

PROBABILIDADE ESTATÍSTICA

PRINCÍPIOS DE REDUÇÃO DE DADOS E ELEMENTOS DA TEORIA DA DECISÃO ESTATÍSTICA

Luis Antonio Fantozzi Alvarez

Universidade de São Paulo

O Problema de Inferência Estatística

AMBIENTE

- Nosso ponto de partida é um espaço mensurável (Ω, Σ) .
 - Ω é o espaço amostral.
 - Σ é o espaço de eventos aos quais podemos atribuir probabilidades (σ -álgebra).
- Pesquisador observa uma amostra, dada pela variável aleatória $\mathbf{X} := \text{Id}_{\Omega}$, cuja lei é dada por uma probabilidade P sobre (Ω, Σ) desconhecida.
- O pesquisador postula uma família de leis de probabilidade $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$ candidatas a terem gerado a amostra.
 - Θ é o espaço de parâmetros que indexa a família.
 - À família \mathcal{P} damos o nome de modelo.
- Dizemos que um modelo está bem-especificado se $\exists \theta_0 \in \Theta, P = P_{\theta_0}$.
- **Notação:** \mathbb{E}_{θ} denota a expectativa de uma variável aleatória com domínio em (Ω, Σ) com respeito a P_{θ} .

AMOSTRAGEM ALEATÓRIA

- Em diversos contextos, o experimento definido por (Ω, Σ, P) tem a seguinte interpretação: pesquisador possui acesso a uma amostra de n unidades sorteadas ao acaso de uma população de interesse, para os quais ele observa um vetor de k características $X_i, i = 1, \dots, n$.
- Se população é **grande**, o conceito adequado para modelar o experimento é o de **amostragem aleatória**.
 - **Amostragem aleatória**: espaço é $(\prod_{i=1}^n (\mathbb{R}^k), \otimes_{i=1}^n \mathcal{B}(\mathbb{R}^k), \otimes_{i=1}^n G)$.
 - G é a lei de probabilidade sobre $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ que reflete a distribuição conjunta das características de interesse na população.
 - Observações $\mathbf{X} = (X_1, \dots, X_n)$ são tais que cada X_i tem a mesma distribuição que G , e as observações X_i são independentes entre si (o valor amostrado na i -ésima posição não exibe associação com o valor amostrado em outra posição $j \neq i$)
 - Quando o experimento é definido por amostragem aleatória, modelo se resume a uma família de distribuições para as características de interesse na população.
 - Em outras palavras, modelo é da forma $\mathcal{P} = \{\otimes_{i=1}^n G : G \in \mathcal{G}\}$, onde \mathcal{G} é uma família de distribuições em $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$

EXEMPLO

- **Exemplo:** pesquisador observa uma variável escalar, e supõe que tenha sido amostrada de uma distribuição normal com média desconhecida e variância unitária.
 - Quem é \mathcal{P} ?
 - E se a variância é desconhecida?
 - E se pesquisador possui n observações independentes?

MODELOS

- Um modelo \mathcal{P} é dito de **dimensão finita** quando o espaço de parâmetros pode ser identificado com um subconjunto de \mathbb{R}^s , $s \in \mathbb{N}$.
 - Amostragem aleatória de uma população normal com média e variância desconhecidas define um modelo cujo espaço de parâmetros é da forma $\mathbb{R} \times \mathbb{R}_+$.
- Um modelo \mathcal{P} é dito de **dimensão infinita** quando ele não tem dimensão finita.
 - Amostragem aleatória de uma população cuja distribuição admite uma densidade com respeito à medida de Lebesgue simétrica em relação à média define um modelo de dimensão infinita.

DEFINIÇÃO DE ESTATÍSTICA

- Uma **estatística** é uma transformação mensurável da amostra, i.e. uma função $T : \Omega \mapsto \mathcal{T}$ em que $T^{-1}(B) \in \Sigma$, para todo $B \in \mathcal{S}$, onde \mathcal{S} é uma σ -álgebra sobre o espaço \mathcal{T} .
 - Nesse caso $T(\mathbf{X})$ define uma variável aleatória com valores em \mathcal{T} .

FAMÍLIA DOMINADA

- No que segue, supomos que existe uma medida μ σ -finita tal que todo elemento $P_\theta \in \mathcal{P}$ admite densidade p_θ com respeito a μ (nesse caso, dizemos que a família \mathcal{P} é *dominada*).
 - Para famílias de distribuições “discretas” sobre o espaço $(\Omega, 2^\Omega)$, com Ω é enumerável, μ é a medida de contagem e p_θ são as f.m.p. definidas por $p_\theta(x) = P_\theta(\{x\})$ para todo $x \in \Omega$.
 - Para famílias de distribuições “contínuas” sobre $(\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s))$, μ é a medida de Lebesgue sobre $\mathcal{B}(\mathbb{R}^s)$ e p_θ são as f.d.p.

CONDIÇÃO TÉCNICA

- No que segue, vamos supor também que, para qualquer estatística, as distribuições condicionais de $\mathbf{X}|T(\mathbf{X})$ sob cada P_θ admitem uma **probabilidade condicional regular** $(A, t) \mapsto P_\theta[\mathbf{X} \in A | T(\mathbf{X}) = t]$.
 - Para cada $A \in \Sigma$, $\omega \mapsto P_\theta[\mathbf{X} \in A | T(\mathbf{X}) = T(\omega)]$ é uma versão da probabilidade condicional $P_\theta[A | \sigma(T(\mathbf{X}))]$; e para cada $\omega \in \Omega$, $A \mapsto P_\theta[A | T(\mathbf{X}) = T(\omega)]$ é uma medida de probabilidade sobre Σ .
 - Essa condição limita a complexidade de (Ω, Σ) .
 - Na maior parte dos casos práticos é satisfeita.
 - Condições suficientes em Durrett (2019) (fora do escopo do curso).
- Exemplo: no caso discreto, podemos tomar $P_\theta[\mathbf{X} \in A | T(\mathbf{X}) = t] = \sum_{a \in A} P_\theta[\{a\} | T(\mathbf{X}) = t]$, onde

$$P_\theta[\{a\} | T(\mathbf{X}) = t] = \begin{cases} \frac{P_\theta(\{a\})}{P_\theta[\{T(\mathbf{X})=t\}]}, & \text{se } T(a) = t \text{ e } P_\theta[\{T(\mathbf{X}) = t\}] > 0 \\ 0, & \text{se } T(a) \neq t \text{ e } P_\theta[\{T(\mathbf{X}) = t\}] > 0 \\ P_\theta(\{a\}), & \text{se } P_\theta[\{T(\mathbf{X}) = t\}] = 0 \end{cases}$$

Suficiência Estatística

ESTATÍSTICA SUFICIENTE

- Uma estatística é tão somente uma transformação (mensurável) da amostra, isto é, uma função $T(\mathbf{X})$ tal que $T^{-1}(A) \in \Sigma$ para todo A na σ -álgebra do contradomínio.

DEFINIÇÃO

Uma estatística T é dita suficiente para θ se a distribuição condicional de $\mathbf{X}|T(\mathbf{X})$ não depende de θ , isto é, se existe H tal que:

$$P_{\theta}[\mathbf{X} \in A | T(\mathbf{X})] = H(A | T(\mathbf{X})), \quad \forall \theta \in \Theta, A \in \Sigma.$$

- Uma vez que conhecemos $T(\mathbf{X})$, não há mais informação adicional sobre θ na amostra.
-

EXEMPLO

Suponha que a amostra consista de duas Bernoullis independentes e identicamente distribuídas, com parâmetro $\theta \in (0, 1)$ desconhecido. Neste caso, a f.m.p. é:

$$P_{\theta}[X_1 = x, X_2 = y] = \theta^x(1 - \theta)^{1-x}\theta^y(1 - \theta)^{1-y}, \quad \forall x, y \in \{0, 1\},$$

de onde segue que, para todo $t \in \{0, 1/2, 1\}$.

$$\begin{aligned} P_{\theta}[X_1 = x, X_2 = y | X_1 + X_2 = 2t] &= \frac{P_{\theta}[X_1 = x, X_2 = y, X_1 + X_2 = 2t]}{P_{\theta}[X_1 + X_2 = 2t]} = \\ &= \frac{\theta^{x+y}(1 - \theta)^{2-x-y} \mathbf{1}\{x + y = 2t\}}{\binom{2}{2t} \theta^{2t} (1 - \theta)^{2-2t}} = \frac{1}{\binom{2}{2t}} \mathbf{1}\{x + y = 2t\}, \end{aligned}$$

de onde concluímos que $(X_1 + X_2)/2$ é suficiente.

TEOREMA DA FATORAÇÃO DE NEYMAN-FISHER

TEOREMA

$T(\mathbf{X})$ é suficiente para θ se, e somente se, existem funções h_θ , $\theta \in \Theta$, e c tais que, para todo $\theta \in \Theta$:

$$p_\theta(\mathbf{x}) = h_\theta(T(\mathbf{x}))c(\mathbf{x}), \quad \mu\text{-q.t.p.}$$

- Critério conveniente para encontrar uma estatística suficiente.
 - No caso discreto, “ μ -q.t.p.” pode ser lido como para todo \mathbf{x} .
 - No caso contínuo, condição pode ser violada num conjunto de medida de Lebesgue zero (por exemplo, conjuntos enumeráveis de pontos).
- Veremos demonstração no caso discreto.
 - Demonstração no caso geral é mais complexo, pois requer resultados adicionais sobre famílias dominadas.

TEOREMA DA FATORAÇÃO (CASO DISCRETO)

No que segue, considere uma família \mathcal{P} sobre o espaço $(\Omega, 2^\Omega)$, onde Ω é enumerável.

DEMONSTRAÇÃO.

\Rightarrow Suponha que $T(\mathbf{X})$ é suficiente. Fixe $x \in \Omega$. Observe que:

$$\begin{aligned} p_\theta(x) &= P_\theta[\{x\}] = P_\theta[\{x\} \cap \{T(\mathbf{X}) = T(x)\}] = \\ &= \mathbb{E}_\theta[P_\theta[\{x\} | T(\mathbf{X})] \mathbf{1}_{\{T(\mathbf{X})=T(x)\}}] = \\ &= \sum_{\omega \in \Omega} P_\theta[\{x\} | T(\mathbf{X}) = T(\omega)] \mathbf{1}_{\{T(\mathbf{X})=T(x)\}}(\omega) p_\theta(\{\omega\}) = \\ &= \underbrace{P_\theta[\{x\} | T(\mathbf{X}) = T(x)]}_{c(x)} \underbrace{\sum_{\omega \in \Omega: T(\omega)=T(x)} p_\theta(\{\omega\})}_{=P_\theta[\{T(\mathbf{X})=T(x)\}]=h_\theta(T(x))} \end{aligned}$$

\Leftarrow Suponha que p_θ é fatorável. Considere $\tilde{P}[A | T(\mathbf{X}) = t] = \frac{\sum_{a \in A} c(a) \mathbf{1}_{\{T(a)=t\}}}{\sum_{\omega \in \Omega: T(\omega)=t} c(\omega)}$ se

$\sum_{\omega \in \Omega: T(\omega)=t} c(\omega) > 0$ e $S(A)$ do contrário, onde S é uma probabilidade arbitrária sobre Ω . Fácil verificar que \tilde{P} define uma probabilidade condicional regular para todo θ . \square

EXEMPLO

Suponha que o pesquisador observe uma amostra iid X_1, X_2, \dots, X_n de uma distribuição exponencial com parâmetro $\lambda > 0$ desconhecido. Neste caso, observe que:

$$\begin{aligned} p_\lambda(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n (\lambda e^{-\lambda x_i} \mathbf{1}_{\{x_i > 0\}}) = \\ &= \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) \mathbf{1}_{\{\min_i x_i > 0\}} = h_\lambda\left(\sum_{i=1}^n x_i\right) c(\mathbf{x}), \end{aligned}$$

de onde concluímos que $(X_1 + X_2 + \dots + X_n)/n$ é estatística suficiente.

ESTATÍSTICA SUFICIENTE MINIMAL

- Observe que o conceito de estatística suficiente só nos informa sobre a capacidade de uma transformação em condensar a informação relevante na amostra sobre θ .
- Este conceito não versa sobre o “tamanho” desta estatística.
 - De fato, a própria amostra, $T(\mathbf{X}) = \mathbf{X}$, é sempre uma estatística suficiente.
- Note que, se T é estatística suficiente e $T = M \circ U$, então U é estatística suficiente (pois $\sigma(T) \subseteq \sigma(U)$).
 - Estatísticas mais “finas” que uma estatística suficiente são estatísticas suficientes.
- Na outra direção, podemos pensar na estatística suficiente mais “grossa” possível.

ESTATÍSTICA SUFICIENTE MINIMAL (CONT.)

DEFINIÇÃO

Uma estatística T é dita suficiente minimal, se:

1. T é suficiente.
2. Para qualquer outra estatística S suficiente, existe M tal que $T = M \circ S$.

LEMA

Considere uma estatística T tal que:

$$T(\mathbf{x}) = T(\mathbf{y}) \iff \mathbf{y} \in D(\mathbf{x}),$$

onde

$$D(\mathbf{x}) = \{\mathbf{y} : p_{\theta}(\mathbf{y}) = p_{\theta}(\mathbf{x})h(x, y) \quad \forall \theta \in \Theta \text{ e algum } h(x, y) > 0\}.$$

Então T é suficiente minimal.

EXEMPLO

Suponha que o espaço amostral é \mathbb{R}_+^n . Considere uma amostra aleatória de $U[0, \theta]$, $\theta > 0$. Neste caso:

$$p_\theta(x_1, \dots, x_n) = \frac{1}{\theta^n} \mathbf{1}\{\max_i x_i < \theta\}.$$

Pelo critério de fatoração, $X_{(n)} := \max_i X_i$ é suficiente. Vamos mostrar que é minimal suficiente usando o lema anterior. Observe que $D(\mathbf{x}) = \{\mathbf{y} : y_{(n)} = x_{(n)}\}$. De fato, se $y_{(n)} \neq x_{(n)}$, ao considerar $\theta' = (x_{(n)} + y_{(n)})/2$, teremos que $0 = p_{\theta'}(\mathbf{x}) < p_{\theta'}(\mathbf{y})$ ou $0 = p_{\theta'}(\mathbf{y}) < p_{\theta'}(\mathbf{x})$. Segue do lema anterior que $X_{(n)}$ é minimal.

ANCILARIDADE

DEFINIÇÃO

Uma estatística é dita ancilar para θ se sua distribuição não depende de θ , i.e., se existe F tal que:

$$P_{\theta}[T(X) \in A] = F[A], \quad \forall \theta \in \Theta, A \in \mathcal{S},$$

onde \mathcal{S} é σ -álgebra acoplada ao contradomínio de T .

EXEMPLO

No modelo linear Gaussiano homocedástico com regressores fixos:

$$\mathbf{y}_{n \times 1} \sim \mathcal{N}(\mathbf{Z}\beta, \sigma^2 \mathbb{I}_n),$$

onde $\text{posto}(\mathbf{Z}) = k$, $\beta \in \mathbb{R}^k$ desconhecido e $\sigma^2 > 0$ conhecido; veremos em Econometria I que a estatística $S = \hat{\mathbf{e}}' \hat{\mathbf{e}}$, onde $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{Z}\hat{\beta}_{\text{MQO}}$, é ancilar.

SUFICIÊNCIA COMPLETA

- Gostaríamos de que uma estatística suficiente fosse independente de estatísticas ancilares, visto que essas não nos trazem informação de θ .
- Conceito apropriado para isto é o de estatística **completa suficiente**.

DEFINIÇÃO

Uma estatística T é dita completa para θ se, para qualquer f mensurável com valores reais:

$$\mathbb{E}_{\theta}[f(T)] = 0, \forall \theta \in \Theta \implies \mathbb{P}_{\theta}[f(T) = 0] = 1, \forall \theta \in \Theta$$

TEOREMA (BASU)

Uma estatística **completa suficiente** para θ é independente de qualquer estatística ancilar de θ .

SUFICIÊNCIA COMPLETA VS. SUFICIÊNCIA MINIMAL

TEOREMA (BAHADUR)

Se U é estatística completa suficiente de dimensão finita, então é suficiente minimal.

- Recíproca **não** é verdadeira: existem estatísticas minimais suficientes de dimensão finita que não são completas (veja Lehmann e Casella, 1998).

Famílias Exponenciais

FAMÍLIA EXPONENCIAL

DEFINIÇÃO

Uma família de distribuições $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ sobre um espaço mensurável (Ω, Σ) é dita uma família exponencial se:

1. Existe uma medida μ tal que cada elemento $P_\theta \in \mathcal{P}$ admite densidade p_θ com respeito a μ .
2. Existem $\eta : \Theta \mapsto \mathbb{R}^s$, $T : \Omega \mapsto \mathbb{R}^s$, $B : \Theta \mapsto \mathbb{R}$ e $h : \Omega \mapsto \mathbb{R}_+$ tais que:

$$p_\theta(\mathbf{x}) = \exp(\eta(\theta)' T(\mathbf{x}) - B(\theta)) h(\mathbf{x})$$

- Diversas distribuições conhecidas pertencem a esta família: Gamma, Chi-Quadrado, Beta, Normal, Poisson, Negativo-Binomial.
- Propriedade útil: se $\mathcal{P}_1, \dots, \mathcal{P}_n$ são famílias exponenciais, então $\mathcal{P}^n := \{\otimes_{i=1}^n P_i : P_i \in \mathcal{P}_i\}$ é uma família exponencial.
 - Consequência: a distribuição amostral de uma amostra aleatória de uma família exponencial também constitui uma família exponencial.

EXEMPLO

Sejam $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ uma amostra aleatória de $N(\mu, \sigma^2)$, onde $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ são desconhecidos. Neste caso, fazendo $\theta = (\mu, \sigma^2)'$, temos:

$$p_{\theta}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x_i - \mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{\mu^2}{2\sigma^2} n\right) \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n, \quad (1)$$

de onde segue que a família é exponencial com

$$T(\mathbf{x}) = (\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i)' \text{ e } \eta(\theta) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)'.$$

IDENTIFICABILIDADE DE FAMÍLIAS DE DISTRIBUIÇÕES

DEFINIÇÃO

Para uma família $\{P_\theta : \theta \in \Theta\}$, dizemos que o parâmetro é identificável se:

$$\theta = \theta' \iff P_\theta = P_{\theta'}$$

- Identificação é o requerimento básico para estimação pontual (mais à frente).
 - Requer a existência de um mapa 1-1 entre parâmetro e distribuição amostral.
- No exemplo anterior, θ é identificado, visto que $f_1(P_\theta) := \mathbb{E}_\theta[X_1] = \mu$ e $f_2(P_\theta) := \mathbb{E}[X_1^2] = \sigma^2 + \mu^2$.
- Por outro lado, se a família paramétrica fosse $\{\mathcal{N}(\theta \vee 0, 1) : \theta \in \mathbb{R}\}$, θ não é identificável.

PARAMETRIZAÇÃO NATURAL DE FAMÍLIA EXPONENCIAL

Note que, da definição de família exponencial:

$$p_{\theta}(\mathbf{x}) = \exp(\eta(\theta)' T(\mathbf{x}) - B(\theta)) h(\mathbf{x}), \quad \theta \in \Theta,$$

segue uma reparametrização **natural**

$$p_{\eta}(\mathbf{x}) = \exp(\eta' T(\mathbf{x}) - B^*(\eta)) h(\mathbf{x}), \quad \eta \in \Xi,$$

onde $\Xi := \{\eta(\theta) : \theta \in \Theta\} \subseteq \mathbb{R}^s$ é o **espaço paramétrico natural**; e $B^*(\eta) = B(\theta)$ para algum (qualquer) $\theta \in \Theta$ tal que $\eta(\theta) = \eta$

DEFINIÇÃO

Uma família exponencial $\{P_{\theta} : \theta \in \Theta\}$ é dita uma família de posto cheio se sua reparametrização natural é tal que η é identificável e Ξ contém uma bola aberta de \mathbb{R}^s .

- No exemplo de amostra aleatória normal, $\Theta = (-\infty, 0) \times \mathbb{R}$, logo a família tem posto cheio.
- $\{N(\mu, \mu^2) : \mu \in \mathbb{R}\}$ não tem posto cheio (família curvada).

SUFICIÊNCIA EM FAMÍLIAS EXPONENCIAIS

COROLÁRIO

Numa família exponencial, $T(\mathbf{X})$ é uma estatística suficiente.

COROLÁRIO

*Numa família exponencial **de posto cheio**, $T(\mathbf{X})$ é uma estatística suficiente minimal.*

TEOREMA

*Numa família exponencial **de posto cheio**, $T(\mathbf{X})$ é uma estatística suficiente completa.*

Elementos de decisão estatística

O PROBLEMA DE DECISÃO ESTATÍSTICA

- Considere um experimento estatístico (Ω, Σ, P) , e um modelo de leis de probabilidade candidatas para P , $\mathcal{P} = \{P : \theta \in \Theta\}$, com $P = P_{\theta_0}$ para algum $\theta_0 \in \Theta$.
- Suponha que o pesquisador esteja interessado em realizar inferência sobre um parâmetro escalar $\psi(\theta_0) \in \mathbb{R}$.
 - Após observar $\mathbf{X} = \text{Id}_{\Omega}$, pesquisador deve produzir estimativa para $\psi(\theta_0)$.
- A **perda** que o pesquisador incorre em afirmar um valor c a $\psi(\theta_0)$ quando $\theta_0 = \theta$ é dada por $L(\theta, c)$, com $L : \Theta \times \mathbb{R} \mapsto \mathbb{R}_+$.
 - Perda quadrática: $L(\theta, c) = (c - \psi(\theta))^2$.
 - Perda absoluta: $L(\theta, c) = |c - \psi(\theta)|$.
- Uma **regra de decisão estatística** ou estimador é uma transformação mensurável $\delta : \Omega \mapsto \mathbb{R}$.
 - Após observar o valor de $\mathbf{X}(\omega)$ realizado, pesquisador reporta a estimativa dada por $\delta(\mathbf{X}(\omega))$.
- A perda esperada ou risco da regra de decisão δ , quando $\theta_0 = \theta$, é dada por:

$$R(\theta, \delta) = \mathbb{E}_{\theta}[L(\theta, \delta(\mathbf{X}))]$$

ESCOLHENDO UMA REGRA DE DECISÃO

- Dado que desconhecemos o valor verdadeiro de θ_0 , uma regra de comunicação ideal δ^* numa classe de regras candidatas Δ deveria ser tal que, para todo $\theta \in \Theta$:

$$\delta^* = \inf_{\delta \in \Delta} R(\theta, \delta),$$

- Infelizmente, se Δ é irrestrita, problema não possuirá, no geral, solução.
 - Por exemplo, sob perda quadrática, para todo $\theta \in \Theta$, a regra de decisão $\delta(\mathbf{X}) = \psi(\theta)$ atinge risco zero sob $\theta_0 = \theta$, o que implica que, a não ser em casos extremos em que $\theta \mapsto \psi(\theta)$ é constante ou \mathbf{X} é perfeitamente informativo sobre $\psi(\theta)$, a regra ótima δ^* não existirá.
- Dada a inexistência, no geral, de um δ^* que minimiza o risco uniformemente sobre Θ quando Δ é irrestrito, a literatura estatística propõe dois conjuntos de soluções alternativas ao problema
 1. Restringir Δ .
 2. Minimizar o pior risco sobre Θ (estimação minimax) ou um risco médio sobre Θ , onde a média se dá com respeito a uma distribuição sobre Θ (estimadores de Bayes).

RESTRINGINDO Δ : ESTIMADORES NÃO VICIADOS

- Uma restrição bastante comum à classe de estimadores é de que eles não possuam viés.
- Formalmente, um estimador δ de um parâmetro $\psi(\theta)$ é **não viciado** numa família $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ de distribuições candidatas se:

$$\mathbb{E}_\theta[\delta(\mathbf{X})] = \psi(\theta), \quad \forall \theta \in \Theta.$$

- Um estimador δ de $\psi(\theta)$ é dito não viciado de variância uniformemente mínima (numa família \mathcal{P}) se:
 1. δ é não viciado.
 2. para qualquer outro estimador $\tilde{\delta}$ de $\psi(\theta)$ não viciado em \mathcal{P} :

$$\mathbb{V}_\theta[\delta(\mathbf{X})] \leq \mathbb{V}_\theta[\tilde{\delta}(\mathbf{X})], \quad \forall \theta \in \Theta$$

CARACTERIZANDO ESTIMADORES NVVUM

Um estimador δ é dito de variância finita em \mathcal{P} se $\mathbb{V}_\theta[\delta(\mathbf{X})] < \infty$ para todo $\theta \in \Theta$.

TEOREMA

Um estimador δ de variância finita e não viciado para $\psi(\theta)$ em \mathcal{P} é não viciado de variância uniformemente mínima em \mathcal{P} se, e somente se, para qualquer estimador U de variância finita e não viciado para zero em \mathcal{P} :

$$\mathbb{E}_\theta[\delta(\mathbf{X})U(\mathbf{X})] = 0, \quad \forall \theta \in \Theta.$$

COROLÁRIO

Se um estimador δ de $\psi(\theta)$ com variância finita é não viciado com variância uniformemente mínima em \mathcal{P} , então ele é único.

FUNÇÕES PERDAS MAIS GERAIS

- Estimadores não viciados de variância uniformemente mínima minimizam o risco, sob perda quadrática, uniformemente na classe de estimadores não viciados.
- Será que é possível encontrar estimadores que minimizem o risco uniformemente, na mesma classe, para outras funções perdas?
 - Resposta dada pelo teorema abaixo.

TEOREMA

Seja δ um estimador não viciado de um parâmetro $\psi(\theta)$ numa família \mathcal{P} , e T uma estatística completa suficiente para \mathcal{P} . Então:

- 1. $\phi(\mathbf{X}) = \mathbb{E}[\delta | T(\mathbf{X})]$ define um estimador não viciado de $\psi(\theta)$, e o único estimador não viciado de $\psi(\theta)$ que é função de T .*
- 2. ϕ minimiza o risco uniformemente na classe de estimadores não viciados de $\psi(\theta)$, para qualquer perda $L(\theta, z)$ convexa no segundo argumento.*
- 3. Se ϕ possui risco finito para algum θ e a perda é estritamente convexa no segundo argumento, então ϕ é o único estimador que minimiza o risco uniformemente na classe de estimadores não viciados de $\psi(\theta)$.*

LIMITE INFERIOR DE CRÁMER-RAO

- Resultados anteriores nos deram dois métodos para se tentar construir um ENNVUM.
 - Encontrar δ resolvendo o “sistema de equações”: $\mathbb{E}_\theta[\delta(\mathbf{X})] = \psi(\theta)$, $\mathbb{E}_\theta[\delta(\mathbf{X})U(\mathbf{X})] = 0$ para todo θ e U não viciado para zero (de variância finita).
 - Encontrar um estimador não viciado, uma estatística completa suficiente e calcular a esperança condicional.
- Uma pergunta alternativa é: se temos um estimador não viciado, será que existe uma condição **suficiente** simples de ser verificada, que quando garantida implica que estimador é ENNVUM?
 - Resposta é afirmativa, e dada pelo limite inferior de Crámer-Rao.
 - Esse limite nos dá a menor variância que um estimador não viciado de um parâmetro em uma família de dimensão finita pode atingir.
 - Limite nem sempre é atingível por um estimador, mas, se for o caso, sabemos que ele é ENNVUM.

LIMITE INFERIOR DE CRÁMER-RAO: DERIVAÇÃO

- Seja $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ uma família (dominada) de dimensão finita, com $\Theta \subseteq \mathbb{R}^s$.
- Suponha que, para todo $\theta \in \text{int}(\Theta)$ e $x \in \Omega$, $\tau \mapsto p_\tau(x)$ é continuamente diferenciável em $\tau = \theta$.
- Considere um estimador δ não viciado de um parâmetro escalar $\psi(\theta)$.
- Tome um ponto $\theta \in \text{int}(\Theta)$.
- Considere uma direção $\mathbf{e} \in \mathbb{R}^s$, $\mathbf{e} \neq \mathbf{0}$.
- Defina a variável aleatória $S_{\theta, \mathbf{e}}(x) = \frac{\nabla_\tau p_{\theta_0}(x) \cdot \mathbf{e}}{p_\theta(x)}$.
- Observe que, como $\int p_\tau(x) \mu(dx) = 1$ para todo τ , se pudermos diferenciar em τ por dentro da integral, segue que $\mathbb{E}_\theta[S_{\theta, \mathbf{e}}(x)] = 0$ e que $\text{cov}_\theta(\delta, S_{\theta, \mathbf{e}}) = \int \delta(x) (\nabla_\tau p_\theta(x) \cdot \mathbf{e}) \mu(dx)$.
- Ademais, como $\int \delta(x) p_\tau(x) \mu(dx) = \psi(\tau)$ para todo τ , se pudermos diferenciar por debaixo da integral, segue que $\text{cov}_\theta(\delta, S_{\theta, \mathbf{e}}) = \nabla_\tau \psi(\theta) \cdot \mathbf{e}$

DERIVAÇÃO (CONT)

- Combinando os fatos anteriores com a desigualdade de Cauchy-Schwarz, temos:

$$\frac{(\partial_\tau \psi(\theta) \cdot \mathbf{e})^2}{\mathbb{V}_\theta[S_{\theta, \mathbf{e}}]} \leq \mathbb{V}_\theta[\delta]$$

onde $\mathbb{V}_\theta[S_{\theta, \mathbf{e}}] = \mathbf{e}' \mathbb{V}_\theta [\partial_\tau \log p_\theta] \mathbf{e}$, e $\mathbb{V}_\theta [\partial_\tau \log p_\theta]$ é a matriz de covariância do vetor aleatório $\partial_\tau \log p_\theta = \frac{\partial_\tau p_\theta}{p_\theta}$.

- Não é difícil mostrar que, se $\mathbb{V}_\theta [\partial_\tau \log p_\theta]$ tem posto cheio, então, maximizando com respeito a \mathbf{e} , chegamos a:

$$\mathbb{V}_\theta[\delta] \geq (\partial_\tau \psi(\theta))' I(\theta)^{-1} (\partial_\tau \psi(\theta)),$$

onde $I(\theta) := \mathbb{V}_\theta [\partial_\tau \log p_\theta]$ é conhecida como informação de Fisher.

- Invertibilidade de $I(\theta)$ está relacionada com identificabilidade da família $\{P_\theta\}$.
- Sob condições adicionais de diferenciabilidade, fácil ver que $I(\theta) = -\mathbb{E}_\theta[\frac{\partial^2}{\partial \tau \partial \tau'} \log p_\theta]$.

Referências

REFERÊNCIAS



Casella, George e Roger L Berger (2001). *Statistical inference*. Duxbury.



Durrett, Rick (2019). *Probability: Theory and Examples*. 5ª ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. DOI: 10.1017/9781108591034.



Lehmann, E e George Casella (1998). *Theory of point estimation*. Springer Science & Business Media.



Lehmann, E e J Romano (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer New York. ISBN: 9780387276052. URL: <https://books.google.com.br/books?id=K6t5qn-SEp8C>.



Schervish, Mark J. (1995). *Theory of Statistics*. Springer New York. DOI: 10.1007/978-1-4612-4250-5. URL: <https://doi.org/10.1007/978-1-4612-4250-5>.