

PROBABILIDADE ESTATÍSTICA  
PRINCÍPIOS DE REDUÇÃO DE DADOS E 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  $(\Omega, \Sigma)$ .
  - $\Omega$  é o espaço amostral.
  - $\Sigma$  é o espaço de eventos aos quais podemos atribuir probabilidades ( $\sigma$ -álgebra).
- Pesquisador observa uma amostra, dada pela variável aleatória  $\mathbf{X} := \text{Id}_{\Omega}$ , cuja lei é dada por uma probabilidade  $P$  sobre  $(\Omega, \Sigma)$  desconhecida.
- O pesquisador postula uma família de leis de probabilidade  $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$  candidatas a terem gerado a amostra.
  - $\Theta$  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:**  $\mathbb{E}_{\theta}$  denota a expectativa de uma variável aleatória com domínio em  $(\Omega, \Sigma)$  com respeito a  $P_{\theta}$ .

# AMOSTRAGEM ALEATÓRIA

- Em diversos contextos, o experimento definido por  $(\Omega, \Sigma, P)$  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 modelagem aleatória, modelo se resume a uma família de distribuições para a característica 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  $\mathbb{R}^s$ ,  $s \in \mathbb{N}$ .
  - Uma amostra aleatória de uma população normal com média e variância desconhecidas define um modelo cujo espaço de parâmetros da forma  $\mathbb{R} \times \mathbb{R}_+$ .
- Um modelo  $\mathcal{P}$  é dito de **dimensão infinita** quando ele não tem dimensão finita.
  - Uma amostra 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  $T : \Omega \mapsto \mathbb{T}$  em que  $T^{-1}(B) \in \Sigma$ , para todo  $B \in \mathcal{T}$ , onde  $\mathcal{T}$  é uma  $\sigma$ -álgebra sobre o espaço  $\mathbb{T}$ .
  - Nesse caso  $T(\mathbf{X})$  define uma variável aleatória com valores em  $\mathbb{T}$ .

# CONDIÇÕES TÉCNICAS

- No que segue, supomo que existe uma medida  $\mu$   $\sigma$ -finita tal que todo elemento  $P_\theta \in \mathcal{P}$  admite densidade  $p_\theta$  com respeito a  $\mu$  (nesse caso, dizemos que a família  $\mathcal{P}$  é *dominada*).
  - Para famílias de distribuições “discretas” sobre o espaço  $(\Omega, 2^\Omega)$  em que  $\Omega$  é enumerável,  $\mu$  é a medida de contagem e  $p_\theta$  são as f.m.p. definidas por  $p_\theta(x) = P_\theta(\{x\})$  para todo  $x \in \Omega$ .
  - Para famílias de distribuições “contínuas” sobre  $(\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s))$ ,  $\mu$  é a medida de Lebesgue sobre  $\mathcal{B}(\mathbb{R}^s)$  e  $p_\theta$  são as f.d.p.
- No que segue, vamos supor também que, para qualquer estatística, as distribuições condicionais de  $\mathbf{X}|T(\mathbf{X})$  sob cada  $P_\theta$  admitem uma **densidade condicional (regular)**  $p_{\mathbf{X}|T(\mathbf{X}),\theta}(\cdot|\cdot)$ .
  - Essa condição limita a complexidade de  $(\Omega, \Sigma)$ .
  - Na maior parte dos casos práticos é satisfeita.
  - Condições suficientes em Durrett (2019, Seção 4.1.3) (fora do escopo do curso).



# 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  $\sigma$ -álgebra do contradomínio.

## DEFINIÇÃO

Uma estatística  $T$  é dita suficiente para  $\theta$  se a distribuição condicional de  $\mathbf{X}|T(\mathbf{X})$  não depende de  $\theta$ , 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  $\theta$  na amostra.
- Sob nossas condições, a definição é equivalente a que a densidade condicional  $p_{\mathbf{X}|T(\mathbf{X}),\theta}$  não dependa de  $\theta$ .

## 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  $t \in \{0, 1/2, 1\}$ .

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

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

# LEMA DA FATORAÇÃO DE NEYMAN-FISHER

## LEMA

*$T(\mathbf{X})$  é suficiente para  $\theta$  se, e somente se, existem funções  $h_\theta$ ,  $\theta \in \Theta$ , 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, “ $\mu$ -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).

## 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  $\theta$ .
- 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

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

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  $\theta$  se sua distribuição não depende de  $\theta$ , 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  $\sigma$ -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  $\theta$ .
- Conceito apropriado para isto é o de estatística **completa suficiente**.

## DEFINIÇÃO

Uma estatística  $T$  é dita completa para  $\theta$  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  $\theta$  é independente de qualquer estatística ancilar de  $\theta$ .

# 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  $(\Omega, \Sigma)$  é dita uma família exponencial se:

1. Existe uma medida  $\mu$  tal que cada elemento  $P_\theta \in \mathcal{P}$  admite densidade  $p_\theta$  com respeito a  $\mu$ .
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.
  - Requer a existência de um mapa 1-1 entre parâmetro e distribuição amostral.
- No exemplo anterior,  $\theta$  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  $\{\mathcal{N}(\theta \vee 0, 1) : \theta \in \mathbb{R}\}$ ,  $\theta$  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  $\eta$  é identificável e  $\Xi$  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.*

# Tópicos adicionais

# MOMENTOS DE $T$

## TEOREMA

*Para uma família exponencial em sua parametrização natural, e  $\eta \in \text{int}(\Xi)$ , temos que, para toda  $h$  tal que*

$$\int_{\Omega} |h(\mathbf{x})| p_{\eta'}(\mathbf{x}) \mu(d\mathbf{x}) < \infty,$$

*para todo  $\eta'$  numa vizinhança de  $\eta$ ; a função:*

$$\psi \mapsto \int_{\Omega} h(\mathbf{x}) p_{\psi}(\mathbf{x}) \mu(d\mathbf{x}),$$

*é diferenciável em  $\eta$ , com derivada dada por diferenciação dentro da integral.*

- Aplicação do resultado: momentos de  $T$ .

## TEOREMA DE RAO-BLACKWELL

- Considere o problema de estimar um parâmetro  $\psi(\theta)$  **escalar** de uma família  $P_\theta$ .
- Seja  $L(\theta; \hat{\delta})$  a perda incorrida em utilizar o estimador  $\hat{\delta}$  para estimar  $\psi(\theta)$ .
- A perda esperada, ou risco, é dada por  $R_\theta(\hat{\delta}) = \mathbb{E}_\theta[L(\theta; \hat{\delta})]$ .

### TEOREMA (RAO-BLACKWELL)

*Suponha que  $z \mapsto L(\theta, z)$  é convexa e não negativa, para todo  $\theta \in \Theta$ . Dado um estimador  $\hat{\delta}$  de  $\psi(\theta)$  e  $T$  uma estatística suficiente integrável de  $\theta$ , temos que:*

1.  $S = \mathbb{E}_\theta[\hat{\delta} | T]$  define um estimador.
2.  $R_\theta(S) \leq R_\theta(\hat{\delta})$  para todo  $\theta \in \Theta$ .
3. Se  $\hat{\delta}$  é não viciado,  $S$  também o é.

# 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>.