ ou $x_{(n)} > \theta_1 > \theta_0$

. $\phi(X) = 0 \Leftrightarrow I_{(0,\theta_1)}(x_{(n)}) < I_{(0,\theta_0)}(x_{(n)})$
N. poss. $\theta_1 > \theta_0$

$$\therefore \phi(x) = \begin{cases} 1, & \theta_0 < x_{(n)} < \theta_1 \\ \gamma, & x_{(n)} < \theta_0 \end{cases}$$

Vamos det. γ a partir de $E_{\theta_0}(\phi(X)) = \alpha \Leftrightarrow$
$P(X_{(n)} > \theta_0) + \gamma P(X_{(n)} < \theta_0) = \alpha \Rightarrow \gamma = \alpha$.
p/ $\theta \leq \theta_0$, temos que
$E_\theta(\phi(x)) = P_\theta(X_{(n)} > \theta_0) + \alpha P_\theta(X_{(n)} < \theta_0) = \alpha$, com $P_\theta(X_{(n)} > \theta_0) = 0$ e $P_\theta(X_{(n)} < \theta_0) = 1$.
Como o teste (a) não dep. de θ_1, com $\theta_1 > \theta_0$, então é UMP p/ testar $H: \theta = \theta_0$ cont. $K: \theta > \theta_0$, mesmo acontece p/ $H: \theta \leq \theta_0$ contra $K: \theta > \theta_0$.

b) Mostre q p/ testar $H: \theta = \theta_0$ contra $K: \theta \neq \theta_0$, o teste $\phi(x) = 1$ quando $x_{(n)} > \theta_0$ ou $x_{(n)} < \theta_0 \alpha^{1/n}$ e $\phi(x) = 0$ c.c. é UMP.
Sug: 3 sit. separadamente.
(i) $H: \theta = \theta_0$ contra $K: \theta = \theta_1 > \theta_0$
(ii) " $K_2: \theta = \theta_1$, com $\theta_0 \alpha^{1/n} < \theta_1 < \theta_0$
(iii) " $K_3: \theta = \theta_1 < \theta_0 \alpha^{1/n}$

Pelo it. (a) $\phi(x) = \begin{cases} 1, & \text{se } X_{(n)} > \theta_0 \\ \alpha, & X_{(n)} < \theta_0 \end{cases}$, é UMP p/ $H: \theta = \theta_0$ ct $K: \theta = \theta_1 > \theta_0$
O poder de ϕ é
$E_{\theta_1}(\phi(X)) = P_{\theta_1}(X_{(n)} > \theta_0) + \alpha P_{\theta_1}(X_{(n)} < \theta_0)$
$= 1 - (\theta_0/\theta_1)^n + \alpha(\theta_0/\theta_1)^n = 1 - (1-\alpha)(\theta_0/\theta_1)^n$
Aqui, o poder de ϕ^* é
$E_{\theta_1}(\phi^*(X)) = P_{\theta_1}(X_{(n)} > \theta_0) + P(X_{(n)} < \theta_0 \alpha^{1/n})$
$\overset{(1)}{=} 1 - (\theta_0/\theta_1)^n + (\theta_0 \alpha^{1/n}/\theta_1)^n$
$= 1 - (1-\alpha)(\theta_0/\theta_1)^n$
Logo, ϕ^* é UMP p/ testar $H: \theta = \theta_0$ cta $K: \theta > \theta_0$.
. Caso ii: $\theta_0 \alpha^{1/n} < \theta_1 < \theta_0$, temos LNP ($\alpha$ ... $\theta_1 < \theta_0$)

$$\phi^{**}(x) = \begin{cases} 1, & p_{\theta_1}(x) > k p_{\theta_0}(x) \\ \gamma, & \quad " \quad = \quad " \\ 0, & \quad " \quad < \quad " \end{cases}, \quad k = (\theta_0/\theta_1)^n \text{ é UMP}$$

. $\phi^{**}(x) = 1 \Leftrightarrow I_{(0,\theta_1)}(x_{(n)}) > I_{(0,\theta_0)}(x_{(n)})$. N poss. pois $\theta_1 < \theta_0$
. $\phi^{**}(x) = \gamma \Leftrightarrow$ " = " $\Leftrightarrow x_{(n)} < \theta_1$ e $x_{(n)} < \theta_0$
. $\phi^{**}(x) = 0 \Leftrightarrow$ " < " $\Leftrightarrow x_{(n)} > \theta_1$ e $x_{(n)} < \theta_0$, $\theta_1 < x_{(n)} < \theta_0$
$x_{(n)} < \theta_1 < \theta_0$ ou $x_{(n)} > \theta_1$ e $x_{(n)} > \theta_0$, $x_{(n)} > \theta_0$

$$\phi^{**}(x) = \begin{cases} \gamma, & x_{(n)} < \theta_1 \\ 0, & x_{(n)} > \theta_1 \end{cases}$$ Agora encont. γ. Usamos $E_{\theta_0}(\phi(X)) = \alpha \Rightarrow$

$\gamma P_{\theta_0}(X_{(n)} < \theta_1) = \alpha \Leftrightarrow \gamma(\theta_1/\theta_0)^n = \alpha \Rightarrow \gamma = \alpha(\theta_0/\theta_1)^n$
O poder de ϕ^{**} é
$E_{\theta_1}(\phi^{**}(x)) = \alpha(\theta_0/\theta_1)^n P_{\theta_1}(X_{(n)} < \theta_1) = \alpha(\theta_0/\theta_1)^n$
Agora, o poder de ϕ^*
$E_{\theta_1}(\phi^*(X)) = P_{\theta_1}(X_{(n)} > \theta_0) + P(X_{(n)} < \theta_0\alpha^{1/n})$
$= 1 - (\theta_0/\theta_1)^n + \left(\frac{\theta_0 \alpha^{1/n}}{\theta_1}\right)^n$
$= \alpha(\theta_0/\theta_1)^n$ $\therefore \phi^*$ é UMP
(iii) Temos $\theta_1 \leq \theta_0 \alpha^{1/n} < \theta_0$, logo o poder de ϕ^* é
$E_{\theta_1}(\phi^*(X)) = P_{\theta_1}(X_{(n)} > \theta_0) + P_{\theta_1}(X_{(n)} < \theta_0 \alpha^{1/n}) = 1 \Rightarrow$
ϕ^* é UMP.

. $X_i \sim \text{Weibull}(c, \theta) \Rightarrow X_i^c/\theta^c \sim \text{Exp}(1)$

. Na Pareto, usar $k = (\gamma_0/\gamma_1)^{-\beta}$

Famílias de Grupo de distribuições: obtida submetendo uma v.a com uma dist. dada a um grupo de transformações.

· Grupo de Transformações: Uma classe de transf. 1-a-1 que é fechada sob composição e inversão.

$g_1 \in J,\ g_2 \in J \Rightarrow g_2 \circ g_1 \in J$ (sob comp)

$g \in J \Rightarrow g^{-1} \in J$ (sob inver.)

· Família de locação: É gerada pela transf. $X = U + a$

$P(X \leq x) = F_U(x-a),\ a \in \mathbb{R}$.

· Família de escala: É gerada pela transf. $X = bU,\ b>0$

$P(X \leq x) = F_U(x/b)$

· Família de locação-escala: $X = a + bU,\ b>0,\ a \in \mathbb{R}$

$P(X \leq x) = F_U((x-a)/b)$

· Família Exponencial:

$$P_\theta(x) = \exp\left\{\sum_{i=1}^{s} \eta_i(\theta) T_i(x) - B(\theta)\right\} h(x)$$

· Forma Canônica (natural)

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

· Família de posto completo: se nem os T's nem os η's satisfazem restrições lineares, e o espaço paramétrico para os η's contém retângulos s-dimens.

· Função Geradora de Momento:

$M_T(u_1, \dots, u_s) = E\{\exp(u_1 T_1 + \dots + u_s T_s)\}$

· Função Geradora de Cumulantes

$$K_T(u_1, \dots, u_s) = \log M_T(u_1, \dots, u_s) = \sum_{r_1, \dots, r_s} K_{r_1, \dots, r_s} \frac{u_1^{r_1} \cdots u_s^{r_s}}{r_1! \cdots r_s!}$$

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

Se X tem forma canônica:

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

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

Estatísticas Suficientes: $T(X)$ é suf. para θ se $f(X|T=t)$ é indep. de θ, para todo t.

Criterio da Fatoração: $P_\theta(x) = g_\theta(T(x))h(x)$, (q.c. $\mathcal{P}$)

Suficiencia Minimal: T é suf. min. se para qualquer est. suf. U, existe uma função H tal que $T = H(U)$ (q.c. $\mathcal{P}$)

Teo: $\mathcal{P}$ família finita de densidades $P_i,\ i=0,1,\dots,k$ com o mismo soporte $\Rightarrow T(x) = \left(\frac{P_1(x)}{P_0(x)}, \dots, \frac{P_k(x)}{P_0(x)}\right)$ é suf. minimal

Cor: uma cond. necessaria e suf. para U ser suf. é que para qualquer θ e θ_0 fixado, $P_\theta(x)/P_{\theta_0}(x)$ seja uma função apenas de $U(x)$

Lem: $\mathcal{P}$ famil. de dist. com suporte comum e $\mathcal{P}_0 \subset \mathcal{P}$. Se T é suf. min. para $\mathcal{P}_0$ e suf para $\mathcal{P}$, então é suf. min. para $\mathcal{P}$.

Teo: sob as condic. do C.F. $T(x) = T(y) \Leftrightarrow y \in D(x)$

$D(x) = \{y \in \mathcal{X} : P_\theta(y) = P_\theta(x) h(x,y),\ \forall \theta \in \Omega\}$ para alguma fun. h não negativa $\Rightarrow T(x)$ é suf. min.

Ancilaridade: $V(x)$ é ancilar se sua dist. não depende de θ. De primeiro ordem se $E_\theta[V(x)]$ não dep. de θ.

Completitividade: T é completa se para qualquer função $f: T \to \mathbb{R}$. $E_\theta[f(T)] = 0,\ \forall \theta \in \Omega \Rightarrow f(T) = 0$, (q.c. $\mathcal{P}$)

Se existe uma est. suf. min. então uma cond. neces. para que uma est. suf. seja comp. é que ela seja suf. min.

T. Basu: se T suf. comp. $\Rightarrow$ qualquer V ancilar é independente de T.

Teo: se $X \sim$ F.E. s-dim. de posto completo $\Rightarrow T(X)$ e suf. comp.

função convexa: ϕ é convexa se

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

$a < x < y < b,\ 0 < \gamma < 1$

ou se $\phi'(x) \leq \phi'(y)$

Desig. de Jensen: se ϕ convexa en I e $P(X \in I) = 1$ e tem esperança finita $\Rightarrow \phi[E(X)] \leq E[\phi(X)]$.

T. de Rao-Blackwell: $X \sim P_\theta \in \mathcal{P}$, T est. suf., δ estim. de $g(\theta)$, $L(\theta, d)$ função de perda estrita/ convexa de d $\Rightarrow$ se δ tem media $< \infty$ y $R(\theta, \delta) = E[L(\theta, \delta(X))] < \infty$ e se $\eta(t) = E[\delta(X)|T=t]$ $\Rightarrow R(\theta, \eta) < R(\theta, \delta)$, a menos que $\delta = \eta$ com prob 1.

Admissibilidade: δ admis. se não existe δ' tal que $R(\theta, \delta') \leq R(\theta, \delta)$.

função de perda quadrática: $L(\theta, \delta) = (\delta - \theta)^2$

função de Risco: $R(\theta, \delta) = E_\theta[L(\theta, \delta)]$

Lema: se δ_0 é qualquer est. não viesado de $g(\theta)$, a totalidad de estim. não viesados é dada por $\delta = \delta_0 - U$, com $E_\theta(U) = 0$, U qualquer est.

ENVVUM: $V_\theta(\delta(X)) \leq V_\theta(\delta^*(X))$, δ^*: qualquer outro est. não viesado de $g(\theta)$.

"O ENVVUM É unico"

Teo: X v.a $\sim P_\theta$, $\theta \in \Omega$, δ un est. em Δ a classe dos est. com $E_\theta(\delta^2) < \infty\ \forall \theta \in \Omega$; U conjunto de todos os est. não viesados de 0 em Δ $\Rightarrow$ uma condição neces. e sufic. para que δ seja ENVVUM de $E(\delta) = g(\theta)$ é que $E_\theta(\delta U) = 0,\ \forall U \in \mathcal{U}$ e todo $\theta \in \Omega$.

Lema: v.a. $X \sim P_\theta \in \mathcal{P}$, T est. suf. comp. para $\mathcal{P}$ $\Rightarrow$ toda função estimavel $g(\theta)$ tem um e somente um est. não viciado que é função de T.

F. Estimavel: se existe um estimador não viesado de $g(\theta)$

Teo: v.a. $X \sim P_\theta \in \mathcal{P}$, T est. suf. comp. para $\mathcal{P}$:

(a) $\forall g(\theta)$ estimavel $\exists$ um est. não viesado que minimiza o risco uniforme/ para qualquer função de perda $L(\theta, d)$

con... em d. Este est. é o ENVVUM.

(b) o ENVVUM em (a) é o unico est. não vies com risco ...desde que seu risco seja finito e L seja estrita/ convexa ... d.

Corol: se $\mathcal{P}$ é uma F.E. de posto completo, s dimensional, então o teo. ant. vale para $\theta = (\theta_1, \dots, \theta_s)$ e $T = (T_1, \dots, T_s)$.

Metodos para obter o ENVVUM:

1. Se T é suf. comp. $\forall$ função estimav. $g(\theta) \Rightarrow$ o ENVVUM é determinado por $E_\theta[\delta(T)] = g(\theta)$ $\forall \theta \in \Omega$.
2. Condicional: $\delta(x)$ não viesado para $g(\theta) \Rightarrow$ o ENVVUM é obtido como $E(\delta(x)|T)$; T suf. comp.

EMV: $\hat{\theta} \in \Omega$ satisfazendo $L(\hat{\theta}; x) = \sup_{\theta \in \Omega} L(\theta; x)$

$\hat{g} = g(\hat{\theta})$ EMV de $g(\theta)$.

Para n grande os EMV's: cosistencia, norm. assintat., eficiência assint.

Lema: $\hat{\theta} = \hat{\theta}(x)$ um EMV de θ, $\hat{\theta}$ único EMV $\Rightarrow$ se $\hat{\theta}$ é uma est. suf. $\Rightarrow$ ele é uma est. suf. minimal.

Lema: se δ é um est. não vies. de $g(\theta): \Theta \to \mathbb{R}$, e se $\delta^* = \delta + b$, onde $b \neq 0$ é o vies e é indep. de θ $\Rightarrow$ δ^* tem risco sob perda quadrática uniforme/ maior. $R_{\delta^*}(\theta) = R_\delta(\theta) + b^2$.

Desigualdade da Informação:

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

$$I(\theta) = E_\theta[\varphi(x,\theta)^2]$$

sob condições adequadas:

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

Teo: δ tem 2º momento finito $\forall \theta \in \Omega$. Uma condição necessario e sufic. para que $Cov_\theta(\delta, \varphi)$ dependa de δ somente através de $g(\theta) = E_\theta(\delta)$ $\forall \theta$ é que $Cov_\theta(U, \varphi) = 0$ $\forall U \in \mathcal{U} = \{U : E_\theta(U) = 0 \text{ e } E_\theta(U^2) < \infty, \forall \theta \in \Omega\}$.

$\Rightarrow$ se δ_1 e δ_2 com $E_\theta(\delta_1) = E_\theta(\delta_2)$ $\forall \theta \Rightarrow Cov(\delta_1, \varphi) = Cov_\theta(\delta_2, \varphi)$

Suposições: (a) Ω interv. aberto. (b) P_θ têm suporte comum tal que $A = \{x, p_\theta(x) > 0\}$ é indep de θ, (c) para qualquer $x \in A$ e $\theta \in \Omega$, $p'_\theta(x) = \partial p_\theta(x)/\partial\theta$ $\exists$ e é $< \infty$

Lema: se (a)-(c) estão satisfeitos e a derivada com respeito a θ do lado esquerdo de $\int p_\theta(x) d\mu(x) = 1$ pode ser obtida diferenciando sob a sinal de integração, então

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

$$I(\theta) = Var_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(x)\right]$$

Também, sob outras condições:

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

Nota: $I(\theta)$ depende da parametrização utilizada. se $\theta = h(\xi)$ então $I^*(h(\xi)) = I(h(\xi)) h'(\xi)^2$.

Teo: $X \sim$ FE unidimens., $\tau(\theta) = E_\theta(T)$, então

$$I[\tau(\theta)] = 1/Var_\theta(T)$$

Teo: $X \sim p_\theta$ e $Y \sim q_\theta$, ind, satisf. (a)-(c) e $E_\theta[\frac{\partial}{\partial\theta} \log p_\theta(x)] = 0$ e $E_\theta[\frac{\partial}{\partial\theta} \log q_\theta(y)] = 0$, se $I_1(\theta)$, $I_2(\theta)$ e $I(\theta)$ são as inf. do θ contidas em X, Y e (X,Y) resp/ então:

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

Corol: válido para $X_1, \dots, X_n$, $I(\theta) = n I_1(\theta)$

Teo: Desig. da Inform: se (a)-(c) valem, $E_\theta[\frac{\partial}{\partial\theta} \log p_\theta] = 0$, $I(\theta) > 0$. Seja δ: $E_\theta(\delta^2) < \infty$, $E_\theta(\delta)$ existe e

$$\frac{\partial}{\partial\theta} E_\theta(\delta) = \int \frac{\partial}{\partial\theta} \delta p_\theta d\mu \quad \text{então}$$

$$Var_\theta(\delta) \geq \left[\frac{\partial}{\partial\theta} E_\theta(\delta)\right]^2 / I(\theta)$$

Nota: se $E_\theta(\delta) = \theta \Rightarrow Var_\theta(\delta) \geq 1/I(\theta)$ $\to$ <u>L.I. Cramer-Rao</u>

se δ estim. de $g(\theta)$ com $E_\theta(\delta) = g(\theta) + b(\theta)$

$\Rightarrow Var(\delta) \geq [b'(\theta) + g'(\theta)]^2 / I(\theta)$

se $X_1, \dots, X_n$ iid, $\delta = \delta(X_1, \dots, X_n) \Rightarrow I(\theta) = n I_1(\theta)$.

· **Caso Multiparamétrico:** $\theta \in \Omega \subset \mathbb{R}^s$

$\Rightarrow I(\theta) = \{I_{ij}(\theta)\}_{s \times s}$ com

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

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

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

$$\Rightarrow I_\theta = E_\theta\left[\frac{\partial}{\partial\underline{\theta}} \log p_\theta(x) \frac{\partial}{\partial\underline{\theta}'} \log p_\theta(x)\right]$$

$$= Cov_\theta\left[\frac{\partial}{\partial\theta} \log p_\theta(x)\right] \to \text{vetor score.}$$

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

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

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

$\Rightarrow I^*(\xi) = J I J'$, $J = \partial\theta' / \partial\xi$

Teo: $\tau = (\tau_1, \dots, \tau_s)$, $\tau_i = E(T_i(x)) \Rightarrow I(\tau) = C^{-1} = Cov_\theta(T)$, $T = (T_1, \dots, T_s)$

Teo: Desig. da Inf: $E_\theta\left[\frac{\partial}{\partial\theta_i} \log p_\theta(x)\right] = 0$ e $Var_\theta(\delta) \geq \alpha' I^{-1}(\theta) \alpha$, $\alpha_i = \partial E_\theta[\delta(x)] / \partial\theta_i$

Def: $f(x|\xi)$ e $L(\xi, d)$ são invariantes por locação se respectiv/ $f(x'|\xi') = f(x|\xi)$ e $L(\xi, d) = L(\xi', d')$ sempre que $x' = x + a$, $\xi' = \xi + a$ e $d' = d + a$.

Def: um estimador satisfazendo $\delta(x_1 + a, \dots, x_n + a) = \delta(x) + a$ sob as transformações $X'_i = X_i + a$, $\xi' = \xi + a$ e $d' = d + a$ é chamado equivariante por locação

Teo: $X = (X_1, \dots, X_n) \sim f(x - \xi)$, $\xi = (\xi_1, \dots, \xi_n)$, δ equiv. por locação, com perda $L(\xi, d) = \rho(d - \xi)$ $\Rightarrow$ o vício, risco e variância de δ são cts (não dependem de ξ).

Def: est. equiv. de risco min. (ERM, MRE)

Lema: δ_0 equiv. por locação, $\Rightarrow$ uma cond. neces. e suf. para δ ser equiv. por locação é que $\delta(x) = \delta_0(x) + u(x)$, $u(x) = u(x + a)$

Lema: $u(x + a) = u(x)$ sse é uma função dos $y_i = x_i - x_n$ ou uma cte se $n = 1$. ($u(x)$ inv. por loc. é ancilar)

Teo: Se δ_0 equiv. por loc. $\Rightarrow$ uma cond. nec. e suf para δ ser equiv por loc é que $\exists$ uma função v de $n-1$ argumentos t.q. $\delta(x) = \delta_0(x) - v(y)$, $\forall x$

Teo: $\underset{\sim}{X} \sim f(\underline{x} - \underline{\xi})$, $Y_i = X_i - X_n$, $L(\xi, d) = \rho(d - \xi)$, δ_0 est. equiv. de risco finito, para cada y existe $v(y) = v^*(y)$ que minimiza $E_0\{\rho[\delta_0(x) - v(y)] | y\}$ $\Rightarrow$ $\exists$ um est. por loc. de risco min. $= \delta^*(x) = \delta_0(x) - v^*(y)$.

Cond: ρ convexa e não monot. $\Rightarrow$ $\exists$ um ERM de ξ e é único se ρ é estric. conv.

Teo: se $L(\xi, d) = (d - \xi)^2 \Rightarrow$

$$\underline{\text{Est. de Pitman}} \leftarrow \delta^*(x) = \frac{\int_{-\infty}^{+\infty} u f(x_1 - u, \dots, x_n - u) du}{\int_{-\infty}^{+\infty} f(x_1 - u, \dots, x_n - u) du}$$

Lema: função de perda quadrat., δ est. equiv. com vício $= b$, cte, e risco $< \infty$. (a) $\delta(x) - b$ é equiv. e tem risco $< \delta(x)$, (b) o único est. ERM é não viesado, (c) se ENVVUM existe e é equiv. $\Rightarrow$ ele é ERM

Resumo prova 1, Avançada 1.

$X_1, \dots, X_n$ a.a de uma família locação escala com função $F((x-a)/b)$ $(a \in \mathbb{R}, b>0)$. Sejam $Y_j = \phi_j(\underline{X})$, $j=1,\dots,n$ que $\phi_j(x_1,\dots,x_n) = \phi_j\left(\frac{x_1-\alpha}{\beta}, \dots, \frac{x_n-\alpha}{\beta}\right)$ $\forall \alpha, \beta$. Então $(Y_1,\dots,Y_n)$ é uma est. ancilar.

— $X_1,\dots,X_n$ de $F((x-a)/b)$ $\Rightarrow$ • se b é conhecido $(x_1-x_i)/b$, $i=2,\dots,n$ são ancilares, • se a é conhec $(x_1-a)/(x_i-a)$, $i=2,\dots,n$ são ancilares, • se a e b são desconh. $(x_1-x_i)/(x_2-x_i)$, $i=3,\dots,n$ são ancilares.

→ $\bar{x}$ e $\Sigma(X_i-\bar{x})^2$ são indep. se X_i são v.a. $N(\theta,\sigma^2)$
$X_{(1)}$ e $\Sigma(X_i - X_{(1)})$ " " $E(a,b)$

→ A subconj do $\mathcal{X}$, P_θ^* é P_θ truncada sobre A
- se T é suf. para $\mathcal{P}$, $\Rightarrow$ é suf. para $\mathcal{P}^*$
- se T é comp. para $\mathcal{P}$, $\Rightarrow$ é comp. para $\mathcal{P}^*$

→ $Var[\delta(X)] = Var[E(\delta(X)|T)] + E[Var(\delta(X)|T)]$

→ se $f_\theta(x) = P_\theta(X=x) = a(x)\theta^x / C(\theta)$, $x = 0,1,\dots$; $a(x) \geq 0$, $\theta > 0$ $\Rightarrow M_X(u) = C(\theta e^u)/C(\theta)$

→ Si $X \sim F.E.$ $\Rightarrow E_\theta(T(X)) = B'(\eta)/\eta'(\theta)$ e
$Var_\theta(T(X)) = B''(\theta)/[\eta'(\theta)]^2 - \eta''(\theta)B'(\theta)/\eta'(\theta)^3$

—||—

→ $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} b^k a^{n-k}$

$\frac{1}{(1-x)^r} = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} x^k$

→ $\Sigma(x_i - \bar{x})^2/\sigma^2 \sim \chi^2_{n-1}$

→ $X_1,\dots,X_n$ a.a de $X \sim E(a,b)$ $\Rightarrow F(x) = 1 - \exp\left(-\frac{x-a}{b}\right)$, $x \geq a$, $X_{(1)} \sim E(a, b/n)$

→ $X_i \overset{iid}{\sim} E(0,1) \Rightarrow Y_i = (n-i+1)[X_{(i)} - X_{(i-1)}] \overset{iid}{\sim} E(0,1)$
$\sum_{i=2}^{n} Y_i = \sum_{i=1}^{n}(x_{(i)} - x_{(1)}) \sim Gama(n-1, 1)$

→ $X \sim Pa(a,b) \Rightarrow \hat{C}_{MV} \sim Pa(na, c)$ e $2na/\hat{a}_{MV} \sim \chi^2_{2(n-1)}$
$V_i = a(\log X_i - \log c) \sim E(0,1)$

→ Se $E(\delta) = g(\theta)$ e $Var(\delta)$ atinge o limite da desig. da inform. $\Rightarrow \delta(x) = g(\theta) \pm \frac{g'(\theta)}{I(\theta)} \frac{\partial}{\partial\theta} \log P_\theta(x)$
$b(\theta) + a(\theta) \frac{\partial}{\partial\theta} \log P_\theta(x)$

→ f_x: família de escala; $f_X(x;\tau) = \frac{1}{\tau^n} f(x/\tau)$, $\tau > 0$, $x \in \mathbb{R}^n$. $\delta(bx) = b^r \delta(x) \Rightarrow \delta$ é equivariante por escala. $\Rightarrow \delta(x) = \delta_0(x)/v(x)$. δ_0 equiv. por escala e $v(cx) = v(x) > 0$ $\forall c > 0$
$v(x) = w(y)$, $y = x/|x_n|$

→ Est. Eq. de Risco mínimo de τ^r:
$$\delta^*(x) = \delta_0(x) \frac{E_1[\delta_0(x)|y]}{E_1[\delta_0^2(x)|y]}$$

→ Est. de Pitman
$$\delta^*(x) = \frac{\int_0^{+\infty} v^{n+r-1} f(vx_1,\dots,vx_n)\,dv}{\int_0^{+\infty} v^{n+2r-1} f(vx_1,\dots,vx_n)\,dv}$$
é um est. equiv. por escala de risco mínimo sob perda $\left[\frac{d-\tau^r}{\tau^r}\right]^2$

função Gama:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{t}\,dt$$

se $n \in \mathbb{Z}^+ \Rightarrow \Gamma(n) = (n-1)!$, $\Gamma(z+1) = z\Gamma(z)$

$$\lim_{z\to 0^+} \Gamma(z) = \lim_{z \to 0^+} \frac{\Gamma(z+1)}{\Gamma(z)} = \infty$$

$$\Gamma(1-z)\Gamma(z) = \frac{\pi}{sen(\pi z)}$$

$$\Gamma(z)\Gamma\left(z+\tfrac{1}{2}\right) = 2^{1-2z}\sqrt{\pi}\,\Gamma(2z)$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

func. trig:

$\cos\theta = sen(\theta + \pi/2)$; $sen\,\theta = \cos(\theta - \pi/2)$
$sen\,\theta = \cos(\theta - \pi/2)$; $\cos\theta = sen(\pi/2 - \theta)$
$sen\,\theta = \cos(\pi/2 - \theta)$

func. digama e trigama:

$$\psi(z) = \frac{d\ln\Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)}$$

$$\psi(z+1) = \psi(z) + 1/z$$

$$\psi(1-x) - \psi(x) = \pi\cot(\pi x)$$

$$\psi_1(z) = \frac{d^2\ln\Gamma(x)}{dz^2} = \frac{d}{dz}\psi(z)$$

$$\psi_1(z+1) = \psi_1(z) - 1/z^2$$

$$\psi_1(1-z) = \psi_1(z) = \pi^2 \csc^2(\pi z)$$

—||—

$$M_T(u) = E(e^{tu}) = E\left[\sum_{k=0}^{\infty} \frac{(tu)^k}{k!}\right]$$

$K_T(u) = \log M_T(u) \Rightarrow M_T(u) = \exp\{k_T(u)\}$
$= 1 + k_T(u) + \frac{[k_T(u)]^2}{2!} + \dots$

$K_T(u) = \sum_{j=0}^{+\infty} \frac{k_j u^j}{j!}$, $k_0 = 0$, $k_1 = E(T)$, $k_2 = Var(T)$

$\hat{p} = \inf\{\alpha : 0 < \alpha < 1 ; T\alpha = 1\}$
" $\{$ " $y \geq C_\alpha\}$
$= \inf_{y \geq C_\alpha} P_{\theta_0}(y \geq C_\alpha)$
$= P_{\theta_0}(Y \geq y)$

OTIMALIDADE DE RISCO MÉDIO:

RISCO Médio: (RISCO de Bayes): seja $R(\theta,\delta)<\infty, \forall\theta$
$r(\Lambda,\delta) = \int R(\theta,\delta)\, d\Lambda(\theta)$, em que $\int d\Lambda(\theta)=1$
depende de δ e da dist. de prob Λ. Não dep. de θ.

Estimador de Bayes: estimador δ^* que minimiza o risco de Bayes

Densidade Marginal de X, $q(x)$ ou $q(x|\lambda)$:
$q(x) = \int f(x|\theta)\, d\Lambda(\theta)$.

Densidade a Posteriori de θ:
$\pi(\theta|x) = f(x|\theta)\pi(\theta)/q(x)$ ou
$\pi(\theta|x,\lambda) = f(x|\theta)\pi(\theta|\lambda)/q(x,\lambda)$

Risco a Posteriori: $E\{L(\Theta,\delta(x))\,|\,X=x\}$

Teorema: suponha que $\Theta\sim\Lambda$ e dado $\Theta=\theta$ suponha que $X\sim P_\theta$. Se quer estimar $g(\Theta)$ com função de perda $L(\theta,d)$, então se
(a) existe um estimador δ_0 com risco finito
(b) para quase todo x (exceto quando $q(x)=0$) existe um $\delta_\Lambda(x)$ que minimiza
$E\{L(\Theta,\delta(x))\,|X=x\}$
então, $\delta_\Lambda(x)$ é um estimador de Bayes.

Corolário: Sob as condições do Teorema, se $L(\theta,d) = w(\theta)[d-g(\theta)]^2$, $w(\theta)>0, \forall\theta\in\Omega$
então $$\delta_\Lambda(x) = \frac{E[w(\Theta)g(\Theta)|X=x]}{E[w(\Theta)|X=x]}$$
se $w(\theta)=1, \forall\theta\in\Omega$,
$$\delta_\Lambda(x) = E[g(\Theta)|X=x]$$

Corolário: se a função de perda é estritamente convexa em d (em particular quadrática), então uma sol. de Bayes 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 q.c.$\mathcal{P}$.

Distribuições a priori conjugadas: priori e a posteriori na mesma família de distribuições.

Teorema: Qualquer estimador de Bayes que é único é admissível [se quaisquer 2 estim. de Bayes são iguais q.c.$\mathcal{P}$].

Est. de Bayes e Suficiência: seja $T=T(x)$ uma est. suf. então $f(x|\theta)=L(\theta|x) = g(t|\theta)h(x)$
$\Rightarrow \pi(\theta|x) = g(t|\theta)\pi(\theta)/\int g(t|\theta')\pi(\theta')d\theta'$
Logo só depende de X através de t.
Todas as "quantidades" Bayesianas são funções de uma Est. suf. minimal.

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

Def: Estimador de Bayes Generalizado: $\delta(x)=\delta_\Lambda(x)$ que minimiza $E\{L(\Theta,\delta(x))\,|X=x\}$ $\forall x$.

Limite de estimadores de Bayes: δ tal que, se existe uma seq. de a prioris Λ_ν e estim. de Bayes δ_{Λ_ν}, tais que $\delta_{\Lambda_\nu}(x)\to\delta(x)$ para quase todo x.
(com respeito a $f(x|\theta)$ qdo $\nu\to+\infty$)

Informação de Kullback-Leibler:
$K[f,g] = \int \log\left[\frac{f(t)}{g(t)}\right] f(t)\,dt \geq 0$
$\Rightarrow$ qto maior K, mais fácil discriminar entre f e g.
Prioris com pouco efeito sobre $\pi(\theta|x) \to K$ grande, os dados tem muita influência sob $\pi(\theta|x)$.
Problema de $K[\pi(\theta|x),\pi(\theta)]$, depende de x.

Informação de Shannon:
$S(\pi) = \int K[\pi(\theta|x),\pi(\theta)]\, q_\pi(x)\,dx$
$q_\pi(x) = \int f(x|\theta)\pi(\theta)d\theta$.

Priori de Jeffreys: $\pi(\theta)\propto[I(\theta)]^{1/2}$, maximiza
$S_n(\pi) = \frac{1}{2}\log\frac{n}{2\pi e} + \int\pi(\theta)\log\frac{I(\theta)^{1/2}}{\pi(\theta)}d\theta + O(1)$
caso vetorial: $\pi(\theta)\propto\sqrt{|I(\theta)|}$

ESTIMAÇÃO MINIMAX:

Estimador Minimax: estimador δ^M de θ que minimiza o risco máximo, ou seja
$$\inf_\delta \sup_\theta R(\theta,\delta) = \sup_\theta R(\theta,\delta^M)$$

Teorema: suponha Λ uma dist. sobre Ω tal que $r(\Lambda,\delta_\Lambda) = \sup_\theta R(\theta,\delta_\Lambda)$, onde δ_Λ é um est. de Bayes com respeito a Λ. Então
i) δ_Λ é minimax
ii) se δ_Λ é a única sol. de Bayes respeito Λ ele é o único est. minimax.
iii) Λ é menos favorável $(r_\Lambda \geq r_{\Lambda'})$

Corolário: se a solução de Bayes tem risco constante ela é minimax.
Corolário: seja $w_\Lambda = \{\theta\in\Omega: R(\theta,\delta_\Lambda) = \sup_{\theta'} R(\theta',\delta_\Lambda)\}$
então δ_Λ é minimax se $\Lambda(w_\Lambda)=1$

Lema: seja δ um est. de Bayes (respect/ NVVUM, minimax, e admisível) de $g(\theta)$ com perda quadrática. Então $a\delta+b$ é estimador de Bayes (respec/ NVVUM, minimax e admis.) de $ag(\theta)+b$.

Def: Una seq. de distribuições a prior $\{\Lambda_n\}$ é menos favorável se para cada dist. a priori Λ; $r_\Lambda \leq r = \lim_{n\to\infty} r_{\Lambda_n}$
onde $r_{\Lambda_n} = \int R(\theta,\delta_n)\, d\Lambda_n(\theta)$.

Teorema: Suponha que $\{\Lambda_n\}$ seja uma seq. de dist. a priori com riscos de Bayes r_{Λ_n} satisfazendo $\lim_{n\to\infty} r_{\Lambda_n} = r$ e que δ seja um est. para o qual $\sup_\theta R(\theta,\delta)=r$, então:
i) δ é minimax
ii) a seq $\{\Lambda_n\}$ é menos favorável.

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

de bayes, então $r_\Delta = \int var[g(\Theta|x)]\, dQ(x)$.
...ular, se a variância a posteriori de $g(\Theta)$ é ...ndente de x, então $r_\Delta = var[g(\Theta)|x]$.

Lema: se δ é minimax de $g(\theta)$ quando $\theta \in \Omega_0 \in \Omega$. ...tão, se $\sup_{\theta\in\Omega_0} R(\theta,\delta) = \sup_{\theta\in\Omega} R(\theta,\delta)$, δ é minimax em Ω.

Método dos momentos: $E[|X|^k] < \infty$, $\mu_j = E(x^j)$, $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$, $\mu_j = h_j(\theta)$, $\hat{\mu}_j = h_j(\hat{\theta})$, $\hat{\theta} = h^{-1}(\hat{\mu}_j)$

Teorema: Y v.a. com média θ e variância σ^2. Então $aY+b$ é um est. inadmisível de θ sob perda quadrática quando: (i) $a>1$ ou (ii) $a<0$ ou (iii) $a=1$ e $b\neq 0$.

Testes de Hipótesis.

Tamanho do teste: $\alpha = \sup_{\theta\in\Omega_H} P_\theta\{x \in S_1\}$, em que S_1 é a região crítica ou de rejeição
Poder do teste: $P_\theta\{x\in S_1\}$ para um $\theta \in \Omega_K$ dado.
· Como função de θ para todo $\theta\in\Omega$ é a <u>função de poder</u> $\beta(\theta)$.

Testes Aleatorizados: rejeita H com prob. ϕ
se ocorre R, H é rejeitada com prob. ϕ.
É total/ especificado por ϕ (função crítica), $0\le\phi(x)\le 1$, $\forall x\in\mathcal{X}$.
Teste Não Aleat: $\phi(x) = \begin{cases} 1 & se\ x\in S_1 \\ 0 & se\ x\in S_0 \end{cases}$
Prob de rejeição: $E_\theta\{\phi(x)\} = \int \phi(x)\, dP_\theta(x)$
O procedimento usual é: maximizar o poder $\beta_\phi(\theta) = E_\theta[\phi(x)]$ sujeito a $E_\theta\{\phi(x)\} \le \alpha$ $\forall\, \theta\in\Omega_H$.

se o teste que maximiza o poder depende de θ $(\in\Omega_K)$ não existe um teste mais poderoso $\forall\ \theta\in\Omega_K$.

Teste Unif. Mais Poderoso, UMP: maximiza o poder $\forall\ \theta\in\Omega_K$.
↳ (ver mais abaixo)

Lema de Neyman-Pearson: P_0 e P_1 dist. de prob. com densidades p_0 e p_1, respect.
(i) <u>Existencia</u>: Para testar $H: p_0$ contra $K: p_1$ existe um teste ϕ e uma cte k tais que:
$$\underbrace{E_0\{\phi(x)\} = \alpha}_{[1]} \quad e \quad \phi(x) = \begin{cases} 1, & qdo\ p_1(x) > k\, p_0(x) \\ 0, & \quad '' \quad < \quad '' \end{cases} \quad [2]$$
(ii) Suficiência: se um teste satisfaz [1] e [2] para algum k → ele é mais poderoso de nível α.
(iii) Necessária: Se ϕ é mais poderoso de nível α então para algum k ele satisfaz [2] q.c. μ.
El também satisfaz [1] a menos que exista um teste de tamanho menor do que α e com poder 1.

Corolário: seja β o poder do teste mais pod. de nível α para testar P_0 contra P_1 → $\alpha<\beta$ a menos que $P_0 = P_1$.

O teste mais poderoso é determinado de forma única por [1] e [2] sempre que $p_1(x) = k\,p_0(x)$ tem medida nula μ.

Teste Unif. mais Poderoso: $H: \theta\in\Omega_H$ e $K: \theta\in\Omega_K$ (composta)
Um teste de H contra K é UMP de nível α se $E_\theta\{\phi(x)\}\le\alpha$, $\forall\,\theta\in\Omega_H$ e se $\forall$ teste ϕ^*, tal que $E_\theta\{\phi^*(x)\}\le\alpha$, $\forall\ \theta\in\Omega_H$, $E_\theta\{\phi(x)\} - E_\theta\{\phi^*(x)\} \ge 0$ $\forall\ \theta\in\Omega_K$.
→ $\theta\in\mathbb{R}$.

Distrib. com razão de verossiml. monótona: A família de densid. $p_\theta(x)$ que se existe uma função $T(x)\to\mathbb{R}$, tais que para quaisquer θ e θ', $\theta<\theta'$ as dist. P_θ e $P_{\theta'}$ são distintas e a razão $p_{\theta'}(x)/p_\theta(x)$ é não decrescente de $T(x)$

Teorema: $\theta\in\mathbb{R}$, X com $p_\theta(x)$ com razão de veros. monot. em $T(x)$:
(i) Para testar $H: \theta\le\theta_0$ contra $K: \theta>\theta_0$ existe um teste UMP que é dado por:
$$\phi(x) = \begin{cases} 1, & qdo\ T(x)>c \\ \gamma, & qdo\ T(x)=c \\ 0, & qdo\ T(x)<c. \end{cases} \quad [3]$$
onde c e γ são determinados por:
$E_{\theta_0}\{\phi(x)\} = \alpha$. [4]
(ii) a função de poder $\beta(\theta) = E_\theta\{\phi(x)\}$ deste teste é estritamente crescente $\forall\ \theta$ para os quais $0<\beta(\theta)<1$.
(iii) $\forall\ \theta'$, o teste determinado por [3] e [4] é UMP para testar H contra K com nível $\alpha' = \beta(\theta')$
(iv) Para qualquer $\theta<\theta_0$ o teste minimiza $\beta(\theta)$ entre todos os testes que satisfazem [4]

Corolário: F.E. unid. se $p_\theta(x) = \exp\{\eta(\theta)T(x) - B(\theta)\}h(x)$
se $\eta(\theta)$ é estrita/ monótona → existe UMP. e
se $\eta(\theta)$ é crescente: $\phi(x) = \begin{cases} 1, & T(x)>c \\ \gamma, & T(x)=c \\ 0, & T(x)<c \end{cases}$ | se $\eta(\theta)$ é decrescente $\phi(x) = \begin{cases} 1, & T(x)<c \\ \gamma, & T(x)=c \\ 0, & T(x)<c \end{cases}$

Nivel descritivo (p-valor) $\hat{p} = \hat{p}(x)$: menor nível de significância para o qual H seria rejeitada para a amostra observada x.
Em testes não aleatorizados: As regiões críticas S_α são aninhadas: $S_\alpha \subset S_{\alpha'}$ se $\alpha<\alpha'$ [5] então
$\hat{p} = \inf\{\alpha : x\in S_\alpha\}$
Em testes aleatorizados: $\phi_\alpha(x) \le \phi_{\alpha'}(x)$ $\forall\ x$ e $\alpha<\alpha'$
→ $\hat{p} = \inf\{\alpha : \phi_\alpha(x) = 1\}$

Lema: $H: \theta\in\Omega_H$. Se as regiões críticas satisfazem [5]
(i) se $\sup_{\theta\in\Omega_H} P_\theta\{x\in S_\alpha\} \le \alpha$ $\forall\ 0<\alpha<1$ [6] então
$P_\theta(\hat{p}\le u) \le u$, $\forall\ 0\le u\le 1$
(ii) se para $\theta\in\Omega_H$, $P_\theta(x\in S_\alpha) = \alpha$ $\forall\ 0<\alpha<1$ [7]
então $P_\theta(\hat{p}\le u) = u$, $\forall\ 0\le u\le 1$.
i.e, $\hat{p}\sim U(0,1)$

Limites de Confiança:
Lim. inf. de conf. uniform/ mais acurado para θ: uma função $\underline{\theta}(x)$, tal que: $P_\theta\{\underline{\theta}(x)\le\theta'\} = \min$, $\forall\ \theta'<\theta$.
sujeito a: $P_\theta\{\underline{\theta}(x)\le\theta\}\ge 1-\alpha$, $\forall\ \theta$ <1>
Estos minimizam $E_\theta\{L(\theta,\underline{\theta})\}$ <2> sujeita a <1>

Uma família de subconj. $S(x)$ do espaço paramet. Ω é dita construir uma famil. de regiões de conf. com coef. de conf. $1-\alpha$, se $P_\theta\{\theta\in S(x)\}\ge 1-\alpha$, $\forall\ \theta\in\Omega$.

$S(x) = \{\theta\in\Omega : \underline{\theta}(x) \le \theta < \infty\}$ Lim. inf. de conf.

Teorema: (i) Para cada $\theta_0\in\Omega$, seja $A(\theta_0)$ a região de aceitação de um teste de nível α de $H(\theta_0): \theta=\theta_0$

... cuja amostra x seja $S(x) = \{\theta : x \in A(\theta), \theta \in \Omega\}$ ... é uma famil. de regiões de conf. para θ com coef. ... confiança $1-\alpha$.

$\forall\, \theta_0$, $A(\theta_0)$ é UMP para testar $H(\theta_0)$ de nível α ... as alternativas $K(\theta_0)$, então para cada $\theta_0 \in \Omega$, ... minimiza $P_\theta\{\theta_0 \in S(x)\}$ $\forall\, \theta \in K(\theta_0)$ entre ... as reg de conf. para θ com coef. de conf $1-\alpha$.

corolário: $p_\theta(x)$, $\theta \in \Omega$ com razão de veros. monót. ... $T(x)$. $F_\theta(t)$ de $T = T(x)$ contínua em cada var. ... e θ quando a outra é fixada.

(i) Existe um L.I.C. unif. mais acurado $\underline{\theta}$ para θ com coef de conf. $1-\alpha$.

(ii) Se x é o valor obs. de X e $t = T(x)$ e se a equação $F_\theta(t) = 1-\alpha$ tem uma solução $\theta = \hat{\theta}$ em Ω, então esta solução é única e $\underline{\theta}(x) = \hat{\theta}$.

Para o limite superior Unif. mais acurado com coef. de conf. $1-\alpha$. $P_\theta\{T \geq t\} = 1-\alpha$ ou $F_\theta(t) = \alpha$.

Intervalos de Confiança: sejam $\underline{\theta}$ e $\bar{\theta}$ limites inf. e sup de conf. com coef. de conf. $1-\alpha_1$ e $1-\alpha_2$, respect/. Dada a amostra X, $F_{\underline{\theta}}(t) = 1-\alpha_1$, $F_{\bar{\theta}}(t) = \alpha_2$ se $\alpha_1 + \alpha_2 < 1$, $\underline{\theta} < \bar{\theta}$.
Os interv. $(\underline{\theta}, \bar{\theta})$ são chamados I.C. para θ com coef de conf. $1-\alpha_1-\alpha_2$.

Se $\bar{\theta}$ e $\underline{\theta}$ são Unif. mais acurado eles minimizam $E_\theta[L_1(\theta,\underline{\theta})]$ e $E_\theta[L_2(\theta,\bar{\theta})]$. Seja $L(\theta;\underline{\theta},\bar{\theta}) = L_1(\theta,\underline{\theta}) + L_2(\theta,\bar{\theta})$. então os intervalos $(\underline{\theta},\bar{\theta})$ minimizam $E\{L(\theta,\underline{\theta},\bar{\theta})\}$ sujeito a $P_\theta\{\underline{\theta} > \theta\} \leq \alpha_1$ e $P_\theta\{\bar{\theta} < \theta\} \leq \alpha_2$.

Quantidade Pivotal: função $h(x,\theta) \to \mathbb{R}$ cuja distrib. não depende de θ.
Seja $F_h(\cdot)$ a f.d.a. de $h(x,\theta)$, estritamente crescente e contínua. Então $S(x) = \{\theta \in \Omega : h(x,\theta) \leq F_h^{-1}(1-\alpha)\}$ é uma reg. de conf. para θ com coef. de conf. $1-\alpha$.

Hipóteses Bilaterais:

Teorema: (i) para testar $H: \theta \leq \theta_1$ ou $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ contra $K: \theta_1 < \theta < \theta_2$ na F.E. unid. com $\eta(\theta)$ estrit/ crescente, existe um teste UMP dado por

$$\phi(x) = \begin{cases} 1 & \text{qdo } C_1 < T(x) < C_2 \quad (C_1 < C_2) \\ \gamma_i & \text{qdo } T(x) = C_i,\ i=1,2. \\ 0 & \text{qdo } T(x) < C_1 \text{ ou } T(x) > C_2 \end{cases}$$

em que os C's e γ's são determ. por:

$$E_{\theta_1}\{\phi(x)\} = E_{\theta_2}\{\phi(x)\} = \alpha \quad \langle * \rangle$$

(ii) Este teste minimiza $E_\theta\{\phi(x)\}$ sujeito a $\langle * \rangle$ $\forall\, \theta < \theta_1$ e $\forall\, \theta > \theta_2$

(iii) Para $0 < \alpha < 1$, a função de poder desse teste tem um máximo em um pto θ_0 entre θ_1 e θ_2 e decresce estrita/ quando θ se afasta de θ_0 em qualquer direção a menos que existam t_1 e t_2 t.q $P_\theta(T(x)=t_1) + P_\theta(T(x)=t_2) = 1$ $\forall\, \theta$.

Testes não viciados: teste ϕ tal que $\beta_\phi(\theta) \leq \alpha$ se $\theta \in \Omega_H$ e $\beta_\phi(\theta) \geq \alpha$ se $\theta \in \Omega_K$.
Sempre que um teste UMP existe, ele é não viciado.

Regiões de Credibilidade: Qualquer conjunto $S(x)$ t.q. $P(\Theta \in S(x) \mid X = x\} \geq 1-\alpha$. $\forall\, x$. é uma reg. de cred. para θ com prob. $1-\alpha$.

Regiões HPD: $S(x) = \{\theta \in \Omega \mid \pi(\theta \mid x) \geq k\}$ com k determinado por $P\{\Theta \in S(x) \mid X = x\} = 1-\alpha$.

Teste da Razão de Verossimilhanças:

$$\lambda(x) = \frac{\sup_{\theta \in \Omega_H} L(\theta)}{\sup_{\theta \in \Omega} L(\theta)}, \quad 0 \leq \lambda(x) \leq 1$$

Um T.R.V. é qualquer teste que rejeita H se e somente se $\lambda(x) < c$, em que $c \in [0,1]$.
Para o nível α: $\sup_{\theta \in \Omega_H} P_\theta(\lambda(x) < c_\alpha) = \alpha$.

Estatística da razão de verossimilhanças:
$$w(x) = 2\{\log L(\hat{\theta}) - \log L(\theta_0)\}$$

Estatística de Wald:
$$W(x) = (\hat{\theta} - \theta_0)^2 I(\hat{\theta})$$

Est. escore (Rao): $S_R(x) = U(\theta_0)^2 I(\theta_0)^{-1}$
$U(\theta) = \partial \log L(\theta)/\partial\theta$. função escore.

Est. gradiente: $S_T(x) = U(\theta_0)(\hat{\theta} - \theta_0)$

i) $\sup_\theta R(\theta,\delta) \geq \int R(\theta,\delta)\, d\Lambda(\theta)$
$\geq \int R(\theta,\delta_\Lambda)\, d\Lambda(\theta) \overset{r(\Lambda,\delta_\Lambda)}{=} \sup_\theta R(\theta,\delta_\Lambda)$

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

Tamanho do teste: $\sup_{\mu \leq \mu_0} E\{T_c(x)\}$