

## PREFÁCIO

A origem deste livro é uma promessa que fiz ao Professor Debrabata Basu de tentar algum dia escrever parte de suas idéias sobre Fundamentos Teóricos da Estatística. Tive o privilégio e a honra de fazer parte do pequeno grupo de estudantes que escreveram teses de doutorado (PhD) sob a supervisão desse extraordinário pensador. Minha admiração pelo Professor Basu não se restringe à sua qualidade como orientador de tese, mas, principalmente, pelo magnífico ser humano. Estas notas são dedicadas a esse Amigo, Orientador e Guru.

Este livro é fruto do trabalho de muitas pessoas por muitos anos. É com orgulho que posso afirmar que, embora tendo trabalhado nos últimos treze anos com o material aqui desenvolvido, muitos dos capítulos foram escritos por meus ex-orientandos. Também sinto orgulho de ter tido o privilégio de ter sido orientador dessas pessoas corajosas por se submeter à minha supervisão em seus respectivos trabalhos de tese.

Os Capítulos 1 a 5 são dedicados ao desenvolvimento do que poderíamos classificar como os fundamentos teóricos da Estatística Fisheriana ou Clássica. Ali apresentamos os conceitos fundamentais de modelo estatístico, modelo dominado, estatística e suficiência clássica. O Capítulo 5 introduz os modelos não dominados, que não são muito divulgados na literatura estatística. Aproveitamos a formação matemática do Alberto Zapata para que esse material fosse escrito de forma sucinta e rigorosa.

Pilar escreveu o Capítulo 6 que trata de um modelo não dominado muito especial conhecido por modelo discreto. É exatamente este modelo que permite colocar a teoria da amostragem como parte do método estatístico. Parte desse material pode ser encontrado apenas nas entrelinhas de alguns artigos do próprio Basu ou na tese de doutorado da Pilar.

Márcia escreveu o Capítulo 7 sobre a utilização dos conceitos introduzidos nos capítulos anteriores na estatística usual. Parte do material desse capítulo pode ser encontrado em sua tese de mestrado.

O Capítulo 8 é um excelente estudo da invariância. Muitos ficarão surpresos com a forma clara e precisa com que Victor escreveu esse material. A forma com que invariância é introduzida é original e não creio que exista similar na literatura.

Os Capítulos 9, 10 e 11 referem-se à Estatística Bayesiana e foram escritos por Daniel. O Capítulo 9 introduz o modelo Bayesiano e muito do material pode ser encontrado tanto no meu trabalho de PhD quanto na tese de doutorado do Daniel. Embora o Capítulo 10 chame atenção pelo estilo elegante de apresentação do Daniel, muito dos resultados de independência condicional foram desenvolvidos em minha tese de doutorado. O Capítulo 11, entretanto, é totalmente original e só pode ser encontrado na tese do Daniel. Não conheço publicação que apresente o problema de Identificação com tanta propriedade como feita pelo Daniel. Este material é parte do primeiro capítulo de sua tese.

Finalmente, o Capítulo 12 é uma comparação entre os conceitos de suficiência. Alberto Zapata escreveu este artigo e apresentou em um congresso de Inferência Bayesiana em São Carlos, em 1991. Creio que material similar só pode ser encontrado na tese de PhD de um ex-companheiro da Florida State University, S.C. Cheng.

Embora tenha desenvolvido o material dessas notas ao longo desses treze anos como professor da disciplina MAE 798 - Teoria Estatística, para o Programa de Doutorado do IME-USP, foram meus alunos que me incentivaram a tornar realidade este livro. Assim, cumprindo a promessa que fiz ao meu Guru, pude ter o “direito” de fazer com que todos trabalhassem duro em cima das idéias do Basu.

A pergunta que surge naturalmente é: “qual a diferença entre o que está desenvolvido aqui e o que já foi apresentado em outros textos com o mesmo nível de formalidade?” A resposta é simples: no lugar de trabalharmos com a definição usual de estatística – função mensurável do Espaço Amostral – definimo-la como sendo qualquer subálgebra (sub- $\sigma$ -álgebra) associada ao Espaço Estatístico Original. Assim, ao invés de trabalharmos com diversos Espaços Estatísticos gerados pelas funções mensuráveis, temos de nos preocupar com apenas um espaço e seus sub-espaços de interesse, definidos pelas estatísticas. O preço da generalidade é compensado pela simplicidade.

Os autores dessas notas desejam que os futuros doutorandos possam usufruí-las no sentido de conseguirem simplificações necessárias aos objetivos de suas pesquisas.

Gostaríamos de agradecer a coordenação do Colóquio Brasileiro de Matemática por nos dar a oportunidade de oferecer o Curso de Teoria Estatística nessa prestigiosa reunião científica. Aos meus alunos da turma de 1992, Antônio José da Silva, Filidor Edilson Vilca Labra e Manuel Galea Rojas, gostaria de agradecer pela paciência em rever toda a digitação dessas notas. Pilar e Márcia coordenaram a coleta dos manuscritos e a formação dos exercícios complementares dos capítulos iniciais e, às duas, agradeço pelo incentivo e por todo esse trabalho.

São Paulo, 30 de abril de 1993

Carlos Alberto de Bragança Pereira

# ÍNDICE

<b>Prefácio</b> .....	i
<b>Capítulo 1</b> - Preliminares .....	1
<b>Capítulo 2</b> - Estatísticas e partições .....	14
<b>Capítulo 3</b> - A suficiência de Fisher e suas derivações .....	22
<b>Capítulo 4</b> - Suficiência de Fisher no caso dominado .....	36
4.1 Introdução .....	36
4.2 O teorema de fatoração de Halmos-Savage .....	42
4.3 Suficiência versus suficiência por par .....	46
<b>Capítulo 5</b> - O caso não dominado .....	50
<b>Capítulo 6</b> - Suficiência de Fisher no caso discreto. Aplicações na teoria de amostragem clássica .....	60
6.1 Suficiência no modelo discreto .....	60
6.2 O papel da suficiência na teoria de amostragem clássica .....	72
<b>Capítulo 7</b> - Aplicações de suficiência .....	79
7.1 Teoria de estimação .....	79
7.2 Teoria de testes de hipóteses .....	85
7.3 Os teoremas de Basu .....	88

<b>Capítulo 8</b> - Relações entre invariância, suficiência e ancilaridade .....	98
8.1 Definições e preliminares .....	98
8.2 Invariância e suficiência .....	101
8.3 Invariância e ancilaridade .....	109
<b>Capítulo 9</b> - O modelo estatístico bayesiano e sua estrutura matemática .....	118
9.1 Construção do modelo bayesiano .....	118
9.2 Dominação do modelo bayesiano .....	121
9.3 Medidas condicionadas adicionais .....	126
9.4 Dominação de modelos marginais e condicionais .....	129
<b>Capítulo 10</b> - Independência condicional e suficiência bayesiana amostral .....	138
10.1 Conceito e propriedades da independência condicional .....	138
10.2 Suficiência bayesiana amostral .....	145
10.3 Completude e suficiência mínima bayesianas .....	151
<b>Capítulo 11</b> - Identificabilidade paramétrica .....	157
11.1 Conceitos e resultados básicos no contexto clássico .....	157
11.2 Suficiência bayesiana paramétrica .....	163
11.3 Identificabilidade bayesiana paramétrica .....	172
<b>Capítulo 12</b> - $F$ -suficiência e $K$ -suficiência .....	177
12.1 Introdução .....	177
12.2 Preliminares .....	178
12.3 Relações entre as noções de $F$ -suficiência e $K$ -suficiência .....	183
<b>Referências</b> .....	188

# CAPÍTULO 1

## PRELIMINARES

Entenderemos por um modelo estatístico (ou estrutura estatística) uma trinca de “entes abstratos”  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , onde  $\mathcal{X}$  é um conjunto de pontos,  $x$ ,  $\mathcal{A}$  é uma  $\sigma$ -álgebra de subconjuntos,  $A$ , de  $\mathcal{X}$  e  $\mathcal{P}$  é uma família (de medidas) de probabilidades,  $P$ , no espaço mensurável  $(\mathcal{X}, \mathcal{A})$ . Nestas notas,  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  estará fixada. No contexto estatístico,  $\mathcal{X}$  é o espaço amostral de todos os resultados possíveis de um experimento (variável ou vetor aleatório por exemplo) cuja distribuição é definida por algum  $P_0 \in \mathcal{P}$  “desconhecido” (a intenção do estatístico é pesquisar qual é este  $P_0$ ). A  $\sigma$ -álgebra  $\mathcal{A}$ , logicamente, representa a classe de todos os eventos  $A$ , para os quais  $P(A)$  são bem definidos  $\forall P \in \mathcal{P}$ . Se a família  $\mathcal{P}$  pode ser escrita como

$$\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$$

onde a correspondência  $\theta : \Theta \rightarrow \mathcal{P}$  é um-à-um (preferivelmente uma bem “simples”), então  $\mathcal{P}$  é, equivalentemente, considerado como o “espaço paramétrico”  $\Theta$ .

Consideraremos simultaneamente, com bastante frequência, mais de uma  $\sigma$ -álgebra de subconjuntos de  $\mathcal{X}$  as quais, por sua vez, estão contidas em  $\mathcal{A}$ . Para efeito de simplificação,  $\mathcal{D}$  é chamada de subálgebra de  $\mathcal{A}$  se  $\mathcal{D}$  é uma sub- $\sigma$ -álgebra de  $\mathcal{A}$ . Se  $\{\mathcal{D}_\delta; \delta \in \Delta\}$  é uma coleção arbitrária de subálgebras de  $\mathcal{A}$ ,  $\bigvee_{\delta \in \Delta} \mathcal{D}_\delta = \sigma\{\mathcal{D}_\delta; \delta \in \Delta\}$  é a menor subálgebra de

$\mathcal{A}$  que contém cada  $\mathcal{D}_\delta$ ,  $\delta \in \Delta$ . O conjunto dos reais e a  $\sigma$ -álgebra de Borel na reta serão denotados por  $R_1$  e  $\mathcal{B}_1$ , respectivamente. Por uma variável aleatória  $f$ , entenderemos uma função  $\mathcal{A}$ -mensurável de  $(\mathcal{X}, \mathcal{A})$  em  $(R_1, \mathcal{B}_1)$ . Se uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  contém

$$f^{-1}(\mathcal{B}_1) = \{f^{-1}(B); \forall B \in \mathcal{B}_1\}$$

então  $f$  é dita ser  $\mathcal{D}$ -mensurável e escreveremos  $f \in \mathcal{D}$ . Seja  $\{f_\delta; \delta \in \Delta\}$  uma coleção arbitrária de variáveis aleatórias (v.a.). Escreveremos algumas vezes  $\sigma\{f_\delta; \delta \in \Delta\}$  no lugar de  $\bigvee_{\delta \in \Delta} f_\delta^{-1}(\mathcal{B}_1)$ , a menor subálgebra que contém as classes  $f_\delta^{-1}(\mathcal{B}_1)$ .

### Exercício 1.1

- (i) Dê um exemplo para mostrar que a união de  $\sigma$ -álgebras não é, necessariamente, uma  $\sigma$ -álgebra.
- (ii) Mostre que  $f^{-1}(\mathcal{B}_1)$  é uma subálgebra de  $\mathcal{A}$ .
- (iii) Mostre que a intersecção de  $\sigma$ -álgebras é uma  $\sigma$ -álgebra.

**Observação:** Lembre-se de que estamos considerando, simultaneamente, diversas medidas de probabilidade no mesmo espaço mensurável  $(\mathcal{X}, \mathcal{A})$ . Esta é a diferença básica do espaço de probabilidade usual  $(\mathcal{X}, \mathcal{A}, P)$  onde apenas uma medida de probabilidade é considerada. Na verdade, o estatístico “clássico” admite que existe em  $\mathcal{P}$  o “verdadeiro” elemento  $P$ , o qual deverá ser investigado (estimado ?), pois ele (estatístico) o desconhece. Nesta seção alguns conceitos bem conhecidos em espaços de probabilidade serão estendidos para as estruturas estatísticas. Isto formará o Instrumento Básico de nosso curso. □

Considere então o modelo estatístico básico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

### Definição 1.1

- (i) Um conjunto  $A \in \mathcal{A}$  é dito ser  $P$ -nulo se  $P(A) = 0$ .
- (ii) Um conjunto  $A \in \mathcal{A}$  é dito ser  $\mathcal{P}$ -nulo se  $A$  é  $P$ -nulo  $\forall P \in \mathcal{P}$ .

Sejam  $f$  e  $g$  duas variáveis aleatórias e sejam  $A$  e  $B$  dois conjuntos quaisquer de  $\mathcal{A}$ .

**Definição 1.2**

- (i)  $f = g [P]$  se  $\{x; f(x) \neq g(x)\}$  é  $P$ -nulo.
- (ii)  $A \equiv B [P]$  se  $P(A\Delta B) = 0$ , onde  $A\Delta B$  é a diferença simétrica entre  $A$  e  $B$ .
- (iii)  $f = g [\mathcal{P}]$  se  $f = g [P] \quad \forall P \in \mathcal{P}$ .
- (iv)  $A \equiv B [\mathcal{P}]$  se  $A \equiv B [P] \quad \forall P \in \mathcal{P}$ .

**Exercício 1.2**

Sejam  $f$  e  $g$  duas variáveis aleatórias,  $I_A$  e  $I_B$  as funções indicadores dos conjuntos  $A$  e  $B$  de  $\mathcal{A}$ , respectivamente:

- (i) Mostre que  $\{x; f(x) \neq g(x)\} \in \mathcal{A}$ .
- (ii) Mostre que  $A \equiv B [\mathcal{P}]$  se e somente se  $I_A = I_B [\mathcal{P}]$ .

Consideremos agora duas subálgebras,  $\mathcal{D}$  e  $\mathcal{E}$  de  $\mathcal{A}$ .

**Definição 1.3**

- (i) Diremos que  $\mathcal{D}$  é  $P$ -essencialmente incluída em  $\mathcal{E}$ ,  $\mathcal{D} \subset \mathcal{E} [P]$ , se  $\forall D \in \mathcal{D}$ ,  $\exists E \in \mathcal{E}$  tal que  $E \equiv D [P]$ .
- (ii)  $\mathcal{D}$  é  $P$ -equivalente a  $\mathcal{E}$ ,  $\mathcal{D} \equiv \mathcal{E} [P]$ , se  $\mathcal{D} \subset \mathcal{E} [P]$  e  $\mathcal{E} \subset \mathcal{D} [P]$ .

**Definição 1.4**

- (i)  $\mathcal{D}$  é  $\mathcal{P}$ -essencialmente incluída em  $\mathcal{E}$ ,  $\mathcal{D} \subset \mathcal{E} [\mathcal{P}]$ , se  $\forall D \in \mathcal{D}$ ,  $\exists E \in \mathcal{E}$  tal que  $E \equiv D [\mathcal{P}]$ .
- (ii)  $\mathcal{D}$  é  $\mathcal{P}$ -equivalente à  $\mathcal{E}$ ,  $\mathcal{D} \equiv \mathcal{E} [\mathcal{P}]$ , se  $\mathcal{E} \subset \mathcal{D} [\mathcal{P}]$  e  $\mathcal{D} \subset \mathcal{E} [\mathcal{P}]$ .



### Exercício 1.3

Mostre que  $\mathcal{D} \subset \mathcal{E} [\mathcal{P}]$  se e somente se para cada v.a.  $f$  que é  $\mathcal{D}$ -mensurável existe uma v.a.  $g$  que é  $\mathcal{E}$ -mensurável tal que  $g = f$   $[\mathcal{P}]$ .

**Observação:** Ocasionalmente escreveremos “ $\subset_{\mathcal{P}}$ ” e “ $\equiv_{\mathcal{P}}$ ” para indicar inclusão  $\mathcal{P}$ -essencial e para indicar  $\mathcal{P}$ -equivalência, respectivamente.

É “evidente” que  $\mathcal{D} \equiv \mathcal{E} [\mathcal{P}]$  implica que  $\mathcal{D} \equiv \mathcal{E} [P]$ ,  $\forall P \in \mathcal{P}$ . O inverso contudo não é verdadeiro visto que, no primeiro caso, a equivalência é independente da escolha de  $P$ , o que não ocorre no segundo.

### Exemplo 1.1

Seja  $\mathcal{X} = R_1$ ,  $\mathcal{A}$  = conjunto de todas as partes de  $\mathcal{X}$  e  $\mathcal{P}$  a família de todas as distribuições degeneradas (em  $R_1$ ). Note que o conjunto vazio,  $\emptyset$ , é o único conjunto  $\mathcal{P}$ -nulo. Não é difícil ver que  $\mathcal{A} \equiv \{\emptyset, R_1\} [P_\theta]$  para todo  $\theta \in R_1$ . Contudo, todo subconjunto próprio  $A$  de  $R_1$  não possui  $\mathcal{P}$ -equivalência, nem com  $\emptyset$  nem com  $R_1$ . Assim,  $\mathcal{A} \not\equiv \{\emptyset, R_1\} [\mathcal{P}]$ ; isto é,  $\mathcal{A}$  não é  $\mathcal{P}$ -equivalente a  $\sigma$ -álgebra trivial.  $\square$

### Exercício 1.4

No Exemplo 1, mostre que  $\mathcal{A} \equiv \{\emptyset, R_1\} [P_\theta] \forall \theta \in R_1$  e que todo subconjunto próprio de  $R_1$  não é  $\mathcal{P}$ -equivalente a  $\emptyset$  ou a  $R_1$ .

**Observação:** Por questões práticas, é natural que o aluno pergunte se não seria possível conseguir um exemplo semelhante ao Exemplo 1, onde tivéssemos uma família  $\mathcal{P}$  de distribuições absolutamente contínuas em  $R_1$ . Voltaremos a discutir esse problema quando estudarmos famílias dominadas.  $\square$

No caso de duas medidas,  $P_1$  e  $P_2$ , escreveremos  $\mathcal{D} \subset \mathcal{E} [P_1, P_2]$  para indicar que  $\forall D \in \mathcal{D}, \exists E \in \mathcal{E}$  tal que  $E \equiv D [P_1]$  e  $E \equiv D [P_2]$ ; isto é,  $E \equiv D [P_1, P_2]$ . Analogamente, com esta ordem parcial, definimos  $\mathcal{D} \equiv \mathcal{E} [P_1, P_2]$ .

### Definição 1.5

Duas subálgebras  $\mathcal{D}$  e  $\mathcal{E}$  são “equivalentes por par” , com respeito a  $\mathcal{P}$ , se para cada subclasse de duas medidas  $\{P_1, P_2\} \subset \mathcal{P}$ ,  $\mathcal{D} \equiv \mathcal{E} [P_1, P_2]$ .

Também aqui, fica claro que  $\mathcal{P}$ -equivalência implica em equivalência por par. Contudo, o inverso é falso como mostra o seguinte exemplo.

### Exemplo 1.2

No Exemplo 1, considere  $\mathcal{P}$  como sendo a família de todas as distribuições discretas em  $R_1$ . Considerando o conjunto vazio,  $\emptyset$ , como sendo enumerável, seja a  $\mathcal{D}$  a subálgebra (ver Exercício 1.5) definida por

$$\mathcal{D} = \{D \in \mathcal{A}; D \text{ ou } D^c \text{ é enumerável}\},$$

onde  $D^c$  indica o complementar de  $D$  ( $\mathcal{X} - D$ ). Vamos provar que  $\mathcal{A}$  e  $\mathcal{D}$  são equivalentes por par. Para isto, fixemos quaisquer subconjuntos  $A \in \mathcal{A}$  e  $\{P_1, P_2\} \subset \mathcal{P}$ . Desde que  $P_1$  e  $P_2$  são medidas discretas de probabilidade, concentram suas massas em um conjunto enumerável  $A_0 = \{a_1, a_2, \dots\}$ . Considere o conjunto  $D = A_0 \cap A$ . Fica então claro que  $D \in \mathcal{D}$  e que  $D \equiv A [P_1, P_2]$ . Assim,  $\forall \{P_1, P_2\} \subset \mathcal{P}$ , temos que  $\mathcal{A} \subset \mathcal{D} [P_1, P_2]$ . Por definição temos que  $\mathcal{D} \subset \mathcal{A}$  e, assim, concluímos que  $\mathcal{A} \equiv \mathcal{D} [P_1, P_2]$  qualquer que seja  $\{P_1, P_2\} \subset \mathcal{P}$ . Contudo,  $\mathcal{A} \not\equiv \mathcal{D} [\mathcal{P}]$  (ver Exercício 1.6).  $\square$

### Exercício 1.5

Mostre que a classe  $\mathcal{D}$ , definida no Exemplo 2, é uma  $\sigma$ -álgebra.

### Exercício 1.6

No Exemplo 1.2, mostre que  $\mathcal{A} \subset \mathcal{B} [\mathcal{P}]$  é falso.

**Observação:** Note que no Exemplo 1.1,  $\mathcal{A}$  e  $\{\emptyset, R_1\}$  não são equivalentes por par. Consequentemente, temos a seguinte conclusão:  $\mathcal{P}$ -equivalência  $\Rightarrow$  Equivalência por par  $\Rightarrow$   $\mathcal{P}$ -equivalência  $\forall P \in \mathcal{P}$ . Contudo, as direções das implicações não podem ser mudadas em geral. É importante ainda notar que, também no Exemplo 1.2, consideramos uma classe “não dominada” de medidas.

### Definição 1.6

Uma v.a.  $f$  é dita ser “ $\mathcal{P}$ -essencialmente  $\mathcal{D}$ -mensurável” se existe uma função,  $g$ ,  $\mathcal{D}$ -mensurável tal que  $g = f [P]$ . Neste caso, escrevemos  $f \in \mathcal{D} [P]$ .

Consideremos agora as seguintes classes de subconjuntos de  $\mathcal{X}$ :

- (i)  $\mathcal{A}_0 = \{\emptyset, \mathcal{X}\}$  ;
- (ii)  $\overline{\mathcal{A}}_0^P = \{A \in \mathcal{A}; P(A) = 0 \text{ ou } P(A^c) = 0\}$  ;
- (iii)  $\overline{\mathcal{A}}_0 = \{A \in \mathcal{A}; P(A) = 0 \text{ ou } P(A^c) = 0 \quad \forall P \in \mathcal{P}\}$  .

**Exercício 1.7**

Mostre que  $\mathcal{A}_0$ ,  $\overline{\mathcal{A}}_0^P$  e  $\overline{\mathcal{A}}_0$  são  $\sigma$ -álgebras e que  $\mathcal{A}_0 \subset \overline{\mathcal{A}}_0 \subset \overline{\mathcal{A}}_0^P$ .

**Definição 1.7**

O complemento de uma subálgebra  $\mathcal{D}$  é a subálgebra  $\overline{\mathcal{D}} = \mathcal{D} \vee \overline{\mathcal{A}}_0$ . Da mesma forma,  $\mathcal{D}$  é dita ser completada se  $\overline{\mathcal{A}}_0 \subset \mathcal{D}$ .

**Exercício 1.8**

Sejam  $\mathcal{D}$  e  $\mathcal{E}$  duas subálgebras e  $f$  uma v.a. Então, prove que:

- (i)  $\overline{\mathcal{D}} = \{A \in \mathcal{A}; A \equiv D [\mathcal{P}] \text{ para algum } D \in \mathcal{D}\}$  ;
- (ii)  $\mathcal{E} \equiv \mathcal{D} [\mathcal{P}] \Leftrightarrow \overline{\mathcal{E}} \equiv \overline{\mathcal{D}}$  ;
- (iii)  $\mathcal{D} \equiv \overline{\mathcal{D}} [\mathcal{P}]$ .
- (iv)  $\mathcal{E} \subset \mathcal{D} [\mathcal{P}] \Leftrightarrow \mathcal{E} \subset \overline{\mathcal{D}}$  ;
- (v)  $f \in \mathcal{D} [\mathcal{P}] \Leftrightarrow f \in \overline{\mathcal{D}}$  .

É natural, no momento, questionarmos se o complemento é mantido quando efetuamos as “operações” entre subálgebras. Os resultados que passamos a discutir, embora sem muito interesse prático, nos fornecerá uma maior intimidade com os conceitos aqui apresentados.

**Proposição 1.1**

Sejam  $\mathcal{D}_1$ ,  $\mathcal{D}_2$ ,  $\mathcal{E}_1$  e  $\mathcal{E}_2$  subálgebras de  $\mathcal{A}$ . Se  $\mathcal{D}_1 \equiv \mathcal{E}_1 [\mathcal{P}]$  e  $\mathcal{D}_2 \equiv \mathcal{E}_2 [\mathcal{P}]$ , então  $\mathcal{D}_1 \vee \mathcal{D}_2 \equiv \mathcal{E}_1 \vee \mathcal{E}_2 [\mathcal{P}]$ .

### Demonstração

Pelo Exercício 1.8 (ii) é suficiente mostrarmos que

$$\overline{\mathcal{D}_1 \vee \mathcal{D}_2} \equiv \overline{\mathcal{E}_1 \vee \mathcal{E}_2} \quad (*)$$

Em vista da definição de completamento e pelo fato da comutatividade da operação  $\vee$ , temos que

$$\overline{\mathcal{D}_1 \vee \mathcal{D}_2} = \overline{\mathcal{D}_1 \vee \mathcal{D}_2}, \quad \overline{\mathcal{E}_1 \vee \mathcal{E}_2} = \overline{\mathcal{E}_1 \vee \mathcal{E}_2}.$$

Como  $\mathcal{D}_i \equiv \mathcal{E}_i[\mathcal{P}]$ ,  $i = 1, 2$ , temos pelo Exercício 1.8 (ii) que  $\overline{\mathcal{D}_i} \equiv \overline{\mathcal{E}_i}$ ,  $i = 1, 2$ . Assim, (\*) é consequência imediata deste fato.  $\square$

O exemplo a seguir vem mostrar que embora a intersecção de subálgebras seja uma subálgebra, o completamento dessa intersecção não necessariamente é a intersecção dos completamentos.

### Exemplo 1.3

Seja  $\mathcal{X} = R_1$ ,  $\mathcal{A} = \mathcal{B}_1$  e  $\mathcal{P}$  = família de medidas  $P$  que são dominadas pela medida de Lebesgue,  $\lambda$ ; isto é,  $P \ll \lambda \forall P \in \mathcal{P}$ . [ $P \ll \lambda$  se  $\forall B \in \mathcal{B}_1$ , tal que  $\lambda(B) = 0$  implicar  $P(B) = 0$ .] Seja  $\mathcal{D} = \{\emptyset, (-\infty, 0), [0, \infty), R_1\}$  e  $\mathcal{E} = \{\emptyset, E, E^c, R_1\}$ , onde  $E$  é o conjunto dos irracionais negativos. Como  $\lambda(E \Delta (-\infty, 0)) = 0$ , temos que  $\mathcal{D} \equiv \mathcal{E} [\mathcal{P}]$ . Então, se  $\mathcal{P}$ -equivalência fosse mantida com a formação de intersecções, teríamos

$$\begin{cases} \mathcal{D} \equiv \mathcal{E} [\mathcal{P}] \\ \mathcal{D} \equiv \mathcal{D} [\mathcal{P}] \end{cases} \Rightarrow \begin{cases} \mathcal{D} \equiv \mathcal{E} \cap \mathcal{D} [\mathcal{P}]; \text{ isto é,} \\ \mathcal{D} \equiv \{\emptyset, R_1\} [\mathcal{P}] \end{cases}$$

o que logicamente é falso.

Por outro lado,  $E \in \overline{\mathcal{D} \cap \mathcal{E}}$ , porém,  $E \notin \overline{\mathcal{D}} \cap \overline{\mathcal{E}} = \overline{\mathcal{A}_0}$ . Isto mostra que o completamento da intersecção não é necessariamente a intersecção dos completamentos.  $\square$

O resultado abaixo vem concluir nossa investigação.

### Proposição 1.2

Sejam  $\mathcal{D}$  e  $\mathcal{E}$  duas quaisquer subálgebras, então  $\overline{\mathcal{D} \cap \mathcal{E}} \subset \overline{\mathcal{D}} \cap \overline{\mathcal{E}}$ .

### Demonstração

Seja  $B \in \overline{\mathcal{D} \cap \mathcal{E}}$ , assim pelo Exercício 1.8 (i),  $\exists A \in \mathcal{D} \cap \mathcal{E}$  tal que  $A \equiv B$  [ $\mathcal{P}$ ]. Assim, desde que  $A \in \mathcal{D}$  e  $A \in \mathcal{E}$ , concluímos que  $B \in \overline{\mathcal{D}}$  e  $B \in \overline{\mathcal{E}}$ , o que finaliza esta demonstração.  $\square$

Com o intuito de diminuir a generalidade de nossa estrutura estatística, vamos considerar uma propriedade topológica que, além de tornar um pouco mais “real” o nosso estudo, vem fornecer características bastante interessantes.

### Definição 1.8

Uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  é dita ser “enumeravelmente gerada” ou “separável” se existe uma classe enumerável,  $\mathbb{D} = \{D_1, D_2, \dots\}$ , de subconjuntos  $D_i$ , de  $\mathcal{D}$  tal que  $\sigma(\mathbb{D}) = \mathcal{D}$  (isto é,  $\mathcal{D}$  é a menor  $\sigma$ -álgebra que contém  $\mathbb{D}$ ). Neste caso,  $\mathbb{D}$  é chamado “gerador” ou “base” de  $\mathcal{D}$ .

### Exemplo 1.4

$\mathcal{D} = \{\emptyset, D, D^c, \mathcal{X}\}$  e  $\mathcal{B}_1$  são separáveis desde que  $\mathcal{D} = \sigma(\{D\})$  e  $\mathcal{B}_1 = \sigma(\{(-\infty, r); r \text{ é racional}\})$ .

### Teorema 1.1

Seja  $\mathcal{D}$  uma subálgebra separável de  $\mathcal{A}$ , então existe uma v.a.  $f$  tal que

$$f^{-1}(\mathcal{B}_1) = \mathcal{D} .$$

### Demonstração

Sejam  $\mathbb{D} = \{D_1, D_2, \dots\}$  tal que  $\mathcal{D} = \sigma(\mathbb{D})$ ,  $I_n = I_{D_n}$  ( $n = 1, 2, \dots$ ) e

$$f = \left\{ \sum_{n=1}^{\infty} \frac{I_n}{3^n} \right\} .$$

Como  $D_n \in \mathcal{D}$ , então  $f \in \mathcal{D}$ ; isto é  $f^{-1}(\mathcal{B}_1) \subset \mathcal{D}$ . Consideremos agora os conjuntos

$$B_i = \left( \sum_{n=1}^{\infty} \left( \frac{a_n}{3^n} \right) : a_i = 1 , \quad a_n = 0 \text{ ou } 1 , \quad \forall n \neq i \right) ,$$

e note que, se  $\forall n, a_n = 0$  ou  $1$ , existe uma correspondência um-a-um entre as seqüências  $(a_1, a_2, \dots)$  e as séries  $\sum_{n=1}^{\infty} \frac{a_n}{3^n}$ .

Como  $D_i = (x : I_i(x) = 1) = f^{-1}(B_i)$  e  $f^{-1}(B_i) \in f^{-1}(\mathcal{B}_1)$ , segue que  $D_i \in f^{-1}(\mathcal{B}_1)$ . Logo  $\mathcal{D} \subset f^{-1}(\mathcal{B}_1) \subset \mathcal{D}$  e portanto  $f^{-1}(\mathcal{B}_1) \equiv \mathcal{D}$ .  $\square$

### Exercício 1.9

Procure esclarecer todos os detalhes da demonstração acima.

### Teorema 1.2

Se  $f$  é uma v.a., então  $f^{-1}(\mathcal{B}_1)$  é separável.

### Demonstração

Observe que  $f^{-1}(\mathcal{B}_1) \subset \mathcal{A}$  e que  $\mathcal{B}_1$  é separável. Seja  $\mathbb{B} = \{B_1, B_2, \dots\}$  uma base de  $\mathcal{B}_1$ . Seja  $\mathcal{D} = \sigma(\{f^{-1}(B_1), f^{-1}(B_2), \dots\})$  que por outro lado está contido em  $f^{-1}(\mathcal{B}_1)$ . Seja  $\mathcal{B}_0 = \{B \in \mathcal{B}_1 : f^{-1}(B) \in \mathcal{D}\}$ , verifique que  $\mathcal{B}_0$  é uma  $\sigma$ -álgebra contendo  $\{B_1, B_2, \dots\}$  e logo  $\mathcal{B}_0 = \mathcal{B}_1$ . Escolha então  $\forall B \in \mathcal{B}_1 = \mathcal{B}_0$  e logo  $f^{-1}(B) \in \mathcal{D}$ . Assim  $f^{-1}(\mathcal{B}_1) \subset \mathcal{D}$ .  $\square$

A seguinte definição possui uma versão geral que depende das medidas de probabilidade do modelo estatístico.

### Definição 1.3

Um conjunto não vazio  $A \in \mathcal{A}$  é chamado de “átomo” de  $\mathcal{A}$  se  $\forall A_0 \subset A$  tal que  $A_0 \in \mathcal{A}$ , então  $A_0 = \emptyset$  ou  $A_0 = A$ .  $\mathcal{A}$  é dita ter uma “estrutura atômica” se seus átomos formam uma partição de  $\mathcal{X}$ .

### Exemplo 1.5

- (i)  $\mathcal{X}$  é o único átomo de  $\mathcal{A}_0$ .
- (ii)  $A$  e  $A^c$  são os átomos de  $\{\emptyset, A, A^c, \mathcal{X}\}$ .
- (iii) Para cada  $x \in R_1$ ,  $\{x\}$  é um átomo de  $\mathcal{B}_1$ . Como  $R_1 = \bigcup_{x \in R_1} \{x\}$ , então  $\mathcal{B}_1$  tem uma estrutura atômica.
- (iv) Seja  $f$  uma v.a., então  $f^{-1}\{x\}, \forall x \in R_1$ , é um átomo de  $f^{-1}(\mathcal{B}_1)$ .

### Teorema 1.3

Se  $\mathcal{A}$  é separável, então  $\mathcal{A}$  tem uma estrutura atômica.

### Demonstração

Seja  $\mathbf{A} = \{A_1, A_2, \dots\}$  uma base de  $\mathcal{A}$  e seja  $\mathbb{B} = \{B : B = \bigcap_{n=1}^{\infty} B_n, \text{ onde } B_n = A_n \text{ ou } A_n^c, n = 1, 2, \dots\}$ . Mostre que os elementos não vazios de  $\mathbb{B}$  formam uma partição de  $\mathcal{X}$ . Seja  $\mathcal{D}$  a classe de todos os conjuntos  $D \in \mathcal{A}$  tais que  $D = \bigcup_{\delta \in \Delta} B_\delta$ , onde  $\Delta$  é um conjunto qualquer de índices e  $B_\delta \in \mathbb{B} \ \forall \delta \in \Delta$ . Verifique que  $\mathcal{D}$  é uma subálgebra de  $\mathcal{A}$  e note que  $\mathcal{D}$  tem uma estrutura atômica, pois cada elemento não vazio de  $\mathbb{B}$  é um átomo de  $\mathcal{D}$ .

Para finalizar essa demonstração, basta provar que  $\mathcal{A} \subset \mathcal{D}$ , pois, logicamente,  $\mathcal{D} \subset \mathcal{A}$ . Assim, observamos que  $A_n \in \mathcal{D}, \forall n = 1, 2, \dots$ , pois  $A_n = \bigcup_{C_n \in \mathbb{B}} C_n$ , onde  $C_n = (\bigcap_{i=1}^{n-1} B_i) \cap A_n \cap (\bigcap_{i=n+1}^{\infty} B_i)$  e  $B_i = A_i$  ou  $A_i^c \ \forall i \neq n$ . Logo,  $\mathbb{B} \subset \mathcal{D}$  e assim  $\mathcal{A} \subset \mathcal{D}$ .  $\square$

### Exercício 1.10

Mostre que a classe  $\mathcal{D}$  definida na demonstração acima é uma  $\sigma$ -álgebra.

Evidentemente, existem  $\sigma$ -álgebras que não são separáveis. O mais surpreendente é que existam  $\sigma$ -álgebras separáveis que contêm subálgebras não separáveis. O exemplo abaixo mostra este fenômeno.

### Exemplo 1.6

Seja  $\mathcal{D} = \{D \in \mathbb{R}_1; D \text{ ou } D^c \text{ é enumerável}\}$ . Vamos supor por absurdo que  $\mathcal{D}$  é separável. Existem  $D_1, D_2, \dots$  tais que  $\mathcal{D} = \sigma(\{D_1, D_2, \dots\})$  e assim  $\mathcal{D}$  tem uma estrutura atômica. Isto é, os elementos não vazios de  $\mathbb{B} = \{B_1 \cap B_2 \cap \dots \cap B_n = D_n \text{ ou } D_n^c\}$  formam uma partição de  $\mathbb{R}_1$  e cada um de seus elementos é um átomo de  $\mathcal{D}$ .

Sem perda de generalidade, podemos admitir que cada  $D_n, n = 1, 2, \dots$ , seja enumerável. Por outro lado,  $\bigcap_{n=1}^{\infty} D_n^c \neq \emptyset$ , caso contrário  $\bigcup_{n=1}^{\infty} D_n = \mathbb{R}_1$  o que é uma contradição, pois  $\mathbb{R}_1$  seria enumerável. Assim,  $\bigcap_{n=1}^{\infty} D_n^c$  é um átomo de  $\mathcal{D}$  provocando a seguinte contradição:

Como  $\bigcup_{n=1}^{\infty} D_n \neq \mathbb{R}_1$ , então consideremos  $D \in \mathcal{D}$  contendo propriamente a  $\bigcup_{n=1}^{\infty} D_n$  assim  $D^c$  é um subconjunto próprio de  $\bigcap_{n=1}^{\infty} D_n^c$  o que obviamente é uma contradição.  $\square$

Note, no exemplo acima, que  $\mathcal{D}$  não é separável, embora  $\mathcal{D} \subset \mathcal{B}_1$  seja separável.

### Exercício 1.11

No exemplo acima, mostre que não há perda de generalidade em supor que  $D_1, D_2, \dots$

são enumeráveis.

## EXERCÍCIOS COMPLEMENTARES

1. Considere o espaço estatístico onde  $\mathcal{X} = R_1$ ,  $\mathcal{A}$  é a classe dos borelianos de  $\mathcal{X}$  e  $\mathcal{P} = \{P_\theta; \theta \in R_1\}$  com  $P_\theta$  sendo a medida degenerada em  $\theta$ . Sendo  $\mathcal{A}_0$  a subálgebra trivial;
  - (i) prove que  $\mathcal{A} = \mathcal{A}_0 [P_\theta], \forall \theta \in R_1$  e que  $\mathcal{A} \neq \mathcal{A}_0 [P]$  e
  - (ii) diga, justificando, se a relação  $\mathcal{A} = \mathcal{A}_0 [P_{\theta_1}, P_{\theta_2}], \forall (\theta_1, \theta_2) \in R_2$ , é válida.
2. Considere o espaço estatístico onde  $\mathcal{X} = R_1$ ,  $\mathcal{A}$  é a classe dos borelianos de  $\mathcal{X}$  e  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ . Determine a classe de conjuntos  $\mathcal{P}$ -nulos quando
  - (i)  $\mathcal{P}$  é a classe de distribuições uniformes em  $(\theta, \theta + 1)$  e  $\Theta = R_1$ ;
  - (ii)  $\mathcal{P}$  é a classe de distribuições uniformes em  $(\theta_1, \theta_2)$  e  $\theta_1 < \theta_2, \theta_1, \theta_2 \in R_1$ .
3. Seja  $(\mathcal{X}, \mathcal{A}, P)$  um espaço de probabilidade e  $G$  o grupo de funções bijetoras mensuráveis de  $\mathcal{X}$  em  $\mathcal{X}$ .
  - (i) Defina a medida de probabilidade  $gP$  sobre  $(\mathcal{X}, \mathcal{A})$  por

$$(gP)(A) = P(g^{-1}(A)) \quad : \quad g \in G, A \in \mathcal{A}$$

e  $\mathcal{P} = \{(gP) : g \in G\}$ .

O grupo  $G$  pode ser considerado como espaço paramétrico? Justifique.

- (ii) Seja  $\mathcal{P}$  um conjunto de medidas de probabilidade definidas sobre  $(\mathcal{X}, \mathcal{A})$  tal que para cada  $P \in \mathcal{P}$  e  $g \in G$ ,  $gP \in \mathcal{P}$ .

Seja  $\mathcal{X}^{(n)} = \mathcal{X} \times \dots \times \mathcal{X}$  e  $\mathcal{A}^{(n)}$  a  $\sigma$ -álgebra produto sobre  $\mathcal{X}^{(n)}$ . Para  $P \in \mathcal{P}$  defina  $P^{(n)}$  sobre  $\mathcal{B}^{(n)}$  como

$$P^{(n)}(A_1 \times A_2 \times \dots \times A_n) = \prod_{i=1}^n P(B_i), \quad B_i \in \mathcal{B}.$$

Além disso, se  $g \in G$ ,  $g$  opera sobre  $\mathcal{X}^{(n)}$  por

$$g(x_1, \dots, x_n) = (gx_1, \dots, gx_n), \quad (x_1, \dots, x_n) \in \mathcal{X}^{(n)}.$$



Se  $\mathcal{P}^{(n)} = \{P^{(n)}/P \in \mathcal{P}\}$ , mostre que  $gP^{(n)} \in \mathcal{P}^{(n)}$  para cada  $P^{(n)} \in \mathcal{P}^{(n)}$ ,  $g \in G$ .

4. No espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , seja  $\mathcal{M}$  a classe dos conjuntos  $\mathcal{P}$ -nulos de  $\mathcal{A}$  e  $\mathcal{D}$  uma subálgebra com complemento  $\overline{\mathcal{D}}$ .

(i) Seja  $\mathcal{N}$  a  $\sigma$ -álgebra gerada por  $\mathcal{M}$ . Mostre que  $\mathcal{N} = \overline{\mathcal{A}_0}$ , onde  $\mathcal{A}_0$  é a  $\sigma$ -álgebra trivial.

(ii) [ Seja  $f : \mathcal{X} \rightarrow R_1$  função limitada, Borel-mensurável. Diga, justificando, se as seguintes afirmações são válidas.

a)  $E_p(f|\mathcal{D}) = E_p(f|\overline{\mathcal{D}})$  q.c.  $P$ .

b)  $E_p(f|\overline{\mathcal{A}_0}) = E_p(f)$  q.c.  $P$ .

5. Considere o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  com  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ . Seja  $\mu$  uma medida de probabilidade sobre o espaço mensurável  $(\Theta, \mathcal{F})$  e defina

$$Q(A) = \int_{\Theta} P_\theta(A) d\mu(\theta), \quad A \in \mathcal{A}.$$

(i) Mostre que  $Q$  é uma medida de probabilidade sobre  $(\mathcal{X}, \mathcal{A})$ .

(ii) Prove que se  $\mathcal{P} \ll \lambda$ , onde  $\lambda$  é uma medida de probabilidade sobre  $(\mathcal{X}, \mathcal{A})$ , então  $Q \ll \lambda$ . O recíproco é verdadeiro? Justifique.

6. Sejam  $P$  e  $Q$  duas medidas de probabilidade definidas sobre o espaço mensurável  $(\mathcal{X}, \mathcal{A})$ . Defina

$$\|P - Q\|_{\mathcal{A}} = 2 \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$$

(i) Se  $\mathcal{C}$  é uma subálgebra de  $\mathcal{A}$ , mostre que  $\|P - Q\|_{\mathcal{C}} \leq \|P - Q\|_{\mathcal{A}}$ . Verifique se  $\|P - Q\|_{\mathcal{C}} = \|P - Q\|_{\overline{\mathcal{C}}}$ .

(ii) Usando o fato que  $\|P - Q\|_{\mathcal{A}} = 2 \sup_{\substack{0 \leq \phi \leq 1 \\ \phi \in \mathcal{A}}} |\int \phi dP - \int \phi dQ|$ , prove que se  $\mathcal{C}$  é uma subálgebra de  $\mathcal{A}$  tal que para cada  $A \in \mathcal{A}$

$$P(A|\mathcal{C}) = Q(A|\mathcal{C}) \quad \text{q.c. } P \text{ e q.c. } Q$$

então  $\|P - Q\|_{\mathcal{C}} = \|P - Q\|_{\mathcal{A}}$ .

7. Uma urna contém  $N_1$  bolas marcadas com o número 1,  $N_2$  com o número 2 e  $N_3$  com o número 3. Embora  $N_1, N_2$  e  $N_3$  sejam desconhecidos, o total de bolas,  $N = N_1 + N_2 + N_3$ , é conhecido. Suponha que uma amostra de tamanho  $n$  ( $n < N$ ) é selecionada, a) com reposição e b) sem reposição, e anota-se o valor do vetor amostral  $(n_1, n_2)$ , onde  $n_1, n_2$  e  $n_3$  são as frequências amostrais de bolas 1, 2, e 3, respectivamente.

Descreva o espaço estatístico induzido pelos vetores amostrais nos casos a) e b). Esses espaços são dominados? Se afirmativo, encontre derivadas de Radon-Nikodym nos casos a) e b).

8. Sejam  $X$  e  $Y$  duas observações independentes de uma normal com média  $\mu \in R_1$  e desvio padrão  $\sigma \in R_1^+$ . Mostre que os espaços estatísticos induzidos por  $(S, D) = (X + Y, X - Y)$  coincide com o produto dos espaços estatísticos induzidos por  $S$  e  $D$ .

9. Um lote contém  $N$  itens e  $s$  defeitos no lote (cada item pode ter de zero até  $s$  defeitos). Seja  $y_i$  o número de defeitos no  $i$ -ésimo item  $\sum_{i=1}^N y_i = s$ . Suponha que as possíveis configurações  $(y_1, \dots, y_N)$  são igualmente prováveis.

a) Descreva o espaço estatístico induzido pelos vetores amostrais  $(y_1, \dots, y_N)$  quando i) os itens e os defeitos são distinguíveis, ii) os defeitos são indistinguíveis.

b) No caso i) em a), suponha que uma amostra de tamanho  $n$  ( $n < N$ ) é selecionada. Seja  $x$  o número de defeitos na amostra. Descreva o modelo estatístico induzido por  $x$ . O que ocorre com este modelo se  $N \mapsto \infty$  e  $s \mapsto \infty$  em tal forma que  $\frac{s}{N} = \lambda$  e  $\frac{n}{N} \mapsto 0$ ?

c) No caso ii) em a), suponha que um item é selecionado. Seja  $y$  o número de defeitos neste item. Descreva o modelo estatístico induzido por  $y$ . O que ocorre com este modelo se  $N \mapsto \infty$  e  $s \mapsto \infty$  em tal forma que  $\frac{s}{N} = \lambda$ ?

d) Os espaços obtidos em a) - c) são dominados? Se afirmativo, encontre as derivadas de Radon-Nikodym.

## CAPÍTULO 2

### ESTATÍSTICAS E PARTIÇÕES

Considere o modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  e considere uma função  $\mathcal{A}$ -mensurável,  $Y$ , de  $(\mathcal{X}, \mathcal{A})$  em um outro espaço mensurável  $(\mathcal{Y}, \mathcal{B})$ . Isto é,

$$\mathcal{A}_Y = Y^{-1}(\mathcal{B}) = \{Y^{-1}(B) : B \in \mathcal{B}\} \subset \mathcal{A} .$$

Qualquer estatística  $Y$  pode ser identificada com a subálgebra  $\mathcal{A}_Y$  gerada por ela (prove, novamente, que  $\mathcal{A}_Y$  é uma  $\sigma$ -álgebra). Correspondendo a cada medida  $P \in \mathcal{P}$ , definimos uma medida  $Q$  em  $(\mathcal{Y}, \mathcal{B})$  da seguinte forma,

$$Q(B) = P(Y^{-1}(B)) \quad \forall B \in \mathcal{B} .$$

Seja  $\mathcal{Q}$  a família de tais medidas induzidas pelos elementos de  $\mathcal{P}$ ; isto é,  $\mathcal{Q}$  é a família induzida por  $\mathcal{P}$ ,  $\mathcal{Q} = \{PY^{-1} : P \in \mathcal{P}\}$ . Assim,  $Y$  está induzindo um novo modelo estatístico  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  e na verdade, o modelo  $(\mathcal{X}, \mathcal{A}_Y, \mathcal{P})$  é o verdadeiro indutor. A seguir, discutiremos o sentido em que consideraremos os modelos  $(\mathcal{X}, \mathcal{A}_Y, \mathcal{P})$  e  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  como equivalentes para o estatístico.

Suponha por um momento que  $Y$  seja uma função sobrejetora; isto é,  $\mathcal{Y} = Y(\mathcal{X})$ . Assim, deve existir uma correspondência um a um entre os elementos de  $\mathcal{A}_Y$  e  $\mathcal{B}$ . Para concordarmos com isto, necessitamos mostrar que  $Y^{-1}(B_1) \neq Y^{-1}(B_2)$  se  $B_1 \neq B_2$ , onde

$B_1, B_2 \in \mathcal{B}$ , isto equivale a mostrar que  $Y^{-1}(B_1) = Y^{-1}(B_2)$  implica em  $B_1 = B_2$ . Lembremos que

$$Y^{-1}(B_1 \Delta B_2) = Y^{-1}(B_1) \Delta Y^{-1}(B_2) = \emptyset .$$

Assim, como  $Y$  é sobrejetora, então  $B_1 = B_2$ .

Para generalizar este resultado, apresentamos a proposição seguinte que elimina a suposição de  $Y$  ser sobrejetora.

### Proposição 2.1

Seja  $\mathcal{Q}$  a classe induzida por  $\mathcal{P}$  através de  $Y$ . Assim, a relação  $A = Y^{-1}(B)$ , onde  $A \in \mathcal{A}_Y$  e  $B \in \mathcal{B}$ , estabelece uma correspondência um-a-um entre as  $\sigma$ -álgebras  $\mathcal{A}_Y$  e  $\mathcal{B}$ , a menos de uma  $\mathcal{Q}$ -equivalência.

### Demonstração

Esta proposição foi provada acima no caso de  $Y$  ser sobrejetora. Assim, vamos supor aqui que  $Y(\mathcal{X})$  é um subconjunto próprio de  $\mathcal{Y}$ . Isto não garante que a imagem de  $Y$ ,  $Y(\mathcal{X})$ , é um conjunto  $\mathcal{B}$ -mensurável.

Sejam  $B_1$  e  $B_2$  dois conjuntos em  $\mathcal{B}$  tal que  $Y^{-1}(B_1) = Y^{-1}(B_2)$ . Assim, não é difícil ver que  $B_1 \Delta B_2 \subset \mathcal{Y} - Y(\mathcal{X})$  e que todo elemento dessa classe é  $\mathcal{Q}$ -nulo, pois, se  $N \in [\mathcal{Y} - Y(\mathcal{X})]$ ,  $Y^{-1}(N) = \emptyset$  e  $\forall Q \in \mathcal{Q}$ ,  $Q(N) = P(Y^{-1}(N)) = P(\emptyset) = 0$ . Assim,  $B_1 \equiv B_2[\mathcal{Q}]$ . Assim, existe uma correspondência “ $\mathcal{Q}$ -essencialmente um-a-um” entre  $\mathcal{A}_Y$  e  $\mathcal{B}$ . □

Seja agora  $g$  uma v.a. definida em  $(\mathcal{Y}, \mathcal{B})$  e  $Y$  uma estatística de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{Y}, \mathcal{B})$ . Observe que  $\forall B_1 \in \mathcal{B}_1$ ,

$$(g \circ Y)^{-1}(B_1) = Y^{-1}(g^{-1}(B_1)) \in \mathcal{A}_Y$$

pois  $g^{-1}(B_1) \in \mathcal{B}$ . Isto é, se  $g \in \mathcal{B}$  e  $Y$  é uma estatística definida em  $(\mathcal{X}, \mathcal{A})$  com valores em  $(\mathcal{Y}, \mathcal{B})$ , então  $g \circ Y \in \mathcal{A}_Y$ . Isto é esquematizado no seguinte diagrama.

$$\begin{array}{ccc} (\mathcal{X}, \mathcal{A}) & \xrightarrow{Y} & (\mathcal{Y}, \mathcal{B}) \\ & & \downarrow g \\ & & (R_1, \mathcal{B}_1) \\ & \searrow^{g \circ Y} & \end{array}$$

A próxima proposição apresenta o inverso deste resultado.

### Proposição 2.2

Com as notações acima descritas, temos que se  $f \in \mathcal{A}_Y$  é uma v.a. definida em  $(\mathcal{X}, \mathcal{A})$ , então existe uma v.a.  $g$  definida em  $(\mathcal{Y}, \mathcal{B})$  tal que  $f = g \circ Y$ .

### Demonstração

Provaremos esse resultado apenas no caso particular onde  $f \in \mathcal{A}_Y$  é uma função simples; isto é,  $f = \sum_{i=1}^k a_i I_{A_i}$ , onde  $\{A_1, A_2, \dots, A_k\} \subset \mathcal{A}_Y$  forma uma partição de  $\mathcal{X}$ . Como  $Y$  é uma estatística, existem  $B_1, \dots, B_k$  em  $\mathcal{B}$  tal que  $A_i = Y^{-1}(B_i)$ ,  $\forall i = 1, 2, \dots, k$ . Defina uma função  $g$  tal que

$$g(y) = \begin{cases} a_i & \text{se } y \in B_i B_1^c B_2^c \cdots B_{i-1}^c \quad \forall i = 1, 2, \dots, k \\ 0 & \text{se } y \in B_1^c B_2^c \cdots B_k^c \end{cases}$$

onde  $B_0 = \emptyset$ .

Não é difícil ver que  $Y^{-1}(B_i B_1^c \cdots B_{i-1}^c) = A_i$  e  $Y^{-1}(B_1^c \cdots B_k^c) = \emptyset$ . Note que os conjuntos  $B_i B_1^c \cdots B_{i-1}^c$ ,  $\forall i = 1, 2, \dots, k$ , juntamente com  $B_1^c B_2^c \cdots B_k^c$  forma uma partição  $\mathcal{B}$ -mensurável de  $\mathcal{Y}$ . Assim,  $g \in \mathcal{B}$  é uma função simples tal que  $f = g \circ Y$ .  $\square$

### Exercício 2.1

Complete a demonstração da proposição acima, para quaisquer v.a.'s.

### Exercício 2.2

Se a v.a.  $f$  tem duas representações diferentes  $g_1 \circ Y$  e  $g_2 \circ Y$ , então  $g_1 = g_2[\mathcal{Q}]$ .

A proposição abaixo relaciona as integrais das v.a.'s  $f$  e  $g$  descritas acima. Além de considerarmos a estatística  $Y$  e as v.a.'s  $f$  e  $g$ , vamos supor (em geral) que  $P$  é uma medida  $\sigma$ -finita em  $(\mathcal{X}, \mathcal{A})$ .

### Proposição 2.3

Seja  $g$  uma v.a. integrável. Então, para todo  $B \in \mathcal{B}$ ,

$$\int_{Y^{-1}(B)} (g \circ Y)(x) P(dx) = \int_B g(y) Q(dy) ,$$

onde,  $\forall B \in \mathcal{B}$ ,  $Q(B) = P(Y^{-1}(B))$ .

### Demonstração

Nos restringiremos ao caso onde  $g$  é uma função indicador de um conjunto  $B_0 \in \mathcal{B}$ . Assim,

$$\begin{aligned}\int_B g(y)Q(dy) &= \int_B I_{B_0}(y)Q(dy) = \int_{BB_0} Q(dy) \\ &= Q(BB_0) = P(Y^{-1}(BB_0)) = P(Y^{-1}(B) \cap Y^{-1}(B_0))\end{aligned}$$

e

$$\int_{Y^{-1}(B)} (I_{B_0} \circ Y)(x)P(dx) = \int_{Y^{-1}(B)} I_{Y^{-1}(B_0)}(x)P(dx) = P(Y^{-1}(B) \cap Y^{-1}(B_0)) . \quad \square$$

### Exercício 2.3

Mostre que  $(I_{B_0} \circ Y)(x) = I_{Y^{-1}(B_0)}(x)$ .

### Exercício 2.4

Complete a demonstração acima para o caso onde  $g$  é uma v.a. integrável qualquer definida em  $(\mathcal{Y}, \mathcal{B})$ .

Com as proposições 2.1, 2.2 e 2.3, podemos agora pensar em  $(\mathcal{X}, \mathcal{A}_Y, \mathcal{P})$  e  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  como sendo equivalentes ou indistinguíveis para todos os nossos propósitos. Assim, livremente identificaremos a estatística  $Y$  com a subálgebra  $\mathcal{A}_Y$  que ela gera.

Em muitos casos, é muito mais simples estabelecer resultados para subálgebras do que estabelecer os resultados correspondentes para as estatísticas. Assim, supondo que em um problema os resultados em questão são verdadeiros para ambos, estatísticas e subálgebras, os resultados para as subálgebras são pelo menos tão úteis quanto os resultados para as estatísticas. Dessa forma, não necessitamos daqui para frente mencionar o espaço  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  para o estudo de  $Y$ , pois teremos em mão os resultados análogos para  $(\mathcal{X}, \mathcal{A}_Y, \mathcal{P})$ .

Para praticarmos a linguagem de subálgebras, consideremos duas estatísticas  $Y$  e  $Y'$ , com as correspondentes estruturas  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  e  $(\mathcal{Y}', \mathcal{B}', \mathcal{Q}')$ , respectivamente. Suponha que exista uma estatística,  $Z$ , definida em  $(\mathcal{Y}', \mathcal{B}')$  tomando valores em  $(\mathcal{Y}, \mathcal{B})$  tal que  $Y(x) \equiv Z(Y'(x)) \forall x \in \mathcal{X}$ .

### Exercício 2.5

Verifique que  $\mathcal{A}_Y \subset \mathcal{A}_{Y'}$ .

Quando  $\mathcal{A}_Y \subset \mathcal{A}_{Y'}$ , escreveremos  $Y \succ Y'$ . No caso em que  $\mathcal{A}_Y \equiv \mathcal{A}_{Y'}$ , escreveremos  $Y \equiv Y'$ .

### Exercício 2.6

Verifique que “ $\succ$ ” define uma ordem parcial entre estatísticas definidas em  $(\mathcal{X}, \mathcal{A})$  e “ $\equiv$ ” define uma relação de equivalência.

Assim, quando desejamos comparar duas estatísticas  $Y$  e  $Y'$ , é mais fácil (conveniente) trabalharmos com os modelos  $(\mathcal{X}, \mathcal{A}_Y, \mathcal{P})$  e  $(\mathcal{X}, \mathcal{A}_{Y'}, \mathcal{P})$  do que com os modelos equivalentes  $(\mathcal{Y}, \mathcal{B}, \mathcal{Q})$  e  $(\mathcal{Y}', \mathcal{B}', \mathcal{Q}')$ . Da mesma forma, as definições e resultados válidos para subálgebras podem ser traduzidos em resultados e definições correspondentes, para qualquer estatística  $Y$ . Necessita-se para isso apenas pensar-se em  $\mathcal{A}_Y$ .

O estudo das partições de  $\mathcal{X}$  completa essa linguagem abstrata em que vamos nos basear.

### Definição 2.1

Por uma partição de  $\mathcal{X}$  queremos dizer uma classe  $\Pi$  (possivelmente não enumerável) de subconjuntos  $\pi_\delta$  de  $\mathcal{X}$  que são mutuamente exclusivos e coletivamente exaustivos. Isto é, seja  $\Delta$  um conjunto de índices tal que

$$\Pi = \{\pi_\delta; \delta \in \Delta\}, \quad \pi_\delta \cap \pi_{\delta'} = \emptyset$$

se  $\delta \neq \delta'$  e  $\bigcup_{\delta \in \Delta} \pi_\delta = \mathcal{X}$ . Em geral, não necessariamente, temos  $\pi_\delta \in \mathcal{A}, \forall \delta \in \Delta$ .

### Exercício 2.7

Se  $Y$  é uma estatística de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{Y}, \mathcal{B})$ , mostre que  $\{Y^{-1}(\{y\}); \forall y \in \mathcal{Y}\}$  forma uma partição de  $\mathcal{X}$ .

### Definição 2.2

Para cada partição  $\Pi$ , definimos  $\mathcal{A}(\Pi)$  como sendo a classe dos conjuntos  $\mathcal{A}$ -mensuráveis que podem ser escritos como uniões de elementos de  $\Pi$ .

### Exercício 2.8

Prove que  $\mathcal{A}(\Pi)$  é uma  $\sigma$ -álgebra.

Esta subálgebra  $\mathcal{A}(\Pi)$  é conhecida como a subálgebra induzida por  $\Pi$ . O exemplo

abaixo mostra que nem toda subálgebra é induzida por alguma partição.

### Exemplo 2.1

Sejam  $\mathcal{X} = R_1$ ,  $\mathcal{A} = \mathcal{B}_1$  e

$$\mathcal{D} = \{D \in \mathcal{B}_1; D \text{ ou } D^c \text{ são enumeráveis}\} .$$

É claro que  $\mathcal{D}$  contém  $\{x\}$ ,  $\forall x \in R_1$ . Assim, toda partição  $\Pi$  que induz  $\mathcal{D}$  deve conter todos os conjuntos unitários de  $R_1$ . Assim,  $\mathcal{A}(\Pi) = \mathcal{B}_1 = \mathcal{A}$ .

### Exercício 2.9

Prove que toda subálgebra de  $\mathcal{A}$  induz uma partição de  $\mathcal{X}$ . (Sugestão: Defina  $\pi_x$  como sendo a intersecção de todos os elementos da subálgebra que contém  $x$ .)

Uma partição  $\Pi$  induzida por uma subálgebra  $\mathcal{D}$  será representada por  $\Pi(\mathcal{D})$ .

### Exercício 2.10

- i) Dê um exemplo para mostrar que nem toda partição é induzida por uma subálgebra.
- ii) Mostre que duas subálgebras diferentes podem induzir a mesma partição e vice-versa.

### Definição 2.3

Se todo elemento de  $\Pi$  é uma união de elementos de  $\Pi'$ , uma outra partição, então diremos que  $\Pi'$  ( $\Pi$ ) é mais fina (mais grossa) que  $\Pi$  ( $\Pi'$ ). Neste caso escrevemos  $\Pi < \Pi'$ .

### Exercício 2.11

Prove que:

- i) Se  $\mathcal{D}_1 \subset \mathcal{D}_2$ , então  $\Pi(\mathcal{D}_1) < \Pi(\mathcal{D}_2)$ .
- ii) Se  $\Pi_1 < \Pi_2$ , então  $\mathcal{A}(\Pi_1) \subset \mathcal{A}(\Pi_2)$ .

### Proposição 2.4

Seja  $\Pi$  uma partição e  $\mathcal{D}$  uma subálgebra.



- (i) Se  $\Pi$  é  $\mathcal{A}$ -mensurável (isto é,  $\Pi \subset \mathcal{A}$ ), então  $\Pi = \Pi(\mathcal{A}(\Pi))$ .
- (ii)  $\mathcal{D}$  está contido em  $\mathcal{A}(\Pi(\mathcal{D}))$ .

### Demonstração

- (i) Observe que  $\forall \pi \in \Pi$  é um átomo de  $\mathcal{A}(\Pi)$ . Contudo, qualquer elemento  $\pi_x$  de  $\Pi(\mathcal{A}(\Pi))$  é a intersecção de todos os elementos de  $\mathcal{A}(\Pi)$  que contém  $x$ . Assim,  $\pi$  só poderá ser um elemento de  $\Pi(\mathcal{A}(\Pi))$ .
- (ii) Escolha  $D \in \mathcal{D}$  e verifique que  $\forall \pi \in \Pi(\mathcal{D})$  satisfaz  $\pi \cap D = \emptyset$ . Assim,  $D$  só poderá ser uma união de elementos de  $\Pi(\mathcal{D})$ ; isto é,  $D \in \mathcal{A}[\Pi(\mathcal{D})]$ .  $\square$

### Exercício 2.12

Mostre que  $\forall D \in \mathcal{D}$  e  $\forall \pi \in \Pi(\mathcal{D}) \Rightarrow \pi \cap D = \emptyset$  ou  $\pi \cap D = \pi$ .

### Proposição 2.5

Uma subálgebra  $\mathcal{D} = \mathcal{A}(\Pi)$  para alguma partição  $\Pi$  se e somente se  $\mathcal{D}$  é fechada para uniões de seus elementos.

### Demonstração

$\Rightarrow$  Segue da definição de  $\mathcal{A}(\Pi)$ .

$\Leftarrow$  Notemos que  $\mathcal{D} = \mathcal{A}[\Pi(\mathcal{D})]$ . Pela hipótese,  $\forall \pi \in \Pi(\mathcal{D}) \subset \mathcal{D}$ , pois  $\mathcal{D}$  é fechada também por intersecções. Seja  $D \in \mathcal{A}[\Pi(\mathcal{D})]$ , então  $D$  é uma união de elementos de  $\Pi(\mathcal{D})$  e então pertence a  $\mathcal{D}$ ; isto é,  $\mathcal{A}[\Pi(\mathcal{D})] \subset \mathcal{D}$ . O resultado segue da proposição anterior.  $\square$

## EXERCÍCIOS COMPLEMENTARES

1. Uma função arbitrária  $f$ , em  $\mathcal{X}$  pode ser definida, genericamente, como uma transformação mensurável entre dois espaços de medida  $(\mathcal{X}, \mathcal{A})$  e  $(\mathcal{Y}, \mathcal{B})$ , onde  $\mathcal{Y}$  é o contradomínio de  $f$  e  $\mathcal{B}$  é a  $\sigma$ -álgebra  $\mathcal{B} = \{B \subset \mathcal{Y} : f^{-1}(B) \in \mathcal{A}\}$ . Neste sentido, a subálgebra induzida por  $f$  é  $\mathcal{A}_f = f^{-1}(\mathcal{B})$ . Prove que nenhuma subálgebra própria de  $\mathcal{A}$ , contendo todos os conjuntos unitários de  $\mathcal{X}$ , pode ser induzida por uma função no sentido acima.

2. Sendo  $\mathcal{D}$  uma subálgebra de  $\mathcal{A}$ , defina em  $\mathcal{X}$  a seguinte relação binária:  $x, y \in \mathcal{X}$ ,  $x \sim y \Leftrightarrow I_D(x) = I_D(y)$ , para cada  $D \in \mathcal{D}$ , onde  $I_D$  é a função indicador de  $D$ .
  - a) Mostre que  $\sim$  é uma relação de equivalência e demonstre que o conjunto das classes de equivalência define a partição induzida por  $\mathcal{D}$ .
  - b) Prove que, se  $\mathcal{D}$  é separável, então cada classe de equivalência é um átomo.
  
3. Considere o modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  com  $\mathcal{A}$  a  $\sigma$ -álgebra das partes. Seja  $\Pi$  uma partição de  $\mathcal{X}$ .
  - a) Mostre que cada parte  $\pi$  de  $\Pi$  é um átomo de  $\mathcal{A}(\Pi)$ .
  - b) Seja  $F : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  tal que  $F$  é  $\mathcal{A}(\Pi)$ -mensurável. Mostre que  $F$  é constante sobre cada parte  $\pi$  de  $\Pi$ .
  
4. Considere o espaço de probabilidade  $(\mathcal{X}, \mathcal{A}, P)$  e  $\mathcal{B}$  uma subálgebra de  $\mathcal{A}$ . Se  $B$  é um átomo de  $\mathcal{B}$ , então  $f^* = E\{f|\mathcal{B}\} \in \mathcal{B}$  é constante em  $B$ . Ilustre este resultado mostrando o seguinte fato: sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e normais padrão (média 0 e variância 1). Seja  $X$  a soma de quadrados e  $Y$  a soma simples dessas variáveis. Mostre que  $X^* = E\{X|Y\} = E\{X|\mathcal{A}_Y\}$  é constante nos átomos de  $\mathcal{A}_Y$ .
  
5. Considere o espaço mensurável  $(\mathcal{X}, \mathcal{A})$ . Seja  $\mathcal{C}$  uma subálgebra de  $\mathcal{A}$  e  $\Pi = \Pi(\mathcal{C})$  a partição induzida por  $\tau$ . Sejam  $\pi_1$  e  $\pi_2$  duas partes diferentes de  $\Pi$ . Mostre que existe  $C \in \mathcal{C}$  tal que  $\pi_1 \subset C$  e  $\pi_2 \cap C = \emptyset$ .
  
6. Sejam  $\mathcal{C}$  e  $\mathcal{D}$  duas subálgebras de  $\mathcal{A}$  e  $\Pi_1 = \Pi(\mathcal{C})$ ,  $\Pi_2 = \Pi(\mathcal{D})$  com  $\Pi_1 \subset \mathcal{C}$  e  $\Pi_2 \subset \mathcal{D}$ . Seja  $\Pi_1 \wedge \Pi_2$  a partição mais fina que é mais grossa que ambas  $\Pi_1$  e  $\Pi_2$ . Verifique que uma partição assim definida sempre existe. Mostre que  $\Pi_1 \wedge \Pi_2 = \Pi(\mathcal{C} \cap \mathcal{D})$ .
  
7. Seja  $F$  uma estatística de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{Y}, \mathcal{B})$  e  $\Pi = \{F^{-1}(y) : y \in \mathcal{Y}\}$ . Observe que  $\Pi$  é uma partição de  $\mathcal{X}$  (Exercício 2.7). Estude condições sob as quais  $\mathcal{A}(\Pi) = \mathcal{A}_F$ .

## CAPÍTULO 3

### A SUFICIÊNCIA DE FISHER E SUAS DERIVAÇÕES

Para introduzirmos a noção de suficiência de Fisher (denotada por  $F$ -suficiência ou simplesmente por suficiência) faremos uma breve discussão do seu significado.

Seja  $\theta \in \Theta$  o “parâmetro” que indexa a distribuição,  $P_\theta$  ( $\theta \in \Theta$ ), de um elemento aleatório observável  $X$  (que toma valores  $x \in \mathcal{X}$ ). Considere  $F = F(X)$  uma estatística definida em  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Segundo Fisher [(1922): On the mathematical foundations of theoretical statistics - *Ph. Trans. of R.S.* - A222, 309-368], a estatística  $F$  é dita ser *suficiente* para  $X$  com respeito a  $\theta$  se a distribuição condicional de  $X$  dado  $F$  é essencialmente a mesma para todo  $\theta \in \Theta$ . A intuição contida nesta definição é descrita nas afirmações abaixo:

1. Tão logo o valor  $F(x)$  de  $F$  é observado, qualquer tentativa, de obter-se detalhes adicionais da “amostra”  $x$  ( $\in \mathcal{X}$ ), corresponderia a observarmos o valor,  $z$ , de um elemento aleatório,  $Z$ , cuja distribuição não é relacionada com  $\theta$ , o “parâmetro desconhecido” (isto é, a distribuição de  $Z$  seria totalmente conhecida).
2. Equivalentemente, após observarmos  $F(x)$ , com um exercício de “pós-aleatorização”, observaríamos o valor  $z$  de um elemento aleatório  $Z$  (cuja distribuição, após obser-

varmos  $F(x)$ , é completamente especificada) e obteríamos o valor  $(F(x), z)$  de  $(F, Z)$  que seria equivalente a observarmos  $X$  (intuitivamente,  $(F, Z)$  teria a mesma distribuição de  $X$ ).

3. Ao observarmos o  $Z$  descrito acima, não estaríamos gerando nenhuma “informação adicional” sobre  $\theta$ . Esta é a razão pela qual os franceses usam a palavra “exaustiva” no lugar de “suficiente”.

Assim, Fisher caracterizou a estatística suficiente como aquela que concentra toda a informação, contida em  $x$ , sobre  $\theta$ .

### Exemplo 3.1

Consideremos aqui o caso de  $n$  provas independentes de Bernoulli, cujo parâmetro é  $p \in (0, 1)$ . Assim,  $X = (X_1, X_2, \dots, X_n)$  onde  $X_i$  toma valores em  $\{0, 1\}$ . Seja  $F(X) = \sum_{i=1}^n X_i$  tomando valores em  $\{0, 1, \dots, n\}$ . Note que  $F$  é suficiente para  $X$  com respeito a  $p$  pois, conhecido  $F$  podemos gerar valores de  $Z = (Z_1, \dots, Z_n)$ , ( $z = (z_1, \dots, z_n)$ ,  $z_i \in \{0, 1\}$ ) através da distribuição  $\frac{1}{\binom{n}{F}}$ . Isto é, a distribuição condicional de  $Z$  dado  $F$ ,  $(Z|F)$ , é  $\frac{1}{\binom{n}{F}}$  que independe do valor de  $p$  e a distribuição de  $(F, Z)$  ou simplesmente  $Z$  é a mesma de  $X$ . Assim, os dois processos: 1 - observar  $X$  e 2 - observar  $F$  e produzir  $Z$ , estão gerando resultados provenientes da mesma distribuição,  $\{p^F(1-p)^{n-F}; p \in (0, 1)\}$ . □

Nos termos abstratos com que estamos tratando o modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ ,  $F$  será vista como uma subálgebra  $\mathcal{A}_F$  de  $\mathcal{A}$ . Vamos denotar por  $\mathcal{L}$  a classe de todas as variáveis aleatórias essencialmente limitadas. Isto é, se  $f$  é um elemento de  $\mathcal{L}$ , então  $f \in \mathcal{A}$  (é  $\mathcal{A}$ -mensurável) e existe  $a \in R_1$  tal que  $\forall P \in \mathcal{P}, P\{x \in \mathcal{X}; |f(x)| > a\} = 0$ . Se  $\mathcal{D}$  é uma subálgebra de  $\mathcal{A}$ , escreveremos  $\mathcal{L}(\mathcal{D})$  para denotar a classe das funções em  $\mathcal{L}$  que são essencialmente  $\mathcal{D}$ -mensuráveis [ $f \in \mathcal{L}(\mathcal{D}) \Rightarrow f \in \overline{\mathcal{D}}$ ].

Recordemos aqui o conceito de esperança condicional de  $f \in \mathcal{L}$  dado uma subálgebra  $\mathcal{D}$ . Fixada uma medida  $P \in \mathcal{P}$ , a esperança condicional de  $f$  dado  $\mathcal{D}$  (com respeito a  $P$ ),  $E_P\{f|\mathcal{D}\}$ , é uma variável aleatória  $f^{*P} \in \mathcal{L}(\mathcal{D})$  tal que,  $\forall g \in \mathcal{L}(\mathcal{D}), E_P\{fg\} = E_P\{(f^{*P})g\}$ . Aqui,  $E_P\{\cdot\}$  é a média calculada com a medida  $P$ .

### Exercício 3.1

Mostre que o operador  $*P$  é uma projeção de  $\mathcal{L}$  em  $\mathcal{L}(\mathcal{D})$ .

### Exercício 3.2

Mostre que no lugar de  $g \in \mathcal{L}(\mathcal{D})$  poderíamos, de forma equivalente, nos restringir às funções indicadores,  $I_D$ , de elementos de  $\mathcal{D}$ .

Introduziremos agora a noção de suficiência para subálgebras.

### Definição 3.1

Uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  é dita ser “suficiente” com respeito ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se, para cada  $f \in \mathcal{L}$ , existir  $f^* \in \mathcal{L}(\mathcal{D})$  tal que,

$$f^* = f^{*P}[P] \quad \forall P \in \mathcal{P} .$$

Note que  $f^*$  é uma versão da esperança condicional que é comum para todo  $P \in \mathcal{P}$ . Por isto,  $f^*$  é chamada de esperança condicional-universal.

### Exercício 3.3

Mostre que obteríamos uma definição equivalente se, no lugar de  $f \in \mathcal{L}$ , considerássemos:

- (i)  $f = I_A[P]$ ,  $\forall A \in \mathcal{A}$  e  $\forall P \in \mathcal{P}$ .
- (ii)  $f \in \mathcal{A}$  e  $f \in \mathcal{L}_1$ , onde  $\mathcal{L}_1$  é a classe de todas as variáveis aleatórias integráveis para  $P \in \mathcal{P}$ .

### Exercício 3.4

Mostre que todas as propriedades da esperança condicional são mantidas para a esperança condicional-universal.

Sugerimos como leitura, para o estudo teórico da esperança condicional, os seguintes artigos, na ordem:

1. Moy, S.T.C. (1954). Characterization of conditional expectation as a transformation on function spaces. *Pacific J. Math.*, 4, 47-63.

2. Bahadur, R.R. (1955). Measurable subspaces and subalgebras. *Proc. Amer. Math. Soc.*, 6, 565-70.

A seguinte proposição apresenta algumas conseqüências imediatas da definição de subálgebra suficiente.

### Proposição 3.1

Sejam  $f \in \mathcal{L}$  e  $h \in \mathcal{L}$  duas variáveis aleatórias e seja  $\mathcal{D}$  uma subálgebra suficiente, então:

- (i)  $f^*$  é essencialmente única. Isto é, se existe  $g \in \mathcal{L}(\mathcal{D})$  tal que  $g = f^{*P}[P] \forall P \in \mathcal{P}$ , então  $f^* = g[\mathcal{P}]$ .
- (ii) Se  $f = h[\mathcal{P}] \Rightarrow f^* = h^*[\mathcal{P}]$ .
- (iii)  $f^* = f[\mathcal{P}] \Leftrightarrow f \in \mathcal{L}(\mathcal{D})$
- (iv)  $\mathcal{D}_0 \equiv \mathcal{D}[\mathcal{P}] \Rightarrow \mathcal{D}_0$  é suficiente.
- (v)  $\overline{\mathcal{D}}$  é suficiente.
- (vi)  $\mathcal{A}$  é suficiente.
- (vii)  $\mathcal{A}_0 = \{\phi, \mathcal{X}\}$  é suficiente  $\Leftrightarrow \mathcal{P}$  possui apenas um elemento.

### Exercício 3.5

Demonstre a Proposição 3.1.

A definição abaixo é conseqüência de toda a discussão apresentada no capítulo anterior. Isto é, uma estatística  $F$  é dita ser suficiente com respeito ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se a subálgebra  $\mathcal{A}_F$ , gerada por  $F$ , é suficiente.

### Definição 3.2

Uma estatística  $F$  de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{Y}, \mathcal{B})$  é suficiente para  $(\mathcal{X}, \mathcal{A})$  com respeito a  $\mathcal{P}$  se, para cada  $h \in \mathcal{L}$ , existe  $f \in \mathcal{L}(\mathcal{B})$  tal que,  $\forall B \in \mathcal{B}$  e  $P \in \mathcal{P}$ ,

$$\int_{F^{-1}(B)} h(x)P(dx) = \int_B f(y)PF^{-1}(dy) .$$

A equivalência entre as definições 3.1 e 3.2 é facilmente entendida com os resultados do Capítulo 2.

**Teorema 3.1**

Sejam  $\mathcal{D}_0$  e  $\mathcal{D}_1$  duas subálgebras de  $\mathcal{A}$  tal que  $\mathcal{D}_0 \subset \mathcal{D}_1$  [ $\mathcal{P}$ ]. Se  $\mathcal{D}_0$  é suficiente com respeito a  $(\mathcal{X}, \mathcal{D}_1, \mathcal{P})$  e  $\mathcal{D}_1$  é suficiente com respeito a  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , então  $\mathcal{D}_0$  é suficiente com respeito a  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

**Demonstração**

Seja  $f \in \mathcal{L}$ . A suficiência de  $\mathcal{D}_1$  implica que  $\exists f_1 \in \mathcal{D}_1$  tal que

$$f_1 = E_P\{f|\mathcal{D}_1\} [P] \quad \forall P \in \mathcal{P} .$$

Como  $\mathcal{D}_0$  é suficiente com respeito a  $(\mathcal{X}, \mathcal{D}_1, \mathcal{P})$ ,  $\exists f_0 \in \mathcal{D}_0$  tal que  $f_0 = E_P\{f_1|\mathcal{D}_0\} [P]$   $\forall P \in \mathcal{P}$ . Assim,  $\forall P \in \mathcal{P}$ ,

$$\begin{aligned} E_P\{f|\mathcal{D}_0\} &= E\{E\{f|\mathcal{D}_1\}|\mathcal{D}_0\} [P] \\ &= E\{f_1|\mathcal{D}_0\} [P] \\ &= f_0 [P] \end{aligned}$$

Isto é,  $f_0$  é uma versão universal da esperança condicional  $E_P\{f|\mathcal{D}_0\} \forall P \in \mathcal{P}$ . □

Em geral, uma estrutura estatística  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  não necessariamente possui uma subálgebra própria,  $\mathcal{D}$ , (isto é,  $\overline{\mathcal{D}} \subsetneq \mathcal{A}$ ) que seja suficiente. Por outro lado, existem casos onde o modelo possui diversas subálgebras (ou estatísticas) suficientes que são essencialmente diferentes. Como a idéia de suficiência é reduzir os “dados” ao máximo sem perder informação sobre o verdadeiro parâmetro (desconhecido),  $P_0$ , é natural investigarmos se existe realmente uma redução extrema. Isto está ligado aos conceitos de suficiência minimal e mínima que discutiremos após a próxima proposição, que por sua vez está altamente ligada à idéia de “emagrecimento” da subálgebra suficiente. O exercício abaixo tem por objetivo fazer com que o aluno recorde o teorema “smooth” de esperanças condicionais.

**Exercício 3.6**

Demonstre a igualdade  $E_P\{f|\mathcal{D}_0\} = E_P\{f_1|\mathcal{D}_0\} [P]$  usada na demonstração do Teorema 3.1.

Daqui para a frente usaremos a notação  $\mathcal{D}_0 \text{ suf } (\mathcal{D}_1; \mathcal{P})$  para indicar que a subálgebra  $\mathcal{D}_0$  (de  $\mathcal{A}$ ) é suficiente com respeito a  $(\mathcal{X}, \mathcal{D}_1, \mathcal{P})$ , onde  $\mathcal{D}_1$  também é uma subálgebra de  $\mathcal{A}$ . Note que não incluímos  $\mathcal{X}$  na notação, pois este está implícito, e que não exigimos nenhuma relação de inclusão entre  $\mathcal{D}_0$  e  $\mathcal{D}_1$ , pois isto se torna desnecessário, como mostra o resultado abaixo.

### Proposição 3.2

Sejam duas subálgebras de  $\mathcal{A}$ ,  $\mathcal{D}_0$  e  $\mathcal{D}_1$ . As duas afirmações abaixo são equivalentes.

- (a)  $\mathcal{D}_0 \text{ suf } (\mathcal{D}_1; \mathcal{P})$ ,
- (b)  $\mathcal{D}_0 \text{ suf } (\mathcal{D}_0 \vee \mathcal{D}_1; \mathcal{P})$ .

### Demonstração

$\Leftarrow$  Note que  $\forall f \in \mathcal{L}(\mathcal{D}_1) \Rightarrow f \in \mathcal{L}(\mathcal{D}_0 \vee \mathcal{D}_1)$ , então (b)  $\Rightarrow$  (a) trivialmente.

$\Rightarrow$  Considere a classe  $\mathbb{D} = \{D \in \mathcal{A}; \exists I_D^* \in \mathcal{D}_0\}$ , onde  $I_D^*$  é a esperança condicional universal de  $I_D$  dado  $\mathcal{D}_0$ . Não é difícil verificar que  $\mathbb{D}$  é um sistema de Dynkin (isto é,  $\mathbb{D}$  é monotônica e fechada por diferenças próprias).

Como  $\mathcal{D}_0 \text{ suf } (\mathcal{D}_1, \mathcal{P})$ ,  $\forall D \in \mathcal{D}_1$ ,  $\exists$  uma função  $I_D^* \in \mathcal{L}(\mathcal{D}_0)$  tal que

$$I_D^* = E_P\{I_D|\mathcal{D}_0\} [P] \quad \forall P \in \mathcal{P} .$$

Note agora que  $\forall D' \in \mathcal{D}_0$  e  $D \in \mathcal{D}_1$ , temos que  $\forall P \in \mathcal{P}$ ,

$$E_P\{I_{D'D}|\mathcal{D}_0\} = E_P\{I_{D'}I_D|\mathcal{D}_0\} = I_{D'}E_P\{I_D|\mathcal{D}_0\} [P] = I_{D'}I_D^* [P] ,$$

onde  $I_{D'}I_D^*$  é independente da escolha de  $P$ . Assim, temos que

$$\{D'D : D' \in \mathcal{D}_0, D \in \mathcal{D}_1\} \subset \mathbb{D} ,$$

onde  $\{D'D : D' \in \mathcal{D}_0, D \in \mathcal{D}_1\}$  é uma classe fechada por intersecções. Pelo Teorema de Dynkin,

$$\mathcal{D}_0 \vee \mathcal{D}_1 = \sigma\{D'D : D' \in \mathcal{D}_0, D \in \mathcal{D}_1\} \subset \mathbb{D} .$$

Isto mostra que  $\mathcal{D}_0 \text{ suf } \{\mathcal{D}_0 \vee \mathcal{D}_1; \mathcal{P}\}$ . □

### Observação

É importante que o leitor consiga interpretar a Proposição 3.2 da seguinte forma:



Se uma estatística  $Y$  é suficiente para uma outra estatística  $Y_1$ , com respeito à família  $\mathcal{P}$ , então, na verdade,  $Y_0$  é suficiente para  $(Y_1, F(Y_0))$ , onde  $F(Y_0)$  é qualquer gerador de uma subálgebra de  $\mathcal{A}_{Y_0}$  (isto é,  $F(Y_0)$  é uma função de  $Y_0$ ).

### Exercício 3.7

Verifique que a classe  $\mathbb{D}$  é um sistema de Dynkin e recorde o Teorema de Dynkin.

Concluiremos este capítulo com alguns conceitos de suficiência que são importantes na prática quando estamos procurando reduções dos dados.

**Definição 3.3** Uma subálgebra  $\mathcal{D}_0$  (de  $\mathcal{A}$ ) é dita ser suficiente minimal com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se  $\mathcal{D}_0 \text{ suf}(\mathcal{A}; \mathcal{P})$  e para toda subálgebra  $\mathcal{D}$  de  $\mathcal{D}_0$  ( $\mathcal{D} \subset \mathcal{D}_0$  [ $\mathcal{P}$ ]) tal que  $\mathcal{D} \text{ suf}(\mathcal{A}; \mathcal{P})$  implicar  $\overline{\mathcal{D}} \equiv \overline{\mathcal{D}_0}$ .

### Definição 3.4

Uma subálgebra  $\mathcal{D}_0$  (de  $\mathcal{A}$ ) é dita ser suficiente mínima com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se, para qualquer subálgebra  $\mathcal{D}$  (de  $\mathcal{A}$ ) tal que  $\mathcal{D} \text{ suf}(\mathcal{A}, \mathcal{P})$ , tivermos  $\overline{\mathcal{D}_0} \subset \overline{\mathcal{D}}$ .

### Observação

Uma estatística  $Y_0$  é dita ser suficiente minimal se não existir nenhuma função própria de  $Y_0$  que seja suficiente.

Uma estatística  $Y_0$  é dita ser suficiente mínima se for função de toda estatística suficiente.

A suficiência mínima nem sempre existe. Contudo, quando existe, suficiência minimal implica suficiência mínima. Isto é, toda estatística suficiente minimal é equivalente a suficiente mínima, quando esta existe.

Estes conceitos serão discutidos nos capítulos seguintes quando teremos casos particulares de famílias  $\mathcal{P}$ .

A definição abaixo é análoga à suficiência em termos de estatísticas.

### Definição 3.5

Uma partição  $\Pi$  de  $\mathcal{X}$  é dita ser suficiente com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se a  $\sigma$ -álgebra  $\mathcal{A}(\Pi)$  induzida por  $\Pi$  é uma subálgebra suficiente com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Fica implícito

que  $\Pi \subset \mathcal{A}$ .

Um outro conceito de extrema importância para simplificar algumas demonstrações futuras é o de suficiência por par.

### Definição 3.6

Uma subálgebra  $\mathcal{D}$  (de  $\mathcal{A}$ ) é dita ser suficiente por par, com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , se, para qualquer subclasse de dois elementos  $\{P_1, P_2\} \subset \mathcal{P}$ , tivermos  $\mathcal{D} \text{ suf}(\mathcal{A}; \{P_1, P_2\})$

É evidente que suficiência implica em suficiência por par. Contudo, o próximo exemplo mostra que não necessariamente a recíproca é verdadeira.

### Exemplo 3.1

Consideremos novamente o Exemplo 1.2 e lembremos que  $\mathcal{D}$  (como definido na ocasião) é suficiente por par, pois

$$\mathcal{D} \equiv \mathcal{A} [\{\mathcal{P}_1, \mathcal{P}_2\}] .$$

Verifiquemos agora que  $\mathcal{D}$  não é suficiente. Na verdade, não existe subálgebra própria de  $\mathcal{A}$  que seja suficiente. Isto é, se  $\mathcal{D}_0 \text{ suf}(\mathcal{A}; \mathcal{P}) \Rightarrow \overline{\mathcal{D}_0} \equiv \mathcal{A}$  ou ainda  $\mathcal{A}$  é essencialmente a única subálgebra suficiente com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Para concluirmos este fato, suporemos por absurdo que  $\exists \mathcal{D}_0 \subsetneq \mathcal{A} [\mathcal{P}]$  tal que  $\mathcal{D}_0 \text{ suf}(\mathcal{A}, \mathcal{P})$ . Escolha um conjunto  $A \in \mathcal{A} - \mathcal{D}_0$  e defina  $P_x$  como sendo uma probabilidade degenerada em  $x \in \mathcal{X}$ . Como  $\mathcal{D}_0 \text{ suf}(\mathcal{A}; \mathcal{P}) \exists h \in \mathcal{L}(\mathcal{D}_0)$  tal que

$$I_A(x) = \int_{\mathcal{X}} I_A(x') P_x(dx') = \int_{\mathcal{X}} h(x') P_x(dx') = h(x) .$$

Como  $x$  é arbitrário,  $h = I_A$ . Assim, desde que  $h \in \mathcal{D}_0 \Rightarrow A \in \mathcal{D}_0$  o que é uma contradição, pois  $A \in \mathcal{A} - \mathcal{D}_0$ . □

### Exercício 3.8

Mostre que no Exemplo 3.1, acima, poderíamos eliminar todas as palavras “essencialmente” e todos os complementos de subálgebras, sem modificar em nada o seu conteúdo.

Os dois exemplos que apresentamos a seguir são bastante famosos e mostram que:

- (1) nem sempre existe uma estatística suficiente mínima - isto é devido a Pitcher (1957);

(2) uma subálgebra  $\mathcal{D}$  que não é suficiente com respeito à  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  pode conter uma subálgebra  $\mathcal{D}_0$  tal que  $\mathcal{D}_0$  suf  $(\mathcal{A}, \mathcal{P})$  - isto é devido a Burkholder (1961).

As referências corretas são

PITCHER, T.S. (1957). Sets of measures not admitting necessary and sufficient statistics or subfields. *Ann. Math. Stat.* 28, 267-268.

BURKHOLDER, D.L. (1961). Sufficiency in the undominated case. *Ann. Math. Stat.* 32, 1191-1200.

Em ambos os exemplos temos  $\mathcal{X} = R_1$  e  $\mathcal{A} = \mathcal{B}_1$ . Além disso, um conjunto não boreliano  $E \neq \emptyset$  é escolhido de tal maneira que  $0 \notin E$  e  $E = -E = \{x; -x \in E\}$ .

### Exemplo 3.2 (Pitcher)

Seja  $\mathcal{P} = \{P_\theta : \theta \in R_1\}$  definido como:

$$P_\theta\{-\theta\} = P_\theta\{\theta\} = \frac{1}{2} \quad \text{se } \theta \in E$$

$$P_\theta\{\theta\} = 1 \quad \text{se } \theta \notin E .$$

Assim,  $\mathcal{A}_0 = \overline{\mathcal{A}}_0 = \{\emptyset, R_1\}$  (isto é,  $\emptyset$  é o único conjunto  $\mathcal{P}$ -nulo).

O nosso objetivo é mostrar que não existe uma subálgebra suficiente mínima.

Suponha por absurdo que  $\mathcal{D}$  é uma subálgebra suficiente mínima. Para cada  $e \in E$ , defina

$$\mathcal{D}_e = \{D \in \mathcal{B}_1; \{-e, e\} \cap D = \{-e, e\} \quad \text{ou} \quad \{e, -e\} \cap D = \emptyset\}$$

Assim:

$$(1) \quad \forall D \in \mathcal{D}_e, e \in D \Leftrightarrow -e \in D ,$$

$$(2) \quad \forall f \in \mathcal{D}_e \Leftrightarrow f(e) = f(-e) .$$

Provaremos agora que  $\mathcal{D}_e$  é suf  $(\mathcal{A}, \mathcal{P})$ .

Para qualquer que seja  $f \in \mathcal{L}$ , defina

$$f_e^*(x) = \begin{cases} f(x) & \text{se } x \in \{-e, e\}^c \\ \frac{f(-x)+f(x)}{2} & \text{se } x \in \{-e, e\} \end{cases} .$$

Assim,  $f_e^* \in \mathcal{L}(\mathcal{D}_e)$ , pois  $f_e^*(-e) = f_e^*(e)$ . Não é difícil provar que

$$(1) \quad \int_D f dP_\theta = \int_D f_e^* dP_\theta \quad \forall D \in \mathcal{D}_e \quad \text{e} \quad \forall P_\theta \in \mathcal{P} .$$

Isto é,  $f_e^*$  é a esperança condicional universal dado a subálgebra  $\mathcal{D}_e$ . Então, como  $\mathcal{D}$  é suficiente mínima,  $\forall e \in E$ ,  $\mathcal{D} \subset \mathcal{D}_e$ . Seja  $f^*$  a versão universal de  $E_\theta(f|\mathcal{D}) \forall \theta \in R_1$ . Então  $f^* \in \mathcal{L}(\mathcal{D}_e) \forall e \in E$ ; isto é,  $f^*(-e) = f^*(e) \forall e \in E$ .

Para  $\theta \notin E$ , note que  $f(\theta) = \int_{\mathcal{X}} f(x)P_\theta(dx) = \int_{\mathcal{X}} f^*(x)P_\theta(dx) = f^*(\theta)$ .

Para  $\theta \in E$ , temos que

$$\begin{aligned} \frac{f(-\theta) + f(\theta)}{2} &= \int_{\mathcal{X}} f(x)P_\theta(dx) = \int_{\mathcal{X}} f^*(x)P_\theta(dx) \\ &= \frac{f^*(-\theta) + f^*(\theta)}{2} = f^*(\theta) \end{aligned}$$

Assim, concluímos que, necessariamente,

$$f^*(x) = \begin{cases} f(x) & \text{se } x \notin E \\ \frac{f(-x)+f(x)}{2} & \text{se } x \in E \end{cases} .$$

Note que  $\mathcal{D} \subset \mathcal{B}_1$  implica que  $\mathcal{D}$  é constituída de conjuntos borelianos. Para concluirmos o absurdo, verifiquemos que  $f^*(x)$  pode não ser Borel-mensurável; isto é,  $f^*$  não necessariamente é um elemento de  $\mathcal{L}$ . Tome a função  $f = I_{(-\infty,0)}$  (função indicador do conjunto  $\{x \in R_1; x < 0\}$  que é um boreliano) e note que

$$f^*(x) = \begin{cases} 0 & \text{se } x \in [0, \infty) \cap E^c \\ 1 & \text{se } x \in (-\infty, 0) \cap E^c \\ \frac{1}{2} & \text{se } x \in E \end{cases} .$$

Se  $f^* \in \mathcal{D} \subset \mathcal{B}_1 \Rightarrow f^{*-1}(\frac{1}{2}) = E \in \mathcal{B}_1$  que é uma contradição, pois  $E$  não é boreliano por hipótese.  $\square$

### Exercício 3.9

Demonstre a igualdade (1) do exemplo acima e demonstre todas as afirmações que tiver dúvida.

### Exemplo 3.3 (Burkholder)

Seja  $\mathcal{P} = \{P_\theta : \theta \in R_1\}$ , onde  $P_\theta\{-\theta\} = P_\theta\{\theta\} = \frac{1}{2} \forall \theta \neq 0$  e  $P_0\{0\} = 1$ . Novamente,  $\mathcal{A}_0 = \overline{\mathcal{A}}_0 = \{\emptyset, R_1\}$ .

Considere agora a subálgebra gerada pelos borelianos simétricos; isto é,

$$\mathcal{D}_0 = \{D \in \mathcal{B}_1; D = -D\} .$$

Usando a mesma técnica do exemplo anterior, provaríamos que  $\mathcal{D}_0 \text{ suf } (\mathcal{A}, \mathcal{P})$ .

Considere agora a partição (de  $R_1$ )

$$\Pi = \{\{x\}, \{e, -e\}; \forall x \in E^c \text{ e } \forall e \in E\} .$$

Note que  $\mathcal{D}_0 \subset \mathcal{D} = \mathcal{A}(\Pi) = \{D \in \mathcal{B}_1; D \cap E = -D \cap E\}$  e provemos que embora  $\mathcal{D}_0$  suf  $(\mathcal{A}, \mathcal{P})$  e  $\mathcal{D}_0 \subset \mathcal{D} \subset \mathcal{A}$ ,  $\mathcal{D}$  não é suf  $(\mathcal{A}, \mathcal{P})$ .

Por contradição, vamos supor que  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P})$ . Então,  $\forall f \in \mathcal{L}, \exists f^* \in \mathcal{L}(\mathcal{D})$  tal que

$$\int_D f(x)P_\theta(dx) = \int_D f^*(x)P_\theta(dx) \quad \forall D \in \mathcal{D} \text{ e } \forall \theta \in R_1 .$$

Assim,  $\forall \theta \in E^c$ , temos  $\{\theta\} \in \mathcal{D}$  e então

$$\frac{1}{2}f(\theta) = \int_{\{\theta\}} f(x)P_\theta(dx) = \int_{\{\theta\}} f^*(x)P_\theta(dx) = \frac{1}{2}f^*(\theta) .$$

Por outro lado, se  $\theta \in E$ , então  $\{-\theta, \theta\} \in \mathcal{D}$  e então:

$$\begin{aligned} \frac{f(-\theta) + f(\theta)}{2} &= \int_{\{-\theta, \theta\}} f(x)P_\theta(dx) = \int_{\{-\theta, \theta\}} f^*(x)P_\theta(dx) \\ &= \frac{f^*(-\theta) + f^*(\theta)}{2} = f^*(\theta) \end{aligned}$$

A igualdade  $\frac{f^*(-\theta) + f^*(\theta)}{2} = f^*(\theta)$  é consequência do fato de que  $\forall \theta \in E$ ,  $\{-\theta, \theta\}$  é um átomo de  $\mathcal{D}$ . Assim,  $\forall g \in \mathcal{L}(\mathcal{D}) \Rightarrow g(\theta) = g(-\theta) \forall \theta \in E$ .

Assim, temos que a versão universal da esperança condicional de  $f(\in \mathcal{L})$  dado  $\mathcal{D}$  é

$$f^*(x) = \begin{cases} f(x) & \text{se } x \notin E \\ \frac{f(-x) + f(x)}{2} & \text{se } x \in E \end{cases} .$$

Se novamente tomamos  $f = I_{(-\infty, 0)}$ , temos;

$$I_{(-\infty, 0)}^*(x) = \begin{cases} 1 & \text{se } x \in (-\infty) \cap E^c \\ 0 & \text{se } x \in [0, \infty) \cap E^c \\ \frac{1}{2} & \text{se } x \in E . \end{cases}$$

Assim, como  $\mathcal{D} \subset \mathcal{B}_1$  e  $f^* \in \mathcal{L}(\mathcal{D})$  teríamos  $f^{*-1}(\frac{1}{2}) = E \subset \mathcal{B}_1$  o que é uma contradição. □

### Exercício 3.10

Prove que  $\mathcal{D}_0$  definida no Exemplo 3.3 é suf  $(\mathcal{A}, \mathcal{P})$ . Demonstre todas as afirmações que tiver dúvidas.

## Observações

1. O Exemplo 3.2 mostra que embora existam diversas estatísticas suficientes minimal, não são necessariamente idênticas e nesse caso não existe uma estatística suficiente mínima.
2. O Exemplo 3.3 mostra que podem existir duas subálgebras suficientes  $\mathcal{D}_0$  e  $\mathcal{D}_2$  onde  $\overline{\mathcal{D}}_0 \subset \overline{\mathcal{D}}_2$ , porém, não necessariamente, uma terceira subálgebra  $\mathcal{D}_1$ , onde  $\overline{\mathcal{D}}_0 \subset \overline{\mathcal{D}}_1 \subset \mathcal{D}_2$  é suficiente.
3. Os exemplos discutidos acima, embora matematicamente corretos, são bastante superficiais. Quando trabalhamos com medidas discretas é intuitivo que devemos trabalhar com  $\mathcal{A} = \sigma$ -álgebra de todas as partes de  $\mathcal{X}$  e não com os borelianos.

Tentaremos nos próximos capítulos nos basear em modelos mais realísticos, onde não possam acontecer inconveniências como estas.

## EXERCÍCIOS COMPLEMENTARES

1. Prove que as Definições 3.1 e 3.2 são equivalentes.
2. Considerando o Exercício 3.3 e a Definição 3.1, mostre que (3.1) é equivalente a:  
**Definição:** A estatística  $T$  é dita suficiente para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se a distribuição condicional de  $X$  dado  $T = t$  é constante em relação a  $\theta$  para todo  $t$ .
3. Sejam  $X_1, X_2, \dots, X_n$  v.a.i.i.d  $\sim U(0, \theta)$ , onde  $\theta > 0$ .
  - (i) Descreva o modelo estatístico induzido por  $X_1, X_2, \dots, X_n$ .
  - (ii) Verifique se a estatística  $\sum_{i=1}^n X_i$  é suficiente para o modelo.
4. Sejam  $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_N$  v.a.i.i.d indexadas pelo parâmetro  $\theta$ .
  - (i) Suponha que  $S$  e  $T$  são estatísticas suficientes para os modelos induzidos por  $X_1, \dots, X_n$  e  $X_{n+1}, \dots, X_N$  ( $N > n$ ), respectivamente. Mostre que  $R = (S, T)$  é uma estatística suficiente para o modelo induzido por  $X_1, \dots, X_N$ .

- (ii) Em (i) suponha que  $S$  e  $T$  são suficientes mínimas para os respectivos modelos. Prove ou desprove: “A estatística suficiente mínima para o modelo induzido por  $X_1, \dots, X_N$  é a estatística suficiente mínima para o modelo induzido por  $(S, T)$ .”
- (iii) Ilustre os resultados acima para a família  $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ .

5. Considere  $E[X|X_1, X_2, \dots] = E[X|\sigma(X_1, X_2, \dots)]$ , onde  $\sigma(X_1, X_2, \dots)$  é a sub- $\sigma$ -álgebra gerada por  $(X_1, X_2, \dots)$ . Queremos provar rigorosamente o seguinte resultado: “Sejam  $X_1, X_2, \dots$  v.a.i.i.d tais que  $E|X_1| < \infty$  e  $S_n = X_1 + \dots + X_n$ . Então  $E(X_1|S_n, S_{n+1}, \dots) = \frac{S_n}{n}$ ”

(i) Sejam  $X, Y$  e  $Z$  v.a. tais que  $E|Y| < \infty$ . Mostre que se  $(X, Y)$  é independente de  $Z$ , então  $E[Y|X] = E[Y|X, Z]$ .

(ii) Mostre que  $E(X_k|S_n) = \frac{S_n}{n}$ ;  $k = 1, \dots, n$ .

(iii) Mostre que  $\sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, X_{n+1}, X_{n+2}, \dots)$ .

(iv) Complete a prova da afirmação utilizando os itens anteriores.

6. Considere o modelo estatístico  $(\mathbb{R}^n, \mathcal{B}, \mathcal{P})$ . Seja  $\pi$  uma permutação sobre  $\{1, 2, \dots, N\}$ . Para  $B \in \mathcal{B}$ , defina  $\pi B = \{(x_{\pi(1)}, \dots, x_{\pi(n)}) : (x_1, \dots, x_n) \in B\}$ . Defina  $(\pi P)(B) = P(\pi B)$  para  $B \in \mathcal{B}$ . Seja  $\mathcal{P} = \{P : \pi P = P \text{ para } \pi \in S(n)\}$ , onde  $S(n)$  é o grupo de permutações de  $\{1, 2, \dots, N\}$ .

Mostre que  $T : (\mathbb{R}^n, \mathcal{B}) \mapsto (\mathbb{R}^n, \mathcal{B})$  definida por  $T(x_1, \dots, x_n) = (x^{(1)}, \dots, x^{(n)})$ , onde  $x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(n)}$  são as estatísticas de ordem, é uma estatística suficiente.

7. Seja  $B$  um evento tal que  $P(B) > 0$ . Defina uma medida  $Q(A) = P(AB)/P(B)$ ,  $\forall A \in \mathcal{A}$ . Sendo  $\mathbf{D}$  uma subálgebra, mostre que

$$Q(A|\mathbf{D}) = P(AB|\mathbf{D})/P(B|\mathbf{D}) .$$

Se  $f$  é uma variável aleatória definida em  $(\mathcal{X}, \mathcal{A}, Q)$ , defina  $E[f|\mathbf{D}]$  em termos da medida  $P$ .

8. Sejam  $\mathbf{D}_1, \mathbf{D}_2$  e  $\mathbf{D}_3$  subálgebras de  $\mathcal{A}$  e defina  $\mathbf{D}_{KL}$  como a menor subálgebra contendo  $\mathbf{D}_K$  e  $\mathbf{D}_L$ ,  $K < L$  e  $L = 1, 2, 3$ . Represente um evento genérico de  $\mathbf{D}_K$  por  $D_K$ . Mostre que as seguintes condições são equivalentes:

(a)  $P(D_3|\mathbf{D}_{12}) = P(D_3|\mathbf{D}_2)$

(b)  $P(D_1D_3|\mathbf{D}_2) = P(D_1|\mathbf{D}_2)P(D_3|\mathbf{D}_2)$

(c)  $P(D_1|\mathbf{D}_{23}) = P(D_1|\mathbf{D}_2)$

9. Mostre que, se  $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$  e  $f \in \mathcal{L}_2(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , então

$$E[(f - f^*)^2] \geq E[(f - f^+)^2],$$

onde  $*$  =  $*$   $\mathcal{C}$  e  $+$  =  $*$   $\mathcal{B}$ . Dê uma interpretação estatística desse resultado.



## CAPÍTULO 4

# SUFICIÊNCIA DE FISHER NO CASO DOMINADO

### 4.1 INTRODUÇÃO

Iniciamos este capítulo com os conceitos básicos de dominância, os quais serão utilizados ao longo destas notas.

Vamos considerar por um momento duas famílias,  $\mathcal{P}_1$  e  $\mathcal{P}_2$ , de medidas de probabilidade em  $(\mathcal{X}, \mathcal{A})$  e uma medida  $\sigma$ -finita,  $\lambda$ , em  $(\mathcal{X}, \mathcal{A})$ . Defina as seguintes  $\sigma$ -álgebras:

$$\overline{\mathcal{A}}_{01} = \{A \in \mathcal{A} : P(A) = 0 \text{ ou } P(A) = 1 \quad \forall P \in \mathcal{P}_1\}$$

$$\overline{\mathcal{A}}_{02} = \{A \in \mathcal{A} : P(A) = 0 \text{ ou } P(A) = 1 \quad \forall P \in \mathcal{P}_2\}$$

$$\overline{\mathcal{A}}_0^\lambda = \sigma\{A \in \mathcal{A} : \lambda(A) = 0\}$$

#### Definição 4.1

- (i) Dizemos que  $\mathcal{P}_1$  é dominada por  $\mathcal{P}_2$  (e escrevemos  $\mathcal{P}_1 \ll \mathcal{P}_2$ ) se  $\overline{\mathcal{A}}_{02} \subset \overline{\mathcal{A}}_{01}$ .
- (ii) Dizemos que  $\mathcal{P}_1$  é equivalente à  $\mathcal{P}_2$  (e escrevemos  $\mathcal{P}_1 \equiv \mathcal{P}_2$ ) se  $\mathcal{P}_1 \ll \mathcal{P}_2$  e  $\mathcal{P}_2 \ll \mathcal{P}_1$ .
- (iii) A família  $\mathcal{P}_1$  é dita ser, simplesmente, dominada, se existir uma medida  $\sigma$ -finita,  $\lambda$ , tal que  $\overline{\mathcal{A}}_0^\lambda \subset \overline{\mathcal{A}}_{01}$ . Neste caso escreveremos  $\mathcal{P}_1 \ll \lambda$  (analogamente, definiríamos  $\mathcal{P}_1 \equiv \lambda$ ).

(iv) Dizemos que  $\mathcal{P}_1$  é dominada internamente se existir  $P_0 \in \mathcal{P}_1$  tal que  $\mathcal{P}_1 \ll P_0$ .

### Exercício 4.1

Seja  $Y$  uma estatística definida em  $(\mathcal{X}, \mathcal{A})$  com contra-domínio em  $(\mathcal{Y}, \mathcal{B})$ . Considere duas medidas  $\sigma$ -finitas,  $\lambda_1$  e  $\lambda_2$ , em  $(\mathcal{X}, \mathcal{A})$ . Prove que se  $\lambda_1 \ll \lambda_2$ , então

$$\lambda_1 Y^{-1} \ll \lambda_2 Y^{-1} .$$

Neste capítulo consideraremos apenas as famílias  $\mathcal{P}$  que são dominadas. Para simplificar a notação e não abandonar a intuição, vamos supor que existe um conjunto de índices,  $\Theta$  (espaço paramétrico), tal que  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ .

Muitos modelos estatísticos, usados na prática, satisfazem a condição de dominância. Por exemplo, a família das distribuições normais em  $(\mathcal{R}_1, \mathcal{B}_1)$  é dominada pela medida de Lebesgue. Neste caso particular, não é difícil ver que a família é dominada internamente, o que não acontece com a família das distribuições uniformes em  $(0, \theta)$ ,  $\theta \in \mathbb{R}_1^+$ . Na verdade, a família das normais é equivalente à medida de Lebesgue. Por outro lado, a família das uniformes em  $(0, \theta)$ ,  $\theta \in \mathbb{R}_1^+$ , é equivalente à medida de Lebesgue restrita à  $\mathbb{R}_1^+$ . [ $\mathbb{R}_1^+$  representa o conjunto dos números reais não negativos.]

Com a suposição de dominância, os objetivos deste capítulo são:

1. Estabelecer o famoso teorema da Fatoração de Halmos e Savage (1949);
2. Mostrar que sempre existe uma subálgebra suficiente mínima. Especificamente, provar que esta subálgebra é gerada pela função de verossimilhança;
3. Demonstrar que os conceitos de suficiência e suficiência por par são equivalentes.

Os exercícios que apresentamos a seguir têm o objetivo de obrigar o leitor a rever alguns conceitos que serão utilizados ao longo destas notas.

### Exercício 4.2

Sejam  $\mu_1$ ,  $\mu_2$  e  $\mu_0$  três medidas em  $(\mathcal{X}, \mathcal{A})$ . A notação  $\mu_1 \perp \mu_2$  indica que  $\mu_1$  e  $\mu_2$  são mutuamente ortogonais (isto é,  $\mu_1 \perp \mu_2 \Leftrightarrow \exists A \in \mathcal{A}$  tal que  $\mu_1(A) = 0$  e  $\mu_2(A^c) = 0$ ). Prove que:

- (i)  $\mu_1 \perp \mu_2$  e  $\mu_2 \perp \mu_0 \Rightarrow \mu_1 + \mu_2 \perp \mu_0$ .
- (ii)  $\mu_1 \ll \mu_0$  e  $\mu_2 \perp \mu_0 \Rightarrow \mu_1 \perp \mu_2$ .
- (iii)  $\mu_1 \ll \mu_0$  e  $\mu_1 \perp \mu_0 \Rightarrow \mu_1 \equiv 0$ .
- (iv) Se  $\mu_1$  é uma medida de probabilidade, então

$$\mu_1 \ll \mu_0 \Leftrightarrow \lim_{\mu_0(A) \rightarrow 0} \mu_1(A) = 0 .$$

### Exercício 4.3

Mostre que o item (iv) do Exercício 4.2 pode falhar se  $\mu_1$  for apenas  $\sigma$ -finita.

### Exercício 4.4

Seja  $f$  uma variável aleatória definida em  $(\mathcal{X}, \mathcal{A})$  e considere uma medida  $\mu$  em  $\mathcal{A}$ . Defina uma outra medida  $\lambda$  em  $\mathcal{A}$  da seguinte forma:

$$\lambda(A) = \int_A |f| d\mu , \quad \forall A \in \mathcal{A} .$$

Mostre que se  $g$  é uma outra variável aleatória definida em  $(\mathcal{X}, \mathcal{A})$ , então,  $\forall A \in \mathcal{A}$ ,

$$\int_A g d\lambda = \int_A g|f| d\mu ,$$

no sentido de que se uma integral existe a outra também existe e são iguais. Intuitivamente,  $\frac{d\lambda}{d\mu} = |f|$  e assim  $d\lambda = |f| d\mu$ .

Iniciaremos as discussões com o caso mais simples onde a família  $\mathcal{P}$ , de probabilidades em  $(\mathcal{X}, \mathcal{A})$ , é dominada internamente. Isto é,  $\exists \theta_0 \in \Theta$  tal que  $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \ll P_{\theta_0}$ . Para cada  $\theta \in \Theta$ , denotaremos por  $q_\theta$  a derivada de Radon-Nikodym (R-N) de  $P_\theta$  com respeito à  $P_{\theta_0}$ . Isto é,  $q_\theta$  é uma versão de  $\frac{dP_\theta}{dP_{\theta_0}}$ .

### Teorema 4.1

Suponha que  $\mathcal{P}$  seja dominada internamente. Assim, uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  é tal que  $\mathcal{D} \text{ suf}(\mathcal{A}, \mathcal{P})$  se, e somente se,  $\forall \theta \in \Theta$ ,  $q_\theta$  é essencialmente  $\mathcal{D}$ -mensurável.

### Demonstração

- (i)  $\mathcal{D} \text{ suf}(\mathcal{A}, \mathcal{P}) \Rightarrow \forall A \in \mathcal{A}, \exists h \in \mathcal{D}$  tal que

$$(4.1) \quad P_0(A \cap D) = \int_D h dP_\theta$$

$\forall D \in \mathcal{D}$  e  $\theta \in \Theta$ . Como  $q_\theta = dP_\theta/dP_{\theta_0}$ ,  $\forall D \in \mathcal{D}$  e  $\theta \in \Theta$ , temos

$$(4.2) \quad P_\theta(A \cap D) = \int_D I_A dP_\theta = \int_D I_A q_\theta dP_{\theta_0}$$

e por outro lado

$$\begin{aligned} \int_D h dP_\theta &= \int_D h q_\theta dP_{\theta_0} = \int_D E_{P_{\theta_0}}(I_A | \mathcal{D}) q_\theta dP_{\theta_0} \\ &= \int_{\mathcal{X}} E_{\theta_0}(I_A | \mathcal{D}) I_D q_\theta dP_{\theta_0} = \int_{\mathcal{X}} E_{\theta_0}(I_A I_D | \mathcal{D}) q_\theta dP_{\theta_0} . \end{aligned}$$

Com a propriedade de auto-adjunção da esperança condicional podemos escrever a última expressão como

$$(4.3) \quad \int_D h dP_\theta = \int_D I_A E\{q_\theta | \mathcal{D}\} dP_{\theta_0} .$$

De (4.2), (4.2) e (4.3), segue que para cada  $A \in \mathcal{A}$  temos

$$(4.4) \quad \int_D I_A q_\theta dP_{\theta_0} = \int_D I_A E_{P_{\theta_0}}\{q_\theta | \mathcal{D}\} dP_{\theta_0} ,$$

para todo  $D \in \mathcal{D}$  e todo  $\theta \in \Theta$ . Fazendo  $D = \mathcal{X}$ , temos que

$$\int_A q_\theta dP_{\theta_0} = \int_A E_{P_{\theta_0}}\{q_\theta | \mathcal{D}\} dP_{\theta_0} ,$$

para todo  $A \in \mathcal{A}$ . Assim,

$$q_\theta = E_{P_{\theta_0}}(q_\theta | \mathcal{D}) [P_{\theta_0}] \quad \text{para todo } \theta \in \Theta .$$

Como  $P_{\theta_0} \gg \mathcal{P}$ , podemos escrever

$$q_\theta = E_{P_{\theta_0}}(q_\theta | \mathcal{D}) [\mathcal{P}] \quad \forall \theta \in \Theta .$$

Isto é,  $q_\theta$  é  $\mathcal{P}$ -essencialmente  $\mathcal{D}$ -mensurável.

(ii) Para provar a parte suficiente, consideremos que  $q_\theta$  seja  $\mathcal{P}$ -essencialmente  $\mathcal{D}$ -mensurável. Assim,  $q_\theta = q_\theta^* [\mathcal{P}]$ , onde  $*$  indica esperança condicional universal. Em particular,  $q_\theta = E\{q_\theta | \mathcal{D}\} [P_{\theta_0}]$ . Assim, fazendo o caminho inverso temos,  $\forall \theta \in \Theta$  e  $D \in \mathcal{D}$ ,

$$\begin{aligned} \int_D I_A q_\theta dP_{\theta_0} &= \int_D I_A q_\theta^* dP_{\theta_0} = \int_D I_A^* q_\theta dP_{\theta_0} \\ &= \int_D h q_\theta dP_{\theta_0} = \int_D h dP_\theta . \end{aligned}$$

### Corolário 4.1

Suponha que  $\mathcal{P}$  seja autodominada. Seja  $\mathcal{D}_0 = \sigma\{q_\theta : \theta \in \Theta\}$  a menor subálgebra que faz  $q_\theta$  (definidas anteriormente) mensurável,  $\forall \theta \in \Theta$ . Então uma subálgebra  $\mathcal{D}$  é suficiente se e só se  $\mathcal{D}_0 \subset \mathcal{D} [\mathcal{P}]$ . Isto é,

$$\mathcal{D} \text{ suf } (\mathcal{A}, \mathcal{P}) \Leftrightarrow \overline{\mathcal{D}}_0 \subset \overline{\mathcal{D}} .$$

**Prova:** Evidentemente  $\mathcal{D}_0 \subset \mathcal{D} [\mathcal{P}]$  se e só se todo  $q_\theta$  é  $\mathcal{P}$ -essencialmente  $\mathcal{D}$ -mensurável e assim a necessidade está provada pelo Teorema 4.1.

Notemos que  $\mathcal{D}_0$  é suficiente pelo Teorema 4.1. Assim,  $\mathcal{D}_0$  é suficiente mínima.

Consideremos a seguir dois exemplos para um melhor entendimento dos resultados acima.

### Exemplo 4.1

Sejam  $x_1, \dots, x_n$  variáveis aleatórias formando um amostra aleatória simples de uma normal com média  $\theta$  desconhecida e variância igual à unidade. Isto é,

$$x_i \sim N(\theta, 1) , \quad -\infty < \theta < \infty , \quad i = 1, 2, \dots, n .$$

Se  $\mathbb{R}_n$  é o espaço euclidiano  $n$ -dimensional,  $\mathcal{B}_n$  é a  $\sigma$ -álgebra de Borel de subconjuntos de  $\mathbb{R}_n$  e  $\mathcal{P} = \{P_\theta : -\infty < \theta < \infty\}$  é uma família de medidas de probabilidade correspondente a  $(x_1, \dots, x_n)$ . Note que  $\mathcal{P} \ll \lambda$ , onde  $\lambda$  é a medida de Lebesgue  $n$ -dimensional [ver Exercício 4.5].

Mostraremos agora que  $\bar{x} = x_1 + x_2 + \dots + x_n$  é uma estatística suficiente. Fixemos  $\theta_0 \in \Theta$ , onde  $\Theta = \{\theta : \theta \in (-\infty, \infty)\}$ . Notemos que para todo  $\theta \in \Theta$

$$q'_\theta = \frac{dP_\theta}{dP_{\theta_0}} = \exp\{n(\theta - \theta_0)\bar{x} - \frac{n}{2}(\theta^2 - \theta_0^2)\}$$

[veja Exercício 4.5].

Como  $q'_\theta$  é uma função um-a-um de  $\bar{x}$ , segue que  $\mathcal{A}_{\bar{x}}$  é a menor  $\sigma$ -álgebra que torna as funções  $q'_\theta$  mensuráveis. Assim,  $\mathcal{A}_{\bar{x}}$  é a subálgebra suficiente mínima. Isto é,  $\bar{x}$  é a estatística suficiente mínima.

### Exercício 4.5

(i) Mostre que a família de normais multivariadas é dominada pela medida de Lebesgue “multivariada”.

Mostre na verdade que esta família é auto-dominada.

(ii) Mostre que a derivada de  $R - N$ ,  $q'_\theta$  é essencialmente igual a

$$\exp\{n(\theta - \theta_0)\bar{x} - \frac{n}{2}(\theta^2 - \theta_0^2)\} .$$

### Exemplo 4.2

No exemplo anterior, em vez da suposição de Normalidade, considere que cada  $x_i$  é uniformemente distribuída no intervalo  $(0, \theta)$ , onde  $\theta$  é desconhecido e  $\Theta = \{\theta; \theta \in (0, \infty)\}$ . Seja  $\mathcal{X} = (0, \infty)^n$ ,  $\mathcal{A}$   $\sigma$ -álgebra de Borel dos conjuntos de  $\mathcal{X}$  e  $\mathcal{P}_0$  a família de uniformes em  $(0, \theta)$ , onde  $\theta \leq \theta_0$ , isto é,  $\mathcal{P}_\theta = \{U(0, \theta); \theta \in (0, \theta_0)\}$ . Não é difícil ver que  $\mathcal{P}_\theta$  é dominada pela distribuição uniforme em  $(0, \theta_0)$  e que para todo  $\theta \in (0, \theta_0)$

$$q'_\theta = \frac{dP_\theta}{dP_{\theta_0}} = \begin{cases} \left(\frac{\theta_0}{\theta}\right)^n I(\theta - t)' & \text{se } t \leq \theta_0 \\ 0 & \text{se } t > \theta_0 \end{cases} ,$$

onde  $t = \max\{x_1, \dots, x_n\}$  e  $I(y)$  é o indicador de  $[0, \infty)$ ; isto é,  $I(y) = 0$  se  $y < 0$  e  $I(y) = 1$  se  $y \geq 0$ . O Exercício 4.7 mostra que  $\mathcal{A}_t$  é a menor subálgebra que torna as funções  $q'_\theta$  mensuráveis essencialmente. Assim  $t$  é a estatística suficiente mínima.

### Exercício 4.6

Mostre que  $q'_\theta$  é na verdade a derivada de  $RN$  de  $P_\theta$  em relação a  $P_{\theta_0}$  no Exemplo 4.2. Mostre que  $\mathcal{P} = \{U(0, \theta); -\infty < \theta < \infty\}$  não é autodominada embora seja dominada pela medida de Lebesgue.

### Exercício 4.7

Mostre que, no Exemplo 4.2, a subálgebra  $\mathcal{A}_t$  é na verdade a menor subálgebra que torna as funções  $q'_\theta$  essencialmente mensuráveis. Considere agora  $q_\theta = \frac{dP_\theta}{d\lambda}$ , onde  $\lambda$  é a medida de Lebesgue. Mostre que  $\mathcal{A}_t$  é a menor  $\sigma$ -álgebra que torna as funções  $q_\theta$  mensuráveis. [Este problema é não trivial.]

## 4.2 O TEOREMA DE FATORAÇÃO DE HALMOS-SAVAGE

O Teorema de Halmos-Savage é de fundamental importância para a estatística. Na verdade, pode ser considerado como o motivador do famoso “princípio da verossimilhança” que produziu uma revolução na estatística. A seguir muitos conceitos serão apresentados antes de enunciarmos propriamente o famoso teorema da fatoração.

### Definição 4.1

O invólucro convexo de uma família  $\mathcal{P}$  de medidas de probabilidade é uma família maior de tais medidas, representada por  $\tilde{\mathcal{P}}$ , cujos elementos são combinações convexas de membros de  $\mathcal{P}$ . Isto é, se  $\tilde{p} \in \tilde{\mathcal{P}}$ , então existe um conjunto  $\{P_1, \dots, P_i, \dots\}$  e um conjunto de reais positivos  $\{a_1, \dots, a_i, \dots\}$  com  $\sum_{i=1}^{\infty} a_i = 1$  e tais que

$$\tilde{p} = \sum_{i=1}^{\infty} a_i p_i .$$

Os resultados a seguir vão permitir que os resultados obtidos sejam estendidos para famílias dominadas, eliminando-se a restrição de auto-domação.

### Lema 4.1

Seja  $\mathcal{D}$  uma subálgebra. Assim,

$$\mathcal{D} \text{ suf } (\mathcal{A}, \mathcal{P}) \Leftrightarrow \mathcal{D} \text{ suf } (\mathcal{A}, \tilde{\mathcal{P}}) .$$

### Prova

A demonstração no sentido  $\Leftarrow$  é imediata, visto que  $\mathcal{P} \subset \tilde{\mathcal{P}}$ .

Notemos que se  $f \in \mathcal{L}_{\infty}(\mathcal{X}, \mathcal{A}, \mathcal{P})$

$$\int_D f^* d\tilde{p} = \sum_{i=1}^{\infty} a_i \int_D f^* dP_i = \sum_{i=1}^{\infty} a_i \int_D f dP_i = \int_D f d\tilde{p} ,$$

para todo  $D \in \mathcal{D}$  e  $\tilde{p} \in \tilde{\mathcal{P}}$ , onde  $f^*$  é a versão comum de  $E_p(f|\mathcal{D})$ ,  $p \in \mathcal{P}$ .

O nosso objetivo agora é mostrar que toda família dominada é equivalente a uma família enumerável e assim uma família auto-dominada. O próximo resultado é conhecido como “Lema da Exaustão”. Aqui,  $\lambda$  é uma medida  $\sigma$ -finita e  $\mathcal{P}$  é dominada por  $\lambda$ .

**Lema 4.2**

Considere o espaço de medida  $(\mathcal{X}, \mathcal{A}, \lambda)$  e defina a ordem parcial  $A \subset_\lambda B$  se  $\lambda(A \cap B^c) = 0$ . Assim, toda coleção  $\mathcal{F}$  de conjuntos de  $\mathcal{A}$  que é fechada por uniões enumeráveis tem um elemento máximo. Isto é,  $\exists F_0 \in \mathcal{F}$  tal que  $F \subset_\lambda F_0 \quad \forall F \in \mathcal{F}$ .

**Prova**

Sem perda de generalidade, considere que  $\lambda$  é uma medida de probabilidade [veja Exercício 4.8]. Assim, o conjunto  $\{\lambda(F); F \in \mathcal{F}\}$  é limitado e então existe pelo menos um limite superior,  $\alpha$ , deste conjunto, digamos  $\alpha = \sup \lambda(F)$ , onde o sup é sobre  $\mathcal{F}$ . Para cada  $n$ ,  $\exists F_n \in \mathcal{F}$  tal que  $\lambda(F_n) > \alpha - \frac{1}{n}$ . Seja  $F_0 = \bigcup_{n=1}^{\infty} F_n$ , então  $F_0 \in \mathcal{F}$  por hipótese. Assim,

$$\alpha - \frac{1}{n} < \lambda(F_0) \leq \alpha \quad \forall n .$$

Fazendo  $n \rightarrow \infty$ , temos  $\lambda(F_0) = \alpha$ .

Para cada  $F \in \mathcal{F}$ , como  $F \cup F_0 = F_0 \cup (F \cap F_0^c)$ , temos que  $\alpha \geq \lambda(F \cup F_0) = \lambda(F_0) + \lambda(F \cap F_0^c) = \alpha + \lambda(F \cap F_0^c)$ . Assim,  $\lambda(F \cap F_0^c) = 0$  para todo  $F \in \mathcal{F}$ ; isto é,  $F \subset_\lambda F_0, \forall F \in \mathcal{F}$ .

**Exercício 4.8**

Mostre que sempre existe uma medida de probabilidade  $\bar{\lambda}$  tal que  $\bar{\lambda} \equiv \lambda$ , onde  $\lambda$  é uma medida  $\sigma$ -finita em  $(\mathcal{X}, \mathcal{A})$ .

**Definição 4.2**

Para cada  $P_\theta \in \mathcal{P} \ll \lambda$ , seja  $p_\theta$  uma versão de  $\frac{dP_\theta}{d\lambda}$ . Um conjunto  $F \in \mathcal{A}$  é “positivo” se  $\lambda(F) > 0$  e para cada  $P_\theta \in \mathcal{P}$  temos  $p_\theta(x) > 0$   $[\lambda]$  em  $F$ .

O resultado abaixo mostra que os resultados de autodominância também valem para o caso dominado.

**Lema 4.3**

Se  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  é dominada por  $\lambda$  e é fechada por combinações convexas enumeráveis, então  $\mathcal{P}$  é auto-dominada.

**Demonstração**



Seja  $\mathcal{F} = \{F; F \text{ é positivo}\}$ . Notemos que  $\mathcal{F}$  é fechada por uniões enumeráveis. Com efeito, seja  $\{F_n\}$  uma seqüência de elementos de  $\mathcal{F}$ ,  $\{P_n\}$  uma seqüência de medidas correspondentes tais que  $\forall n, P\{A_n\} > 0$  e  $F = \bigcup_{n=1}^{\infty} F_n$ . Considere a medida  $P = \sum_{i=1}^{\infty} a_n P_n$ , onde  $a_n > 0$  e  $\sum_{n=1}^{\infty} a_n = 1$ . Então  $p = \frac{dP}{d\lambda} = \sum_{n=1}^{\infty} a_n p_n > 0$   $[\lambda]$  em  $F$ . Assim,  $F \in \mathcal{F}$ .

Pelo Lema 4.2,  $\mathcal{F}$  tem um elemento máximo, digamos  $F_0$ , com a medida correspondente  $P_0$ . Mostraremos agora que  $P_\theta(F_0^c) = 0 \forall \theta \in \Theta$ . Suponha por absurdo que  $\exists \theta' \in \Theta$  tal que  $P_{\theta'}(F_0^c) > 0$ . Seja  $B = \{x; p_{\theta'}(x) > 0 \text{ e } x \in F_0^c\}$ . Assim,

$$0 < P_{\theta'}(F_0^c) = P_{\theta'}(B) + P_{\theta'}(F_0^c \cap B^c)$$

e

$$P_{\theta'}(F_0^c \cap B^c) = \int_{F_0^c \cap B^c} p_{\theta'} d\lambda \leq 0,$$

pois  $p_{\theta'} \leq 0$  em  $F_0^c \cap B^c$ . Concluimos assim que  $P_{\theta'}(B) > 0$  e  $\lambda(B) > 0$ . Consequentemente,  $B$  é um conjunto positivo de  $F_0^c$ .

Assim,  $F_0 \cup B$  se torna um conjunto positivo maior do que  $F_0$ , o que é uma contradição.

Para finalizar, vamos mostrar que  $\mathcal{P} \ll P_0$ . Seja  $N$  tal que  $P(N) = 0$ . Como  $0 = P_0(N) \geq P_0(N \cap F_0) = \int_{N \cap F_0} p_0 d\lambda$  e  $p_0(x) > 0$   $[\lambda]$  em  $F_0$ , temos que  $\lambda(N \cap F_0) = 0$ . Consequentemente,  $P_\theta(N) = P_\theta(N \cap F_0) + P_\theta(N \cap F_0^c) = 0 \forall \theta \in \Theta$ .  $\square$

De acordo com o Lema 4.3, podemos concluir que se  $\mathcal{P}$  é dominado, então  $\tilde{\mathcal{P}}$  é auto-dominado, pois  $\tilde{\mathcal{P}}$  é, além de dominado, fechado por combinações convexas enumeráveis. Existe assim uma medida  $Q \in \tilde{\mathcal{P}}$  tal que  $\tilde{\mathcal{P}} \ll Q$  e então  $\mathcal{P} \ll Q$ . Como  $Q = \sum_{n=1}^{\infty} a_n P_n$ , onde  $P_n \in \mathcal{P}$  e  $\sum_{n=1}^{\infty} a_n = 1$  com  $a_n > 0$ , podemos ver que  $\{P_1, P_2, \dots\} \equiv \mathcal{P}$ . Demonstramos assim o seguinte teorema.

### Teorema 4.2

Suponha que  $\mathcal{P}$  seja uma família dominada. Existe assim  $\tilde{P} \in \tilde{\mathcal{P}}$  tal que  $\mathcal{P} \ll \tilde{P}$ . Equivalentemente, existe uma subfamília enumerável  $\mathcal{P}_0 = \{P_1, P_2, \dots\}$  tal que  $\mathcal{P} \equiv \mathcal{P}_0$ .

A conseqüência imediata deste teorema é a substituição da condição de auto-dominância pela dominância nos resultados anteriores. Assim, enunciaremos os seguintes resultados.

### Teorema 4.3

Suponha que  $\mathcal{P}$  seja dominada. Uma condição necessária e suficiente para que  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P})$  é que exista uma medida  $\tilde{p} \in \tilde{\mathcal{P}}$  tal que  $\mathcal{P} \ll \tilde{\mathcal{P}}$  e  $\forall P \in \mathcal{P}$ ,  $p = \frac{dP}{d\tilde{p}}$  seja  $\mathcal{P}$ -essencialmente  $\mathcal{D}$ -mensurável.

### Corolário 4.2

Suponha que  $\mathcal{P}$  seja dominada. Como no Teorema 4.3, seja  $p = \frac{dP}{d\tilde{p}} \forall P \in \mathcal{P}$  e defina  $\mathcal{D}_0 = \sigma\{p; \forall P \in \mathcal{P}\}$ . Assim,  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P}) \Leftrightarrow \mathcal{D}_0 \subset \mathcal{D} [\mathcal{P}]$  e  $\mathcal{D}_0$  é a subálgebra suficiente mínima.

### Exercício 4.9

Demonstre o Teorema 4.3 e o Corolário 4.2.

Estamos aptos agora a apresentar o famoso Teorema da Fatoração de Halmos-Savage.

### Teorema 4.4

Uma condição necessária e suficiente para que uma subálgebra  $\mathcal{D}$  seja suficiente é que,  $\forall \theta \in \Theta$ , existe uma função não negativa  $\mathcal{D}$ -mensurável,  $g_\theta$ , e uma função não negativa  $\mathcal{A}$ -mensurável,  $h$ , tal que  $p_\theta = \frac{dP_\theta}{d\lambda} = g_\theta h [\lambda]$ .

### Demonstração

$\mathcal{D}$  suficiente  $\Rightarrow \exists \tilde{p} \in \tilde{\mathcal{P}}$  tal que  $\mathcal{P} \ll \tilde{\mathcal{P}}$  e  $\forall \theta \in \Theta$ ,  $g_\theta = \frac{dP_\theta}{d\tilde{P}}$  seja essencialmente  $\mathcal{D}$ -mensurável. Tomando  $h = \frac{d\tilde{P}}{d\lambda}$ , temos que

$$p_\theta = \frac{dP_\theta}{d\lambda} = \frac{dP_\theta}{d\tilde{P}} \frac{d\tilde{P}}{d\lambda} = g_\theta h [\lambda] .$$

Suponha agora que  $\forall \theta \in \Theta$ ,  $p_\theta = g_\theta h [\lambda]$ . Considere  $\tilde{P} \in \tilde{\mathcal{P}}$  tal que  $\mathcal{P} \ll \tilde{\mathcal{P}}$ . Assim,

$$\frac{d\tilde{P}}{d\lambda} = \sum_{n=1}^{\infty} a_n \frac{dP_{\theta_n}}{d\lambda} = \sum_{n=1}^{\infty} a_n \frac{dP_{\theta_n}}{d\tilde{P}} \frac{d\tilde{P}}{d\lambda} = h \sum_{n=1}^{\infty} a_n g_{\theta_n} = h \tilde{g} .$$

Defina agora,  $\forall \theta \in \Theta$ ,

$$q_\theta = \frac{g_\theta}{\tilde{g}} I_E ,$$

onde  $E = \{x; \tilde{g}(x) > 0\}$ . Evidentemente,  $q_\theta$  é  $\mathcal{D}$ -mensurável  $\forall \theta \in \Theta$ . A suficiência de  $\mathcal{D}$  será verdadeira se provarmos que  $q_\theta$  é uma versão de  $\frac{dP_\theta}{d\tilde{P}}$ ,  $\forall \theta \in \Theta$ . Com efeito,  $\forall A \in \mathcal{A}$ ,

temos que

$$\begin{aligned} \int_A q_\theta d\tilde{P} &= \int_{A \cap E} \frac{g_\theta}{\tilde{g}} d\tilde{P} = \int_{A \cap E} \frac{g_\theta}{\tilde{g}} \tilde{g} h d\lambda \\ &= \int_{A \cap E} g_\theta h d\lambda = \int_{A \cap E} p_\theta d\lambda = P_\theta(A \cap E) = P_\theta(A) . \end{aligned}$$

Para obter a última igualdade, lembremos que  $E^c$  é  $\tilde{P}$  nulo e assim  $P_\theta$  nulo,  $\forall \theta \in \Theta$ .  $\square$

Em linguagem estatística, podemos enunciar o resultado como “uma condição necessária e suficiente para que uma estatística  $T$  seja suficiente é que exista uma função não negativa  $h$  em  $\mathcal{X}$  e um conjunto de funções,  $\{g_\theta; \theta \in \Theta\}$ , não negativas de  $T$  tal que  $p_\theta = g_\theta h [\lambda]$   $\forall \theta \in \Theta$ ”.

Os importantes resultados a seguir são conseqüências imediatas do Teorema da Fatoração.

#### **Teorema 4.5**

Sejam  $\mathcal{C}$  e  $\mathcal{D}$  duas subálgebras de  $\mathcal{A}$  tais que  $\mathcal{C} \subset \mathcal{D} [\lambda]$ , onde  $\lambda$  é a medida  $\sigma$ -finita que domina  $\mathcal{P}$ . Se  $\mathcal{C}$  é suficiente, então  $\mathcal{D}$  também é suficiente.

A versão estatística deste resultado é a seguinte:

Se  $T$  é suficiente e  $\mathcal{P}$ -essencialmente função de outra estatística  $S$ , então  $S$  também é suficiente.

O resultado seguinte mostra que o Teorema 4.5 é o inverso do Teorema 3.1.

#### **Corolário 4.3**

Sejam  $\mathcal{C}$  e  $\mathcal{D}$  duas subálgebras de  $\mathcal{A}$  tais que  $\mathcal{C} \subset \mathcal{D} [\lambda]$ , onde  $\lambda$  é a medida  $\sigma$ -finita que domina  $\mathcal{P}$ , então  $\mathcal{C}$  suf  $(\mathcal{D}, \mathcal{P})$  e  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P})$  se, e somente se,  $\mathcal{C}$  suf  $(\mathcal{A}, \mathcal{P})$ .

#### **Exercício 4.10**

Demonstre o Teorema 4.5 e o Corolário 4.3.

### **4.3 SUFICIÊNCIA VERSUS SUFICIÊNCIA POR PAR**

Considere o modelo  $(\mathcal{X}, \mathcal{A}, \{P_1, P_2\})$ , onde  $\{P_1, P_2\} \subset \mathcal{P}$  [ $\mathcal{P}$  é uma família não necessariamente dominada]. Seja  $p = \frac{dP_1}{dP}$ , onde  $P = \frac{1}{2}(P_1 + P_2)$ . Notemos que  $2 - p$  é uma

versão de  $\frac{dP_2}{dP}$  e que uma subálgebra  $\mathcal{D}$  é suficiente com respeito ao modelo  $(\mathcal{X}, \mathcal{A}, \{P_1, P_2\})$  se e só se  $p \in \{P_1, P_2\}$ -essencialmente  $\mathcal{D}$ -mensurável. Defina  $\mathcal{D}(\{P_1, P_2\}) = \sigma\{p\}$  que é suficiente com respeito a  $(\mathcal{X}, \mathcal{A}, \{P_1, P_2\})$ . Pelo Corolário 4.2, temos que  $\mathcal{D} \text{ suf } \{\mathcal{A}, \{P_1, P_2\}\}$  se e só se  $\mathcal{D}(\{P_1, P_2\}) \subset \mathcal{D} [P_1, P_2]$ . Temos assim o critério para a suficiência por par.

**Lema 4.4**

A subálgebra  $\mathcal{D}$  é suficiente por par com respeito ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se e só se  $\forall \{P_1, P_2\} \subset \mathcal{P}$ , a derivada  $p = \frac{dP_1}{dP}$ , onde  $P = \frac{1}{2}(P_1 + P_2)$ , é  $\{P_1, P_2\}$ -essencialmente  $\mathcal{D}$ -mensurável; isto é,  $\mathcal{D}(\{P_1, P_2\}) \subset \mathcal{D} [P_1, P_2]$ .

Seja  $\mathcal{D}^*$  a menor subálgebra que contém  $\{\mathcal{D}(\{P_1, P_2\}); \forall \{P_1, P_2\} \subset \mathcal{P}\}$ . De acordo com o Lema 4.4,  $\mathcal{D}^*$  é suficiente por par com respeito ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Como  $\mathcal{D}^* \subset \mathcal{D} [P_1, P_2] \forall \{P_1, P_2\} \subset \mathcal{P}$ , então  $\mathcal{D}$  é suficiente por par. Não é verdade em geral que suficiência por par implica em suficiência, como já foi visto anteriormente. Entretanto, os dois conceitos são equivalentes no caso dominado, como pode ser visto nos resultados a seguir. Inicialmente tratamos de famílias enumeráveis de probabilidade.

**Lema 4.5**

Se  $\mathcal{P} = \{P_1, P_2, \dots\}$  é enumerável, então  $\mathcal{D} \text{ suf } (\mathcal{A}, \mathcal{P})$  se, e só se,  $\mathcal{D}$  é suficiente por par em relação a  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

**Demonstração**

Seja  $f \in \mathcal{L}_\infty$  e  $\{P_i, P_j\} \subset \mathcal{P}$ . Como  $\mathcal{D}$  é suficiente por par,  $\exists f_{ij}^* \in \mathcal{D}$  que é versão comum de  $E_{P_i}(f|\mathcal{D})$  e  $E_{P_j}(f|\mathcal{D})$ . (Notemos que  $f_{ij}^* = f_{ji}^* [P_i, P_j]$ .) Para mostrar que  $\mathcal{D} \text{ suf } (\mathcal{A}, \mathcal{P})$ , necessitamos encontrar uma versão universal para  $E_P\{f|\mathcal{D}\} \forall P \in \mathcal{P}$ . Seja  $\varphi_i = \sup_j f_{ij}^*$ ,  $\psi_j = \inf_i f_{ij}^*$  e  $f^* = \inf_i \sup_j f_{ij}^*$ . Para  $i$  fixado, as funções  $f_{ij}^*$ ,  $j = 1, 2, \dots$  são versões de  $E_{P_i}\{f|\mathcal{D}\}$  e conseqüentemente  $P_i$ -equivalentes. Como o supremo de um conjunto enumerável de funções  $P_i$ -equivalentes é também  $P_i$ -equivalente a essas funções, segue que  $\varphi_i$  é também uma versão de  $E_{P_i}(f|\mathcal{D})$ . Assim,  $\psi_j$  é também uma versão de  $E_{P_j}(f|\mathcal{D})$ . Então,  $\psi_n = \varphi_n [P_n]$ ,  $\forall n = 1, 2, \dots$ . Observando que  $\psi_n \leq f^* \leq \varphi_n$ ,  $\forall n$ , podemos concluir que  $f^*$  é a versão universal desejada de  $E_P(f|\mathcal{D})$ .  $\square$

Notemos que a função  $f^0 = \sup_j \inf_i f_{ij}^*$  é também uma versão universal de  $E_P(f|\mathcal{D})$ .

As demonstrações dos resultados que apresentamos a seguir são deixados como exercício e complementam os objetivos do presente capítulo. No próximo capítulo apresentaremos os problemas criados pela falta de dominação da família  $\mathcal{P}$ .

#### **Teorema 4.6**

Se  $\mathcal{P}$  é dominada e  $\mathcal{D}$  é suficiente por par em relação a  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , então  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P})$ .

#### **Lema 4.6**

Se  $\mathcal{P}$  é dominada e  $\mathcal{C} \subset \mathcal{D} [P_1, P_2], \forall \{P_1, P_2\} \subset \mathcal{P}$ , então  $\mathcal{C} \subset \mathcal{D} [\mathcal{P}]$ . Assim, se  $\mathcal{C}$  é suficiente por par,  $\mathcal{D}$  suf  $(\mathcal{A}, \mathcal{P})$ .

#### **Teorema 4.7**

Suponha que  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  é um modelo estatístico arbitrário (não necessariamente dominado). Se  $\mathcal{C}$  é suficiente por par e  $\mathcal{C} \subset \mathcal{D} [\mathcal{P}]$ , então  $\mathcal{D}$  é também suficiente por par.

#### **Exercício 4.11**

Demonstre o Teorema 4.6, o Lema 4.6 e o Teorema 4.7.

## **EXERCÍCIOS COMPLEMENTARES**

1. Considere as v.a.  $Y_1, Y_2$  independentes e identicamente distribuídas exponenciais de parâmetro 1. Uma observação do vetor  $(X_1, X_2) = (\theta_1 Y_1, \theta_2 Y_2)$  é considerada para inferência sobre os parâmetros (de interesse e desconhecidos)  $\theta_1$  e  $\theta_2$ , os quais são números reais satisfazendo  $\theta_1, \theta_2 > 0$ .
  - i) Descreva o modelo estatístico induzido pelo vetor  $(X_1, X_2)$ .
  - ii) Mostre que  $T_1 : \mathbb{R}^2 \rightarrow [0, \infty[$  definida por  $T(x_1, x_2) = |x_1|$  é suficiente para o modelo em (i), no caso de  $\theta_1$  ser conhecido. Conclua também que  $T_2 : \mathbb{R}^2 \rightarrow [0, \infty[$  definida por  $T(x_1, x_2) = |x_2|$  é suficiente para o modelo em (i), no caso de  $\theta_2$  ser conhecido.

- iii) Verifique se  $(T_1, T_2)$  é suficiente para o modelo no caso de ambos,  $\theta_1$  e  $\theta_2$  serem desconhecidos.
2. Considere o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  onde  $\mathcal{P} = \{P_i : i = 0, \dots, k\} \ll \nu$ ,  $\nu$  medida  $\sigma$ -finita. Seja  $p_i = \frac{dP_i}{d\nu}$ , e defina-se o suporte de  $P_i$  pelo conjunto  $\{x : p_i(x) > 0\}$ ,  $i = 0, \dots, k$ . Suponha que as  $P_i$ ,  $i = 0, \dots, k$  têm suporte comum. Mostre que a estatística  $T = (\frac{p_1}{p_0}, \dots, \frac{p_k}{p_0})$  é suficiente minimal em relação ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .
3. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com função de densidade  $f(\cdot, \boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \Theta$ . Encontre a estatística suficiente minimal quando
- i)  $f(x, \theta) = \theta^{-1} I_{(0, \theta)}(x)$ ,  $\theta > 0$ ;
  - ii)  $f(x, \theta) = (2\theta)^{-1} I_{(-\theta, \theta)}(x)$ ,  $\theta > 0$ ;
  - iii)  $f(x, \theta_1, \theta_2) = (\theta_2 - \theta_1)^{-1} I_{(\theta_1, \theta_2)}(x)$ ,  $\theta > 0$ ;
  - iv) em cada um dos casos i) a iii) encontre a distribuição condicional de  $X_1, \dots, X_n$  dada a estatística suficiente minimal.

## CAPÍTULO 5

### O CASO NÃO-DOMINADO

Mesmo que muitos modelos estatísticos satisfaçam a condição de dominação, existem outros que não. Por exemplo, o trabalho estatístico não-paramétrico, especialmente, é rico em famílias de medidas de probabilidade não-dominadas. Parece apropriado, portanto, examinar a noção de suficiência no caso não-dominado.

Começamos estabelecendo uma condição necessária para que uma subálgebra seja suficiente no caso em que  $\mathcal{A}$  é enumeravelmente gerada (ou separável).

#### Teorema 5.1

Suponhamos que  $\mathcal{A}$  é enumeravelmente gerada. Se  $\mathcal{C}$  é uma subálgebra suficiente, então existe uma subálgebra suficiente enumeravelmente gerada  $\mathcal{D}$  tal que  $\mathcal{C} = \mathcal{D} [\mathcal{P}]$ .

#### Demonstração

Observemos que  $\mathcal{A} = \sigma(\mathcal{F})$ , onde  $\mathcal{F}$  é uma coleção enumerável em  $\mathcal{A}$ . Sem perda de generalidade, podemos assumir que  $\mathcal{F}$  é um  $\pi$ -sistema. Para cada  $F \in \mathcal{F}$  existe uma função  $\mathcal{C}$ -mensurável  $h_F$  tal que

$$P(F \cap C) = \int_C h_F dP ,$$

para todo  $C \in \mathcal{C}$  e  $P \in \mathcal{P}$ . Definimos  $\mathcal{D} = \sigma\{h_F; F \in \mathcal{F}\}$ . É claro que  $\mathcal{D}$  é enumeravelmente gerada e que  $\mathcal{D} \subset \mathcal{C}$ .

Seja  $\mathcal{K}$  a classe de conjuntos  $K \in \mathcal{A}$  tais que existem funções  $\mathcal{D}$ -mensuráveis  $g_K$  satisfazendo

$$(5.1) \quad P(K \cap C) = \int_C g_K dP ,$$

para todo  $C \in \mathcal{C}$  e  $P \in \mathcal{P}$ . Então,  $\mathcal{F} \subset \mathcal{K} \subset \mathcal{A}$ . É fácil verificar que  $\mathcal{K}$  é um  $\lambda$ -sistema. Portanto  $\mathcal{K} \supset \lambda(\mathcal{F}) = \sigma(\mathcal{F}) = \mathcal{A}$ , donde  $\mathcal{K} = \mathcal{A}$ . Da definição de  $\mathcal{K}$  e do fato de que  $\mathcal{D} \subset \mathcal{C}$ , segue-se que  $\mathcal{D}$  é suficiente.

Agora demonstraremos que  $\mathcal{C} = \mathcal{D} [\mathcal{P}]$ . É suficiente demonstrar que  $\mathcal{C} \subset \mathcal{D} [\mathcal{P}]$ . Escolhemos e fixamos  $C \in \mathcal{C}$ . Então,  $c \in \mathcal{A} = \mathcal{K}$  e portanto existe uma função  $\mathcal{D}$ -mensurável  $g_C$  satisfazendo (5.1), isto é,  $g_C$  é uma versão  $\mathcal{D}$ -mensurável de  $E_P(I_C|\mathcal{C})$ ,  $P \in \mathcal{P}$ . Mas,  $E_P(I_C|\mathcal{C}) = I_C [\mathcal{P}]$  para toda  $P \in \mathcal{P}$ . Logo,  $g_C = I_C [\mathcal{P}]$ . Fazendo  $D = \{x; g_C(x) = 1\}$ , temos que  $D \in \mathcal{D}$  e  $C = D [\mathcal{P}]$ . Consequentemente,  $\mathcal{C} \subset \mathcal{D} [\mathcal{P}]$ .

Observemos que se o conjunto vazio é o único conjunto  $\mathcal{P}$ -nulo, então  $\overline{\mathcal{D}} = \mathcal{D}$ . Portanto, o seguinte resultado é uma consequência imediata do Teorema 5.1 e o Exercício 1.8(ii).

### Corolário 5.1

Suponhamos que  $\mathcal{A}$  é enumeravelmente gerada e que o conjunto vazio é o único conjunto  $\mathcal{P}$ -nulo. Se  $\mathcal{C}$  é suficiente, então  $\mathcal{C}$  é enumeravelmente gerada.

O resultado seguinte é consequência imediata da aplicação do Teorema 1.1 junto com o Teorema 5.1.

### Corolário 5.2

Suponhamos que  $\mathcal{A}$  é enumeravelmente gerada. Se  $\mathcal{C}$  é uma subálgebra suficiente, então existe uma v.a.  $f$  tal que  $f^{-1}(\mathcal{B}_1) = \mathcal{C} [\mathcal{P}]$ .

### Teorema 5.2

Suponhamos que  $\{\mathcal{C}_n\}$  é uma seqüência de subálgebras suficientes

- (i) se  $\mathcal{C}_n \supset \mathcal{C}_{n+1}$  para todo  $n$ , então  $\bigcap_{n=1}^{\infty} \mathcal{C}_n$  é suficiente,
- (ii) se  $\mathcal{C}_n \subset \mathcal{C}_{n+1}$  para todo  $n$ , então  $\bigvee_{n=1}^{\infty} \mathcal{C}_n$  é suficiente.



### Demonstração

Para provar (i), escolhamos e fixamos uma função  $\mathcal{A}$ -mensurável limitada  $f$ . Para cada  $n$ , existe uma função  $\mathcal{C}_n$ -mensurável  $f_n^*$  tal que  $f_n^* = E_P(f|\mathcal{C}_n) [P]$  para toda  $P \in \mathcal{P}$ . Definimos

$$f^*(x) = \begin{cases} \lim_{n \rightarrow \infty} f_n^*(x) & , \text{ se o limite existe,} \\ 0 & , \text{ em caso contrário.} \end{cases}$$

Desde que  $\mathcal{C}_n \supset \mathcal{C}_{n+1}$  para todo  $n$ , segue-se que  $f^*$  é  $\bigcap_{n=1}^{\infty} \mathcal{C}_n$ -mensurável. Pelo teorema de continuidade para a esperança condicional (por exemplo, ver o teorema 4.3 de Doob, J.K., 1953, Stochastic processes, John Wiley and Sons, New York), temos que  $\lim_{n \rightarrow \infty} f_n^* = E_P(f|\bigcap_{n=1}^{\infty} \mathcal{C}_n) [P]$  para todo  $P \in \mathcal{P}$ . Logo  $f^* = E_P(f|\bigcap_{n=1}^{\infty} \mathcal{C}_n) [P]$  para todo  $P \in \mathcal{P}$  e portanto  $\bigcap_{n=1}^{\infty} \mathcal{C}_n$  é suficiente.

A prova da parte (ii) é similar. □

### Teorema 5.3

Se  $\mathcal{C}$  e  $\mathcal{D}$  são subálgebras suficientes, então  $\mathcal{C} \cap \overline{\mathcal{D}}$  é suficiente.

### Demonstração

Sejam  $\mathcal{F}_{2n-1} = \mathcal{C}$  e  $\mathcal{F}_{2n} = \mathcal{D}$  para todo  $n$ . Escolhamos e fixamos uma função  $\mathcal{A}$ -mensurável limitada  $f$ . Definimos  $\{f_n^*\}$  indutivamente como segue. Seja  $f_1^*$  uma versão comum de  $E_P(f|\mathcal{F}_1)$ ,  $P \in \mathcal{P}$ . Supondo que  $f_{n-1}^*$  tem sido definida, seja  $f_n^*$  uma versão comum de  $E_P(f_{n-1}^*|\mathcal{F}_n)$ ,  $P \in \mathcal{P}$ . Uma tal seqüência  $\{f_n^*\}$  existe, desde que  $\mathcal{F}_n$  é uma subálgebra suficiente para todo  $n$ . Definimos

$$f^*(x) = \begin{cases} \lim_{n \rightarrow \infty} f_{2n-1}^*(x) & , \text{ se o limite existe,} \\ 0 & , \text{ em caso contrário;} \end{cases}$$

$$f^0(x) = \begin{cases} \lim_{n \rightarrow \infty} f_{2n}^*(x) & , \text{ se o limite existe,} \\ 0 & , \text{ em caso contrário.} \end{cases}$$

É claro que  $f^*$  é  $\mathcal{C}$ -mensurável e  $f^0$  é  $\mathcal{D}$ -mensurável. Pelo teorema 3 de Burkholder e Chow (1961) (Iterates of conditional expectation operators. Proc. Amer. Math. Soc. 12, 490-495), segue-se que, para cada  $P \in \mathcal{P}$ ,

$$\lim_{n \rightarrow \infty} f_n^* = E_P(f|\overline{\mathcal{C}}^P \cap \overline{\mathcal{D}}^P) [P] ,$$

onde  $\overline{\mathcal{C}}^P$  e  $\overline{\mathcal{D}}^P$  denotam os  $P$ -completamentos de  $\mathcal{C}$  e  $\mathcal{D}$ , respectivamente. Logo,  $f^* = E_P(f|\overline{\mathcal{C}}^P \cap \overline{\mathcal{D}}^P) [P]$ , para todo  $P \in \mathcal{P}$  e  $\{x; f^*(x) \neq f^0(x)\} \in \overline{\mathcal{A}}_0 = \{A \in \mathcal{A}; P(A) = 0\}$

$\forall P \in \mathcal{P}$  ou  $P(A^c) = 0 \forall P \in \mathcal{P}$ . Portanto,  $f^* - f^0$  é  $\overline{\mathcal{A}}_0$ -mensurável. Então  $f^* = (f^* - f^0) + f$ , sendo a soma de uma função  $\overline{\mathcal{A}}_0$ -mensurável e uma função  $\mathcal{D}$ -mensurável, é  $\overline{\mathcal{D}}$ -mensurável. Segue-se que  $f^*$  é  $\mathcal{C} \cap \overline{\mathcal{D}}$ -mensurável, desde que é  $\mathcal{C}$ -mensurável e  $\overline{\mathcal{D}}$ -mensurável. Mais ainda, para todo  $P \in \mathcal{P}$ , temos

$$\begin{aligned} f^* &= E_P(f^* | \mathcal{C} \cap \overline{\mathcal{D}}) [P] \\ &= E_P[E_P(f | \overline{\mathcal{C}}^P \cap \overline{\mathcal{D}}^P) | \mathcal{C} \cap \overline{\mathcal{D}}] [P] \\ &= E_P(f | \mathcal{C} \cap \overline{\mathcal{D}}) [P], \end{aligned}$$

desde que  $\mathcal{C} \cap \overline{\mathcal{D}} \subset \overline{\mathcal{C}}^P \cap \overline{\mathcal{D}}^P$ . Consequentemente,  $\mathcal{C} \cap \overline{\mathcal{D}}$  é suficiente.  $\square$

### Observações

- É fácil ver que  $\overline{\mathcal{D}} \cap \mathcal{D} = \mathcal{C} \cap \overline{\mathcal{D}} = \overline{\mathcal{C}} \cap \overline{\mathcal{D}} [P]$ . Portanto, pelo Teorema 5.3 juntamente com a Proposição 3.1 (vi), segue-se que se  $\mathcal{C}$  e  $\mathcal{D}$  são subálgebras suficientes, então  $\overline{\mathcal{C}} \cap \mathcal{D}$  e  $\overline{\mathcal{C}} \cap \overline{\mathcal{D}}$  são também suficientes.
- Os Teoremas 5.2 e 5.3 são demonstrados facilmente no caso em que  $\mathcal{P}$  é dominada. Para demonstrar o Teorema 5.3, por exemplo, seja  $\mathcal{C}_0$  uma subálgebra suficiente mínima. Então  $\mathcal{C}_0 \subset \mathcal{C} [P]$  e  $\mathcal{C}_0 \subset \mathcal{D} [P]$  e portanto  $\mathcal{C}_0 \subset \overline{\mathcal{C}} \cap \overline{\mathcal{D}}$ . Consequentemente, pelo Teorema 4.5,  $\overline{\mathcal{C}} \cap \overline{\mathcal{D}}$  é suficiente e  $\overline{\mathcal{C}} \cap \mathcal{D}$  e  $\mathcal{C} \cap \overline{\mathcal{D}}$  são também suficientes.
- A intersecção de duas subálgebras suficientes, em geral, não é necessariamente suficiente, como mostram os seguintes exemplos.

### Exemplo 5.3

Sejam  $\mathcal{X} = R_1$  e  $\mathcal{A} = \mathcal{B}_1$ . Denotemos por  $\mathcal{C}$  a subálgebra gerada pela estatística

$$T(x) = \text{sgn } x = \begin{cases} -1 & , \text{ se } x < 0, \\ +1 & , \text{ se } x \geq 0, \end{cases}$$

isto é,  $\mathcal{C} = \{\emptyset, (-\infty, 0), [0, \infty), \mathcal{X}\}$ . Escolhemos duas funções não-negativas com valores reais  $\varphi$  e  $\psi$  satisfazendo

$$\int_0^\infty \varphi(x) dx = 1 = \int_{-\infty}^c \psi(x) dx .$$

Denotemos por  $\lambda$  a medida de Lebesgue e definimos para cada  $\theta \in (0, 1)$ ,

$$p_\theta = \frac{dP_\theta}{d\lambda} = \begin{cases} \theta\psi(x) & , \text{ se } x < 0, \\ (1 - \theta)\varphi(x) & , \text{ se } x \geq 0. \end{cases}$$

Em conformidade, consideramos a família de medidas de probabilidades  $\mathcal{P} = \{P_\theta; 0 < \theta < 1\}$ , definida por

$$P_\theta(A) = \int_A p_\theta d\lambda, \text{ para todo } A \in \mathcal{A}.$$

Pelo teorema de fatorização a estatística  $T$  é suficiente e portanto  $\mathcal{C}$  é suficiente. Seja  $\mathcal{D} = \{\emptyset, E, E^c, \mathcal{X}\}$ , onde  $E$  é o conjunto de números irracionais negativos. Desde que  $\mathcal{C} = \mathcal{D} [\mathcal{P}]$ , como foi mostrado no Exemplo 1.3, segue-se que  $\mathcal{D}$  é suficiente. Porém,  $\mathcal{C} \cap \mathcal{D} = \{\emptyset, \mathcal{X}\}$  não pode ser suficiente, pela Proposição 3.1 (vii).

#### Exemplo 5.4

Sejam  $\mathcal{X} = R_2$  (o espaço euclidiano bidimensional).  $\mathcal{A} = \mathcal{B}_2$  (a  $\sigma$ -álgebra de subconjuntos borelianos de  $R_2$ ) e  $\mathcal{P}$  a família de todas as medidas de probabilidade  $P$  sobre  $\mathcal{A}$  satisfazendo  $P(D) = 1$ , onde  $D = \{(x_1, x_2) \in \mathcal{X}; x_1 = x_2\}$ . Seja  $\mathcal{C}$  a subálgebra induzida pela estatística  $T(x_1, x_2) = x_1$ , isto é,  $\mathcal{C} = \{B \times R_1; B \in \mathcal{B}_1\}$ . Definimos  $F = \{x_1; (x_1, x_2) \in (A_1 \times A_2) \cap D\}$ , onde  $A_1 \times A_2$  é um conjunto fixado de  $\mathcal{A}$ .

É fácil ver que

$$P[(A_1 \times A_2) \cap (B \times R_1)] = \int_{B \times R_1} I_{F \times \mathbb{R}_1} dP,$$

para todo  $B \times \mathbb{R}_1 \in \mathcal{C}$  e  $P \in \mathcal{P}$ . Portanto,  $\mathcal{C}$  é suficiente.

Similarmente, seja  $\mathcal{D} = \{R_1 \times B; B \in \mathcal{B}_1\}$  a subálgebra induzida por  $S(x_1, x_2) = x_2$ . Então  $\mathcal{D}$  é suficiente. Mas  $\mathcal{C} \cap \mathcal{D} = \{\emptyset, \mathcal{X}\}$  não é suficiente.

#### Corolário 5.3

Se  $\mathcal{C}_0$  é uma subálgebra suficiente minimal, então é suficiente mínima.

#### Demonstração

Seja  $\mathcal{C}$  uma subálgebra suficiente. Pelo Teorema 5.3,  $\mathcal{C}_0 \cap \bar{\mathcal{C}}$  é suficiente. Desde  $\mathcal{C}_0 \cap \bar{\mathcal{C}} \subset \mathcal{C}_0$  e a suficiência minimal de  $\mathcal{C}_0$ , segue-se que  $\mathcal{C}_0 \subset \mathcal{C}_0 \cap \bar{\mathcal{C}} [\mathcal{P}]$ . Logo,  $\mathcal{C}_0 \subset \bar{\mathcal{C}} [\mathcal{P}]$ , isto é,  $\mathcal{C}_0 \subset \bar{\bar{\mathcal{C}}} = \bar{\mathcal{C}}$ . Portanto,  $\mathcal{C}_0 \subset \mathcal{C} [\mathcal{P}]$ . Consequentemente,  $\mathcal{C}_0$  é suficiente mínima.  $\square$

#### Corolário 5.4

Seja  $\{\mathcal{C}_n\}$  uma seqüência de subálgebras suficientes. Então  $\bigcap_{n=1}^{\infty} \bar{\mathcal{C}}_n$  é suficiente.

### Demonstração

Por indução e pelo Teorema 5.3,  $\mathcal{D}_n = \bigcap_{k=1}^n \overline{\mathcal{C}}_k$  é suficiente para cada  $n$ . Desde que  $\mathcal{D}_n \supset \mathcal{D}_{n+1}$ , para todo  $n$ , segue-se do Teorema 5.2 (i) que  $\bigcap_{n=1}^{\infty} \overline{\mathcal{C}}_n = \bigcap_{n=1}^{\infty} \mathcal{D}_n$  é suficiente.  $\square$

### Observações

- A contraparte não enumerável do Corolário 5.4 não é verdadeira em geral, isto é, pode existir uma coleção não-enumerável  $\{C_\alpha; \alpha \in \Lambda\}$  de subálgebras suficientes tal que  $\bigcap_{\alpha \in \Lambda} \overline{\mathcal{C}}_\alpha$  não é suficiente. A existência de uma tal coleção pode ser provada como segue. Suponhamos o contrário. Seja  $\{C_\gamma; \gamma \in \Gamma\}$  a classe de todas as subálgebras suficientes. Então  $C_0 = \bigcap_{\gamma \in \Gamma} \overline{\mathcal{C}}_\gamma$  é suficiente. Mais ainda, se  $C$  é qualquer subálgebra suficiente, então  $C \in \{C_\gamma; \gamma \in \Gamma\}$  e portanto  $C_0 = \bigcap_{\gamma \in \Gamma} \overline{\mathcal{C}}_\gamma \subset \overline{C}$ , isto é,  $C_0 \subset C [\mathcal{P}]$ . Logo,  $C_0$  é suficiente mínima. Consequentemente, sempre existiria uma subálgebra suficiente mínima, contradizendo o fato mostrado no Exemplo 3.2 (Pitcher).
- No caso dominado, o Corolário 5.4 é verdadeiro para uma coleção não-enumerável de subálgebras suficientes, isto é, se  $\mathcal{P}$  é dominada e  $\{C_\alpha; \alpha \in \Lambda\}$  é uma coleção arbitrária de subálgebras suficientes, então  $\bigcap_{\alpha \in \Lambda} \overline{\mathcal{C}}_\alpha$  é suficiente. Para ver isto, seja  $C_0$  a subálgebra suficiente mínima. Então,  $C_0 \subset C_\alpha [\mathcal{P}]$  para todo  $\alpha \in \Lambda$ , isto é,  $C_0 \subset \overline{\mathcal{C}}_\alpha$  para todo  $\alpha \in \Lambda$ . Logo,  $C_0 \subset \bigcap_{\alpha \in \Lambda} \overline{\mathcal{C}}_\alpha$ . Portanto, segue-se do Teorema 4.5 que  $\bigcap_{\alpha \in \Lambda} \overline{\mathcal{C}}_\alpha$  é suficiente.

### Teorema 5.4

Se  $\mathcal{C}$  é uma subálgebra suficiente e  $A \in \mathcal{A}$ , então  $\sigma(\mathcal{C} \cup \{A\})$  é suficiente.

### Demonstração

Observemos primeiro que

$$\sigma(\mathcal{C} \cup \{A\}) = \{(C_1 \cap A) \cup (C_2 \cap A^c) ; C_1 \in \mathcal{C}, C_2 \in \mathcal{C}\} .$$

Como fazemos usualmente, para qualquer função  $\mathcal{A}$ -mensurável  $\mathcal{P}$ -integrável  $f$ , denotamos por  $f^*$  a versão comum de  $E_P(f|\mathcal{C})$ ,  $P \in \mathcal{P}$ . Escolhemos e fixamos  $E \in \mathcal{A}$ .

Definimos

$$g_1(x) = \begin{cases} (I_A \cdot I_E)^*(x) | I_A^*(x) & , \text{ se } I_A^*(x) \neq 0, \\ 0 & , \text{ se } I_A^*(x) = 0 \end{cases}$$

$$g_2(x) = \begin{cases} (I_A - I_E)^*(x) | I_{A^c}^*(x) & , \text{ se } I_{A^c}^*(x) \neq 0, \\ 0 & , \text{ se } I_{A^c}^*(x) = 0 \end{cases} .$$

Claramente,  $g_1$  e  $g_2$  são  $\mathcal{C}$ -mensuráveis. Portanto,  $g = I_A g_1 + I_{A^c} g_2$  é uma função  $\sigma(\mathcal{C} \cup \{A\})$ -mensurável. Observemos que  $I_A g = I_A I_A g_1 + I_A I_{A^c} g_2 = I_A g_1$  e  $I_{A^c} g = I_{A^c} I_A g_1 + I_{A^c} I_{A^c} g_2 = I_{A^c} g_2$ . Desde que  $0 \leq I_A I_E \leq I_A$ , temos que  $0 \leq (I_A I_E)^* \leq I_A^* [\mathcal{P}]$  e portanto  $0 \leq g_1 \leq 1 [\mathcal{P}]$  e  $I_A^* g_1 = (I_A I_E)^* [\mathcal{P}]$ . Resultados similares valem para  $I_{A^c}$  e  $g_2$ . Seja  $D \in \sigma(\mathcal{C} \cup \{A\})$ , isto é,  $D = (C_1 \cap A) \cup (C_2 \cap A^c)$  para alguns  $C_1, C_2 \in \mathcal{C}$ . Então, para todo  $P \in \mathcal{P}$ , temos

$$(i) \quad \begin{aligned} \int_{C_1 \cap A} g dP &= \int_{C_1} I_A g dP = \int_{C_1} I_A g_1 dP \\ &= \int_{C_1} (I_A g_1)^* dP = \int_{C_1} I_A^* g_1 dP = \int_{C_1} (I_A I_E)^* dP \\ &= \int_{C_1} I_A I_E dP = \int_{C_1 \cap A} I_E dP \end{aligned}$$

Similarmente, temos

$$(ii) \quad \int_{C_2 \cap A^c} g dP = \int_{C_2 \cap A^c} I_E dP$$

Somando (i) e (ii), obtemos  $\int_D I_E dP = \int_D g dP$  para todo  $D \in \sigma(\mathcal{C} \cup \{A\})$  e  $P \in \mathcal{P}$ . Consequentemente,  $\sigma(\mathcal{C} \cup \{A\})$  é suficiente.  $\square$

### Teorema 5.5

Se  $\mathcal{C}$  é uma subálgebra suficiente e  $\mathcal{D}$  é uma subálgebra enumeravelmente gerada, então  $\mathcal{C} \vee \mathcal{D}$  é suficiente.

### Demonstração

Seja  $\{A_1, A_2, A_3, \dots\}$  uma coleção enumerável tal que  $\mathcal{D} = \sigma(\{A_1, A_2, A_3, \dots\})$  e seja  $\mathcal{C}_0 = \mathcal{C}$ . Definimos  $\{\mathcal{C}_n\}_{n=0}^\infty$  indutivamente como segue. Supondo que  $\mathcal{C}_{n-1}$  tem sido definida, fazemos  $\mathcal{C}_n = \sigma(\mathcal{C}_{n-1} \cup \{A_n\})$ . Pelo Teorema 5.4, cada  $\mathcal{C}_n$  é uma subálgebra suficiente. É fácil verificar que  $\bigvee_{n=0}^\infty \mathcal{C}_n = \mathcal{C} \vee \mathcal{D}$ . Desde que  $\mathcal{C}_n \subset \mathcal{C}_{n+1}$  para todo  $n \geq 0$ , segue-se do Teorema 5.2 (ii) que  $\mathcal{C} \vee \mathcal{D}$  é suficiente.  $\square$

### Corolário 5.5

Se  $C$  é uma subálgebra enumeravelmente gerada contendo uma subálgebra suficiente, então  $C$  é suficiente.

### Demonstração

Suponhamos que  $\mathcal{D}$  é uma subálgebra suficiente e que  $\mathcal{D} \subset C$ . Segue-se do Teorema 5.5 que  $C = \sigma(C) = \sigma(C \cup \mathcal{D}) = C \vee \mathcal{D} = \mathcal{D} \vee C$  é suficiente.  $\square$

### Teorema 5.6

Suponhamos que  $\mathcal{A}$  é enumeravelmente gerada. Se  $C$  e  $\mathcal{D}$  são subálgebras suficientes, então  $C \vee \mathcal{D}$  é suficiente.

### Demonstração

Pelo Teorema 5.1, existem duas subálgebras suficiente enumeravelmente geradas  $C_1$  e  $\mathcal{D}_1$  tais que  $C = C_1 [\mathcal{P}]$  e  $\mathcal{D} = \mathcal{D}_1 [\mathcal{P}]$ , e portanto  $C \vee \mathcal{D} = C_1 \vee \mathcal{D}_1 [\mathcal{P}]$ . Pelo Teorema 5.5 segue-se que  $C_1 \vee \mathcal{D}_1$  é suficiente. Portanto  $C \vee \mathcal{D}$  também é suficiente.  $\square$

### Corolário 5.6

Suponhamos que  $\mathcal{A}$  é enumeravelmente gerada. Se  $\{\mathcal{C}_n\}$  é uma seqüência de subálgebras suficientes, então  $\bigvee_{n=1}^{\infty} \mathcal{C}_n$  é suficiente.

### Demonstração

Por indução, segue-se do Teorema 5.6 que cada  $\mathcal{D}_n = \bigvee_{k=1}^n \mathcal{C}_k$  é suficiente. Desde que  $\mathcal{D}_n \subset \mathcal{D}_{n+1}$  para todo  $n$ , pelo Teorema 5.2 (ii), temos que  $\bigvee_{n=1}^{\infty} \mathcal{C}_n = \bigvee_{n=1}^{\infty} \mathcal{D}_n$  é suficiente.  $\square$

### Observação

Sejam  $\mathcal{C}$  e  $\mathcal{D}$  duas subálgebras suficientes. Até aqui temos demonstrado que  $\mathcal{C} \vee \mathcal{D}$  é também suficiente, se qualquer uma das seguintes condições é satisfeita:

- (i) a família  $\mathcal{P}$  é dominada (pelo Teorema 4.5),
- (ii)  $\mathcal{C}$  ou  $\mathcal{D}$  é enumeravelmente gerada (pelo Teorema 5.5),
- (iii)  $\mathcal{A}$  é enumeravelmente gerada (pelo Teorema 5.6).

Em geral, porém,  $\mathcal{C} \vee \mathcal{D}$  não é necessariamente suficiente, como mostra o seguinte exemplo.

**Exemplo 5.5**

Seja  $\mathcal{X} = \{(x_1, x_2) \in \mathbb{R}_2; |x_1| = |x_2|, x_1 \neq 0\}$ . Sejam  $\mathcal{C} = \sigma(\{(x_1, x_2), (x_1, -x_2)\}; x_1 = x_2 \neq 0\}$  e  $\mathcal{D} = \sigma(\{(x_1, x_2), (-x_1, x_2)\}; x_1 = x_2 \neq 0\}$ . Sejam  $D = \{(x_1, x_2) \in \mathcal{X}; x_1 = x_2\}$  e  $\mathcal{A} = \sigma[(\mathcal{C} \vee \mathcal{D}) \cup \{D\}]$ . Para cada  $x = (x_1, x_2) \in \mathcal{X}$ , seja  $P_x$  a medida de probabilidade sobre  $\mathcal{A}$  distribuindo probabilidade  $1/4$  sobre cada um dos pontos  $(x_1, x_2)$ ,  $(x_1, -x_2)$ ,  $(-x_1, x_2)$  e  $(-x_1, -x_2)$ . Consideremos  $\mathcal{P} = \{P_x; x \in \mathcal{X}\}$ .

É fácil verificar as seguintes afirmações:

- (i) um conjunto  $C$  está em  $\mathcal{C}$  se e somente se existe um conjunto enumerável  $S \subset D$  tal que  $\bigcup_{(x_1, x_2) \in S} \{(x_1, x_2), (x_1, -x_2)\}$  é igual a  $C$  ou  $C^c$ ;
- (ii) um conjunto  $E$  está em  $\mathcal{D}$  se, e somente se, existe um conjunto enumerável  $S \subset D$  tal que  $\bigcup_{(x_1, x_2) \in S} \{(x_1, x_2), (-x_1, x_2)\}$  é igual a  $E$  ou  $E^c$ ;
- (iii) um conjunto  $B$  está em  $\mathcal{C} \vee \mathcal{D}$  se, e somente se,  $B$  ou  $B^c$  é enumerável;
- (iv) se  $(x_1, x_2) \in \mathcal{X}$ , então  $\{(x_1, x_2)\} \in \mathcal{C} \vee \mathcal{D}$ ;
- (v)  $D \notin \mathcal{C} \vee \mathcal{D}$ .
- (vi)  $\mathcal{A} = \{(B_1 \cap D) \cup (B_2 \cap D^c); B_1, B_2 \in \mathcal{C} \vee \mathcal{D}\}$ ;
- (vii) se  $\mathcal{A}_0 = \{\emptyset, \mathcal{X}\}$ , então  $\overline{\mathcal{A}}_0 = \mathcal{A}_0 = \{\emptyset, \mathcal{X}\}$ .

Para provar que  $\mathcal{C}$  é suficiente, escolhemos e fixamos  $A = (B_1 \cap D) \cup (B_2 \cap D^c) \in \mathcal{A}$ . Definimos  $g_1(x_1, x_2) = I_{B_1 \cap D}(x_1, x_2) + I_{B_1 \cap D}(x_1, -x_2)$ ,  $g_2(x_1, x_2) = I_{B_2 \cap D^c}(x_1, x_2) + I_{B_2 \cap D^c}(x_1, -x_2)$  e  $g = (g_1 + g_2)/2$ . É claro que se  $B_1$  é enumerável, então  $\{(x_1, x_2); g_1(x_1, x_2) \neq 0\}$  é enumerável, e se  $B_2^c$  é enumerável, então  $\{(x_1, x_2); g_2(x_1, x_2) \neq 0\}$  é enumerável. Portanto, desde que  $g_1(x_1, x_2) = g_1(x_1, -x_2)$ , em qualquer caso  $g_1$  é  $\mathcal{C}$ -mensurável. Similarmente,  $g_2$  é  $\mathcal{C}$ -mensurável. Segue-se que  $g$  é  $\mathcal{C}$ -mensurável. Se  $C \in \mathcal{C}$  e  $P \in \mathcal{P}$ , então é claro que

$$\int_C I_A(x_1, x_2) dP = \int_C I_A(x_1, -x_2) dP .$$

Então, para todo  $C \in \mathcal{C}$  e  $P \in \mathcal{P}$ , temos

$$\int_C I_A dP = \frac{1}{2} \int_C I_A(x_1, x_2) dP + \frac{1}{2} \int_C I_A(x_1, -x_2) dP$$



$$\begin{aligned}
&= \int_{\mathcal{C}} \frac{1}{2} [I_A(x_1, x_2) + I_A(x_1, -x_2)] dP \\
&= \int_{\mathcal{C}} g dP .
\end{aligned}$$

Consequentemente,  $\mathcal{C}$  é suficiente. Similarmente, prova-se que  $\mathcal{D}$  é suficiente. Porém,  $\mathcal{C} \vee \mathcal{D}$  não é suficiente, pois em caso contrário existiria uma função  $\mathcal{C} \vee \mathcal{D}$ -mensurável  $h$  satisfazendo

$$\int_B I_D dP = \int_B h dP ,$$

para todo  $B \in \mathcal{C} \vee \mathcal{D}$  e  $P \in \mathcal{P}$ . Escolhendo  $B = \{x\}$  e  $P = P_x$ , temos

$$\frac{1}{4} I_D(x) = \int_{\{x\}} h dP_x = \frac{1}{4} h(x) ,$$

para qualquer  $x \in \mathcal{X}$ . Portanto,  $h = I_D$ . Isto implica que  $D \in \mathcal{C} \vee \mathcal{D}$ , o que é falso.  $\square$

## CAPÍTULO 6

# SUFICIÊNCIA DE FISHER NO CASO DISCRETO. APLICAÇÕES NA TEORIA DE AMOSTRAGEM CLÁSSICA

Neste capítulo estudaremos uma classe especial de modelos não dominados. Será mostrado, ao longo do capítulo, que em tal classe existe sempre uma subálgebra suficiente mínima. Além disto, uma caracterização para tal subálgebra será fornecida.

Através da formalização da teoria de amostragem clássica no contexto da estrutura estatística  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  discutiremos algumas aplicações desses conceitos.

### 6.1 SUFICIÊNCIA NO MODELO DISCRETO

#### Definição 6.1

O modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  é dito discreto se as seguintes condições são satisfeitas:

1. Cada  $P \in \mathcal{P}$  é uma medida de probabilidade discreta;
2. O conjunto vazio é o único conjunto  $\mathcal{P}$ -nulo;
3.  $\mathcal{A}$  é a  $\sigma$ -álgebra de todas as partes de  $\mathcal{X}$ .

## Observações

1. A condição 3 evita que surjam inconveniências, como por exemplo a não existência de uma subálgebra suficiente mínima (Pitcher).
2. A condição 2 não é restritiva, pois sempre é possível remover os conjuntos  $\mathcal{P}$ -nulos de  $\mathcal{X}$  e trabalhar com o espaço amostral reduzido.

## Exercício 6.1

Mostre que se no modelo discreto  $\mathcal{X}$  ou  $\mathcal{P}$  são enumeráveis, então o modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  é dominado. Exiba as medidas dominantes.

Ao longo deste capítulo vamos considerar  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  como o modelo estatístico discreto. Dado que quando  $\mathcal{X}$  ou  $\mathcal{P}$  são enumeráveis o modelo é dominado, assumiremos no que segue que tanto  $\mathcal{X}$  como  $\mathcal{P}$  são não enumeráveis. Os resultados a serem estabelecidos baseiam-se fundamentalmente na caracterização de partições suficientes. Como veremos adiante no modelo discreto existe uma correspondência biunívoca entre subálgebras suficientes e partições suficientes.

## Notação

Para cada  $x \in \mathcal{X}$ ,  $\pi_x$  denota a parte da partição  $\Pi$  que inclui  $x$ . Escreveremos  $P(x)$  para a  $P$ -medida do conjunto  $\{x\}$ .  $S_P = \{x \in \mathcal{X} : P(x) > 0\}$  denota o suporte de  $P$ . Também,  $\mathcal{P}_x = \{P \in \mathcal{P} : P(x) > 0\}$ . Note que a condição 2 na Definição 6.1 implica em  $\mathcal{P}_x \neq \emptyset$ .

O famoso teorema da fatoração de Basu e Ghosh (1967) é provado a seguir.

## Teorema 6.1

Uma condição necessária e suficiente para que a partição  $\Pi$  seja suficiente é que exista uma função real  $g$ , definida sobre  $\mathcal{X}$ , tal que

$$(6.1) \quad P(x) = g(x)P(\pi_x), \quad \text{para todo } x \in \mathcal{X}, \quad P \in \mathcal{P}.$$

### Demonstração

**Necessidade:** Pela Definição 3.5,  $\Pi$  é suficiente se e somente se a  $\sigma$ -álgebra induzida  $\mathcal{A}(\Pi)$  é suficiente. Portanto, para cada  $A \in \mathcal{A}$ , existe uma função  $h_A : \mathcal{A}(\Pi)$  mensurável, tal que

$$(1) \quad P(A \cap B) = \int_B h_A dP, \quad \text{para todo } B \in \mathcal{A}(\Pi) \text{ e } P \in \mathcal{P}.$$

Em particular, se  $A = \{x\}$  e  $B = \pi_x$ , obtemos

$$(2) \quad P(x) = \int_{\Pi_x} h_{\{x\}} dP, \quad \text{para todo } P \in \mathcal{P}.$$

Já que todos os subconjuntos de  $\mathcal{X}$  são mensuráveis, cada parte  $\pi$  de  $\Pi$  é um átomo de  $\mathcal{A}(\Pi)$  e  $h_A$  é constante sobre cada parte  $\pi$  de  $\Pi$  (Exercício complementar 3, Capítulo 2). Definindo  $g(x)$  como o valor constante de  $h_{\{x\}}$  sobre  $\pi_x$ , obtemos a partir de (2) a identidade (6.1).

**Suficiência:** Suponhamos que existe uma função  $g$  sobre  $\mathcal{X}$ , para a qual a identidade (6.1) é verdadeira. Então, pela condição 3 na Definição 6.1, temos que  $g(x) > 0$ , para cada  $x \in \mathcal{X}$ .

Agora, seja  $\pi$  uma parte de  $\Pi$ . A partir da identidade 6.1, temos que para cada  $x \in \pi$

$$(3) \quad P(x) = g(x)P(\pi), \quad \text{para cada } P \in \mathcal{P}.$$

Seja  $P \in \mathcal{P}$ , tal que  $P(x) > 0$  e, portanto, tal que  $P(\pi) > 0$ , já que  $x \in \pi$ . Logo, se  $y \in \pi$ ,  $P(y) > 0$ , pois  $g$  é positiva. Portanto,  $\pi \subset S_P$ . Consequentemente,  $\pi$  é enumerável, pois  $P$  é uma medida de probabilidade discreta.

Assim, somando ambos os lados de (3) sobre todo  $x$  em  $\pi$ , temos que

$$\sum_{x \in \pi} P(x) = \sum_{x \in \pi} g(x)P(\pi),$$

implicando em

$$(4) \quad \sum_{x \in \pi} g(x) = 1.$$

Agora, para cada  $A \in \mathcal{A}$ , defina a função  $h_A$  da seguinte forma:

$$(5) \quad h_A(x) = \sum_{y \in A \cap \Pi_x} g(y).$$

É claro que  $0 \leq h_A \leq 1$ . Também, não é difícil ver que  $h_A$  é constante em cada parte  $\pi$  de  $\Pi$  e então  $\mathcal{A}(\Pi)$  mensurável. Vamos mostrar que para cada  $P \in \mathcal{P}$ ,  $h_A = E_P(I_A | \mathcal{A}(\Pi))$ . Denote por  $h_A(\pi)$  o valor constante de  $h_A$  sobre  $\pi$ . Note que  $h_A(\pi) = \sum_{y \in A \cap \pi} g(y)$ . Assim, se  $B \in \mathcal{A}(\Pi)$ , temos que

$$\begin{aligned}
 \int_B h_A dP &= \sum_{x \in B} h_A(x) P(x) \quad (\text{já que } P \text{ é discreto}) \\
 &= \sum_{\pi \subset B} h_A(\pi) P(\pi) \quad (\text{já que } B \text{ é união de partes de } \Pi) \\
 &= \sum_{\pi \subset B} \sum_{y \in A \cap \pi} g(y) P(\pi) \\
 &= \sum_{\pi \subset B} \sum_{y \in A \cap \pi} P(y) \quad (\text{da hipótese}) \\
 &= \sum_{y \in A \cap B} P(y) = P(A \cap B) .
 \end{aligned}$$

Isto completa a prova. □

### Observação

Note que o Teorema 6.1 também vale se  $\mathcal{X}$  ou  $\mathcal{P}$  é enumerável.

### Exercício 6.2

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  o modelo estatístico discreto e  $F$  uma estatística definida sobre  $\mathcal{X}$ , verifique a veracidade ou falsidade da seguinte afirmação: “ $F$  é uma estatística suficiente se, e somente se, a partição induzida por  $F$  é suficiente”.

### Exercício 6.3

Seja  $\mathcal{X} = \{0, 1\}^n$ ,  $n \in \mathbb{N}$  e  $\mathcal{P}_1 = \{P_\theta : \theta \in (0, 1)\}$  com

$$P_\theta(x_1, \dots, x_n) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} .$$

Seja  $F : \mathcal{X} \rightarrow \{0, 1, \dots, n\}$  tal que  $F(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ . Mostre, utilizando o Teorema 6.1, que a partição induzida por  $F$  é suficiente. Compare, com o Exemplo 3.1.

### Exercício 6.4

Seja  $\mathcal{X} = \{0, 1\}^n$ ,  $n \in \mathbb{N}$  e  $\mathcal{A}$  a  $\sigma$ -álgebra das partes. Uma medida de probabilidade  $P$  definida sobre  $\mathcal{A}$  é dita permutável se para cada  $(x_1, \dots, x_n) \in \mathcal{X}$

$$P(x_1, \dots, x_n) = P(x_{\pi(1)}, \dots, x_{\pi(n)})$$

para toda permutação  $\pi$  de  $\{1, \dots, n\}$ . Seja  $\mathcal{P}_2 = \{P \text{ definida sobre } \mathcal{A} : P \text{ permutável}\}$ . Mostre que

- i) A família  $\mathcal{P}_1$  do Exercício 6.3 está contida na classe de medidas de probabilidade permutáveis definidas sobre  $\mathcal{A}$ .
- ii) Mostre que a estatística  $F$ , definida no Exercício 6.3, induz uma partição suficiente para o modelo estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_2)$ .
- iii) Mostre que a família  $\mathcal{P}_2$  é a maior família para a qual a partição  $\Pi$ , induzida por  $F$ , é suficiente.

### Exercício 6.5

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  modelo estatístico discreto. Duas funções de verossimilhança  $L_x(\cdot)$  e  $L_y(\cdot)$ ,  $x, y \in \mathcal{X}$ , são ditas equivalentes ( $L_x(\cdot) \sim L_y(\cdot)$ ) se existe uma constante  $c > 0$  tal que  $L_x(\cdot) = cL_y(\cdot)$ .

- i) Seja  $\mathcal{R}$  uma relação de  $\mathcal{X}$  em  $\mathcal{X}$  definida por:  $x\mathcal{R}y$  se, e somente se,  $L_x(\cdot) \sim L_y(\cdot)$ . Mostre que  $\mathcal{R}$  é uma relação de equivalência.
- ii) Mostre que as classes de equivalências induzidas por  $\mathcal{R}$  formam uma partição suficiente.
- iii) Mostre que se  $\Pi$  é uma partição suficiente, então para todo  $\pi \in \Pi$  e  $x, y \in \pi$  tem-se  $L_x(\cdot) \sim L_y(\cdot)$ .

No Capítulo 4 (modelo dominado) vimos que se  $\mathcal{C}$  é uma subálgebra suficiente e  $\mathcal{C} \subset \mathcal{D}[\mathcal{P}]$ , então  $\mathcal{D}$  é suficiente. Um resultado similar vale para partições suficientes no modelo discreto.

### Corolário 6.1

Sejam  $\Pi$  e  $\Pi'$  duas partições de  $\mathcal{X}$ . Se  $\Pi'$  é mais fina que  $\Pi$  e  $\Pi$  é suficiente, então  $\Pi'$  é suficiente.

### Demonstração

Sejam  $\pi_x, \pi'_x$  as partes de  $\Pi$  e  $\Pi'$  que incluem  $x$ , respectivamente. Dado que  $\Pi'$  é mais fina que  $\Pi$  temos que  $\pi'_x \subset \pi_x$  para todo  $x \in \mathcal{X}$ . Agora, como  $\Pi$  é suficiente, temos a partir do Teorema 6.1 que

$$\frac{P(x)}{P(\pi'_x)} = \frac{P(x)|P(\pi_x)}{P(\pi'_x)|P(\pi_x)} = \frac{g(x)}{\sum_{y \in \pi'_x} g(y)}, \quad \text{para todo } P \in \mathcal{P}_x.$$

Assim, definindo  $\tilde{g}(x) = \frac{g(x)}{\sum_{y \in \pi'_x} g(y)}$ , temos que  $P(x) = \tilde{g}(x)P(\pi'_x)$  para todo  $P \in \mathcal{P}_x$ . Agora, se  $P \notin \mathcal{P}_x$ , então  $P(\pi_x) = 0$ , pois  $g(x) > 0$ . Portanto  $P(\pi'_x) = 0$  se  $P \notin \mathcal{P}_x$ , pois  $\pi'_x \subset \pi_x$ . Consequentemente,  $P(x) = \tilde{g}(x)P(\pi'_x)$ , para todo  $x \in \mathcal{X}$  e  $P \in \mathcal{P}$ .  $\square$

### Exercício 6.6

Seja  $\mathcal{R}$  uma relação de  $\mathcal{X}$  em  $\mathcal{X}$  definida por:  $x\mathcal{R}y$  se, e somente se,  $\mathcal{P}_x = \mathcal{P}_y$  e  $P(x)/P(y)$  é uma constante em  $P$ , para todo  $P \in \mathcal{P}_x$ . Verifique que  $\mathcal{R}$  é uma relação de equivalência. Compare com o Exercício 6.5.

### Exercício 6.7

Considere o modelo estatístico discreto  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Seja  $\bar{L}_x(\cdot) = \frac{L_x(\cdot)}{\sup_{\theta} L_x(\theta)}$  a função de verossimilhança padronizada. Seja  $F$  uma estatística que leva  $x \mapsto \bar{L}_x(\cdot)$ . Mostre que  $F$  induz uma partição suficiente. Compare com o Exercício 6.5.

Os Exercícios 6.5, 6.6 e 6.7 fornecem uma maneira de achar uma partição suficiente. Parece natural a partir dos resultados anteriores perguntarmos se existe uma partição suficiente  $\Pi^*$ , tal que toda partição suficiente  $\Pi$  seja mais fina que  $\Pi^*$ . Se tal partição  $\Pi^*$  existe, esta é chamada de partição suficiente mínima.

O teorema a seguir garante a existência de uma partição suficiente mínima.

### Teorema 6.2

Existe uma partição suficiente mínima.

### Demonstração

Seja  $\pi_x^* = \{y : y \sim x\}$ , onde  $\sim$  é a relação de equivalência definida por  $x\mathcal{R}y$  se, e somente se,  $\mathcal{P}_x = \mathcal{P}_y$  e  $\frac{P(x)}{P(y)}$  é uma constante em  $P$ , para todo  $P \in \mathcal{P}_x$ . Então segue desta

definição que  $P(\pi_x^*)/P(x)$  é constante em  $P$ . Portanto definindo

$$g(x) = \frac{P(x)}{P(\pi_x^*)},$$

obtemos a identidade

$$P(x) = g(x)P(\pi_x^*)$$

para todo  $P \in \mathcal{P}_x$ .

Agora, se  $P \notin \mathcal{P}_x$ , então  $P(\pi_x^*) = 0$ , pois  $\mathcal{P}_x = \mathcal{P}_y$ . Logo,  $P(x) = g(x)P(\pi_x^*)$ , para todo  $x \in \mathcal{X}$  e  $P \in \mathcal{P}$ . Assim, pelo Teorema 6.1, temos que  $\Pi^* = \{\pi_x^*\}$  é uma partição suficiente.

Por outro lado, se  $\Pi$  é uma partição suficiente, então  $P(x)/P(\pi_x)$  é constante em  $P$ , para todo  $P \in \mathcal{P}_x$  e se  $y \in \pi_x$ , então  $\mathcal{P}_x = \mathcal{P}_y$ , pois  $g$  é positiva. Consequentemente, se  $y \in \pi_x$ , então  $x \sim y$ . Portanto  $\pi_x \subset \pi_x^*$ , concluindo-se a partir disto que  $\Pi^*$  é mais grossa que  $\Pi$ .  $\square$

Note que a prova do Teorema 6.2 fornece uma caracterização da partição suficiente mínima.

### Exercício 6.8

Mostre que as partições suficientes nos exercício 3, 4 e 7 são suficientes mínimas.

### Exercício 6.9

Seja  $\mathcal{X} = \mathbb{R}$  e  $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}\}$  a família de distribuições degeneradas. Encontre a partição suficiente mínima.

### Exercício 6.10

Seja  $\mathcal{X} = \mathbb{R}$  e  $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}^N\}$ , com  $P_\theta$  definida por

$$P_\theta(x) = \frac{1}{N} \sum_{i=1}^N I_{\{\theta_i\}}(x).$$

Encontre a partição suficiente mínima e interprete o resultado.

### Exercício 6.11

Uma urna contém 100 bolas idênticas que estão numeradas consecutivamente com  $\theta + 1, \theta + 2, \dots, \theta + 100$ , onde  $\theta \in \mathbb{Z}$ . Considere o seguinte experimento: dez bolas



são selecionadas ao acaso sem reposição. Formule o modelo estatístico associado a este experimento e encontre a partição suficiente mínima. Refaça o experimento quando a seleção é com reposição. Comente o resultado.

Voltemos agora nossa atenção para subálgebras suficientes. Em geral, não é verdadeiro que toda subálgebra suficiente seja induzida por uma partição suficiente. O próximo resultado estabelece que no modelo discreto toda subálgebra suficiente é assim induzida.

### Teorema 6.3

Toda subálgebra suficiente é induzida por uma partição suficiente.

### Demonstração

Seja  $\mathcal{C}$  uma subálgebra suficiente e seja  $\Pi = \{\pi\}$  a partição induzida por  $\mathcal{C}$ , isto é,  $\Pi = \Pi(\mathcal{C})$ . Vamos mostrar que  $\mathcal{C} = \mathcal{A}(\Pi(\mathcal{C}))$ . Seja  $\pi$  uma parte típica de  $\Pi$  e seja  $h$  a função de probabilidade condicional de  $\pi$  dado  $\mathcal{C}$ . Dado que  $\mathcal{C}$  é suficiente,  $h = E_P(I_\pi/\mathcal{C})$  é constante em  $P$ . Agora,  $h$  é  $\mathcal{C}$ -mensurável, portanto,  $\mathcal{A}(\Pi(\mathcal{C}))$  mensurável, já que  $\mathcal{C} \subset \mathcal{A}(\Pi(\mathcal{C}))$  (Proposição 2.4). Logo,  $h(x)$  é uma constante, digamos  $\alpha$ , sobre  $\pi$ . Seja  $C = h^{-1}(\{\alpha\})$ . Notando que  $C \in \mathcal{C}$  e  $\pi \subset C$ , temos que

$$P(\pi) = P(\pi \cap C) = \int_C h dP = \alpha P(C) ,$$

para todo  $P \in \mathcal{P}$ , com  $\alpha > 0$ , pois  $P(\pi) > 0$  para algum  $P \in \mathcal{P}$ .

Seja  $\pi_1$  outra parte de  $\Pi$ , então  $h(x)$  é uma constante, digamos  $\beta$ , sobre  $\pi_1$ . Vamos mostrar que  $\beta \neq \alpha$ . Suponhamos que  $h(x) = \alpha$  para todo  $x \in \pi_1$ , isto é,  $\pi_1 \subset C = h^{-1}(\{\alpha\})$ . Dado que  $\pi$  e  $\pi_1$  são partes diferentes da partição induzida por  $\mathcal{C}$ , existe  $D \in \mathcal{C}$  que separa  $\pi$  e  $\pi_1$ , isto é,  $\pi_1 \subset D$  e  $\pi \cap D = \emptyset$  (Exercício complementar 5, Cap. 2). Então, para todo  $P \in \mathcal{P}$ , temos que

$$0 = P(\pi \cap C \cap D) = \int_{D \cap C} h dP = \alpha P(D \cap C) \geq \alpha P(\pi_1) ,$$

pois  $\pi_1 \subset C \cap D$ . Como  $\alpha > 0$ , a desigualdade acima implica em  $P(\pi_1) = 0$ , para todo  $P \in \mathcal{P}$ , em contradição com a condição 2, na definição do modelo discreto. Logo,  $\pi = C = h^{-1}(\{\alpha\}) \in \mathcal{C}$ . Consequentemente,  $\Pi(\mathcal{C}) \subset \mathcal{C}$ .

Agora, seja  $D$  uma união arbitrária de partes de  $\Pi$ . Vamos mostrar que  $D \in \mathcal{C}$ . Seja  $g$  a função de probabilidade condicional de  $D$  dado  $\mathcal{C}$ . Já que  $\mathcal{C}$  é suficiente,  $g$  é

constante em  $P$ . Seja  $\pi$  uma parte típica de  $\pi$  e seja  $\gamma$  o valor constante de  $g$  sobre  $\pi$  ( $g$  é  $\mathcal{A}(\Pi(\mathcal{C}))$ -mensurável).

Como  $\pi \in \mathcal{C}$ , temos que, para todo  $P \in \mathcal{P}$ ,

$$\gamma P(\pi) = \int_{\pi} g dP = P(D \cap \pi) = \begin{cases} P(\pi) & \text{se } \pi \subset D \\ 0 & \text{em caso contrário.} \end{cases}$$

Como  $P(\pi) > 0$ , para algum  $P$ , a identidade acima implica em

$$\gamma = \begin{cases} 1 & \text{se } \pi \subset D \\ 0 & \text{em caso contrário.} \end{cases}$$

Isto prova que  $g$  é a função indicadora do conjunto  $D$  e então  $D \in \mathcal{C}$ . Logo,  $\mathcal{A}(\Pi(\mathcal{C})) \subset \mathcal{C}$ . □

Os resultados a seguir são consequência imediata dos Teoremas 6.2 e 6.3.

### Corolário 6.2

Existe uma subálgebra suficiente mínima.

### Demonstração

Seja  $\mathcal{C}$  uma subálgebra suficiente. A partir do Teorema 6.3, temos que  $\mathcal{C}$  é induzida por uma partição suficiente  $\Pi$ , isto é,  $\mathcal{C} = \mathcal{A}(\Pi)$ . Como  $\Pi$  é suficiente, segue-se do Teorema 6.2 que  $\Pi^*$  é mais grossa que  $\Pi$ , onde  $\Pi^*$  é a partição suficiente mínima. Logo,  $\mathcal{A}(\Pi^*) \subset \mathcal{C}$  e portanto  $\mathcal{A}(\Pi^*)$  é uma subálgebra suficiente mínima. □

### Corolário 6.3

Se  $\mathcal{C}$  e  $\mathcal{D}$  são duas subálgebras suficientes, então  $\mathcal{C} \cap \mathcal{D}$  é suficiente.

### Prova

A partir do Teorema 6.3, sabemos que  $\mathcal{C}$  e  $\mathcal{D}$  são induzidas por partições suficientes  $\Pi_1$  e  $\Pi_2$ , respectivamente. Seja  $\Pi_1 \wedge \Pi_2$  a partição mais fina que é mais grossa que ambas  $\Pi_1$  e  $\Pi_2$  (tal partição sempre existe). Não é difícil ver que  $\mathcal{C} \cap \mathcal{D}$  é induzida por  $\Pi_1 \wedge \Pi_2$ . Agora, a partir do Teorema 6.2, temos que existe uma partição suficiente mínima  $\Pi^*$ . Portanto,  $\Pi^*$  é mais grossa que  $\Pi_1 \wedge \Pi_2$ . A prova segue a partir disto e do Corolário 6.1. □

### Obaservação

Note que no modelo não dominado os resultados acima não são verdadeiros em geral.

### Exercício 6.12

Nos Exercícios 6.8 e 6.9, descreva a subálgebra suficiente mínima.

### Exercício 6.13

Mostre que pode existir uma subálgebra  $\mathcal{C}$ , contendo uma partição suficiente, mas  $\mathcal{C}$  não é suficiente.

### Exercício 6.14

Estude a veracidade ou falsidade da seguinte afirmação: “no modelo discreto, se  $\mathcal{D}$  é uma subálgebra suficiente e  $\mathcal{C}$  outra subálgebra tal que  $\mathcal{D} \subset \mathcal{C}$ , então  $\mathcal{C}$  é suficiente”.

### Exercício 6.15

Seja  $\Pi^*$  a partição suficiente mínima e  $\mathcal{F} = \{A \in \mathcal{A} : A = \bigcup_{i \in I} \Pi_i^* \text{ ou } A^c = \bigcup_{i \in I} \Pi_i^*, \text{ para algum conjunto } I \text{ enumerável}\}$ . Mostre que  $\mathcal{F}$  é a  $\sigma$ -álgebra gerada por  $\Pi^*$ . Mostre que  $\mathcal{F}$  não pode ser suficiente.

### Exercício 6.16

Seja  $\Pi^*$  a partição suficiente mínima e  $S_p = \{x : P(x) > 0\}$ . Mostre que  $S_p$  é reunião enumerável de partes de  $\Pi^*$ . Conclua que  $S_p$  é reunião enumerável de partes de  $\Pi$ , onde  $\Pi$  é uma partição suficiente.

### Teorema 6.4

Seja  $\Pi^*$  a partição suficiente mínima e  $\mathcal{F}$  a  $\sigma$ -álgebra gerada por  $\Pi^*$ , então  $\mathcal{F}$  é suficiente por par.

### Demonstração

É suficiente mostrar que para  $P_1, P_2 \in P$  e  $Q = P_1 + P_2$ ,  $\frac{dP_1}{dQ}$  é  $\mathcal{F}$ -mensurável (Lema 4.4). Seja  $A = \{x : P_1(x)P_2(x) > 0\}$  e  $B = \{x : P_1(x) > 0 \text{ e } P_2(x) = 0\}$ . Então, temos

$$\frac{dP_1}{dQ}(x) = \begin{cases} \frac{1}{1+P_2(x)/P_1(x)} & \text{se } x \in A \\ 1 & \text{se } x \in B \\ 0 & \text{em caso contrário} \end{cases} .$$

Já que  $A = S_{P_1} \cap S_{P_2}$  e  $B = S_{P_1} \cap S_{P_2}^c$ , decorre do Exercício 6.16 que  $A, B \in \mathcal{F}$ . Consequentemente,  $(A \cup B)^c \in \mathcal{F}$  e, portanto,  $\{x : \frac{dP_1}{dQ}(x) \geq 1\} \in \mathcal{F}$  e  $\{x : \frac{dP_1}{dQ}(x) \leq 0\} \in \mathcal{F}$ . Resta mostrar então que  $E_\alpha = \{x : 0 < \frac{dP_1}{dQ}(x) \leq \alpha\} \in \mathcal{F}$ , para cada  $0 < \alpha < 1$ . Mostraremos no que segue que  $E_\alpha$  é reunião enumerável de partes de  $\Pi^*$ .

Seja  $\pi^* \in \Pi^*$  tal que  $\pi^* \subset A$ . Segue da definição de  $\pi^*$  que para qualquer par de pontos  $x, y \in \pi^*$ ,  $\frac{P_1(x)}{p_1(y)} = \frac{P_2(x)}{p_2(y)}$  ou  $\frac{P_2(x)}{P_1(x)} = \frac{P_2(y)}{P_1(y)}$ . Isto implica que  $\frac{P_2(x)}{P_1(x)}$  é constante sobre cada  $\pi^* \in \Pi^*$ . Então  $E_\alpha$  é reunião de partes de  $\Pi^*$ , mas  $E_\alpha \subset A$  e  $A$  é reunião enumerável de partes de  $\Pi^*$ , portanto  $E_\alpha$  também é reunião enumerável de partes de  $\Pi^*$ . Assim,  $E_\alpha \in \mathcal{F}$  para cada  $0 < \alpha < 1$ . Consequentemente,  $\frac{dP_1}{dQ}$  é  $\mathcal{F}$ -mensurável e portanto  $\mathcal{F}$  é suficiente por par.  $\square$

### Observação

A partir do Exercício 6.15 e do Teorema 6.4, conclui-se que suficiêcia por par e suficiêcia não são noções equivalentes no modelo discreto (não dominado).

### Exercício 6.17

Nos Exercícios 6.8 e 6.9, descreva a subálgebra suficiente por par.

### Corolário 6.4

Seja  $\Pi$  uma partição suficiente e  $\mathcal{C}$  a subálgebra gerada por  $\Pi$ . Então  $\mathcal{C}$  é suficiente por par.

### Demonstração

É suficiente mostrar que  $\mathcal{F} \subset \mathcal{C}$  (Teorema 4.7). Para ver isto notemos primeiro que cada parte  $\pi^*$  de  $\Pi^*$  é enumerável, pois  $\pi^* \subset S_P$ , para algum  $P \in \mathcal{P}$ . Por outro lado,  $\Pi^*$  é mais grossa que  $\Pi$ , pois  $\Pi^*$  é a partição suficiente mínima. Portanto  $\Pi^*$  é uma reunião enumerável de partes de  $\Pi$ . Assim, cada  $F \in \mathcal{F}$  ou seu complementar é uma reunião enumerável de partes de  $\Pi$ . Logo,  $\mathcal{F} \subset \mathcal{C}$ .  $\square$

Para completarmos o estudo sobre suficiêcia no modelo discreto, um teorema da fatoração para subálgebras, análogo ao Teorema 6.1, será provado. O lema a seguir é de utilidade para tal propósito.

**Lema 6.1**

Se  $\mathcal{C}$  é uma subálgebra contendo  $S_P$  para todo  $P \in \mathcal{P}$ , então  $\Pi(\mathcal{C}) \subset \mathcal{C}$ .

**Demonstração**

Seja  $\pi \in \Pi(\mathcal{C})$ . Vamos mostrar primeiro que  $\pi \subset S_P$  para algum  $P \in \mathcal{P}$ . Para ver isto, escolha e fixe  $x \in \pi$ . A partir da condição 3 na definição do modelo discreto, segue que existe algum  $P \in \mathcal{P}$  tal que  $x \in S_P$ . Dado que  $S_P \in \mathcal{C}$  e  $\pi = \bigcap \{C \in \mathcal{C} : x \in C\}$ , temos que  $\pi$  está contido em  $S_P$ . Assim, cada parte de  $\pi(\mathcal{C})$  é disjunta de  $S_P$  ou está inteiramente contida em  $S_P$ , para algum  $P \in \mathcal{P}$ . Isto mostra que  $S_P$  deve ser reunião enumerável de partes de  $\pi(\mathcal{C})$ , pois  $S_P$  é enumerável.

Para mostrar que  $\pi \in \mathcal{C}$ , escreva  $S_P = \pi \cup (\bigcup_{n=1}^{\infty} \pi_n)$ , onde  $\pi_n \in \Pi(\mathcal{C})$  para todo  $n$ . Agora, para cada  $n$ , existe um conjunto  $C_n \in \mathcal{C}$  tal que  $\pi \subset C_n$  e  $C_n \cap \pi_n = \emptyset$ . Seja  $C = \bigcap_{n=1}^{\infty} C_n$ , então  $\pi \subset C$  e  $C \cap \pi_n = \emptyset$  para todo  $n$ . Assim,

$$\begin{aligned} C \cap S_P &= C \cap \left[ \pi \cup \left( \bigcup_{n=1}^{\infty} \pi_n \right) \right] \\ &= (C \cap \pi) \cup \left( \bigcup_{n=1}^{\infty} (C \cap \pi_n) \right) \\ &= \pi \cup \emptyset = \pi . \end{aligned}$$

Logo,  $\pi \in \mathcal{C}$ , isto é,  $\Pi(\mathcal{C}) \subset \mathcal{C}$ . □

**Exercício 6.18**

Seja  $\mathcal{C}$  uma subálgebra contendo  $S_P$  para todo  $P \in \mathcal{P}$  e  $\Pi(\mathcal{C})$  uma partição suficiente. Mostre que  $\mathcal{C}$  é suficiente por par.

Sugestão: mostre que  $\Pi^* = \Pi(\mathcal{F})$ . Utilize o Exercício 6.16 e o Lema 6.1 para concluir que  $\mathcal{F} \subset \mathcal{C}$ .

**Teorema 6.5**

Seja  $\mathcal{C}$  uma subálgebra contendo  $\Pi(\mathcal{C})$ , com  $\Pi(\mathcal{C})$  suficiente. Então, para cada  $P \in \mathcal{P}$ , existe uma função não negativa  $\mathcal{C}$ -mensurável  $h_P$  e uma função  $g$  não negativa, independente de  $P$  tal que

$$(6.2) \quad P(x) = h_P(x)g(x)$$

para todo  $x \in \mathcal{X}$ . Reciprocamente, se cada  $P \in \mathcal{P}$  é fatorado como em (6.2), então  $\Pi(\mathcal{C})$  está contido em  $\mathcal{C}$  e  $\Pi(\mathcal{C})$  é suficiente.

### Demonstração

Para provar a primeira parte do teorema, escolhamos  $P \in \mathcal{P}$ . Como  $\Pi(\mathcal{C})$  é suficiente, temos a partir do Teorema 6.1 que existe  $g$  tal que  $P(x) = g(x)P(\pi_x)$  para todo  $x \in \mathcal{X}$  e  $P \in \mathcal{P}$ , onde  $\pi_x$  é a parte de  $\Pi(\mathcal{C})$  que inclui  $x$ . Definindo  $h_P(x) = P(\pi_x)$ ,  $x \in \mathcal{X}$ , e notando que  $S_p$  é reunião enumerável de partes de  $\Pi(\mathcal{C})$ , conclui-se a primeira parte da prova. De fato,  $h_p(x) = 0$  se  $x \in S_p^c$ , pois  $g(x) > 0$ . Portanto  $h_p$  é mensurável com respeito a qualquer subálgebra contendo  $S_p$ , em particular, com respeito a  $\mathcal{C}$ , pois  $\Pi(\mathcal{C}) \subset \mathcal{C}$  e  $S_p$  é reunião enumerável de partes de  $\Pi(\mathcal{C})$ .

A segunda parte do teorema é provada como segue. Dado que  $\mathcal{C} \subset \mathcal{A}(\Pi(\mathcal{C}))$  (Proposição 2.4) e que cada parte de  $\Pi(\mathcal{C})$  é um átomo de  $\mathcal{A}(\Pi(\mathcal{C}))$ , temos que  $h_P$  é constante em cada parte de  $\Pi(\mathcal{C})$ . Seja  $\pi \in \Pi(\mathcal{C})$ . A partir da condição 2, na Definição 6.1, temos que  $g(x) > 0$  para todo  $x \in \mathcal{X}$  e  $P(\pi) > 0$  para algum  $P \in \mathcal{P}$ . Para tal  $P$  temos, somando a ambos os lados da identidade 6.2, que

$$\sum_{x \in \pi} P(x) = h_P(x) \sum_{x \in \pi} g(x) = \alpha \sum_{x \in \pi} g(x),$$

onde  $\alpha$  é o valor constante de  $h_P$  sobre  $\pi$ . Portanto,

$$\alpha = \frac{P(\pi)}{\sum_{x \in \pi} g(x)}.$$

Fazendo  $g_1(x) = \frac{g(x)}{\sum_{y \in \pi} g(y)}$ , temos que  $P(x) = g_1(x)P(\pi)$ . Logo,  $\Pi(\mathcal{C})$  é suficiente. Agora, a partir da identidade 6.2, temos que  $S_P = \{x : P(x) > 0\} = \{x : h_p(x) > 0\} \in \mathcal{C}$ , para todo  $P \in \mathcal{P}$ , pois  $h_P$  é  $\mathcal{C}$ -mensurável. Consequentemente, pelo Lema 6.1,  $\Pi(\mathcal{C}) \subset \mathcal{C}$ . □

## 6.2 O PAPEL DA SUFICIÊNCIA NA TEORIA DE AMOSTRAGEM CLÁSSICA

As principais características dos procedimentos utilizados na amostragem clássica podem ser resumidas como segue:

- i) existe uma população finita de  $N$  unidades, cujos membros são rotulados pelos inteiros  $1, 2, \dots, N$ . Tal população é denotada usualmente por  $\mathcal{P} = \{1, 2, \dots, N\}$ ;
- ii) associada à  $i$ -ésima unidade em  $\mathcal{P}$ , existe uma característica (ou vetor de características)  $X_i$  desconhecida. O estado desconhecido da natureza é  $\boldsymbol{\theta} = (X_1, \dots, X_N)$  e o pesquisador conhece o conjunto  $\Theta$  de possíveis estados da natureza;
- iii) o problema de amostragem surge quando o estatístico planeja ganhar informação a respeito de alguma quantidade populacional  $\tau = \tau(\boldsymbol{\theta})$  observando as características de um subconjunto  $s = \{i_1, i_2, \dots, i_n\}$  ( $n < N$ ) de unidades selecionadas de  $\mathcal{P}$  (para alguns tipos de planejamento, por exemplo, amostragem com reposição, é conveniente pensar  $s$  como seqüência);
- iv) o processo de seleção é definido por uma função de probabilidade  $p$  definida sobre  $\mathcal{J} = \{s : s \subset \mathcal{P}\}$ . O par  $(\mathcal{J}, p)$  é chamado de planejamento amostral. A escolha do plano usualmente depende de considerações de custos e um certo conhecimento a priori  $C$  (por exemplo, informação de variáveis auxiliares de cada uma das unidades em  $\mathcal{P}$ );
- v) o dado  $d = \{(i, x_i) : i \in s\}$ , onde  $x_i$  é a característica observada na  $i$ -ésima unidade da população, é obtido após a execução do planejamento amostral. O dado  $d$  é utilizado para se fazer inferências a respeito de alguma quantidade populacional  $\tau = \tau(\boldsymbol{\theta})$ ;
- vi) as inferências estão ligadas ao planejamento amostral, referindo-se fundamentalmente à variância e vício de estimadores (funções de  $d$ ) sob o plano amostral  $(\mathcal{J}, p)$ .

Até aqui pode ser notado que a teoria de amostragem clássica considera o método utilizado para selecionar a amostra como a principal fonte de aleatorização que define a estrutura estocástica para inferência. Sob esta abordagem, raramente são mencionados conceitos de inferência estatística como suficiência, verossimilhança, completividade etc. (veja por exemplo “Sampling techniques” de W.G. Cochran), embora exista um espaço paramétrico  $\Theta$  bem definido e uma fonte de aleatorização sobre o espaço  $\mathcal{X}$  de possíveis dados  $d$ .

Basu (1967, 1971) foi um dos primeiros a formalizar a teoria de amostragem clássica através da estrutura estatística  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , notando que aleatorização torna possível considerar o conjunto  $s$  e então o dado  $d$  com um elemento aleatório. Seguindo Basu (1967), o espaço amostral  $\mathcal{X}$  induzido pelas características da amostragem clássica é definido por  $\mathcal{X} = \{(i, x_i) : i \in s\}; s \in \mathcal{S}, (x_1, \dots, x_N) \in \Theta\}$ . Notemos que cada  $d \in \mathcal{X}$  informa exatamente algumas das  $X$ -coordenadas do vetor  $\theta$ . Seja  $\Theta_d$  o conjunto de pontos no espaço paramétrico consistentes com  $d$ , isto é,  $\Theta_d = \{(X_1, X_2, \dots, X_N) \in \Theta : X_i = x_i, \text{ para cada } i \in \mathcal{J}\}$ . Logo, se  $\theta \in \Theta_d$ , a probabilidade de se obter o dado  $d$  é exatamente a probabilidade de seleção do subconjunto  $s$ , i.é

$$P_{\theta}(d) = \begin{cases} p(s/\theta) & \text{se } \theta \in \Theta_d \\ 0 & \text{se } \theta \notin \Theta_d . \end{cases}$$

Note que escrevemos  $p(s/\theta)$  para denotar a probabilidade de seleção do subconjunto  $s \in \mathcal{S}$ . De fato, o plano de seleção poderia depender de  $\theta$ . Contudo, em situações típicas de amostragem  $p(s/\theta)$  é constante em  $\theta$ , para cada  $\theta \in \Theta$ . Tendo isto em consideração, podemos escrever

$$P_{\theta}(d) = \begin{cases} p(s) & \text{se } \theta \in \Theta_d \\ 0 & \text{se } \theta \notin \Theta_d . \end{cases}$$

Isto leva à seguinte caracterização de um modelo de amostragem.

### Definição 6.2

O modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , com  $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$  é chamado modelo de amostragem, se o modelo é discreto e se  $P_{\theta}(d)$  é uma constante para cada  $\theta \in \Theta_d$ , onde

$$\Theta_d = \{\theta \in \Theta : P_{\theta}(d) > 0\} .$$

Note que o modelo de amostragem não expressa nenhuma relação entre as unidades amostradas e não amostradas a não ser pelo fato de que estas últimas poderiam ter sido selecionadas com probabilidade positiva. A função de verossimilhança é não informativa, pois todas as componentes não observadas de  $\theta$  têm a mesma verossimilhança. Por exemplo, suponha que  $N = 4$  e que uma amostra aleatória simples de tamanho 1 é selecionada. Suponha também que a unidade 1 foi selecionada com valor observado  $X_1 = 10$ . Neste caso, a função de verossimilhança é dada por

$$L(\theta|\{(1, 10)\}) = \begin{cases} \frac{1}{4} & \text{se } \theta \in \Theta_{\{(1,10)\}} \\ 0 & \text{se } \theta \notin \Theta_{\{(1,10)\}} , \end{cases}$$



onde  $\Theta_{\{(1,10)\}} = \{(X_1, X_2, X_3, X_4) \in \mathbb{R}^4 : X_1 = 10\}$ . Note que a função de verossimilhança é constante e positiva para cada  $\theta \in \Theta_{\{(1,10)\}}$  (homeomorfo a  $\mathbb{R}^3$ ). Uma das conseqüências deste fato é que o estimador de máxima verossimilhança não é único.

### Exercício 6.19

Considere uma população finita  $\mathcal{P} = \{1, 2, \dots, N\}$ . Seja  $X_i$  a característica associada (real valorada) à  $i$ -ésima unidade populacional. Suponha que para se obter informação sobre uma quantidade populacional  $\tau = \tau(\theta)$  o seguinte planejamento amostral é levado a efeito: “selecione a unidade 1 e observe  $X_1$ . Se  $X_1$  é maior que  $b$  (conhecido), então escolha a unidade  $N$ ”. Para este planejamento explicita a estrutura estatística  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  no sentido da Definição 6.2.

### Exercício 6.20

Considere o modelo de amostragem da Definição 6.2. Um estimador não viciado de  $\tau(\theta)$  é uma estatística  $F$  que satisfaz a identidade

$$E_{\theta}(F) = \mathcal{C}(\theta), \quad \text{para cada } \theta \in \Theta.$$

Seja  $T_{\mathbf{b}} = \sum_{i=1}^N b_i I_i X_i$ , onde  $\mathbf{b} \in \mathbb{R}^N$  e

$$I_i = \begin{cases} 1 & \text{se } i \in s \\ 0 & \text{se } i \notin s \end{cases}.$$

- i) Encontre as expressões para  $E_{\theta}(T_{\mathbf{b}})$  e  $\text{Var}_{\theta}(T_{\mathbf{b}})$ .
- ii) Encontre expressões para  $b_i$  de tal forma que  $T_{\mathbf{b}}$  seja um estimador não viciado de  $\tau(\theta) = \sum_{i=1}^N X_i$ .
- iii) Seja  $\theta_0 = (X_{01}, \dots, X_{0N})$  um ponto arbitrário em  $\Theta = \mathbb{R}^N$  e  $T_0 = \sum_{i=1}^N \Pi_i^{-1} I_i (X_i - X_{i0}) + \sum_{i=1}^N X_{i0}$ , onde  $\Pi_i = P_{\theta_0}(I_i = 1)$ . Mostre que  $T_0$  é um estimador não viciado de  $\tau(\theta) = \sum_{i=1}^N X_i$  e calcule  $\text{Var}_{\theta_0}(T_0)$ . Compare com  $\text{Var}_{\theta_0}(T_{\mathbf{b}})$ , onde  $T_{\mathbf{b}}$  é o estimador encontrado em ii). Comente o resultado.

### Teorema 6.6

Se  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  com  $\mathcal{P} = \{P_{\theta} : \theta \in \Theta\}$  é um modelo de amostragem, então a partição suficiente mínima é a partição induzida pela função que leva  $d \mapsto \Theta_d$ .

**Prova**

Segue da Definição 6.2 e do Teorema 6.2.

□

### Observação

Note que se  $d, d' \in \mathcal{X}$ , então  $\Theta_d = \Theta_{d'}$  se, e somente se,  $d = d'$ . Logo,  $\Pi^* = \{\{d\} : d \in \mathcal{X}\}$ . Portanto, a partição induzida pela função  $F$  definida por  $F(d) = \{x_i : i \in s\}$  não é suficiente. De fato, sem a informação dos rótulos não é possível especificar a quais coordenadas de  $\theta$  correspondem os  $n$  valores observados.

O leitor terá oportunidade de mostrar no próximo capítulo que a estatística (partição) suficiente mínima no Teorema 6.6 não é completa. Isto tem importantes conseqüências com relação à existência de estimadores não viciados de variância uniformemente mínima (Basu, 1971).

### EXERCÍCIOS COMPLEMENTARES

1. Considere uma população formada por 100 elementos. Seja  $\theta = (Y_1, Y_2, \dots, Y_{100})$  o vetor de características populacionais e  $\tau(\theta) = \sum_{i=1}^{100} Y_i$  a quantidade populacional de interesse. Suponha que sabemos que existe  $i_0 \in \{1, 2, \dots, 100\}$  tal que  $Y_{i_0}$  é da ordem  $10^{10}$  (o valor de  $i_0$  é desconhecido) e que  $0 \leq Y_i \leq 1$  se  $i \neq i_0$ .

Com o objetivo de obter informação a respeito de  $\tau(\theta)$ , uma amostra de tamanho  $n$ ,  $s = \{i_1, \dots, i_n\}$ , é selecionada ao acaso e sem reposição da população. Suponha que você deve escolher entre as duas seguintes estatísticas para estimar  $\tau(\theta)$ .

$$T_1(d) = 100 \frac{\sum_{i \in s} Y_i}{n},$$

e

$$T_2(d) = \begin{cases} Y_{i_0} + 99 \frac{\sum_{\substack{i \in s \\ i \neq i_0}} Y_i}{n-1} & \text{se } i_0 \in s \\ 10^{10} + 99 \frac{\sum_{i \in s} Y_i}{n} & \text{se } i_0 \notin s \end{cases}$$

- a) Determine  $E_{\theta}(T_i(d))$  e  $E_{\theta}(T_i(d) - \tau(\theta))^2$ ,  $i = 1, 2$ .
  - b) Qual dos dois estimadores parece mais apropriado? Comente.
  - c) Suponha que você pode escolher entre uma amostra que contenha  $i_0$  e outra que não a contenha. Qual das duas amostras você escolheria e porquê?
2. Suponha que uma máquina produza  $N$  itens em um particular dia. Seja  $Y_i$  o indicador de item defeituoso, isto é,  $Y_i = 1$  se o  $i$ -ésimo artigo produzido pela máquina

é defeituoso e  $Y_i = 0$  em caso contrário. Sabe-se que a partir do momento em que a máquina produz o primeiro item defeituoso, ela continuará produzindo só itens defeituosos. No final do dia interessa estimar o número total de itens defeituosos produzidos pela máquina. Se  $\boldsymbol{\theta} = (Y_1, Y_2, \dots, Y_N)$ , então  $\tau(\boldsymbol{\theta}) = \sum_{i=1}^N Y_i = N - \max\{i \in \{1, \dots, N\} : Y_i = 0\}$ .

- i) Mostre que existe uma correspondência biunívoca entre  $\tau(\boldsymbol{\theta})$  e  $\boldsymbol{\theta}$ .
- ii) Encontre o estimador (es) de máxima verossimilhança de  $\tau(\boldsymbol{\theta})$  baseado numa amostra de tamanho  $n$ , selecionada mediante um planejamento amostral  $p$ , qualquer. Ilustre o resultado com  $N = 100$  e uma amostra de tamanho 4 que forneceu os seguintes dados:  $d = \{(17, 0), (24, 0), (40, 1), (73, 1)\}$ .
- iii) Considere o seguinte planejamento amostral: selecione a unidade  $\left[\frac{N}{2}\right]$  (onde  $[\cdot]$  denota a função maior inteiro). Se  $Y_{\left[\frac{N}{2}\right]} = 1$  selecione a unidade  $\left[\frac{N}{4}\right]$ , em caso contrário ( $Y_{\left[\frac{N}{2}\right]} = 0$ ) selecione a unidade  $\left[3\frac{N}{4}\right]$ . Se a unidade  $\left[\frac{N}{4}\right]$  foi selecionada e  $Y_{\left[\frac{N}{4}\right]} = 1$ , então selecione a unidade  $\left[\frac{N}{8}\right]$  em caso contrário ( $Y_{\left[\frac{N}{4}\right]} = 0$ ) selecione a unidade  $\left[3\frac{N}{8}\right]$ . Se  $Y_{\left[3\frac{N}{4}\right]}$  foi selecionada e  $Y_{\left[3\frac{N}{4}\right]} = 1$  selecione a unidade  $\left[5\frac{N}{8}\right]$ , em caso contrário ( $Y_{\left[3\frac{N}{4}\right]} = 0$ ) selecione a unidade  $\left[7\frac{N}{8}\right]$ , e assim sucessivamente. Formule o modelo estatístico associado a este procedimento e encontre o estimador (es) de máxima verossimilhança de  $\tau(\boldsymbol{\theta})$ . Ilustre o procedimento com  $N = 100$  e compare com os resultados obtidos em ii).

## REFERÊNCIAS

- Basu, D. and Ghosh, J.K. (1967). Sufficient statistics from a finite universe. *Bull. Int. Stat. Inst.*, 42, BK 2, 850-859.
- Basu, D. (1969). Role of the sufficiency and likelihood principles in sampling survey theory. *Sankhyā*, Parte 4, ser. A, vol. 31, pp. 441-454.
- Basu, D. (1971). An essay on the logical foundations of survey sampling. In: *Foundations of statistical inference*. Godambe and Sprott, editors. Hol, Rinehart and Winston. Toronto.

Basu, D. (1978b). On the relevance of randomization in data analysis (with discussions).  
In *Survey sampling and measurement*, N.Y. Namboodini (ed.) Academic Press: New  
York, 267-339. [Chapter XIV].

## CAPÍTULO 7

### APLICAÇÕES DE SUFICIÊNCIA

O conceito de estatística suficiente é essencial para a teoria clássica de inferência estatística. Nas seções (7.1) e (7.2) deste capítulo serão discutidas algumas aplicações de suficiência nas teorias de estimação e teste de hipóteses. Na seção (7.3) apresentamos os teoremas de Basu (1955, 1958). Estes resultados nos permitem, entre outras coisas, identificar a independência entre duas estatísticas. As noções de ancilaridade e completividade serão de utilidade para este propósito.

Ao longo deste capítulo assumiremos uma estrutura paramétrica para a família de medidas de probabilidades, dada por  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  onde  $\Theta$  é o espaço paramétrico.

#### 7.1 TEORIA DA ESTIMAÇÃO

Considere  $\tau(\theta)$  uma função paramétrica assumindo valores reais,  $\tau : \Theta \rightarrow \mathbb{R}$ . Seja  $T$  uma estatística a qual denominaremos estimador de  $\tau(\theta)$ . Para cada valor observado  $x \in \mathcal{X}$ ,  $T(x) = t$  é denominada estimativa de  $\tau(\theta)$ . Nosso objetivo é estimar  $\tau(\theta)$  (constante desconhecida) através do valor  $t$ . É claro que temos uma perda nesse processo, que será dada por uma função não negativa  $W(t, \theta)$ . Para cada valor  $\theta$  fixado,  $W(t, \theta)$  é  $\mathcal{A}_T$ -mensurável e portanto  $\mathcal{A}$ -mensurável. O risco de um estimador será medido pela

esperança da função perda, quando esta esperança existir.

$$(7.1) \quad \gamma_T(\theta) = E_{P_\theta}[W(T, \theta)] .$$

Considere  $\mathcal{T}$  a classe dos estimadores de  $\tau(\theta)$  tal que (7.1) está bem definido. Nesta classe é possível fazer a comparação dos estimadores através de seus riscos. Como veremos a seguir, segundo este critério, um estimador ótimo deve ser uma estatística suficiente.

### Teorema 7.1

Seja  $W$  uma função perda, convexa. Se  $\mathcal{C}$  é uma subálgebra suficiente, então para todo  $T \in \mathcal{T}$ , existe  $T_0 \in \mathcal{T}$   $\mathcal{C}$ -mensurável, tal que

$$\gamma_{T_0}(\theta) \leq \gamma_T(\theta) , \quad \forall \theta \in \Theta .$$

### Prova

Considere  $T \in \mathcal{T}$  e defina

$$T_0 = E_{P_\theta}[T|\mathcal{C}] .$$

Como  $\mathcal{C}$  é suficiente,  $T_0$  não depende de  $\theta$ . Portanto,  $T_0$  está bem definido como estimador. Pelo fato de  $W$  ser convexa, utilizando a desigualdade de Jensen, temos

$$W(T_0, \theta) = W(E[T|\mathcal{C}], \theta) \leq E[W(T, \theta)|\mathcal{C}] .$$

Aplicando a esperança dos dois lados da desigualdade, temos

$$\gamma_{T_0}(\theta) = E[W(T_0, \theta)] \leq E[E(W(T, \theta)|\mathcal{C})] = \gamma_T(\theta) .$$

Este teorema é conhecido como teorema de Rao-Blackwell. Nossa próxima etapa é apresentar o teorema de Lehmann-Scheffé (1950), que garante a existência de um estimador ótimo. Para tal, é necessário antes introduzir alguns novos conceitos.

### Definição 7.1

Um estimador  $T$  é não viesado para  $\tau(\theta)$  se

$$E_{P_\theta}(T) = \tau(\theta) , \quad \forall \theta \in \Theta .$$

Seja  $\mathcal{T}_0$  a classe dos estimadores não viesados para  $\tau(\theta)$  tal que  $E_{P_\theta}(T^2) < \infty$ .

### Definição 7.2

Um estimador  $T_0$  é denominado Estimador Não Viesado de Variância Uniformemente Mínima (ENVVUM) de  $\tau(\theta)$  se para todo  $T \in \mathcal{T}_0$ ,

$$E_{P_\theta}(T_0 - \tau(\theta))^2 \leq E_{P_\theta}(T - \tau(\theta))^2, \quad \forall \theta \in \Theta.$$

Isso significa que considerando a perda quadrática  $W(T, \theta) = (T - \theta)^2$ , o ENVVUM é o estimador de menor risco na classe dos estimadores não viesados.

### Definição 7.3

Uma subálgebra  $\mathcal{C}$  é (limitadamente) completa se para toda função  $g$  (essencialmente limitada)  $\mathcal{C}$ -mensurável, tal que

$$E_{P_\theta}[g] = 0, \quad \forall \theta \in \Theta,$$

então

$$g = 0 \text{ } [\mathcal{P}].$$

Uma estatística  $T$  é (limitadamente) completa se  $\mathcal{A}_T$  é (limitadamente) completa.

### Exemplo 7.1

Considere o modelo  $(\mathbb{N}, \mathbb{P}(\mathbb{N}), \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta : \theta \in \{1/4, 1/2\}\}$  e  $P_\theta$  é Binomial com  $n = 2$  e  $p = \theta$  e  $\mathbb{P}(\mathbb{N})$  é o conjunto das partes dos números naturais. Vamos mostrar que  $\mathbb{P}(\mathbb{N})$  não é completa.

Seja  $g$  uma função essencialmente limitada, tal que

$$\begin{aligned} \int_{\mathcal{X}} g dP_\theta &= 0, \quad \theta \in \{1/4, 1/2\} \\ \Rightarrow g(0)(1 - \theta)^2 + 2g(1)(1 - \theta) + g(2)\theta^2 &= 0 \\ \Rightarrow g(0)(1 - 2\theta + \theta^2) + 2g(1)(\theta - \theta^2) + g(2)\theta^2 &= 0 \\ \Rightarrow [g(0) - 2g(1) + g(2)]\theta^2 + [g(1) - g(2)]2\theta + g(0) &= 0 \end{aligned}$$

Considere  $g(0) = 1$ ,  $g(1) = -2$  e  $g(2) = 3$  de modo que as raízes da equação acima sejam  $\frac{1}{4}$  e  $\frac{1}{2}$ , então

$$(7.2) \quad 8\theta^2 - 6\theta + 1 = 0, \quad \forall \theta \in \{1/4, 1/2\}$$



Logo,  $\mathbb{P}(\mathbb{N})$  não é limitadamente completa e portanto não é completa.

Vamos alterar um pouco o problema considerando agora  $\mathcal{P} = \{P_\theta : \theta \in \{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}\}\}$ . Neste caso, não é mais possível garantir a validade de (7.2); isto é, não existe  $g \neq 0$  tal que a equação do segundo grau se anule. Logo, para este novo modelo  $\mathbb{P}(\mathbb{N})$  é completa.

Observe que a noção de completa está diretamente ligada à família de medidas de probabilidades. A palavra completa está associada a complementação do espaço paramétrico. Talvez o mais adequado seja dizer que a família de medidas de probabilidades é completa.

### Teorema 7.2

Seja  $\mathcal{T}_0$  uma classe não vazia de estimadores não viesados para  $\tau(\theta)$  tal que  $E_{P_\theta}[T^2] < \infty, \forall T \in \mathcal{T}_0$ . Se  $\mathcal{C}$  é uma subálgebra suficiente e completa, então existe um ENVVUM para  $\tau(\theta)$   $\mathcal{C}$ -mensurável, que é essencialmente único.

### Prova

Considere  $T_1 \in \mathcal{T}_0$  e defina

$$T_0 = E_{P_\theta}(T_1|\mathcal{C}) ,$$

claramente  $T_0 \in \mathcal{T}_0$ .

Pelo Teorema 7.1,  $\forall T \in \mathcal{T}_0, \exists T^* = E_{P_\theta}[T|\mathcal{C}]$  tal que

$$E_{P_\theta}(T^* - \tau(\theta))^2 \leq E_{P_\theta}(T - \tau(\theta))^2 , \quad \forall \theta \in \Theta .$$

Por construção,  $T^*$  é  $\mathcal{C}$ -mensurável e não viesado. Portanto,

$$E_{P_\theta}[T_0 - T^*] = 0 , \quad \forall \theta \in \Theta .$$

Como  $\mathcal{C}$  é completa, segue que

$$T_0 = T^* [\mathcal{P}] .$$

Logo,  $\forall T \in \mathcal{T}_0$ ,

$$E_{P_\theta}[T_0 - \mathcal{I}(\theta)]^2 \leq E_{P_\theta}[T - \mathcal{I}(\theta)]^2 , \quad \forall \theta \in \Theta .$$

Portanto,  $T_0$  é ENVVUM para  $\mathcal{I}(\theta)$ . □

O último teorema desta seção, também devido a Lehman-Scheffé, propicia um novo critério para a determinação de uma subálgebra suficiente mínima.

### Teorema 7.3

Se  $\mathcal{C}$  é uma subálgebra suficiente e limitadamente completa, então  $\mathcal{C}$  é suficiente mínima.

#### Prova

Considere  $\mathcal{C}$  uma subálgebra suficiente e limitadamente completa. Devemos mostrar que  $\mathcal{C} \subset \mathcal{D} [\mathcal{P}]$ ,  $\forall \mathcal{D}$  suficiente.

Considere  $C \in \mathcal{C}$  e  $g = E[I_C|\mathcal{D}]$  e  $h = E[g|\mathcal{C}]$ . Temos que

$$\int_{\mathcal{X}} I_C dP_\theta = \int_{\mathcal{X}} g dP_\theta = \int_{\mathcal{X}} h dP_\theta, \quad \forall \theta \in \Theta .$$

Logo,

$$\int_{\mathcal{X}} (I_C - h) dP_\theta = 0, \quad \forall \theta \in \Theta .$$

$I_C - h$  é uma função  $\mathcal{C}$ -mensurável e limitada. Por hipótese,  $\mathcal{C}$  é limitadamente completa, então

$$I_C = h [\mathcal{P}] .$$

Agora, pela desigualdade de Jensen,

$$E(g^2) = E(E(I_C/\mathcal{D})^2) \leq E(E(I_C^2/\mathcal{D})) = E(I_C^2)$$

e

$$E(h^2) = E(E(g/\mathcal{C})^2) \leq E(E(g^2/\mathcal{C})) = E(g^2) .$$

Logo,

$$E(I_C^2) \geq E(g^2) \geq E(h^2) . \tag{7.2}$$

Mas,  $I_C = h [\mathcal{P}]$ , portanto, de (7.2), temos que

$$E(I_C^2) = E(g^2) = E(h^2) . \tag{7.3}$$

Por outro lado, pelas propriedades da esperança condicional, temos que

$$\begin{aligned} E(I_C^2) &= E((I_C - E(I_C/\mathcal{D}))^2) + E(E(I_C/\mathcal{D})^2) \\ &= E((I_C - g)^2) + E(g^2) . \end{aligned}$$

Consequentemente, a partir de (7.3) segue que

$$E((I_C - g)^2) = 0$$

e portanto  $I_C = g [\mathcal{P}]$ .

Seja  $D = \{x : g(x) = 1\}$ . Então  $D \in \mathcal{D}$  e  $C = D [\mathcal{P}]$ , concluindo-se assim a prova.  $\square$

O exercício a seguir mostra que a recíproca do teorema acima não é verdadeira.

### Exercício 7.1

Considere o modelo de amostragem definido no Capítulo 6. Seja  $T$  uma estatística de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{X}, \mathcal{A})$  definida por  $T(d) = d$  (isto é,  $T$  é a estatística suficiente minimal). Mostre que  $T$  não é completa.

### Exercício 7.2

Considere, novamente, o modelo de amostragem do Capítulo 6. Seja  $T = T(d)$  um estimador (assumindo valores reais) não viesado do parâmetro  $\mathcal{I}(\theta)$  (assumindo valores reais), e  $\theta_0 \in \Theta$  um ponto arbitrário fixado. Mostre que sempre é possível encontrar um estimador  $T_0$  não viesado tal que  $E_{\theta_0}(T_0 - \mathcal{I}(\theta_0))^2 = 0$ .

Comente a respeito da existência de estimadores não viesados de variância uniformemente mínima (ENNVUM) no modelo de amostragem.

(Sugestão: defina  $T_0 = T(d) - T(d_0) + \mathcal{I}(\theta_0)$ , onde  $d_0$  é o dado associado a  $\theta_0$ .)

### Exercício 7.3

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com distribuição comum  $U[0, \theta]$ ,  $\theta > 0$ . Prove que  $T = \left(1 + \frac{1}{n}\right) \max_{1 \leq i \leq n} \{X_i\}$  é um estimador não viesado de  $\theta$  cuja variância não é superior à do estimador definido por  $S = 2\bar{X}$ .

### Exercício 7.4

Sejam  $X_1, \dots, X_m; Y_1, \dots, Y_n$  independentes e normalmente distribuídas  $N(\mu, \sigma_1^2)$  e  $N(\mu, \sigma_2^2)$ , respectivamente ( $\sigma_1^2 > 0$ ,  $\sigma_2^2 > 0$ ,  $\mu \in \mathbb{R}$ ). Seja  $Z = (X_1, \dots, X_m, Y_1, \dots, Y_n)$  e  $(\mathcal{Z}, \mathcal{A}, \mathcal{P})$  o modelo estatístico induzido por  $Z$ . Mostre que  $T = (\sum_{i=1}^m X_i, \sum_{i=1}^m X_i^2, \sum_{j=1}^n Y_j, \sum_{j=1}^n Y_j^2)$  é uma estatística suficiente que não é completa.

### Exercício 7.5

Considere o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ . Seja  $\mathcal{C}$  uma subálgebra suficiente e limitadamente completa. Prove que se  $f \in \mathcal{L}_{00}(\mathcal{X}, \mathcal{A}, P_\theta)$  tal que  $E_\theta(f) = 0$ ,  $\forall \theta \in \Theta$ , então  $E_\theta(f, g) = 0$ ,  $\forall g \in \mathcal{L}_{00}(\mathcal{X}, \mathcal{C}, P_\theta)$ ,  $\forall \theta \in \Theta$ .

### Exercício 7.6

Seja  $\mathcal{U}$  a classe de distribuições uniformes sobre intervalos finitos e seja  $\mathcal{P}$  a classe de combinações convexas de um número finito de distribuições em  $\mathcal{U}$  (isto é,  $P \in \mathcal{P}$  se  $P = \sum_{i=1}^n \alpha_i P_i$ ,  $\sum_{i=1}^n \alpha_i = 1$ ,  $\alpha_i > 0$  para cada  $i = 1, \dots, n$  e  $P_i \in \mathcal{U}$ ). Se  $X_1, \dots, X_n$  são variáveis aleatórias independentes com lei comum  $P \in \mathcal{P}$ . Mostre que  $T = (X^{(1)}, \dots, X^{(N)})$  o vetor de estatísticas de ordem é suficiente e completo.

## 7.2 TEORIA DE TESTE DE HIPÓTESES

Nesta seção o objetivo é estudar procedimentos para testar hipóteses a respeito do parâmetro  $\theta$ , do tipo abaixo

$$H_0 : \theta \in \Theta_0 \subset \Theta; \quad \text{hipótese nula,}$$

$$H_1 : \theta \in \Theta_0 \setminus \Theta; \quad \text{hipótese alternativa.}$$

Considere a função teste  $\varphi : \mathcal{X} \rightarrow [0, 1]$  que fornece a probabilidade de rejeição de  $H_0$  para cada valor observado  $x \in \mathcal{X}$ .

### Exemplo 7.2

Considere a estrutura  $(R_1, \mathcal{B}_1, \mathcal{P})$ , onde  $\mathcal{P} = \{P_{\mu, \sigma} : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$  e  $P_{\mu, \sigma}$  é a medida produto de  $n$  Normais independentes e identicamente distribuídas com média  $\mu$  e variância  $\sigma^2$ .

$$\begin{cases} H_0 : \mu = 0 & , \sigma > 0 \\ H_1 : \mu \neq 0 & , \sigma > 0 \end{cases}$$
$$\varphi(\mathbf{x}) = \begin{cases} 1 & , \text{ se } \left| \frac{\tilde{x}}{r} \right| \geq 3, 14 \\ 0 & , \text{ caso contrário,} \end{cases}$$

onde  $\tilde{x}$  é a mediana amostral e  $r$  a amplitude de variação amostral.

O poder de um teste estatístico é dado por  $E_{P_\theta}[\varphi]$ . No exemplo (testes não aleatorizados) o poder do teste é a probabilidade de um valor observado  $x$  pertencer à região crítica (região de rejeição da hipótese  $H_0$ ).

### Teorema 7.4

Se  $\mathcal{C}$  é uma subálgebra suficiente, então para toda função teste  $\varphi$   $\mathcal{A}$ -mensurável existe uma função teste  $\psi$   $\mathcal{C}$ -mensurável, tal que

$$E_{P_\theta}[\psi] = E_{P_\theta}[\varphi] , \quad \forall \theta \in \Theta .$$

A prova do Teorema 7.4 fica como exercício para o leitor. Sugestão: faça  $\psi = E[\varphi|\mathcal{C}]$ .

Com relação ao exemplo anterior, sabemos que  $(\overline{\mathcal{X}}, S)$  é suficiente para a família normal. Portanto, existe uma função teste baseada em  $(\overline{\mathcal{X}}, S)$  com o mesmo poder de  $\varphi$ .

Os lemas a seguir serão de utilidade para estabelecer uma recíproca do Teorema 7.5.

### Lema 7.1

Seja  $\mu$  uma medida finita em  $(\mathcal{X}, \mathcal{A})$  e  $\{A_\alpha : \alpha \in \Delta\}$  uma coleção arbitrária de conjuntos disjuntos de  $\mathcal{A}$ . Então,  $S = \{\alpha : \mu(A_\alpha) > 0\}$  é enumerável.

### Prova

Considere  $\mu(\mathcal{X}) = K < \infty$  e  $S_n = \{\alpha : \mu(A_\alpha) > \frac{1}{n}\}$ ,  $n \in \mathbb{N}$ . Vamos mostrar que  $S_n$  tem no máximo  $(nK - 1)$  elementos e portanto é finita.

Suponha que  $S_n$  contenha  $nK$  elementos. Utilizando o fato dos  $A_{\alpha_i}$ 's serem disjuntos, temos

$$\mu\left(\bigcup_{i=1}^{nK} A_{\alpha_i}\right) = \sum_{i=1}^{nK} \mu(A_{\alpha_i}) > n \cdot K \frac{1}{n} = K ,$$

o que contradiz a hipótese inicial.

Para finalizar, basta considerar  $S = \bigcup_{n=1}^{\infty} S_n$ . □

### Lema 7.2

Se  $E$  é um conjunto enumerável de números reais, então  $E^c$  é denso.

### Prova

Dado  $r \in \mathbb{R}$ , vamos mostrar que  $\forall \epsilon > 0, \exists y \in E^c$  tal que  $r - \epsilon < y < r + \epsilon$ .

Suponha que não exista tal ponto  $y$ . Neste caso,  $(r - \epsilon, r + \epsilon) \subset E$ , o que contradiz a enumerabilidade de  $E$ .

Portanto,  $r \in \overline{E^c}$ , onde  $\overline{E^c}$  denota o fechamento de  $E^c$ . Isto significa que  $\mathbb{R} = \overline{E^c}$ .  $\square$

### Teorema 7.5

Seja  $\mathcal{P}$  uma família dominada. Se para qualquer função teste  $\varphi$   $\mathcal{A}$ -mensurável existe uma função  $\psi$   $\mathcal{C}$ -mensurável tal que

$$E_{P_\theta}(\psi) = E_{P_\theta}(\varphi), \quad \forall \theta \in \Theta;$$

então  $\mathcal{C}$  é suficiente.

### Prova

Considere  $P_1$  e  $P_2$  em  $\mathcal{P}$  e  $Q = \frac{1}{2}(P_1 + P_2)$ . Para  $k > 0$ , definimos o teste  $\varphi$  como

$$\varphi(x) = \begin{cases} 1 & \text{se } \frac{dP_1}{dQ}(x) > k \\ 0 & \text{se } \frac{dP_2}{dQ}(x) < k. \end{cases}$$

Pelo lema fundamental de Neyman-Pearson,  $\varphi$  é um teste mais poderoso de nível  $\alpha = E_Q(\varphi)$ . Por hipótese, existe um teste  $\psi$   $\mathcal{C}$ -mensurável com o mesmo nível  $\alpha$  e com o mesmo poder. Portanto, aplicando o lema de Neyman-Pearson novamente temos que  $\varphi = \psi[Q]$ .

Seja  $B_k = \{x : \frac{dP_1}{dQ}(x) = k\}$ . Segue do Lema 7.1 que  $K = \{k : Q(B_k) > 0\}$  é enumerável.

Para todo  $k \in K$ , temos que  $Q(B_k) = 0$  e portanto o conjunto  $\{x : \frac{dP_1}{dQ}(x) < k\}$  é  $\mathcal{C}$ -mensurável,  $Q$ -essencialmente.

Pelo Lema 7.2 temos que  $K^c$  é denso. Então, para todo  $k \in K$ , existe uma seqüência crescente  $\{K_n\}_{n \geq 1}$  de  $K^c$  que converge para  $k$  e portanto o conjunto

$$\left\{x : \frac{dP_1}{dQ}(x) < k\right\} = \bigcup_{n=1}^{\infty} \left\{x : \frac{dP_1}{dQ}(x) < k_n\right\}$$

é  $\mathcal{C}$ -mensurável,  $Q$ -essencialmente.

Pelo Lema 4.4 temos que  $\mathcal{C}$  é suficiente por par. Como por hipótese o modelo é dominado, segue que  $\mathcal{C}$  é suficiente.

### 7.3 OS TEOREMAS DE BASU

O que denominamos teoremas de Basu são três teoremas que relacionam os conceitos de suficiência, ancilaridade e independência estatística. D. Basu estabeleceu condições sob as quais dois desses conceitos implicam num terceiro.

Uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  é ancilar se, para todo  $D \in \mathcal{D}$ ,  $P_\theta(D)$  é constante,  $\forall P_\theta \in \mathcal{P}$ .

A independência estatística difere da probabilística pelo fato de trabalharmos com uma família de medidas de probabilidades. Assim, diremos que duas subálgebras  $\mathcal{C}$  e  $\mathcal{D}$  são estatisticamente independentes se,  $\forall C \in \mathcal{C}$

$$P_\theta(C|\mathcal{D}) = P_\theta(C) [\mathcal{P}] .$$

Duas estatísticas  $T$  e  $S$  são estatisticamente independentes se e somente se  $\mathcal{A}_T$  e  $\mathcal{A}_S$  forem estatisticamente independentes.

#### Exercício 7.7

Mostre que  $\mathcal{C}$  e  $\mathcal{D}$  são estatisticamente independentes se, e somente se,

$$P_\theta(C \cap D) = P_\theta(C)P_\theta(D) [\mathcal{P}] , \quad \forall C \in \mathcal{C}; \forall D \in \mathcal{D} .$$

#### Exemplo 7.3

Considere  $(R_1, \mathcal{B}_1, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta : \theta \in \mathbb{Z}\}$  tal que  $P_\theta(x) = I_{[\theta, \theta+1[}(x)$ .

Considere  $Y(x) = [x]$  o maior inteiro de  $x$ .

Se  $A \in \mathcal{B}_1$  e  $D_y = \{x \in \mathbb{R} : Y(x) = y\}$ , onde  $y \in \mathbb{Z}$ , então

$$P_\theta(A \cap D_y) = P_\theta(A)I_{\{\theta=y\}} = P_\theta(A)P_\theta(D_y) .$$

Logo,  $X$  e  $Y$  são estatisticamente independentes. No entanto, observando  $X = 3.2$ , o que podemos dizer a respeito de  $Y$ ? Claramente,  $Y = 3$ . Mas,  $X$  e  $Y$  não são independentes?

A independência entre as estatísticas  $X$  e  $Y$  é condicionada ao conhecimento de  $\theta$ , isto é, ela é válida para uma dada medida de probabilidade fixada  $P_\theta$ . O que ocorre usualmente em problemas estatísticos é o desconhecimento de  $\theta$  e portanto variáveis independentes podem fornecer informações relevantes, uma a respeito da outra.

**Teorema 7.6** (Basu, 1955)

Seja  $\mathcal{C}$  uma subálgebra limitadamente completa e suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Se  $\mathcal{D}$  é uma subálgebra ancilar, então  $\mathcal{C}$  e  $\mathcal{D}$  são estatisticamente independentes.

**Prova**

Considere  $D \in \mathcal{D}$ . Pela suficiência de  $\mathcal{C}$  sabemos que existe uma  $f = E[I_D|\mathcal{C}]$  tal que

$$P_\theta(D) = \int_{\mathcal{X}} f dP_\theta, \quad \forall \theta \in \Theta.$$

Pela ancilaridade de  $\mathcal{D}$ ,  $P_\theta(D) = k$  (constante em relação a  $\theta$ ). Portanto,

$$\int_{\mathcal{X}} (f - k) dP_\theta = 0, \quad \forall \theta \in \Theta,$$

onde  $(f - k)$  é  $\mathcal{C}$ -mensurável. Pela completividade de  $\mathcal{C}$ , temos que

$$f = k [\mathcal{P}].$$

Isto é,

$$P_\theta(D|\mathcal{C}) = E_\theta[I_D|\mathcal{C}] = P_\theta(D) [\mathcal{P}], \quad \forall D \in \mathcal{D}. \quad \square$$

**Exemplo 7.4**

Considere  $(R_n, \mathcal{B}_n, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}\}$  é a família de medidas de probabilidade Normal  $n$ -variada gerada por  $n$  observações independentes Normais com média  $\theta$  e variância 1.

Vamos mostrar que, nesse caso, as estatísticas  $\bar{X}$  (média amostral) e  $S^2$  (variância amostral) são estatisticamente independentes.

O modelo é dominado pela medida de Lebesgue e sua função densidade é

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= (2\pi)^{-n/2} e^{-\sum_{i=1}^n \frac{(x_i - \theta)^2}{2}} \\ &= (2\pi)^{-n/2} e^{-\frac{1}{2}[\sum_{i=1}^n x_i^2 - 2\theta n\bar{X} - n\theta^2]} \\ &= (2\pi)^{-n/2} e^{-\sum_{i=1}^n \frac{x_i^2}{2}} e^{\frac{\theta n\bar{X} - n\theta^2}{2}}. \end{aligned}$$

Pelo teorema da fatoração,  $\bar{X}$  é suficiente.

Seja  $\mathcal{P}^{\bar{X}}$  a família de medidas induzida por  $\bar{X}$ . Para  $Q_\theta \in \mathcal{P}^{\bar{X}}$ , sabemos que  $Q_\theta$  tem distribuição  $N(\theta, \frac{1}{n})$ .



Seja  $f$  uma função Borel mensurável, tal que

$$\int_{\mathbb{R}} f dQ_{\theta} = 0 \Rightarrow \int_{-\infty}^{+\infty} f(y)[2\pi n]^{-2} e^{-\frac{(y-\theta)^2}{2n}} dy = 0 .$$

Fazendo uma transformação de variáveis e utilizando a unicidade da Transformada de Laplace, temos que  $f = 0$   $[\mathcal{P}^{\bar{X}}]$ . Logo,  $\bar{X}$  é completa.

A ancilaridade de  $S^2$  provém dos seguintes fatos:

$$\text{i) } (n-1)S^2 = \sum_{i=1}^n (x_i - \bar{X})^2 = \sum_{i=1}^n (x_i - \theta)^2 - n(\bar{X} - \theta)^2;$$

ii)  $(\bar{X} - \theta)$  e  $(x_i - \theta)$ ,  $\forall i = 1, 2, \dots, n$ , têm distribuição  $N(0, 1)$ .

Portanto,  $(\bar{X} - \theta)^2$  tem distribuição Qui-quadrado com 1 grau de liberdade e  $\sum_{i=1}^n \frac{(x_i - \theta)^2}{n}$  tem distribuição Qui-quadrado com  $n$  graus de liberdade. De (i) temos que  $\frac{(n-1)}{n}S^2$  tem distribuição Qui-quadrado com  $(n-1)$  graus de liberdade, logo  $S^2$  é ancilar.

A independência de  $\bar{X}$  e  $S^2$  é aplicação do Teorema 7.6.

### Exemplo 7.5

Considere  $X_1, X_2, \dots, X_{n+1}$  variáveis aleatórias estatisticamente independentes com distribuição Gama  $(\alpha_i, \beta)$ ,  $i = 1, 2, \dots, n+1$ , respectivamente. Sejam,

$$S = \sum_{i=1}^{n+1} X_i \quad \text{e} \quad Y_k = \frac{X_k}{S}, \quad k = 1, 2, \dots, n+1 .$$

Neste caso, o vetor  $\mathbf{Y} = (Y_1, \dots, Y_n)$  tem distribuição de Dirichlet com parâmetro  $(\alpha_1, \alpha_2, \dots, \alpha_{n+1})$ .

Fixado  $\boldsymbol{\alpha}_0 = (\alpha_1, \alpha_2, \dots, \alpha_{n+1})$ , é fácil verificar que  $S$  é suficiente e completa para  $\beta$ . É claro que  $\mathbf{Y}$  é ancilar para  $\beta$ . Logo,  $\mathbf{Y}$  e  $S$  são estatisticamente independentes.

### Exemplo 7.6

Considere  $(R_n, \mathcal{B}_n, \mathcal{P})$ , onde  $\mathcal{P}$  é a família de Normais  $n$ -variadas com covariância zero gerada por  $n$  Normais com média  $\mu$  e variância  $\sigma^2$ .

Sejam  $\bar{X}$  e  $S^2$  as estatísticas média e variância amostrais, respectivamente. Além disso, considere  $\bar{y}$  uma variável aleatória  $N(0, \frac{1}{n})$  e  $s^2$  uma v.a. tal que  $(n-1)s^2$  tem distribuição Qui-quadrado com  $(n-1)$  graus de liberdade.

Vamos definir as variáveis  $y_1, y_2, \dots, y_n$  pela seguinte transformação:

$$\frac{y_i - \bar{y}}{s} = \frac{X_i - \bar{X}}{S}, \quad \text{para } i = 1, 2, \dots, n.$$

Nosso objetivo é mostrar que  $y_1, y_2, \dots, y_n$  são estatisticamente independentes com distribuição comum  $N(0, 1)$ .

Não é difícil mostrar, utilizando o teorema da fatoração e a unicidade da transformada de Laplace que  $(\bar{X}, S^2)$  é suficiente e completa para  $(\mu, \sigma^2)$ .

Seja  $W_{K-1} = (Z_1, Z_2, \dots, Z_{K-1})$ , para  $K = 2, 3, \dots, n$ , onde  $Z_i = \frac{X_i - \bar{X}}{S}$ ,  $i = 1, 2, \dots, n$ .

Observamos que  $W_{K-1}$  é invariante escala e locação e portanto é ancilar para  $(\mu, \sigma^2)$  que são, respectivamente, parâmetros locação e escala (Proposição 8.3.2). Logo, pelo Teorema 7.6,  $W_{K-1}$  e  $(\bar{X}, S^2)$ , para  $k = 2, 3, \dots, n$ , são estatisticamente independentes. Além disso,  $W_{K-1}$  é independente de  $X_K$  e  $Z_K = \frac{X_K - \bar{X}}{S}$ . Portanto,  $W_{K-1}$  é estatisticamente independente de  $Z_K$ , para  $K = 2, 3, \dots, n$ . Logo,  $Z_1, Z_2, \dots, Z_n$  são estatisticamente independentes.

Por construção, temos que

$$y_i = Z_i s + \bar{y}.$$

Se observarmos  $s$  e  $\bar{y}$ ,  $x_i$  é uma função linear de  $Z_i$ ,  $i = 1, 2, \dots, n$ . Logo, as  $y_i$ 's são estatisticamente independentes e claramente são identicamente distribuídas.

Finalmente, devemos mostrar que cada  $y_i$  tem distribuição Normal com média zero e variância 1.

Observe que

$$\sum_{i=1}^n a_i y_i = \bar{y} \sum_{i=1}^n a_i,$$

onde  $\bar{y}$  é  $N(0, \frac{1}{n})$ . Portanto, para toda seqüência de números reais  $\{a_i\}_{i=1}^n$  temos que  $\sum_{i=1}^n a_i y_i$  tem distribuição Normal. Logo, pelo Teorema de Cramer-Wald, temos que cada  $y_i$  deve ser Normal. Se considerarmos  $a_i = 1, \forall i$ , concluímos finalmente que  $y_i$  é  $N(0, 1)$ ,  $\forall i$ .

### Exercício 7.8

Verifique ou discuta porque não são válidas as seguintes relações:

- (a) No modelo Normal  $n$ -variado com covariância zero originário de variáveis  $N(\mu, \sigma^2)$ , considere  $\bar{X}$  a média amostral e  $S$  o desvio padrão amostral. As estatísticas  $\frac{\bar{X}}{S}$  e  $S$

são estatisticamente independentes.

(b)  $X_1, X_2, \dots, X_n$  são variáveis independentes e identicamente distribuídas  $N(0, 1)$ .

Neste caso,

$$t_1 = \frac{X_2}{\sqrt{X_1^2}}, \quad t_2 = \frac{X_3}{\sqrt{\frac{X_1^2 + X_2^2}{2}}}, \quad \dots, \quad t_{n-1} = \frac{X_n}{\sqrt{\frac{X_1^2 + \dots + X_{n-1}^2}{n-1}}}$$

são mutuamente independentes.

### Exercício 7.9

Considere o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ . Seja  $\mathcal{D}$  uma subálgebra ancilar e independente da subálgebra  $\mathcal{J} = \sigma(\mathcal{A} - \mathcal{D})$  para todo  $\theta \in \Theta$ . Mostre que  $\mathcal{J}$  é uma subálgebra suficiente.

### Definição 7.4

Duas medidas de probabilidades  $P_1$  e  $P_2$  em  $(\mathcal{X}, \mathcal{A})$  são sobrepostas se, para cada  $A \in \mathcal{A}$  tal que  $P_1(A) = 1$ , então  $P_2(A) > 0$ .

Observe que tal propriedade é simétrica.

### Definição 7.5

Uma família  $\mathcal{P}$  de medidas de probabilidades é conectada se para todo  $P_\theta, P_\alpha$  em  $\mathcal{P}$ , existem  $P_{\theta_1}, P_{\theta_2}, \dots, P_{\theta_n}$  em  $\mathcal{P}$ , com  $\theta_1 = \theta; \theta_n = \alpha$  e  $n \in \mathbb{N}$ , tais que  $P_{\theta_i}$  e  $P_{\theta_{i+1}}$  são sobrepostas para  $i = 1, 2, \dots, n-1$ .

Observe que duas medidas sobrepostas devem ter uma intersecção não vazia entre seus suportes.

### Exemplo 7.7

Considere o modelo  $(R_1, \mathcal{B}_1, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta : \theta \in \mathbb{Z}\}$  e  $P_\theta$  é uniforme em  $[\theta, \theta + 1]$ . Seja  $A = [0, 1]$ , temos que  $P_\theta(A) = 1$  e  $P_\theta(A) = 0, \forall \theta \in \mathbb{Z}, \theta \neq 0$ .

Logo, não existe  $\theta \in \mathbb{Z} - \{0\}$  tal que  $P_\theta(A) > 0$ . Portanto, a família não é conectada.

Se alterarmos o espaço paramétrico para  $\Theta = \{k/2 : k \in \mathbb{Z}\}$  a família será conectada. Observe a intersecção entre os suportes das medidas  $P_\theta$  e  $P_{\theta+1/2}, \forall \theta \in \Theta$ .

### Exemplo 7.8 (Basu, 1958)

Sejam  $\mathcal{C}$  e  $\mathcal{D}$  subálgebras tal que  $\mathcal{C}$  é suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  e estatisticamente independente de  $\mathcal{D}$ . Se  $\mathcal{P}$  é conectada, então  $\mathcal{D}$  é ancilar.

**Prova**

Seja  $D \in \mathcal{D}$ . Pela suficiência de  $\mathcal{C}$  podemos escrever

$$P_\theta(D \cap C) = \int_B f dP_\theta, \quad \forall \theta \in \Theta, \quad \forall C \in \mathcal{C},$$

onde  $f$  é  $\mathcal{C}$ -mensurável e não depende de  $\theta$ .

Por outro lado, pela independência entre  $\mathcal{C}$  e  $\mathcal{D}$ , temos que

$$P_\theta(D \cap C) = P_\theta(D)P_\theta(C) = \int_C P_\theta(D) dP_\theta, \quad \forall \theta \in \Theta, \quad \forall C \in \mathcal{C}.$$

Portanto,

$$P_\theta(D) = f [P_\theta]. \tag{7.4}$$

Considere  $P_{\theta_1}$  e  $P_{\theta_2}$  medidas sobrepostas. Neste caso, existe  $x_0 \in A_{\theta_1} \cap A_{\theta_2}$ , onde  $A_{\theta_i}$  é o suporte de  $P_{\theta_i}$ ,  $i = 1, 2$ . Então, para pelo menos um ponto  $x_0$  vale a igualdade (7.4) e portanto

$$P_{\theta_1}(D) = P_{\theta_2}(D).$$

Considere agora que  $P_{\theta_1}$  e  $P_{\theta_n}$  são conectadas. Neste caso, existem  $P_{\theta_2}, P_{\theta_3}, \dots, P_{\theta_{n-1}}$  tais que  $P_{\theta_i}$  e  $P_{\theta_{i+1}}$  são sobrepostas,  $i = 1, 2, \dots, n - 1$ . Portanto,

$$P_{\theta_i}(D) = P_{\theta_{i+1}}(D), \quad \forall i = 1, 2, \dots, n - 1.$$

Logo,  $P_{\theta_1}(D) = P_{\theta_n}(D)$ . Isso implica na ancilaridade de  $A$ . Como  $A$  é um conjunto arbitrário em  $\mathcal{D}$ , temos que  $\mathcal{D}$  é ancilar. □

**Definição 7.6**

$A \in \mathcal{A}$  é um conjunto separador se,  $\forall \theta \in \Theta$ ,  $P_\theta(A) = 0$  ou  $P_\theta(A) = 1$  e para pelo menos um par  $\theta_1, \theta_2 \in \Theta$ , temos que  $P_{\theta_1}(A) = P_{\theta_2}(A^c) = 1$ .

**Teorema 7.8** (Thomas e Koehn, 1975)

Seja  $\mathcal{C}$  uma subálgebra suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Toda subálgebra  $\mathcal{D}$  estatisticamente independente de  $\mathcal{C}$  é ancilar se, e somente se, não existe um conjunto separador em  $\mathcal{A}$ .

**Prova**

Vamos mostrar que existe uma  $\mathcal{C}$  não ancilar  $\Leftrightarrow$  existe um conjunto  $A \in \mathcal{A}$ , separador.  
 $(\Leftarrow)$  Suponha que exista  $A$  separador.

Sejam  $\varphi = \{\theta \in \Theta : P_\theta(A) = 1\}$  e  $S(x) = I_A(x)$ ,  $\forall x \in \mathcal{X}$ . Então

$$\begin{aligned} P_\theta(S(x) = 1) &= 1, \quad \text{se } \theta \in \varphi \text{ e} \\ P_\theta(S(x) = 0) &= 1, \quad \text{se } \theta \in \varphi^c. \end{aligned}$$

Para  $\theta$  fixado,  $S$  é essencialmente constante, o que implica na independência entre  $\mathcal{A}_S$  e qualquer outra subálgebra. Em particular, podemos ter  $\mathcal{D} = \mathcal{A}_S$  estatisticamente independentes de  $\mathcal{C}$ . No entanto,  $\mathcal{D}$  não é ancilar.

$(\Rightarrow)$  Suponha que exista  $\mathcal{D}$  não ancilar. Neste caso, existe  $D \in \mathcal{D}$  tal que para algum par  $\theta_1, \theta_2 \in \Theta$ ,

$$P_{\theta_1}(D) \neq P_{\theta_2}(D). \quad (7.5)$$

Pela suficiência de  $\mathcal{C}$  e independência de  $\mathcal{D}$ , temos

$$P_\theta(B) = f[\mathbb{P}_\theta],$$

onde  $f$  é  $\mathcal{C}$ -mensurável.

Considere  $A_\theta = \{x \in \mathcal{X} : P_\theta(B) = f(x)\}$  de modo que  $P_\theta(A_\theta) = 1$ . De (7.5) temos que existem  $\theta_1$  e  $\theta_2 \in \Theta$  tais que  $A_{\theta_1} \cap A_{\theta_2} = \emptyset$ . Seja  $\varphi = \{\theta \in \Theta : A_\theta = A_{\theta_1}\}$ . Observe que  $\varphi^c$  e  $\varphi$  são conjuntos não vazios. Então

$$\begin{aligned} \theta \in \varphi &\Rightarrow P_\theta(A_{\theta_1}) = 1, \\ \theta \in \varphi^c &\Rightarrow P_\theta(A_{\theta_1}^c) \geq P_\theta(A_\theta) = 1 \Rightarrow P_\theta(A_{\theta_1}^c) = 1. \end{aligned}$$

Portanto,  $A_{\theta_1}$  é um conjunto separador. □

É possível estabelecer a equivalência entre os teoremas (7.7) e (7.8) nos casos do modelo dominado e do modelo discreto. Para tal vejamos os dois próximos lemas.

### Lema 7.3

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  um modelo discreto. A família  $\mathcal{P}$  é conectada se, e somente se, não existe um conjunto separador.

### Prova

Vamos mostrar que existe um  $A$  separador se, e somente se,  $\mathcal{P}$  não é conectado.

( $\Rightarrow$ ) Considere  $A$  um conjunto separador. Sejam

$$\begin{aligned}\mathcal{P}_1 &= \{P_\theta \in \mathcal{P} : P_\theta(A) = 1\} \\ \mathcal{P}_2 &= \{P_\theta \in \mathcal{P} : P_\theta(A) = 0\} .\end{aligned}$$

Observe que todos elementos de  $\mathcal{P}_1$  são sobrepostos e o mesmo vale para  $\mathcal{P}_2$ . No entanto, para  $P_{\theta_1} \in \mathcal{P}_1$  e  $P_{\theta_2} \in \mathcal{P}_2$ , tais medidas não podem ser sobrepostas. Logo, o modelo não é conectado.

( $\Leftarrow$ ) Suponha que  $\mathcal{P}$  não é conectado.

Para cada  $\theta_1, \theta_2$ , defina a seguinte relação de equivalência,

$$P_{\theta_1} \sim P_{\theta_2} \Leftrightarrow P_{\theta_1}, P_{\theta_2} \text{ são conectadas.}$$

Fixado  $P_1 \in \mathcal{P}$ , considere  $\mathcal{P}_1 = \{P_\theta \in \mathcal{P} : P_\theta \sim P_1\}$ . Observe que  $\mathcal{P}_2 = \mathcal{P} - \mathcal{P}_1 \neq \emptyset$ . Além disso, sejam

$$\Gamma_1 = \{\theta : P_\theta \in \mathcal{P}_1\} \quad \text{e} \quad \Gamma_2 = \{\theta : P_\theta \in \mathcal{P}_2\} .$$

Seja  $A_\theta$  o suporte associado à medida  $P_\theta$ , isto é,  $P_\theta(A_\theta) = 1$  e  $\forall B \subset A_\theta, P_\theta(B) > 0$ , então

$$A_{\theta_1} \cap A_{\theta_2} = \emptyset, \quad \forall \theta_1 \in \Gamma_1, \quad \forall \theta_2 \in \Gamma_2 .$$

Defina,

$$A = \bigcup_{\theta_1 \in \Gamma_1} A_{\theta_1} .$$

Como o modelo é discreto, a sigma-álgebra associada é a das partes e portanto  $A$  é mensurável. E ainda,

$$\begin{aligned}P_{\theta_1}(A) &= 1, \quad \forall \theta_1 \in \Gamma_1 \quad \text{e} \\ P_{\theta_2}(A) &= 0, \quad \forall \theta_2 \in \Gamma_2 .\end{aligned}$$

Logo,  $A$  é um conjunto separador. □

#### Lema 7.4

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  um modelo dominado. A família  $\mathcal{P}$  é conectada se, e somente se, não existe um conjunto separador.

### Prova

Exercício. Sugestão: use o fato da existência de uma família enumerável  $\mathcal{P}_0 \subset \mathcal{P}$  tal que  $\mathcal{P}_0 \equiv \mathcal{P}$ , no caso de  $\mathcal{P}$  ser dominada.

Considerando os lemas anteriores, observe-se que a condição da família conectada dada pelo Teorema 7.7 é uma condição necessária e suficiente para garantir a ancilaridade no caso dos modelos dominado e discreto.

Finalmente, apresentaremos o último teorema de Basu, no qual ancilaridade e independência implicam em suficiência.

### Teorema 7.9

Seja  $\mathcal{D}$  uma subálgebra ancilar e estatisticamente independente da subálgebra  $\mathcal{C}$ . Se  $\mathcal{C} \vee \mathcal{D}$  é suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , então  $\mathcal{C}$  é suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

### Prova

Considere  $H = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ .

Observe que  $H$  é fechado por intersecções finitas e  $\mathcal{C} \vee \mathcal{D} = \sigma(H)$ .

Seja  $\mathcal{H} = \{A : A \in \mathcal{C} \vee \mathcal{D} \text{ e } \int_B I_A dP_A = \int_B E[I_A|\mathcal{C}] dP_\theta, \forall B \in \mathcal{C}, \forall \theta \in \Theta\}$ .

Não é difícil mostrar que  $\mathcal{H}$  é um sistema Dinkyn. Além disso,  $H \subset \mathcal{H}$  como veremos a seguir.

Para  $A \in H$ ,  $\exists C \in \mathcal{C}$  e  $D \in \mathcal{D}$  tal que  $A = C \cap D$ ,

$$\int_B I_A dP_\theta = \int_B I_{C \cap D} dP_\theta = \int_{\mathcal{X}} I_{(B \cap C) \cap D} dP_\theta, \quad B \in \mathcal{C}.$$

Considere  $B^* = B \cap C \in \mathcal{C}$ . Da independência entre  $\mathcal{C}$  e  $\mathcal{D}$  e da ancilaridade de  $D$ , resulta

$$\begin{aligned} \int_B I_A dP_\theta &= \int_{\mathcal{X}} I_{B^* \cap D} dP_\theta = P_\theta(B^*)P_\theta(D) = \int_{\mathcal{X}} (\alpha - I_{B^*}) dP_\theta \\ &= \int_B \alpha I_C dP_\theta, \quad \forall \theta \in \Theta, \quad \forall B \in \mathcal{C}, \end{aligned}$$

onde  $\alpha = I_D$  é constante em relação a  $\theta$ .

A função  $\alpha I_C$  é  $\mathcal{C}$ -mensurável é uma versão da esperança condicional  $E[I_A|\mathcal{C}]$ . Logo,  $H \subset \mathcal{H}$ .

Pelo teorema de sistemas Dinkyn, temos que  $\mathcal{C} \vee \mathcal{D} \subset \mathcal{H}$  e portanto  $\mathcal{C} \vee \mathcal{D} = \mathcal{H}$ . Logo,  $\mathcal{C}$  é suficiente para  $(\mathcal{X}, \mathcal{C} \vee \mathcal{D}, \mathcal{P})$ .

Por hipótese,  $\mathcal{C} \vee \mathcal{D}$  é suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Portanto,  $\mathcal{C}$  é suficiente para  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  (Teorema 3.1).

Note que o Exercício 7.9 é uma consequência direta deste último teorema.

## REFERÊNCIAS

1. Bahadur, R.R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.*, **25**, 423–462.
2. Bahadur, R.R. and Lehmann, E.L. (1955). Two comments on sufficiency and statistical decision functions. *Ann. Math. Statist.*, **26**, 139–141.
3. Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā, A*, **15**, 377–380.
4. Basu, D. (1958). On statistics independent of a sufficient statistic. *Sankhyā, A*, **20**, 223–226.
5. Basu, D. (1959). The family of ancillary statistics. *Sankhyā, A*, **21**, 247–256.
6. Basu, D. and Tiwari, R.C. (1980). *Note on Dirichlet process*. Tallahassee, FSU, 26p. (FSU Statistics Report, M 536)
7. Basu, D. (1982). Basu theorems. *Encyclopedia of Statistical Sciences*, **1**, 193–196, New York: Wiley.
8. Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statist.*, **18**, 105–110.
9. Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika*, **48**, 1 and 2, p. 41.
10. Koehn, U. and Thomas, D.L. (1975). On statistics independent of sufficient statistic: Basu's lemma. *The Amer. Statist.*, **29**, 40–42.
11. Lehmann, E.L. (1959). *Testing statistical hypotheses*. John Wiley and Sons, New York.
12. Lehmann, E.L. and Scheffé (1950). Completeness, similar regions, and unbiased estimation. Part I. *Sankhyā, A*, **10**, 305–340.
13. Pathak, P.K. (1975). Note on Basu's lemma. New Mexico, UNM, 5p. - Technical Report, 308.
14. Rao, C.R. (1949). Sufficient statistics and minimum variance unbiased estimates. *Proc. Cambridge Phil. Soc.*, **45**, 213–218.



## CAPÍTULO 8

# RELAÇÕES ENTRE INVARIÂNCIA, SUFICIÊNCIA E ANCILARIDADE

### 8.1 DEFINIÇÕES E PRELIMINARES

Neste capítulo introduziremos o conceito de invariância no contexto da Teoria Estatística. Nosso objetivo é estudar suas relações com outros conceitos tais como suficiência fisheriana e ancilaridade.

Antes de dar alguma motivação estatística a respeito, precisamos de algumas definições e notações.

Como sempre  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  será nosso espaço estatístico.

#### Definição 8.1

Dizemos que a transformação  $g : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{A})$  preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  se:

- $g$  é 1-1 bimensurável
- a família de medidas de probabilidade induzida por  $g$ , coincide com  $\mathcal{P}$ , isto é:

$$Pg^{-1} = P \quad \forall P \in \mathcal{P}$$

Chamamos de  $\mathcal{G}$  à classe de tais transformações  $g$ , é claro que  $\mathcal{G} \neq \emptyset$ , pois a função identidade é um elemento de  $\mathcal{G}$ .

### Exemplo 8.1

Considerar  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A} = \mathcal{B}_n$  e  $\mathcal{P}$  a família de medidas de probabilidade induzida pelas v.a.  $X_1, X_2, \dots, X_n$  independentes e todas com a mesma distribuição  $N(0, \sigma^2)$ , isto é, se  $P_{\sigma^2} \in \mathcal{P}$  tem-se que

$$P_{\sigma^2}(B) = \int_B \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp \left\{ \frac{-1}{2\sigma^2} \sum_1^n x_i^2 \right\} d\lambda_n(\mathbf{x}) , \quad B \in \mathcal{B}_n, \quad \sigma^2 > 0$$

onde  $\lambda_n$  é a medida de Lebesgue em  $\mathbb{R}^n$ ; então algumas transformações  $g$  que pertencem a  $\mathcal{G}$  são:

$g(\mathbf{x}) = (|x_1|\Phi(x_1), \dots, |x_n|\Phi(x_n))$ , onde  $\mathbf{x} = (x_1, \dots, x_n)$ , sendo  $\Phi$  uma função ímpar que assume apenas os valores 1 e  $-1$ ;

$g(\mathbf{x}) = A\mathbf{x}$  onde  $A$  é uma  $n \times n$ -matriz ortogonal.

### Exercício 8.1

Considere a classe de funções  $\mathcal{G}$  como foi definida anteriormente, munida com a operação “ $\circ$ ”, composição de funções. Mostre que  $(\mathcal{G}, \circ)$  é um grupo de transformações.

### Definição 8.2

Seja  $A \in \mathcal{A}$ , e  $g \in \mathcal{G}$ .

O conjunto  $A$  é dito  $g$ -invariante se  $g^{-1}A = A$  e a subálgebra  $\mathcal{A}(g) = \{A \in \mathcal{A} : g^{-1}A = A\}$  é chamada de subálgebra de conjuntos  $g$ -invariantes.  $\overline{\mathcal{A}}(g)$  denotará o completamento de  $\mathcal{A}(g)$ .

O conjunto  $A$  é dito essencialmente  $g$ -invariante se  $A \in \overline{\mathcal{A}}(g)$ .

O conjunto  $A$  é dito quase  $g$ -invariante se  $P(A\Delta g^{-1}A) = 0$ ,  $\forall P \in \mathcal{P}$ , onde  $A\Delta B = (A \setminus B) \cup (B \setminus A)$ , isto é, a diferença simétrica entre os conjuntos  $A$  e  $B$ .

### Exercício 8.2

- Mostre que efetivamente  $\mathcal{A}(g)$  é uma subálgebra de  $\mathcal{A}$ .
- Mostre que os conceitos de essencialmente  $g$ -invariante e quase- $g$ -invariante são equivalentes

Baseados na Definição 8.2, consideramos três subálgebras fundamentais para nosso estudo.

( $\alpha$ )  $\mathcal{A}(\mathcal{G}) = \bigcap_{g \in \mathcal{G}} \mathcal{A}(g)$ , chamada de subálgebra  $\mathcal{G}$ -invariante.

( $\beta$ )  $\overline{\mathcal{A}}(\mathcal{G})$  o  $\mathcal{P}$ -completamento de  $\mathcal{A}(\mathcal{G})$ , chamada de subálgebra essencialmente  $\mathcal{G}$ -invariante.

( $\gamma$ )  $\tilde{\mathcal{A}}(\mathcal{G}) = \bigcap_{g \in \mathcal{G}} \overline{\mathcal{A}}(g)$ , chamada de subálgebra quase- $\mathcal{G}$ -invariante.

Notemos que  $\mathcal{A}(\mathcal{G}) \subset \overline{\mathcal{A}}(\mathcal{G}) \subset \tilde{\mathcal{A}}(\mathcal{G})$ . Com algumas hipóteses sobre  $\mathcal{G}$ , é possível provar (ver Lehmann, 1986, capítulo 6) que  $\overline{\mathcal{A}}(\mathcal{G}) = \tilde{\mathcal{A}}(\mathcal{G})$ , mas na seqüência mostraremos com um exemplo que  $\overline{\mathcal{A}}(\mathcal{G})$  pode ser bem menor em relação a  $\tilde{\mathcal{A}}(\mathcal{G})$ .

As definições anteriores também podem ser consideradas para funções  $f$  definidas sobre o espaço  $\mathcal{X}$ , e assumindo valores num espaço  $\mathcal{Y}$ , munido da subálgebra  $\mathcal{B}$

( $\alpha'$ ) a função  $f$  é  $\mathcal{G}$ -invariante, se  $f = f \circ g, \forall g \in \mathcal{G}$ ;

( $\beta'$ ) a função  $f$  é essencialmente  $\mathcal{G}$ -invariante, se  $f \sim$  alguma função  $\mathcal{G}$ -invariante;

( $\gamma'$ ) a função  $f$  é quase  $\mathcal{G}$ -invariante se  $f \sim f \circ g, \forall g \in \mathcal{G}$ , onde  $f \sim h$  denota  $f = h[\mathcal{P}]$ .

### Proposição 8.1

A função  $f$  satisfaz as definições ( $\alpha'$ ), ( $\beta'$ ) ou ( $\gamma'$ ) se e somente se  $f$  é mensurável em relação à correspondente subálgebra definida em ( $\alpha$ ), ( $\beta$ ) ou ( $\gamma$ ).

#### Prova

( $\alpha'$ )  $\Rightarrow$  ( $\alpha$ )

$f$   $\mathcal{G}$ -invariante  $\Leftrightarrow f = f \circ g \forall g \in \mathcal{G}$  então se  $B \in \mathcal{B}$   $f^{-1}(B) = g^{-1}(f^{-1}(B)) \forall g \in \mathcal{G}$   
 $\Leftrightarrow f^{-1}(B) \in \mathcal{A}(g) \forall g \in \mathcal{G} \Leftrightarrow f \in \mathcal{A}(\mathcal{G})$ .

( $\alpha$ )  $\Rightarrow$  ( $\alpha'$ )

$f = 1_A \in \mathcal{A}(\mathcal{G}) \Leftrightarrow 1_A \in \mathcal{A}(g) \forall g \in \mathcal{G} \Leftrightarrow A \in \mathcal{A}(g) \forall g \in \mathcal{G} \Leftrightarrow g^{-1}(A) = A \forall g \in \mathcal{G}$ .

Mas,  $1_A \circ g = 1_{g^{-1}(A)} \forall g \in \mathcal{G}$  portanto  $1_A \circ g = 1_A \forall g \in \mathcal{G}$ .

Considerar  $f \in \mathcal{A}(\mathcal{G})$ ,  $f \geq 0$  função simples; logo  $f \in \mathcal{A}(\mathcal{G})$ ,  $f \geq 0$  e finalmente  $f = f^+ - f^- \in \mathcal{A}(\mathcal{G})$  para concluir.  $\square$

### Exercício 8.3

- (a) Complete a demonstração da Proposição 8.1, isto é,  $(\beta) \Leftrightarrow (\beta')$  e  $(\gamma) \Leftrightarrow (\gamma')$ .
- (b) Mostre que  $\mathcal{A}(\mathcal{G}) \subset \overline{\mathcal{A}}(\mathcal{G}) \subset \tilde{\mathcal{A}}(\mathcal{G})$ .

## 8.2 INVARIÂNCIA E SUFICIÊNCIA

Suponhamos que temos dois sistemas diferentes de coordenadas para o resultado de uma experiência estatística, se um resultado é registrado como  $x$  sob o primeiro sistema, o mesmo resultado é registrado como  $g(x)$  no segundo sistema. Suponhamos também que as duas variáveis estatísticas  $x$  e  $g(x)$  possuem o mesmo domínio  $\mathcal{X}$ , a mesma família de eventos  $\mathcal{A}$ , e a mesma classe de medidas de  $\mathcal{P}$ . Além disso, se  $P \in \mathcal{P}$  está associada a  $x$ , então a mesma medida de probabilidade  $P$  também está associada a  $g(x)$ . O segundo sistema de coordenadas pode ser representado matematicamente como uma transformação  $g$  que preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

Assim, se  $g$  é uma transformação que preserva o modelo, o princípio de invariância nos leva a reduzir o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  a um modelo mais simples  $(\mathcal{X}, \mathcal{A}(g), \mathcal{P})$ , onde  $\mathcal{A}(g)$  é a subálgebra de conjuntos  $g$ -invariantes. De outra forma, o princípio de invariância estabelece que toda função de decisão deve ser  $g$ -invariante ou  $\mathcal{A}(g)$ -mensurável. Consideremos agora  $\mathcal{G}$ , a classe de todas transformações  $g$  que preservam o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Podemos exigir que toda regra de decisão seja  $\mathcal{G}$ -invariante, isto é,  $g$ -invariante  $\forall g \in \mathcal{G}$ ? Existindo exemplos (ver Exemplo 8.2) onde  $\mathcal{A}(\mathcal{G}) = \{\emptyset, \mathcal{X}\}$  então a resposta a nossa pergunta é negativa, não devemos reduzir  $\mathcal{A}$  a  $\mathcal{A}(\mathcal{G})$  ou a  $\overline{\mathcal{A}}(\mathcal{G})$ . Um compromisso lógico, com o princípio, deveria ser reduzir  $\mathcal{A}$  à subálgebra de conjuntos quase  $\mathcal{G}$ -invariantes,  $\tilde{\mathcal{A}}(\mathcal{G})$ , assim estamos exigindo que a função de decisão – nosso procedimento de inferência – seja quase  $\mathcal{G}$ -invariante.

Por outra parte, o princípio de suficiência é outro princípio de redução que é frequentemente utilizado na inferência estatística. Neste caso a subálgebra  $\mathcal{A}$  é reduzida a  $\mathcal{S}$ , onde  $\mathcal{S}$  é uma subálgebra suficiente. No caso de existir uma subálgebra suficiente mínima  $\mathcal{M}$ ,  $\mathcal{A}$  é reduzida a  $\mathcal{M}$ , isto é, reduzimos o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  ao modelo  $(\mathcal{X}, \mathcal{M}, \mathcal{P})$ .

Qual das duas reduções, invariância ou suficiência é mais extensiva? Em outras palavras, qual a relação entre  $\tilde{\mathcal{A}}(\mathcal{G})$  e  $\mathcal{M}$ ? Isto será o assunto desta seção, primeiramente

daremos dois exemplos ilustrativos.

### Exemplo 8.2

Consideremos  $\mathcal{X} = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}_1$  e  $\mathcal{P}$  a família de distribuições contínuas sobre  $\mathbb{R}$ .

Seja  $\mathcal{G} = \{g : \mathbb{R} \rightarrow \mathbb{R} / \{x : g(x) \neq x\} \text{ é finito } \}$ .

O leitor pode verificar que cada  $g \in \mathcal{G}$  preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  e que  $\mathcal{A}(\mathcal{G}) = \{\emptyset, \mathcal{X}\}$ . Logo  $\mathcal{A}(\mathcal{G})$  não é suficiente, a menos que  $\mathcal{P}$  seja um conjunto unitário.

Por outro lado, sendo  $A \Delta g^{-1}(A) = A \setminus g^{-1}(A)$ , pois  $g^{-1}(A) = A \setminus F$ , onde  $F$  é um subconjunto de  $\{x \in \mathbb{R} : g(x) \neq x\}$ , tem-se que  $P(A \Delta g^{-1}(A)) = 0$  para todo  $A \in \mathcal{A}$  e toda função  $g \in \mathcal{G}$ . Conclui-se que  $\tilde{\mathcal{A}}(\mathcal{G}) = \mathcal{A}$ , que obviamente é suficiente.

Este exemplo mostra também que  $\mathcal{A}(\mathcal{G})$  pode ser bem menor em relação a  $\tilde{\mathcal{A}}(\mathcal{G})$ .  $\square$

### Exemplo 8.3

Consideremos  $\mathcal{X} = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}_{[0,1]}$  e  $\mathcal{P}$  constituído só pela medida de Lebesgue em  $[0, 1]$ .

Seja  $\mathcal{G} = \{g_c : [0, 1] \rightarrow [0, 1] / 0 < c < 1\}$  onde  $g_c(x) = (x + c) \bmod 1 \forall x \in [0, 1]$ , isto é:

$$g_c(x) = \begin{cases} x + c & , \text{ se } x + c \leq 1 \\ x + c - 1 & , \text{ se } x + c > 1 \text{ onde } 0 < c < 1. \end{cases}$$

É claro que se  $A \subset [0, c)$  então  $g_c^{-1}(A) \subset [c, 1)$ , e quando  $A \subset (c, 1)$  tem-se que  $g_c^{-1}(A) \subset [0, c)$ . No caso de  $A$  ser da forma  $A = A_1 \cup A_2$ , onde  $A_1 \subset [0, c)$  e  $A_2 \subset [c, 1)$  tem-se que  $g_c^{-1}(A) \subset A$ , a menos que  $A_1 = [0, c)$  e  $A_2 = [c, 1)$  em cujo caso  $A = [0, 1]$ , logo  $g_c^{-1}([0, 1]) = [0, 1]$ . Assim,  $\mathcal{A}(\mathcal{G}) = \{\emptyset, [0, 1]\}$ , e sendo  $\mathcal{P}$  unitário, segue que  $\mathcal{A}(\mathcal{G})$  é uma subálgebra suficiente.

Basu (1965,1969) mostrou relações existentes entre as subálgebras suficiente e invariante. Também observou que esses resultados podem ser deduzidos do Teorema Ergódico de Birkhoff. Na seqüência apresentamos em detalhe os teoremas pertinentes.

Começamos com um lema auxiliar, cuja prova é deixada como exercício para o leitor.

#### Lema 8.1

Seja  $S$  uma subálgebra suficiente e limitadamente completa.

Se  $\varphi \in \mathcal{A}$  é limitada [na verdade,  $\varphi \in \mathcal{L}_\infty(\mathcal{X}, \mathcal{A}, \mathcal{P})$ ] com  $E_P\{\varphi\} = 0, \forall P \in \mathcal{P}$ , então

$$\forall f \in S \text{ e limitada [na verdade } f \in \mathcal{L}_\infty(\mathcal{X}, S, \mathcal{P})]$$

tem-se que  $E_P\{\varphi f\} = 0, \forall P \in \mathcal{P}$ .

**Prova**

Exercício 8.4. □

**Teorema 8.1**

Seja  $\mathcal{S}$  é uma subálgebra suficiente e limitadamente completa então  $\mathcal{S} \subset \tilde{\mathcal{A}}(\mathcal{G})$ .

**Prova**

Sejam  $S \in \mathcal{S}, g \in \mathcal{G}$  e  $S_0 = g^{-1}(S)$  então  $P(S) = P(S_0) \forall P \in \mathcal{P}$ .

Escrevamos  $1_S$  função indicadora do conjunto  $S$ , logo  $E_P(1_S - 1_{S_0}) = 0 \forall P \in \mathcal{P}$ . Aplicando o Lema 8.1 com  $\varphi = 1_S - 1_{S_0}$  e  $f = 1_S$  obtemos  $E_P\{(1_S - 1_{S_0}) \cdot 1_S\} = 0 \forall P \in \mathcal{P}$ , isto é,  $P(S) = P(S \cap S_0) \forall P \in \mathcal{P}$ .

Agora, calculamos  $P(S \Delta S_0)$  para concluir o resultado.

$$\begin{aligned} P(S \Delta S_0) &= P(S) + P(S_0) - 2P(S \cap S_0) \\ &= 2[P(S) - P(S \cap S_0)], \text{ pois } P(S) = P(S_0) \\ &= 0 \end{aligned}$$

Assim,  $\forall S \in \mathcal{S}$  os conjuntos  $S$  e  $g^{-1}(S)$  são  $\mathcal{P}$ -equivalentes  $\forall g \in \mathcal{G}$ , portanto  $\mathcal{S} \subset \bigcap_{g \in \mathcal{G}} \bar{\mathcal{A}}(g) = \tilde{\mathcal{A}}(\mathcal{G})$ . □

**Observação 8.1**

O teorema anterior estabelece que o princípio de suficiência reduz mais que o princípio de invariância.

No que segue consideramos o *caso dominado*, que merece uma atenção especial.

Sejam  $T : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{B})$  uma estatística,  $P$  e  $Q$  medidas de probabilidade induzidas por  $T$  em  $(\mathcal{Y}, \mathcal{B})$ .

Suponhamos que  $P \ll Q$ , e logo  $PT^{-1} \ll QT^{-1}$ . Denotemos por  $f = \frac{dP}{dQ}, h = \frac{dPT^{-1}}{dQT^{-1}}$ , as respectivas versões das derivadas de Radon-Nikodym. A função  $hT$  (definida como  $hT(x) = h(T(x))$ ) é  $\mathcal{A}_T$ -mensurável e satisfaz a seguinte relação

**Lema 8.2**

$$hT = E_Q(f|A_T) .$$

**Prova**

Seja  $B \in \mathcal{B}$  e logo

$$\int_{T^{-1}(B)} hT dQ = \int_B h dQT^{-1} = \int_B dPT^{-1} = \int_{T^{-1}(B)} dP = \int_{T^{-1}(B)} f dQ . \quad \square$$

**Corolário**

Se  $T^{-1}(\mathcal{B}) = \mathcal{A}[Q]$  então  $f = hT[Q]$  .  $\square$

O teorema seguinte mostra que no caso dominado tem-se que  $\tilde{\mathcal{A}}(\mathcal{G})$  é suficiente.

**Teorema 8.2**

Se  $\mathcal{P} \ll \mu$ , onde  $\mu$  é  $\sigma$ -finita, então  $\tilde{\mathcal{A}}(\mathcal{G})$  é suficiente.

**Prova**

Dado que  $\mathcal{P} \ll \mu$ , sendo  $\mu$   $\sigma$ -finita, é conhecido (cf. capítulo 3) que existe uma seqüência  $(P_n)_n$  em  $\mathcal{P}$  tal que se  $Q = \sum_{n \in \mathbb{N}} c_n P_n$ , onde  $c_n > 0$  e  $\sum_{n \in \mathbb{N}} c_n = 1$ , então  $\mathcal{P} \ll Q$ .

Seja  $f_P = \frac{dP}{dQ}$ ,  $P \in \mathcal{P}$ , e mostremos que  $f_P \in \tilde{\mathcal{A}}(\mathcal{G})$ .

Agora,  $Pg^{-1} = P \forall P \in \mathcal{P}$  e então  $Qg^{-1} = Q$ , portanto  $Pg^{-1} \ll Qg^{-1}$ .

Aplicando o Lema 8.2 com  $f = f_P$  e  $T = g$ , temos que:

$$f_P \circ g = \frac{dP}{dQ} \circ g = \frac{dPg^{-1}}{dQg^{-1}} \circ g = E_Q(f_P/\mathcal{A}_g) = E_Q(f_P/\mathcal{A}) = f_P[Q] ,$$

pois  $g^{-1}(\mathcal{A}) = \mathcal{A}$ , onde  $\mathcal{A}_g$  é subálgebra induzida pela transformação  $g$ .

Segue que  $f_P \circ g = f_P[Q]$ , mas  $P \ll Q \forall P \in \mathcal{P}$ , logo  $f_P \circ g = f_P[P] \forall P \in \mathcal{P}$ ,  $\forall g \in \mathcal{G}$ , isto é,  $f_P \in \tilde{\mathcal{A}}(\mathcal{G})$   $\square$

Os argumentos seguintes mostram que os Teoremas 8.1 e 8.2 podem ser deduzidos do Teorema Ergódico de Birkhoff (1931), o qual estabelecemos a continuação, mas na linguagem da Teoria Estatística.

### Teorema 8.3

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  um espaço estatístico, e seja  $f \in \mathcal{A}$ ,  $f$  limitada [na verdade  $f \in \mathcal{L}_\infty(\mathcal{X}, \mathcal{A}, \mathcal{P})$ ].

Sejam  $g : \mathcal{X} \rightarrow \mathcal{X}$ , uma transformação que preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  e  $\mathcal{A}(g)$  a correspondente subálgebra de conjuntos invariantes.

Seja  $(f_n)_n$  a seqüência de funções definida sobre  $\mathcal{X}$ , onde  $f_n(x) = \frac{1}{n+1} \sum_{k=0}^n (f \circ g^k)(x)$ , onde  $g^k = g \circ g^{k-1}$ ,  $k = 1, \dots, n$ , sendo  $g^0$  a função identidade.

Então  $f_n \rightarrow f^* = E_P\{f/\mathcal{A}(g)\} [P] \forall P \in \mathcal{P}$ .

#### Prova

Ver por exemplo Mañé (1983). □

O teorema anterior estabelece a existência de uma versão universal de  $E_P\{f/\mathcal{A}(g)\}$ ,  $\forall P \in \mathcal{P}$ , que corresponde ao limite ( $P$ -q.c.) da seqüência  $(f_n)_n$ . Portanto, se  $\mathcal{M}$  é uma subálgebra suficiente mínima para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  então  $\mathcal{M} \subset \mathcal{A}(g) \forall g \in \mathcal{G}$  e logo  $\overline{\mathcal{M}} \subset \overline{\mathcal{A}}(g) \forall g \in \mathcal{G}$ . Tendo-se o seguinte resultado.

### Teorema 8.4

Seja  $\mathcal{M}$  uma subálgebra suficiente mínima então

$$\overline{\mathcal{M}} \subset \tilde{\mathcal{A}}(\mathcal{G}) = \bigcap_{g \in \mathcal{G}} \overline{\mathcal{A}}(g) . \quad \square$$

Agora voltamos aos Teoremas 8.1 e 8.2.

No Teorema 8.1, considerando que  $\mathcal{S}$  é suficiente e limitadamente completa então  $\mathcal{S}$  é suficiente mínima e portanto esse teorema é um corolário do Teorema 8.4.

No Teorema 8.2, tratando-se do caso dominado, o fato  $\mathcal{M} \subset \tilde{\mathcal{A}}(\mathcal{G})$  implica que  $\tilde{\mathcal{A}}(\mathcal{G})$  é suficiente.

## 8.2.1 UMA APLICAÇÃO EM ESPAÇOS ESTATÍSTICOS NORMAIS

Esta seção é dedicada a estabelecer alguns resultados em espaços estatísticos normais. Primeiramente caracterizamos a subálgebra  $\mathcal{A}(\mathcal{G})$  de conjuntos  $\mathcal{G}$ -invariantes em termos



da chamada função invariante maximal. Para detalhes desse conceito o leitor é referido ao livro de Berger (1985, Capítulo 6)

### Definição 8.3

Seja  $\mathcal{G}$  um grupo de transformações de  $\mathcal{X}$  em  $\mathcal{X}$ .

Uma função  $T : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{B})$   $\mathcal{A}$ - $\mathcal{B}$  mensurável é dita invariante maximal se

- $T$  é invariante sob  $\mathcal{G}$ , e se satisfaz
- $T(x) = T(y) \Rightarrow y = g(x)$  para alguma  $g \in \mathcal{G}$ .

O teorema seguinte caracteriza as funções invariantes através de funções invariantes maximais.

### Teorema 8.5

Seja  $T : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{B})$  uma função invariante maximal sob  $\mathcal{G}$ . Então uma condição necessária e suficiente para que  $S : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{B})$  seja uma função invariante é que  $S = \varphi \circ T$  para alguma transformação  $\varphi$   $\mathcal{B}$ -mensurável.

### Prova

Suficiência.

Se  $S(x) = \varphi(T(x)) \forall x \in \mathcal{X}$  então  $S(g(x)) = \varphi(T(g(x))) = \varphi(T(x)) = S(x), \forall x \in \mathcal{X}$ .

Necessidade

Se  $S$  é invariante sob  $\mathcal{G}$  e  $T(x) = T(y) \Rightarrow y = g(x)$  para alguma  $g \in \mathcal{G}$ , então  $S(x) = S(g(x)) = S(y)$ , e isto implica que  $S$  é uma função de  $T$  □

O teorema seguinte estabelece que a subálgebra de conjuntos  $\mathcal{G}$ -invariantes  $\mathcal{A}(\mathcal{G})$  corresponde à subálgebra induzida pela função invariante maximal sob  $\mathcal{G}$ .

### Teorema 8.6

Seja  $\mathcal{G}$  um grupo de transformações que preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

Seja  $T$  uma função invariante maximal sob  $\mathcal{G}$ .

Então  $\mathcal{A}_T = \mathcal{A}(\mathcal{G})$ .

### Prova

$$\mathcal{A}_T \subset \mathcal{A}(\mathcal{G})$$

Seja  $T^{-1}(B) \in \mathcal{A}_T$ , sendo  $T$  invariante sob  $\mathcal{G}$ , segue que  $g^{-1}(T^{-1}(B)) = (T \circ g)^{-1}(B) = T^{-1}(B)$ ,  $\forall g \in \mathcal{G}$ . Portanto  $T^{-1}(B) \in \mathcal{A}(\mathcal{G})$ .

$$\mathcal{A}(\mathcal{G}) \subset \mathcal{A}_T$$

Se  $f \in \mathcal{A}(\mathcal{G})$ , isto é  $f$  é uma função invariante sob  $\mathcal{G}$  e  $T$  uma função invariante maximal sob  $\mathcal{G}$ , o Teorema 8.5 garante que  $f = \varphi \circ T$  para alguma transformação  $\varphi \in \mathcal{B}$ , isto é  $f \in \mathcal{A}_T$ . A conclusão segue considerando  $f = 1_A$ , a função indicadora do conjunto  $A \in \mathcal{A}(\mathcal{G})$  □

### Exemplo 8.4

Aplicação em espaços estatísticos normais.

Consideramos  $X_1, \dots, X_n$  variáveis aleatórias independentes com distribuição comum  $N(0, \sigma^2)$ .

$$\mathcal{G}_0 = \{g : \mathbb{R}^n \rightarrow \mathbb{R}^n / g(\mathbf{x}) = A\mathbf{x}, A'A = AA' = I_n\},$$

onde  $A'$  é a matriz transposta de  $A$  e  $I_n$  é a  $n \times n$ -matriz identidade, isto é,  $\mathcal{G}_0$  é o grupo das transformações lineares ortogonais.

### Exercício 8.5

Mostre que  $\mathcal{G}_0$  é um grupo.

Neste caso  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A} = \mathcal{B}_n$  e  $\mathcal{P} = \{P_{\sigma^2} : \sigma^2 > 0\} \ll \lambda_n$ , onde  $\frac{dP_{\sigma^2}}{d\lambda_n} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n x_i^2\right\}$ , sendo  $\lambda_n$  a medida de Lebesgue em  $\mathbb{R}^n$ .

É claro que se  $g \in \mathcal{G}_0$  então  $g$  preserva o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  pois,

$$\begin{aligned} P_{\sigma^2}(g^{-1}(B)) &= \int_{g^{-1}(B)} dP_{\sigma^2} \\ &= \int_B dP_{\sigma^2} \circ g^{-1} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_B \exp\left\{\frac{-1}{2\sigma^2} \|g^{-1}(\mathbf{x})\|^2\right\} d\lambda_n(\mathbf{x}) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_B \exp\left\{\frac{-1}{2\sigma^2} \|\mathbf{x}\|^2\right\} d\lambda_n(\mathbf{x}) \\ &= P_{\sigma^2}(B), \quad \forall B \in \mathcal{B}_n, \forall \sigma^2 > 0, \end{aligned}$$

onde  $\|\cdot\|$  denota a norma euclidiana em  $\mathbb{R}^n$ , a qual é preservada por transformações ortogonais.

### Exercício 8.6

Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}$ , definida por  $T(\mathbf{x}) = \|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$ , onde  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ .

Use o Teorema da Fatoração para mostrar que  $T$  é uma estatística suficiente para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .

### Exercício 8.7

Mostre que a família  $\mathcal{P} = \{P_{\sigma^2} : \sigma^2 > 0\}$  é autodominada.

Seja  $\sigma_0^2 > 0$  tal que  $\mathcal{P} \ll P_{\sigma_0^2}$ , e  $q_{\sigma^2} = \frac{dP_{\sigma^2}}{dP_{\sigma_0^2}}$  então  $q_{\sigma^2}$  é uma função 1-1 de  $T$ , logo conclua que  $\mathcal{A}_T = \mathcal{C}_0 = \sigma\{q_{\sigma^2} : \sigma^2 > 0\}$  e portanto  $\mathcal{A}_T$  é uma subálgebra suficiente mínima.

Mostremos que  $T$  é invariante maximal sob  $\mathcal{G}_0$ .

É claro que  $T$  é invariante, pois  $T(g(\mathbf{x})) = \|g(\mathbf{x})\|^2 = \|\mathbf{x}\|^2 \forall \mathbf{x} \in \mathcal{X}, \forall g \in \mathcal{G}_0$ .

Agora,  $T(\mathbf{x}) = T(\mathbf{y}) \Leftrightarrow \|\mathbf{x}\|^2 = \|\mathbf{y}\|^2$  e sejam  $\{\frac{\mathbf{x}}{\|\mathbf{x}\|}, v_2, \dots, v_n\}$  e  $\{\frac{\mathbf{y}}{\|\mathbf{y}\|}, v'_2, \dots, v'_n\}$  bases ortonormais de  $\mathbb{R}^n$  e  $g$  a transformação linear definida por  $g\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) = \frac{\mathbf{y}}{\|\mathbf{y}\|}$ ,  $g(v_i) = v'_i, i = 1, \dots, n$ . Segue que  $g$  é uma transformação ortogonal.

O Teorema 8.5 garante que  $\mathcal{A}_T = \mathcal{A}(\mathcal{G}_0)$ , e do Exercício 8.7 temos que se  $\mathcal{M}$  é uma subálgebra suficiente mínima então  $\mathcal{M} = \mathcal{A}(\mathcal{G}_0)$ .

Agora consideremos  $\mathcal{G}$  a classe de todas as transformações que preservam o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  então do Teorema 8.4  $\overline{\mathcal{M}} \subset \tilde{\mathcal{A}}(\mathcal{G})$ .

Mas, é possível mostrar que  $\mathcal{A}(\mathcal{G}_0) = \tilde{\mathcal{A}}(\mathcal{G}_0) [\mathcal{P}]$  (cf. Lehmann, 1986, capítulo 6; ver também Bondar e Milnes, 1981) e logo  $\mathcal{M} = \tilde{\mathcal{A}}(\mathcal{G}_0) [\mathcal{P}]$ , implicando que  $\overline{\mathcal{M}} = \tilde{\mathcal{A}}(\mathcal{G}_0)$ .

Finalmente dado que  $\tilde{\mathcal{A}}(\mathcal{G}) \subset \tilde{\mathcal{A}}(\mathcal{G}_0)$  então  $\tilde{\mathcal{A}}(\mathcal{G}) = \overline{\mathcal{M}}$ , isto é, o completamento de qualquer versão da subálgebra suficiente mínima coincide com a subálgebra de conjuntos quase  $\mathcal{G}$ -invariantes.

### Exercício 8.8

Considere o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , onde  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A} = \mathcal{B}_n$  e  $\mathcal{P} = \{P_{\boldsymbol{\mu}, \sigma^2} : \boldsymbol{\mu} \in \mathbb{R}^n; \sigma^2 > 0\}$ , onde  $P_{\boldsymbol{\mu}, \sigma^2}(B) = \int_B \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_i)^2\right\} d\lambda_n(\mathbf{x})$ , onde  $\mathbf{a} = (a_1, \dots, a_n)$ .

Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}^2$ , definida por  $T(\mathbf{x}) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$ ,

Considere  $\mathcal{G}$  a classe de todas as transformações que preservam o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , então  $\mathcal{A}_T$  é suficiente mínima,  $T$  é invariante maximal e  $\overline{\mathcal{A}}_T = \tilde{\mathcal{A}}(\mathcal{G})$  (para detalhes, ver Basu, 1969).

### 8.3 INVARIÂNCIA E ANCILARIDADE.

Esta seção é dedicada fundamentalmente a estudar as relações existentes entre os conceitos de invariância e ancilaridade. Primeiramente introduzimos o conceito de ancilaridade no contexto da Teoria Estatística e alguns exemplos em espaços estatísticos dominados são dados. Vemos também que utilizando um Teorema de Basu (1955, 1958) é possível estabelecer de maneira muito simples independência entre estatísticas definidas sobre certos modelos estatísticos paramétricos.

Seja  $T : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{B})$  uma estatística, e seja  $\mathcal{P}$  uma família de medidas de probabilidade sobre  $(\mathcal{X}, \mathcal{A})$ , isto é,  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  constitui um modelo estatístico.

É claro que  $T$  gera de maneira natural uma família de medidas de probabilidade  $\mathcal{Q}$  sobre  $(\mathcal{Y}, \mathcal{B})$ , tal que

$$Q(B) = PT^{-1}(B) = (P \circ T^{-1})(B) \quad \forall B \in \mathcal{B}, Q \in \mathcal{Q}.$$

Naturalmente, um problema clássico é dados  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ ,  $(\mathcal{Y}, \mathcal{B})$  e  $T$ , determinar a família  $\mathcal{Q}$ . Mas, o conceito de ancilaridade tem a ver com um problema específico: Qual a classe de estatísticas  $T$  tal que  $\mathcal{Q}$  está formada por apenas uma medida de probabilidade? De existir essa classe de estatísticas, os elementos dessa classe são ditos estatísticas ancilares.

Especificamente, se  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ , quais as estatísticas  $T$  tais que  $P_\theta T^{-1} = Q \forall \theta \in \Theta$ ? Seria nossa pergunta no caso de considerar um modelo estatístico paramétrico.

Um outro assunto de nosso interesse é mostrar que em muitos modelos paramétricos, os quais estão caracterizados por parâmetros de locação e/ou parâmetros de escala, a pro-

priedade de ancilaridade de uma determinada estatística pode ser estabelecida de maneira direta, utilizando alguns argumentos de invariância sobre certos grupos de transformações.

A definição rigorosa do conceito de ancilaridade é dada a seguir

#### Definição 8.4

Sejam  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  um modelo estatístico, e  $T$  uma estatística de  $(\mathcal{X}, \mathcal{A})$  em  $(\mathcal{Y}, \mathcal{B})$ . A estatística  $T$  é ancilar em relação ao modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , se  $T$  induz apenas uma medida de probabilidade no espaço  $(\mathcal{Y}, \mathcal{B})$ .

Em particular se  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ ,  $T$  é ancilar se  $P_\theta T^{-1} = Q \forall \theta \in \Theta$ , isto é, a medida induzida pela estatística  $T$  é constante em relação ao parâmetro  $\theta$ .

Naturalmente a subálgebra induzida por  $T$ :  $\mathcal{A}_T$ , é ancilar, se  $T$  é uma estatística ancilar.

Um conjunto  $A \in \mathcal{A}$  é dito ancilar, se  $P_\theta(A)$  é constante em relação ao parâmetro  $\theta$ .

Os exemplos seguintes tratam com a caracterização de conjuntos e estatísticas ancilares num modelo estatístico dominado.

#### Exemplo 8.5

Caracterização de conjuntos ancilares em espaços estatísticos normais. Suponhamos que  $X$  é uma variável aleatória normal com média 0 e variância  $\sigma^2$ , isto é, estamos considerando o espaço estatístico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , sendo  $\mathcal{X} = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}_1$  e  $\mathcal{P} = \{P_{\sigma^2} : \sigma^2 > 0\}$ , onde  $P_{\sigma^2}(B) = \int_B \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left\{\frac{-x}{2\sigma^2}\right\} d\lambda(x)$ ,  $B \in \mathcal{B}_1$ ,  $\sigma^2 > 0$ .

Nosso objetivo é caracterizar os conjuntos ancilares  $A \in \mathcal{B}_1$ , isto é, os conjuntos  $A \in \mathcal{B}_1$ , tais que  $P_{\sigma^2}(A) = \alpha$ , onde  $\alpha$  é uma constante em relação a  $\sigma^2$ .

Mostremos que  $\alpha$  só pode assumir os valores 0,  $\frac{1}{2}$  ou 1 e que no caso  $\alpha = \frac{1}{2}$ ,  $A$  é um conjunto *skew-simétrico* em relação à origem, isto é, um e só um dos fatos seguintes ocorre,  $x \in A$  ou  $-x \in A$

#### Exercício 8.9

Use o Teorema de Fatoração para mostrar que a estatística  $S(x) = |x|$  é suficiente.

$S$  é completa, isto é, a família de medidas induzida por  $S$  é completa. Notamos que  $S : (\mathbb{R}, \mathcal{B}_1, \mathcal{P}) \rightarrow (\mathbb{R}, \mathcal{B}_1, \mathcal{Q})$  onde se  $Q_{\sigma^2} \in \mathcal{Q}$  então  $Q_{\sigma^2}(B) = \int_B \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-x}{2\sigma^2}\right\} d\lambda(x)$ ,  $B \in \mathcal{B}_1$ ,  $\sigma^2 > 0$ .

Seja  $g \in \mathcal{B}_1$ ,

$$E_{Q_{\sigma^2}}(g) \stackrel{=}{=}_{\forall \sigma^2 > 0} 0 \Rightarrow \int_{[0, \infty)} g(x) \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-x}{2\sigma^2}\right\} d\lambda(x) = 0, \quad \forall \sigma^2 > 0.$$

fazendo  $x^2 = u$  e  $s = \frac{1}{2\sigma^2} > 0$  temos que  $\int_{[0, \infty)} \frac{g(\sqrt{u})}{2\sqrt{u}} \exp\{-su\} d\lambda(u) = 0 \forall s > 0$ . Pela unicidade (essencial) da transformada de Laplace conclui-se que  $g = 0$   $[\lambda]$  e dado que  $\mathcal{Q} \ll \lambda$  então  $g = 0$   $[\mathcal{Q}]$ .

Seja  $A \in \mathcal{B}_1$  um conjunto ancilar, tal que  $P_{\sigma^2}(A) = \alpha \forall \sigma^2 > 0$ . Um Teorema de Basu (1955) garante que  $A \perp \mathcal{A}_S$  e logo  $P_{\sigma^2}(A/S) = P_{\sigma^2}(A/\mathcal{A}_S) = \alpha \forall \sigma^2 > 0$

Mas,

$$P_{\sigma^2}(A/|X| = u) = \begin{cases} 0 & \text{se } u, -u \notin A \\ \frac{1}{2} & \text{se } u \in A \text{ e } -u \notin A \\ 1 & \text{se } u, -u \in A \end{cases}$$

e logo um conjunto ancilar  $A$  não trivial tem probabilidade  $\alpha = \frac{1}{2}$ , e  $A$  deve ser *skew-simétrico* em relação à origem.

Inversamente, se  $A$  é um conjunto *skew-simétrico* em relação à origem então  $P_{\sigma^2}(A) = \frac{1}{2} \forall \sigma^2 > 0$ , isto é,  $A$  é ancilar.

No exemplo anterior podemos considerar um modelo estatístico mais geral, isso é estabelecido na proposição seguinte.

### Proposição 8.2

Seja  $(\mathbb{R}, \mathcal{B}_1, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta \in \mathbb{R}\} \ll \lambda$ ,  $\lambda$   $\sigma$ -finita e  $\mathcal{P}$  é uma família simétrica em relação à origem, isto é,  $P_\theta(A) = P_\theta(-A) \forall A \in \mathcal{B}_1, \theta \in \mathbb{R}$ .

Suponhamos que a estatística  $S$  de  $\mathbb{R}$  em  $[0, \infty)$ , definida por  $S(x) = |x|$  é suficiente e limitadamente completa para  $(\mathbb{R}, \mathcal{B}_1, \mathcal{P})$ .

Então se  $A \in \mathcal{B}_1$  é um conjunto ancilar não-trivial, tem-se que  $P_\theta(A) = \frac{1}{2}, \forall \theta$  e  $A$  é essencialmente um conjunto *skew-simétrico* em relação à origem.

Inversamente, se  $A \in \mathcal{B}_1$  é um conjunto *skew-simétrico* em relação à origem então  $A$  é ancilar com  $P_\theta(A) = \frac{1}{2} \forall \theta$  □

### Exercício 8.10

No modelo descrito na proposição 8.3.1, mostre que uma estatística  $T$  é ancilar para esse modelo se e somente se  $T$  induz essencialmente uma partição de  $\mathbb{R}$  em dois conjuntos *skew-simétricos* em relação à origem.

Na seqüência apresentamos alguns resultados relativos a invariância e ancilaridade, em modelos com parâmetros de locação e/ou de escala. Especificamente, mostramos que a propriedade de ancilaridade de uma determinada estatística em relação ao parâmetro de locação e/ou escala pode ser estabelecida utilizando argumentos de invariância sobre certos grupos de transformações de locação e/ou escala.

Consideramos  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A} = \mathcal{B}_n$  e  $\mathcal{P} = \{P_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ , onde

$$\frac{dP_{\mu,\sigma}}{d\lambda_n}(\mathbf{x}) = \frac{1}{\sigma^n} f\left(\frac{x_1 - \mu}{\sigma}, \dots, \frac{x_n - \mu}{\sigma}\right), \quad \mathbf{x} = (x_1, \dots, x_n),$$

sendo  $f(\mathbf{x}) = \frac{dP}{d\lambda_n}(\mathbf{x})$ , onde  $P$  é uma medida de probabilidade sobre  $(\mathcal{X}, \mathcal{A})$ .

Nas condições anteriores,  $\mu$  é dito *parâmetro de locação* e  $\sigma$  é dito *parâmetro de escala*, e o modelo anterior é dito *modelo de locação-escala*.

### Proposição 8.3

Seja  $\mathcal{G} = \{g_{a,b} : \mathbb{R}^n \rightarrow \mathbb{R}^n / a \in \mathbb{R}, b > 0\}$ , uma família de transformações, onde  $g_{a,b}(x_1, \dots, x_n) = \left(\frac{x_1 - a}{b}, \dots, \frac{x_n - a}{b}\right)$ ;  $a \in \mathbb{R}, b > 0$ .

Então  $(\mathcal{G}, \circ)$  é um grupo de transformações, onde  $\circ$  denota composição de funções.

### Prova

Notemos que  $g_{a,b} \circ g_{c,d} = g_{ad+bc,bd}$ ;  $a, d \in \mathbb{R}$ ;  $b, d > 0$ . É conhecido que a composição de funções é associativa.

O elemento neutro de  $\mathcal{G}$  é  $g_{0,1}$ , isto é, a função identidade. Finalmente,  $g_{a,b}^{-1}$ , a inversa de  $g_{a,b}$  é dada por  $g_{a,b}^{-1}(\mathbf{x}) = g_{\frac{-a}{b}, \frac{1}{b}}(\mathbf{x}) = (bx_1 + a, \dots, bx_n + a)$ , onde  $\mathbf{x} = (x_1, \dots, x_n)$ .  $\square$

### Observação 8.2

O grupo de transformações  $\mathcal{G} = \{g_{a,b} : \mathbb{R}^n \rightarrow \mathbb{R}^n / a \in \mathbb{R}, b > 0\}$ , onde  $g_{a,b}(x_1, \dots, x_n) = \left(\frac{x_1 - a}{b}, \dots, \frac{x_n - a}{b}\right)$ ;  $a \in \mathbb{R}, b > 0$ , é chamado *grupo de locação-escala*.

Analogamente, o grupo formado pelas transformações  $g_b(x_1, \dots, x_n) = \left(\frac{x_1}{b}, \dots, \frac{x_n}{b}\right)$ ,  $b > 0$ , é dito *grupo de escala*.

### Teorema 8.7

Consideremos o modelo de locação-escala  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ , onde  $\mathcal{P} = \{P_{\mu, \sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ , sendo

$$\frac{dP_{\mu, \sigma}}{d\lambda_n}(\mathbf{x}) = \frac{1}{\sigma^n} f\left(\frac{x_1 - \mu}{\sigma}, \dots, \frac{x_n - \mu}{\sigma}\right), \quad f(\mathbf{x}) = \frac{dP(\mathbf{x})}{d\lambda_n},$$

sendo  $P$  uma medida de probabilidade sobre  $(\mathbb{R}^n, \mathcal{B}_n)$ .

Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  uma estatística invariante sob o grupo de locação-escala  $\mathcal{G} = \{g_{a,b} : a \in \mathbb{R}, b > 0\}$ , isto é,  $T$  é uma função  $\mathcal{G}$ -invariante para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ . Então  $T$  é uma estatística ancilar para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ .

### Prova

Sejam  $B \in \mathcal{B}_n$  e  $Q = P_{\mu, \sigma} T^{-1}$ ;  $\mu \in \mathbb{R}, \sigma > 0$ , a medida induzida pela estatística  $T$  segue

$$\begin{aligned} Q(B) &= P_{\mu, \sigma}(T^{-1}(B)) = \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(g_{\mu, \sigma}(\mathbf{x})) d\lambda_n(\mathbf{x}) \\ &= \int_{(T \circ g_{\mu, \sigma})^{-1}(B)} \frac{1}{\sigma^n} f(g_{\mu, \sigma}(\mathbf{x})) d\lambda_n(\mathbf{x}) = \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(\mathbf{x}) d\lambda_n(\mathbf{x}) g_{\mu, \sigma}^{-1}(\mathbf{x}) \\ &= \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(\mathbf{x}) d\lambda_n(\sigma x_1 + \mu, \dots, \sigma x_n + \mu) = \int_{T^{-1}(B)} f(\mathbf{x}) d\lambda_n(\mathbf{x}) \end{aligned}$$

portanto  $Q(B)$  é constante em relação a  $(\mu, \sigma)$ ;  $\mu \in \mathbb{R}, \sigma > 0, \forall B \in \mathcal{B}_n$ .  $\square$

Em algumas ocasiões, o estatístico está interessado apenas no parâmetro de locação [respectivamente parâmetro de escala] o o parâmetro presente no modelo é o chamado *parâmetro nuisance*. Nesses casos uma pequena modificação no teorema anterior é feita, o qual é estabelecido na proposição seguinte.

### Proposição 8.4

(a) Consideremos o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ , onde  $\mathcal{P} = \{P_{\mu, \sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ , sendo

$$\frac{dP_{\mu, \sigma}}{d\lambda_n}(\mathbf{x}) = f_{\sigma}(x_1 - \mu, \dots, x_n - \mu), \quad \text{e} \quad f_{\sigma}(\mathbf{x}) = \frac{dP_{\sigma}}{d\lambda_n}(\mathbf{x}),$$

$P_{\sigma}$  medida de probabilidade sobre  $(\mathbb{R}^n, \mathcal{B}_n)$ , isto é, trata-se de um modelo de locação com parâmetro *nuisance*.



Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  uma estatística invariante sob o grupo de locação  $\mathcal{G} = \{g_a : \mathbb{R}^n \rightarrow \mathbb{R}/a \in \mathbb{R}\}$ , onde  $g_a(x_1, \dots, x_n) = (x_1 - a, \dots, x_n - a)$ ,  $a \in \mathbb{R}$ .

Então fixado  $\sigma = \sigma_0 > 0$ , tem-se que  $T$  é uma estatística ancilar para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0})$ , onde  $\mathcal{P}_{\sigma_0} = \{P_{\mu, \sigma_0} : \mu \in \mathbb{R}\}$ .

(b) Consideremos o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ , onde  $\mathcal{P} = \{P_{\mu, \sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ , sendo

$$\frac{dP_{\mu, \sigma}}{d\lambda_n}(\mathbf{x}) = f_{\mu}\left(\frac{x_1}{\sigma}, \dots, \frac{x_n}{\sigma}\right), \quad \text{e} \quad f_{\mu}(\mathbf{x}) = \frac{dP_{\mu}}{d\lambda_n}(\mathbf{x}),$$

$P_{\mu}$  medida de probabilidade sobre  $(\mathbb{R}^n, \mathcal{B}_n)$ , isto é, trata-se de um modelo de escala com parâmetro *nuisance*.

Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  uma estatística invariante sob o grupo de escala  $\mathcal{G} = \{g_b : \mathbb{R}^n \rightarrow \mathbb{R}/b > 0\}$ , onde  $g_b(x_1, \dots, x_n) = (\frac{x_1}{b}, \dots, \frac{x_n}{b})$ ,  $b > 0$ .

Então, fixado  $\mu = \mu_0$ , tem-se que  $T$  é uma estatística ancilar para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\mu_0})$ , onde  $\mathcal{P}_{\mu_0} = \{P_{\mu_0, \sigma} : \sigma > 0\}$ .

## Prova

Análoga à do Teorema 8.7. □

O exemplo seguinte é um resultado clássico na Estatística.

## Exemplo 8.6

Consideramos uma amostra aleatória simples  $X_1, \dots, X_n$  proveniente de uma distribuição  $N(\mu, \sigma^2)$ , isto é, as variáveis aleatórias  $X_1, \dots, X_n$  são estatisticamente independentes com distribuição comum  $N(\mu, \sigma^2)$ .

Vamos mostrar que  $\bar{X} \perp\!\!\!\perp S^2 / (\mu, \sigma^2)$ , onde  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  é a média amostral, e  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  é a variância amostral.

Neste caso,  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A} = \mathcal{B}_n$ ,  $\mathcal{P} = \{P_{\mu, \sigma^2} : \mu \in \mathbb{R}, \sigma^2 > 0\}$  onde  $P_{\mu, \sigma^2}(B) = \int_B \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} d\lambda_n(\mathbf{x})$ ,  $B \in \mathcal{B}_n$ ;  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ .

Fixamos  $\sigma^2 = \sigma_0^2 > 0$  e consideramos o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$ , onde  $\mathcal{P} = \{P_{\mu, \sigma^2} : \mu \in \mathbb{R}\}$ .

- (1) Do Teorema de Fatoração decorre que  $T(x_1, \dots, x_n) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  é uma estatística suficiente para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$ .

Para mostrar que  $T(x_1, \dots, x_n) = \bar{x}$  é completa para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\sigma_0^2})$ , devemos provar que

$$\forall f \in \mathcal{A}_T \text{ tal que } E_{P_{\mu, \sigma_0^2}}(f) = 0, \forall \mu \Rightarrow f = 0 \text{ } [\mathcal{P}_{\sigma_0^2}]$$

ou equivalentemente,

$$\forall \varphi \in \mathcal{B}_1 \text{ tal que } E_{Q_{\mu, \sigma_0^2}}(\varphi) = 0, \forall \mu \Rightarrow \varphi = 0 \text{ } [Q_{\mu, \sigma_0^2}] \forall \mu$$

onde  $Q_{\mu, \sigma_0^2} = P_{\mu, \sigma_0^2} T^{-1}$ ,  $\mu \in \mathbb{R}$ , a medida induzida pela estatística  $T$ .

Sendo  $Q_{\mu, \sigma_0^2}$  a medida normal com média  $\mu$  e variância  $\frac{\sigma_0^2}{n}$ , então

$$\frac{dQ_{\mu, \sigma_0^2}}{d\lambda}(t) = \frac{1}{\sqrt{2\pi \frac{\sigma_0^2}{n}}} \exp \left\{ \frac{-n}{2\sigma_0^2} (t - \mu)^2 \right\}$$

logo,  $E_{Q_{\mu, \sigma_0^2}}(\varphi) = 0, \forall \mu$ , implica que

$$\begin{aligned} \int_{\mathbb{R}} \frac{n\varphi(t)}{2\sigma_0^2} \exp \left\{ \frac{-n}{2\sigma_0^2} (t - \mu)^2 \right\} d\lambda(t) &= 0 \quad \forall \mu \\ \Rightarrow \int_{\mathbb{R}} \psi(t) \exp \left\{ \frac{n\mu}{2\sigma_0^2} t \right\} d\lambda(t) &= 0 \quad \forall \mu \end{aligned}$$

sendo  $\psi(t) = \exp \left\{ \frac{-n}{2\sigma_0^2} (t^2 + \mu^2) \right\}$ ,  $t \in \mathbb{R}$ , da unicidade (essencial) da transformada de Laplace bilateral, segue que  $\psi = 0$   $[\lambda]$  e logo  $\varphi = 0$   $[\lambda]$ , e portanto  $\varphi = 0$   $[Q_{\mu, \sigma_0^2}] \forall \mu$ .

Assim a estatística  $T(x_1, \dots, x_n) = \bar{x}$  é suficiente e completa para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$ .

- (2) A estatística  $S^2(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  é ancilar para o modelo  $(\mathbb{R}, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$ .

Notamos que  $\frac{dP_{\mu, \sigma_0^2}}{d\lambda_n}(\mathbf{x}) = f_{\sigma_0^2}(x_1 - \mu, \dots, x_n - \mu)$ , onde  $f_{\sigma_0^2}(z_1, \dots, z_n) = \frac{1}{(2\pi\sigma_0^2)^{\frac{n}{2}}} \exp \left\{ \frac{-1}{2\sigma_0^2} \sum_{i=1}^n z_i^2 \right\}$  que corresponde à densidade de uma amostra  $Z_1, \dots, Z_n$  de uma distribuição  $N(0, \sigma_0^2)$ .

Logo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$  é um modelo de locação com parâmetro nuisance  $\sigma_0^2$ , o qual foi fixado.

Mostremos que a estatística  $S^2$  é invariante sob o grupo de locação  $\mathcal{G} = \{g_a : \mathbb{R}^n \rightarrow \mathbb{R}^n / a \in \mathbb{R}\}$ , onde  $g_a(x_1, \dots, x_n) = (x_1 - a, \dots, x_n - a)$ ,  $a \in \mathbb{R}$ .

$$\begin{aligned} S^2(g_a(x_1, \dots, x_n)) &= S^2(x_1 - a, \dots, x_n - a) = \frac{1}{n-1} \sum_{i=1}^n (x_i - a - (\bar{x} - a))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = S^2(x_1, \dots, x_n) \end{aligned}$$

Da Proposição 8.4(a) decorre-se que  $S^2$  é ancilar para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P}_{\sigma_0^2})$ .

- (3) Usando um Teorema de Basu (1955), temos que fixado  $\sigma^2 = \sigma_0^2 > 0$ ,  $\bar{X} \perp\!\!\!\perp S^2/\mu$ , e como isto vale para todo  $\sigma_0^2 > 0$  fixo, então  $\bar{X} \perp\!\!\!\perp S^2/(\mu, \sigma^2)$ , isto é,  $\bar{X}$  e  $S^2$  são estatisticamente independentes para o modelo  $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ .

Finalizamos este capítulo, deixando para o leitor um exercício similar para ser desenvolvido.

### Exercício 8.11

Seja  $X_1, \dots, X_n$  uma amostra aleatória simples proveniente de uma distribuição Gamma com parâmetros  $\alpha$  (forma) e  $\beta$  (escala), positivos.

Definamos  $S_i = \sum_{j=1}^i X_j$  e  $R_k = S_k/S_{k-1}$ ;  $i = 1, \dots, n$ ,  $k = 1, \dots, n-1$ .

Mostre que  $R_1, \dots, R_{n-1} \perp\!\!\!\perp S_n/(\alpha, \beta)$ , acompanhando o seguinte esquema.

- (a) Determine o modelo estatístico induzido pelas variáveis aleatórias  $X_1, \dots, X_n$ , digamos  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , onde  $\mathcal{P} = \{P_{\alpha, \beta} : \alpha, \beta > 0\}$ .
- (b) Fixado  $\alpha = \alpha_0 > 0$ , mostre que a estatística  $S_n = \sum_{j=1}^n X_j$  é uma estatística suficiente e completa para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\alpha_0})$  onde  $\mathcal{P}_{\alpha_0} = \{P_{\alpha_0, \beta} : \beta > 0\}$ .
- (c) Fixado  $\alpha = \alpha_0 > 0$ , mostre que a estatística  $R_k = S_k/S_{k+1}$  é ancilar para o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\alpha_0})$ .
- (d) Conclua a demonstração.

### REFERÊNCIAS

- D. Basu (1955). On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377-380
- D. Basu (1958). On statistics independent of sufficient statistics. *Sankhyā*, **20**, 223–226.
- D. Basu (1959). The family of ancillary statistics. *Sankhyā*, **21**, 227–256.
- D. Basu (1969). On sufficiency and invariance. *Essays in Probability & Statistics*, 61–84.
- J.O.Berger (1985). *Statistical decision theory and bayesian analysis*. 2.ed. New York: Springer-Verlag.
- G.D.Birkhoff (1931). Proof of the ergodic theorem. *Proc. Nat'l.Acad.Sci. USA*, **17**, 656–660.
- J.V.Bondar and P.Milnes (1981). Amenability: A survey of Statistical applications of Hunt-Stein and related conditions on group. *Z.Wahrsch.veruw. Gebiete*, **57**, 103–128.
- E.L.Lehmann (1986). *Testing statistical hypotheses*. 2.ed. New York: Wiley
- R.Mañé (1983). *Introdução à teoria ergódica*. IMPA-CNPq. Projeto Euclides.

## CAPÍTULO 9

# O MODELO ESTATÍSTICO BAYESIANO E SUA ESTRUTURA MATEMÁTICA

### 9.1 CONSTRUÇÃO DO MODELO BAYESIANO

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  o modelo estatístico “clássico” cuja família de medidas de probabilidades está especificada parametricamente, i.e.,  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ . Dotemos  $\Theta$  de uma  $\sigma$ -álgebra  $\mathcal{B}$  de modo que as funções  $\theta \rightarrow P_\theta(A)$ ,  $\forall A \in \mathcal{A}$ , sejam mensuráveis. Deste modo,  $P_\theta(A) : \Theta \times \mathcal{A} \rightarrow [0, 1]$  passa a representar uma função de transição de  $(\Theta, \mathcal{B})$  para  $(\mathcal{X}, \mathcal{A})$ . Seja ainda definida em  $(\Theta, \mathcal{B})$  a medida de probabilidade a priori  $\nu$ .

Denotemos por  $\Omega = \Theta \times \mathcal{X}$  o produto cartesiano dos espaços paramétrico e amostral e por  $\mathcal{F} = \mathcal{B} \times \mathcal{A}$  a chamada  $\sigma$ -álgebra produto, gerada pela classe de retângulos mensuráveis  $\mathcal{J} = \{B \times A : B \in \mathcal{B}, A \in \mathcal{A}\}$ .

#### Exercício 9.1

Prove que  $\mathcal{J}$  é fechada sob intersecções finitas (i.e., é uma classe  $\pi$ ) e que  $\mathcal{F}$  é gerada pela álgebra de uniões finitas de elementos disjuntos de  $\mathcal{J}$ .

O recurso ao teorema generalizado da medida produto (vide, e.g., Ash, R.B. (1972); *Real Analysis and Probability*; Academic Press, New York; pp. 97-100) permite-nos

definir no espaço produto  $(\Omega, \mathcal{F})$  uma medida de probabilidade  $\mu$  tal que

$$(9.1) \quad \forall F \in \mathcal{F}, \quad \mu(F) = \int_{\Theta} P_{\theta}(F^{\theta}) \nu(d\theta)$$

onde  $F^{\theta} = \{x \in \mathcal{X} : \omega = (\theta, x) \in F\}$  é a secção de  $F$  em  $\theta$ .

A medida produto, usualmente denotada por  $\mu = \nu \times P_{\theta}$ , e cuja restrição a  $\mathcal{J}$  é definida por

$$(9.2) \quad \forall B \times A \in \mathcal{J}, \quad \mu(B \times A) = \int_B P_{\theta}(A) \nu(d\theta),$$

é única no sentido em que é idêntica a qualquer outra medida em  $(\Omega, \mathcal{F})$  que com ela coincide em  $\mathcal{J}$ .

O espaço de probabilidade obtido,  $(\Omega, \mathcal{F}, \mu)$ , define o chamado modelo estatístico bayesiano. Neste espaço, as observações  $x$  e o parâmetro  $\theta$  podem ser visualizados através das funções mensuráveis  $\bar{x}(\omega) = x : (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{A})$  e  $\bar{\theta}(\omega) = \theta : (\Omega, \mathcal{F}) \rightarrow (\Theta, \mathcal{B})$ , ou equivalentemente, através das respectivas subálgebras induzidas,  $\mathcal{F}(\bar{x}) = \{\Theta \times A : A \in \mathcal{A}\}$  e  $\mathcal{F}(\bar{\theta}) = \{B \times \mathcal{X} : B \in \mathcal{B}\}$ . Deste modo, a  $\sigma$ -álgebra produto pode exprimir-se por  $\mathcal{F} = \mathcal{F}(\bar{\theta}) \vee \mathcal{F}(\bar{x}) \equiv \sigma(\{\bar{B} \cup \bar{A} : B \in \mathcal{B}, A \in \mathcal{A}\})$ , onde  $\bar{B} = B \times \mathcal{X}$  e  $\bar{A} = \Theta \times A$ .

Da expressão (9.2) conclui-se que  $\nu(B) = \mu(\bar{B})$ ,  $\forall B \in \mathcal{B}$ , pelo que a medida de probabilidade a priori pode ser identificada com a restrição,  $\mu^{\bar{\theta}}$ , de  $\mu$  a  $\mathcal{F}(\bar{\theta})$ . Por outro lado, a restrição,  $\mu^{\bar{x}}$ , de  $\mu$  a  $\mathcal{F}(\bar{x})$  identifica a medida de probabilidade marginal em  $(\mathcal{X}, \mathcal{A})$

$$(9.3) \quad P(A) = \int_{\Theta} P_{\theta}(A) \nu(d\theta)$$

que é assim dominada pela família  $\mathcal{P}$ .

O fato de  $P_{\theta}(A)$  definir uma função de transição em  $\Theta \times \mathcal{A}$  permite-nos obter a função de transição associada em  $\Omega \times \mathcal{A}$

$$P_{\bar{\theta}(\omega)}^*(A) = P_{\theta}(A) \quad \text{para } \omega : \bar{\theta}(\omega) = \theta$$

e, deste modo, (9.2) pode ser reescrita como

$$(9.4) \quad \mu(B \times A) = \int_B P_{\bar{\theta}(\omega)}^*(A) \mu^{\bar{\theta}}(d\omega).$$

Por definição da esperança condicional  $E(I_A | \mathcal{F}(\bar{\theta})) = \mu(\bar{A} | \mathcal{F}(\bar{\theta}))$ , temos

$$(9.5) \quad \begin{aligned} \forall B \in \mathcal{B}, \quad \mu(B \times A) &= \mu(\bar{B} \cap \bar{A}) = \int_B I_{\bar{A}}(\omega) \mu(d\omega) \\ &= \int_B \mu(\bar{A} | \mathcal{F}(\bar{\theta})) \mu^{\bar{\theta}}(d\omega) \end{aligned}$$

A comparação de (9.4) e (9.5) permite concluir que para todo  $A \in \mathcal{A}$ ,

$$P_{\bar{\theta}(\omega)}^*(A) = \mu(\bar{A}|\mathcal{F}(\bar{\theta}))(\omega) [\mu^{\bar{\theta}}] ,$$

ou seja, que  $P_{\bar{\theta}}(A)$  está associada a uma versão regular da restrição a  $\mathcal{F}(\bar{x})$  da probabilidade condicional dado  $\mathcal{F}(\bar{\theta})$ .

A construção de  $P_{\bar{\theta}}(A)$  como uma função de transição em  $\Theta \times \mathcal{A}$  tem assim implícito que o espaço mensurável  $(\mathcal{X}, \mathcal{A})$  é tal que admite a existência de uma probabilidade condicional regular para  $\bar{x}$  dado  $\mathcal{F}(\bar{\theta})$ , i.é., de uma função  $\mu(\cdot|\mathcal{F}(\bar{\theta}))(\omega)$  que, para cada  $\omega$  fixo, seja uma medida de probabilidade em  $\mathcal{F}(\bar{x})$ . Este requisito é satisfeito nas condições usuais em que  $(\mathcal{X}, \mathcal{A})$  é um espaço euclidiano, entendendo-se esta expressão no sentido de  $\mathcal{X}$  ser um subconjunto de  $\mathbb{R}^n$ , para algum  $n$ , e  $\mathcal{A}$  a correspondente  $\sigma$ -álgebra de Borel (veja-se, e.g., Ash (1982), op. cit., sec. 6.6).

Deste modo, sendo  $(\Theta, \mathcal{B})$  igualmente um espaço euclidiano, denotemos por  $\nu_{\bar{x}(\omega)}(B)$  a função de transição em  $\Omega \times \mathcal{B}$  obtida tomando uma versão regular da probabilidade condicional  $\mu(\bar{B}|\mathcal{F}(\bar{x}))$ , e seja

$$\nu_x(B) = \nu_{\bar{x}(\omega)}^*(B) \quad \text{para } \omega : \bar{x}(\omega) = x$$

a correspondente função de transição em  $\mathcal{X} \times \mathcal{B}$ . Então, para todo  $B \times A \in \mathcal{J}$

$$\begin{aligned} \mu(B \times A) &= \int_A I_{\bar{B}}(\omega) \mu(d\omega) = \int_A \mu(\bar{B}|\mathcal{F}(\bar{x}))(\omega) \mu^{\bar{x}}(d\omega) \\ (9.6) \quad &= \int_A \nu_x(B) P(dx) \end{aligned}$$

Pela unicidade estabelecida no teorema generalizado da medida produto, conclui-se que

$$(9.7) \quad \forall F \in \mathcal{F} , \quad \mu(F) = \int_{\mathcal{X}} \nu_x(F^x) P(dx) ,$$

onde  $F^* = \{\theta \in \Theta : \omega = (\theta, x) \in F\}$  é a secção de  $F$  em  $x$ . Por outras palavras, a medida produto  $\mu = \nu \times P_{\bar{\theta}}$  admite também, nas condições mencionadas, a decomposição  $\mu = P \times \nu_x$ , característica que denominaremos de regularidade do modelo bayesiano, em consonância com a terminologia de Florens et al. (1990); Elements of Bayesian Statistics; Marcel Dekker, Inc., New York.

Sob as asas de um modelo bayesiano regular fica assim definido um novo modelo, a denominada família de medidas de probabilidade a posteriori,  $(\Theta, \mathcal{B}, Q)$ , onde  $Q = \{\nu_x, x \in \mathcal{X}\}$ , constituindo o dual bayesiano de  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ . Note-se ainda que  $Q$  domina  $\nu$ , pois

$$(9.8) \quad \forall B \in \mathcal{B}, \quad \nu(B) = \int_{\mathcal{X}} \nu_x(B) P(dx)$$

### Exercício 9.2

No contexto de um modelo bayesiano regular, encare  $\mu(B \times A)$  como uma medida de probabilidade em  $(\mathcal{X}, \mathcal{A})$  para  $B$  fixo (e denote-a por  $P_B$ ) e como uma medida de probabilidade em  $(\Theta, \mathcal{B})$  para  $A$  fixo (e denote-a por  $\nu_A$ ).

Prove que:

- a)  $P_B \ll P$ , sendo  $\nu_x(B)$  uma versão da respectiva derivada de Radon-Nikodym;
- b)  $\nu_A \ll \nu$ , sendo  $P_\theta(A)$  uma versão da respectiva derivada de Radon-Nikodym.

## 9.2 DOMINAÇÃO DO MODELO BAYESIANO

Consideremos que  $\mathcal{P}$  é dominada por uma medida  $\sigma$ -finita  $m$  em  $(\mathcal{X}, \mathcal{A})$  e seja  $f(\theta, x)$  uma versão da derivada de Radon-Nikodym  $dP_\theta(x)/dm(x)$ , que é não negativa e  $\mathcal{A}$ -mensurável para cada  $\theta \in \Theta$ .

Neste quadro, (9.2) exprime-se por

$$(9.9) \quad \forall B \times A \in \mathcal{J}, \quad \mu(B \times A) = \int_B \int_A f(\theta, x) m(dx) \nu(d\theta).$$

A aplicação do teorema de Fubini a (9.9) com vista à definição de  $\mu$  em  $\mathcal{F}$  exige a  $\mathcal{F}$ -mensurabilidade de  $f(\theta, x)$ , condição que ainda não foi expressamente garantida. Uma resposta é dada pelo teorema de Doob que enunciamos em seguida (para a sua demonstração veja-se, e.g. Dellacherie and Meyer (1982); Probabilities and Potential B-Theory of Martingales; North-Holland [cap. 5]).



### Teorema de Doob

Seja  $(\mathcal{X}, \mathcal{A})$  um espaço mensurável separável,  $(\Theta, \mathcal{B})$  um espaço mensurável arbitrário e  $\{P_\theta(A)\}$  uma família de funções de transição em  $\Theta \times \mathcal{A}$  dominada pela medida  $\sigma$ -finita  $m$ . Existe então uma função  $\mathcal{B} \times \mathcal{A}$ -mensurável,  $f(\theta, x)$ , tal que para cada  $\theta \in \Theta$   $f(\theta, x)$  é uma versão de  $dP_\theta(x)/dm(x)$ .

Deste modo, a suposição de  $(\mathcal{X}, \mathcal{A})$  ser euclidiano garante, pela separabilidade de  $\mathcal{A}$ , que seja possível selecionar uma versão de  $dP_\theta(x)/dm(x)$   $\mathcal{F}$ -mensurável. Admitimos então que  $(\mathcal{X}, \mathcal{A})$  é euclidiano e que  $f(\theta, x)$  em (9.9) é  $\mathcal{F}$ -mensurável, o que assegura também a sua  $\mathcal{B}$ -mensurabilidade para cada  $x \in \mathcal{X}$ . Pelo teorema de Fubini, (9.9) corresponde a

$$(9.10) \quad \forall B \times A \in \mathcal{J}, \quad \mu(B \times A) = \int_{B \times A} f(\theta, x)(\nu \times m)(d\theta \times dx)$$

Definindo agora em  $\mathcal{F}$  a medida finita

$$(9.11) \quad \mu^*(F) = \int_F f(\theta, x)(\nu \times m)(d\theta \times dx)$$

por (9.10),  $\mu^*(F) = \mu(F)$  em  $\mathcal{J}$ . Por um argumento em termos do teorema de Dynkin (mostre-se que  $\mathcal{R} \equiv \{F \in \mathcal{F} : \mu^*(F) = \mu(F)\} = \mathcal{F}$ ), prova-se que (9.11) define efetivamente a medida  $\mu$  no caso de dominação de  $\mathcal{P}$  em causa.

Esta definição mostra ainda que  $\mu$  é dominada pela medida produto  $\nu \times m$ , sendo  $f(\theta, x)$  igualmente uma versão da derivada de Radon-Nikodym  $d\mu/d(\nu \times m)$ .

O teorema de Fubini permite-nos reescrever (9.3) como

$$(9.12) \quad P(A) = \int_A \left[ \int_\Theta f(\theta, x)\nu(d\theta) \right] m(dx)$$

revelando que a função  $\mathcal{A}$ -mensurável  $g(x) = \int_\Theta f(\theta, x)\nu(dx)$  presente em (9.12) é uma versão de  $dP/dm$  (note-se que  $P \ll m$ ), e conseqüentemente de  $d(\nu \times P)/d(\nu \times m)$ . Sublinhe-se que  $g(x) = 0$  implica  $f(\theta, x) = 0$   $[\nu]$ .

Nestas condições de dominação de  $\mu$  por  $\nu \times m$ , prova-se que  $\mu \ll \nu \times P$  (veja-se Exercício 9.3 em seguida), característica que traduz a dominação do modelo bayesiano  $(\Omega, \mathcal{F}, \mu)$ . Deste modo, a derivada de Radon-Nikodym de  $\mu$  com respeito à medida produto  $\Pi = \nu \times P$  pode ser definida por

$$(9.13) \quad h(\theta, x) = f(\theta, x)/g(x) \quad \text{se } g(x) > 0$$

e quando  $g(x) = 0$ ,  $d\mu/d\Pi$  pode ser arbitrariamente definida como igual a 1.

### Exercício 9.3

Mostre que sob a condição de  $(\mathcal{X}, \mathcal{A})$  ser euclidiano:

- A dominação de  $\mathcal{P}$  garante a dominação do modelo bayesiano;
- A dominação de  $\mu$  (por  $\Pi = \nu \times P$ ) assegura a dominação  $\nu$ -essencial de  $\mathcal{P}$  e a regularidade do modelo bayesiano.

### Solução

a) Sob a condição de  $\mathcal{P}$ , definida num espaço euclidiano, ser dominada pela medida  $\sigma$ -finita  $m$ , com correspondente função densidade  $\frac{dP_\theta(x)}{dm(x)} = f(\theta, x)$ , pretende mostrar-se que

$$\forall F \in \mathcal{F}, \quad (\nu \times P)(F) = 0 \Rightarrow \mu(F) = 0.$$

Por definição da medida produto  $\Pi = \nu \times P$  e pelo teorema de Fubini

$$\Pi(F) = \int_{\Theta} P(F^\theta) \nu(d\theta) = \int_{\mathcal{X}} \int_{\Theta} I_{F^\theta}(x) \nu(d\theta) P(dx)$$

onde  $F^\theta = \{x : (\theta, x) \in F\}$ . Como  $dP(x)/dm(x) = g(x)$  segue-se que

$$\Pi(F) = \int_{\mathcal{X}} g(x) \left[ \int_{\Theta} I_{F^\theta}(x) \nu(d\theta) \right] m(dx).$$

Assim,  $\Pi(F) = 0$  implica que

$$g(x) \int_{\Theta} I_{F^\theta}(x) \nu(d\theta) = 0$$

exceto num conjunto  $\mathcal{X}_0 \in \mathcal{A}$ , digamos, com  $m(\mathcal{X}_0) = 0$ . Portanto,  $\forall x \in \mathcal{X} - \mathcal{X}_0 \cap \{g(x) > 0\}$ ,  $I_{F^\theta}(x) = 0$   $[\nu]$  e, por outro lado,  $\forall x \in \mathcal{X} - \mathcal{X}_0 \cap \{g(x) = 0\}$ ,  $f(\theta, x) = 0$   $[\nu]$ . Em suma,  $I_{F^\theta}(x) f(\theta, x) = 0$   $[\nu]$ ,  $\forall x \in \mathcal{X} - \mathcal{X}_0$ . Como já sabemos que nas condições assumidas  $P_\theta(F^\theta) = \int_{\mathcal{X}} I_{F^\theta}(x) f(\theta, x) m(dx)$ , para todo  $F \in \mathcal{F}$  e  $\theta \in \Theta$  tem-se pelo teorema de Fubini

$$\mu(F) = \int_{\mathcal{X} - \mathcal{X}_0} \left[ \int_{\Theta} I_{F^\theta}(x) f(\theta, x) \nu(d\theta) \right] m(dx) = 0 \quad \text{Q.E.D.}$$

### Nota 9.1

Observe-se que por definição das medidas produto  $\mu$  e  $\Pi$  em  $\mathcal{J}$ , elas verificam, independentemente da dominação considerada de  $\mathcal{P}$ , a relação  $\mu \ll \Pi$ , quer em  $\mathcal{J}$  quer na álgebra gerada por  $\mathcal{J}$  (constituída por uniões finitas de elementos disjuntos de  $\mathcal{J}$ ).

Pode provar-se que a relação  $\mu \ll \Pi$  mantém-se em  $\mathcal{F}$  se  $(\mathcal{X}, \mathcal{A})$  e  $(\Theta, \mathcal{B})$  forem euclidianos, pelo que, nestas condições, o modelo bayesiano é dominado sem que  $\mathcal{P}$  seja necessariamente dominada – veja-se o Lema 3.4 de Picci (1974); Theory of sufficient statistics and structure analysis of transition probabilities with application to identifiability; Report CNR-LADSEB 74/06, Padova, Italy.

b) A hipótese agora é  $\mu \ll \Pi$ . Sendo  $h(\theta, x)$  uma versão da respectiva derivada de Radon-Nikodym tem-se pelo teorema de Fubini que

$$\begin{aligned} \forall B \times A \in \mathcal{J}, \quad \mu(B \times A) &= \int_B \int_A h(\theta, x) P(dx) \nu(d\theta) \\ &= \int_A \int_B h(\theta, x) \nu(d\theta) P(dx) . \end{aligned}$$

A comparação da primeira integral definidora de  $\mu(B \times A)$  com (9.2) produz

$$P_\theta(A) = \int_A h(\theta, x) P(dx) \quad [\nu]$$

onde o conjunto  $\nu$ -nulo é em geral dependente do conjunto  $A$ . Contudo, pelo carácter euclidiano de  $(\mathcal{X}, \mathcal{A})$  pode mostrar-se que a expressão acima definidora de  $P_\theta(A)$  é válida para todo  $A \in \mathcal{A}$  desde que  $\theta \in \Theta - N$ , onde  $N \in \mathcal{B}$  é um conjunto  $\nu$ -nulo independente de  $A$  – mas possivelmente dependente da versão escolhida de  $d\mu/d\Pi$  (veja-se para o efeito Picci (1974), op. cit., teorema A.4 do apêndice).

Resulta então que  $P$  domina a sub-família de  $\mathcal{P}$ ,  $\{P_\theta : \theta \in \Theta - N\}$  e, portanto, domina  $\nu$ -quase sempre (e é sempre dominada por)  $\mathcal{P}$ .

Por outro lado, a dominação de  $\mu$  por  $\Pi = \nu \times P$  permite definir  $\mu = \nu \times P_\theta = P \times \nu_x$ , tomando

$$\begin{aligned} P_\theta(A) &= \int_A h(\theta, x) P(dx) , \quad \forall A \in \mathcal{A} \\ \nu_x(B) &= \int_B h(\theta, x) \nu(d\theta) , \quad \forall B \in \mathcal{B} , \end{aligned}$$

que representam assim versões regulares das probabilidades  $\mu$  condicionais a  $\mathcal{F}(\bar{\theta})$  e a  $\mathcal{F}(\bar{x})$ , respectivamente (restringidas a  $\mathcal{F}(\bar{x})$  e a  $\mathcal{F}(\bar{\theta})$ , conforme o caso).  $\square$

Do exercício anterior deduz-se igualmente que a família  $Q$ , quando definida num espaço mensurável euclidiano, é  $P$ -essencialmente dominada por  $\nu$ , já que

$$(9.14) \quad \nu_x(B) = \int_B h(\theta, x) \nu(d\theta)$$

para todo  $x \in \mathcal{X} - M$ , onde  $P(M) = 0$ . Deste modo, as derivadas de Radon-Nikodym,  $dP_\theta/dP$  e  $d\nu_x/d\nu$ , coincidem  $\nu \times P$ -quase em toda parte.

Quando  $\nu$  é absolutamente contínua em relação a uma medida  $\sigma$ -finita  $\lambda$  em  $(\Theta, \mathcal{B})$ , com função densidade  $q(\theta)$ , tem-se então  $\nu_x \ll \lambda, \forall x \in \mathcal{X} - M$  com

$$(9.15) \quad \begin{aligned} q(\theta, x) &= \frac{d\nu_x(\theta)}{d\lambda(\theta)} = \frac{d\nu_x(\theta)}{d\nu(\theta)} q(\theta) \quad [\lambda] \\ &= \begin{cases} \frac{f(\theta, x)q(\theta)}{g(x)} & , \text{ se } g(x) > 0 \\ q(\theta) & , \text{ c.c.} \end{cases} \end{aligned}$$

traduzindo a conhecida expressão da função densidade a posteriori resultante da aplicação do teorema de Bayes.

### Nota 9.2

Se a medida a priori  $\nu$  for apenas  $\sigma$ -finita (é o caso das chamadas distribuições a priori impróprias), a medida produto  $\mu = \nu \times P_\theta$  será igualmente  $\sigma$ -finita de acordo com o teorema da medida produto generalizado. Contudo, pode acontecer que  $P$  não seja  $\sigma$ -finita (se o for, no caso dominado,  $d\mu/d(\nu \times P)$  é bem definida e a distribuição a posteriori é própria) e, nesse caso, não se garante a decomposição  $\mu = P \times \nu_x$  nem a definição de  $\nu \times P$ . Além disso,  $dP/dm$  não é  $m$ -essencialmente finita e nos pontos em que é infinita a medida  $\int_B f(\theta, x)\nu(d\theta)$  também é infinita, pelo que a medida a posteriori não é bem definida. Para mais detalhes veja-se, e.g., Mouchart (1976); A note on Bayes theorem; *Statistica*, 36(2), 349-357.

### Exemplo 9.1

Seja  $\mathcal{X} = \{0, 1\}^n$ ,  $\mathcal{A}$  o conjunto das partes de  $\mathcal{X}$  e  $\mathcal{P} = \{P_\theta : \theta \in \Theta = (0, 1)\}$  tal que

$$P_\theta(A) = \int_A \theta^{\sum_1^n x_i} (1 - \theta)^{n - \sum_1^n x_i} m\left(\prod_{i=1}^n dx_i\right)$$

onde  $m$  é a medida de contagem em  $(\mathcal{X}, \mathcal{A})$ . Seja  $\mathcal{B}$  a  $\sigma$ -álgebra de Borel de  $\Theta$  no qual está definida a seguinte medida de probabilidade a priori

$$\nu(B) = \int_B \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \lambda(d\theta), \quad a, b > 0$$

onde  $\lambda$  é a medida de Lebesgue em  $(\Theta, \mathcal{B})$ . A medida de probabilidade marginal em  $(\mathcal{X}, \mathcal{A})$  é definida pela função de probabilidade

$$g(x) = \frac{B(a + \sum_1^n x_i, b + n - \sum x_i)}{B(a, b)}, \quad x = (x_1, \dots, x_n) \in \mathcal{X}$$

e a medida produto em  $(\Omega, \mathcal{F})$ , representando a distribuição conjunta de  $(\theta, x)$ , é definida por

$$\begin{aligned} \mu(B \times A) &= \int_{B \times A} \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \frac{B(a, b)}{B(a + \sum x_i, b + n - \sum x_i)} (\nu \times P)(d\theta \times \prod_i dx_i) \\ &= \int_{B \times A} \frac{\theta^{a + \sum x_i - 1} (1 - \theta)^{b + n - \sum x_i - 1}}{B(a, b)} (\lambda \times m)(d\theta \times \prod_i dx_i). \end{aligned}$$

A medida de probabilidade a posteriori é então definida por

$$\begin{aligned} \nu_x(B) &= \int_B \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \frac{B(a, b)}{B(a + \sum x_i, b + n - \sum x_i)} \nu(d\theta) \\ &= \int_B q(\theta, x) \lambda(d\theta) \end{aligned}$$

onde  $q(\theta, x)$  é a função densidade da distribuição  $B(a + \sum x_i, b + n - \sum x_i)$ .

### 9.3 MEDIDAS CONDICIONADAS ADICIONAIS

Para evitar complicações com a eventual inexistência de versões regulares de probabilidades no espaço produto condicionais em subálgebras de  $\mathcal{F}$ , admitiremos doravante que  $(\mathcal{X}, \mathcal{A})$  e  $(\Theta, \mathcal{B})$  são espaços euclidianos. As funções de transição,  $P_\theta(A)$  em  $\Theta \times \mathcal{A}$  e  $\nu_x(B)$  em  $\mathcal{X} \times \mathcal{B}$ , serão por vezes designadas por  $P_B(A)$  e  $\nu_A(B)$ , para destacar, respectivamente, a sua  $\mathcal{B}$ -mensurabilidade para cada  $A \in \mathcal{A}$  e a sua  $\mathcal{A}$ -mensurabilidade para cada  $B \in \mathcal{B}$ .

Consideremos agora as funções mensuráveis  $\phi : (\Theta, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_1)$  e  $t : (\mathcal{X}, \mathcal{A}) \rightarrow (T, \mathcal{A}_1)$ , indutoras das subálgebras  $\mathcal{B}(\phi) \subset \mathcal{B}$  e  $\mathcal{A}(t) \subset \mathcal{A}$ , respectivamente. Elas definem as transformações compostas  $\bar{\phi} = \phi \circ \bar{\theta}$  e  $\bar{t} = t \circ \bar{x}$  que induzem, respectivamente as subálgebras  $\mathcal{F}(\bar{\phi}) \subset \mathcal{F}(\bar{\theta})$  e  $\mathcal{F}(\bar{t}) \subset \mathcal{F}(\bar{x})$ .

Por definição das probabilidades condicionais em  $(\Omega, \mathcal{F})$  dado  $\mathcal{F}(\bar{t})$  e  $\mathcal{F}(\bar{\phi})$ , respectivamente, temos para todo  $B \in \mathcal{B}$  e  $A \in \mathcal{A}$

$$(9.16) \quad \forall A_1 \in \mathcal{A}(t), \quad \mu[B \times (A \cap A_1)] = \int_{A_1} \mu[B \times A | \mathcal{F}(\bar{t})](\omega) \mu^{\bar{t}}(d\omega)$$

$$(9.17) \quad \forall B_1 \in \mathcal{B}(\phi), \quad \mu[(B \cap B_1) \times A] = \int_{B_1} \mu[B \times A | \mathcal{F}(\bar{\phi})](\omega) \mu^{\bar{\phi}}(d\omega)$$

onde  $\mu^{\bar{t}}$  (resp.  $\mu^{\bar{\phi}}$ ) denota a restrição de  $\mu$  a  $(\Omega, \mathcal{F}(\bar{t}))$  (resp.  $(\Omega, \mathcal{F}(\bar{\phi}))$ ).

As restrições a  $\mathcal{F}(\bar{x})$  e a  $\mathcal{F}(\bar{\theta})$  de uma versão regular de  $\mu(\cdot | \mathcal{F}(\bar{t}))$  conduzem, respectivamente, à medida preditiva (a priori) e à medida a posteriori condicionais em  $t$  que denotamos pelas funções de transição  $P_t(A) : T \times \mathcal{A} \rightarrow [0, 1]$  e  $\nu_t(B) : T \times \mathcal{B} \rightarrow [0, 1]$ .

Deste modo, pela propriedade de “smoothing” da esperança condicional, tem-se que a probabilidade condicional em (9.16) pode ser expressa por

$$(9.18) \quad \begin{aligned} \mu[B \times A | \mathcal{F}(\bar{t})] &= E[I_{\bar{B}}(\bar{\theta}) I_{\bar{A}}(\bar{x}) | \mathcal{F}(\bar{t})] \\ &= E\{I_{\bar{A}}(\bar{x}) E[I_{\bar{B}}(\bar{\theta}) | \mathcal{F}(\bar{x})] | \mathcal{F}(\bar{t})\} \\ &= E\{I_{\bar{A}}(\bar{x}) \mu[\bar{B} | \mathcal{F}(\bar{x})] | \mathcal{F}(\bar{t})\} \\ &= \int_A \nu_x(B) P_t(dx) \end{aligned}$$

Consideremos agora o condicionamento na subálgebra induzida pelo vetor  $(\bar{\theta}, \bar{t})$ ,  $\mathcal{F}(\bar{\theta}, \bar{t}) = \mathcal{F}(\bar{\theta}) \vee \mathcal{F}(\bar{t}) = \mathcal{B} \times \mathcal{A}(t)$ . Por definição da restrição a  $\mathcal{F}(\bar{x})$  da probabilidade condicional  $\mu$  dado  $\mathcal{F}(\bar{\theta}, \bar{t})$ , obtém-se para todo  $A \in \mathcal{A}$

$$\forall B_1 \in \mathcal{B}, \quad A_1 \in \mathcal{A}(t), \quad \mu[B_1 \times (A \cap A_1)] = \int_{B_1 \cap \bar{A}_1} \mu[\bar{A} | \mathcal{F}(\bar{\theta}, \bar{t})](\omega) \mu^{\bar{\theta}, \bar{t}}(d\omega)$$

onde  $\mu^{\bar{\theta}, \bar{t}}$  denota, na linha da notação que tem vindo a ser seguida, a restrição de  $\mu$  a  $(\Omega, \mathcal{F}(\bar{\theta}, \bar{t}))$ . Tomando uma versão regular dessa probabilidade condicional obtemos a função de transição  $P_{\theta, t}(A) : (\Theta \times T) \times \mathcal{A} \rightarrow [0, 1]$  que, pelo teorema de Fubini, verifica a relação

$$(9.19) \quad \forall B_1 \in \mathcal{B}, \quad A_1 \in \mathcal{A}(t), \quad \mu[B_1 \times (A \cap A_1)] = \int_{B_1} \int_{A_1} P_{\theta, t}(A) P_{\theta}^t(dx) \nu(d\theta) .$$

Com base nesta nova função de transição, a probabilidade condicional em (9.16) pode ser alternativamente expressa por

$$(9.20) \quad \begin{aligned} \mu[B \times A | \mathcal{F}(\bar{t})] &= E\{I_{\bar{B}}(\bar{\theta}) E[I_{\bar{A}}(\bar{x}) | \mathcal{F}(\bar{\theta}, \bar{t})] | \mathcal{F}(\bar{t})\} \\ &= E\{I_{\bar{B}}(\bar{\theta}) \mu[\bar{A} | \mathcal{F}(\bar{\theta}, \bar{t})] | \mathcal{F}(\bar{t})\} \\ &= \int_B P_{\theta, t}(A) \nu_t(d\theta) . \end{aligned}$$

As expressões (9.18) e (9.20) revelam que

$$\mu_{\mathcal{F}(\bar{t})} = P_t \times \nu_x = \nu_t \times P_{\theta,t} ,$$

i.e., a regularidade do modelo condicional  $(\Omega, \mathcal{F}, \mu_{\mathcal{F}(\bar{t})})$ .

A clarificação do significado da função de transição  $P_{\theta,t}(A)$  pode ser operada através da reexpressão do primeiro membro de (9.19) como

$$\begin{aligned} \mu[B_1 \times (A \cap A_1)] &= \int_{B_1} I_{\bar{A} \cap \bar{A}_1}(\omega) \mu(d\omega) = \int_{B_1} P_\theta(A \cap A_1) \nu(d\theta) \\ (9.21) \qquad \qquad \qquad &= \int_{B_1} \int_{A_1} P_\theta[A|\mathcal{A}(t)] P_\theta^t(dx) \nu(d\theta) \end{aligned}$$

pela definição da probabilidade amostral condicional  $P_\theta$  dado  $\mathcal{A}(t)$ .

Comparando (9.19) com (9.21) conclui-se, dada a arbitrariedade de  $B_1 \in \mathcal{B}$ , que

$$(9.22) \qquad \forall A \in \mathcal{A}, A_1 \in \mathcal{A}(t), P_\theta(A \cap A_1) = \int_{A_1} P_{\theta,t}(A) P_\theta^t(dx) \quad [\nu]$$

Quando  $\mathcal{A}(t)$  é separável (o que é garantido se  $t$  é uma variável aleatória), existe uma versão  $\mathcal{F}(\bar{\theta}, \bar{t})$ -mensurável de  $P_\theta(A|\mathcal{A}(t))$ ,  $\forall A \in \mathcal{A}$ , quando enquadrada no espaço produto. Tomando essa versão em (9.21), a sua comparação com (9.22), implica que

$$\forall A \in \mathcal{A}, P_{\theta,t}(A) = P_\theta[A|\mathcal{A}(t)] \quad [P_\theta^t]$$

para  $\nu$ -essencialmente todo  $\theta \in \Theta$ .

Em suma, a versão regular da probabilidade condicional  $\mu[\cdot|\mathcal{F}(\bar{\theta}, \bar{t})]$  restringida a  $\mathcal{F}(\bar{x})$  identifica-se  $\mu$ -essencialmente com a distribuição amostral condicional a  $t$ .

Seguindo um argumento análogo, é possível definir probabilidades a priori, probabilidades amostrais e probabilidades a posteriori, todas condicionais a  $\phi$ , e, a partir delas, clarificar o modelo condicional  $(\Omega, \mathcal{F}, \mu_{\mathcal{F}(\bar{\phi})})$ . Tal é o propósito do próximo exercício.

#### Exercício 9.4

Considere uma versão regular da probabilidade condicional  $\mu$  dado  $\mathcal{F}(\bar{\phi})$ , definida em (9.17).

- a) Defina a medida de probabilidade a priori e a medida de probabilidade amostral, ambas condicionais a  $\phi$ , e denote-as pelas funções de transição  $\nu_\phi(B) : \Phi \times \mathcal{B} \rightarrow [0, 1]$  e  $P_\phi(A) : \Phi \times \mathcal{A} \rightarrow [0, 1]$ .

b) Mostre que  $\mu_{\mathcal{F}(\bar{\phi})} = \nu_{\phi} \times P_{\theta}$ , i.e.,

$$\mu[B \times A | \mathcal{F}(\bar{\phi})] = \int_B P_{\theta}(A) \nu_{\phi}(d\theta)$$

c) Defina a função de transição  $\nu_{\phi,x}(B) : (\Phi \times \mathcal{X}) \rightarrow [0, 1]$  através da restrição a  $\mathcal{F}(\bar{\theta})$  de uma versão regular de probabilidade condicional  $\mu$  dado  $\mathcal{F}(\bar{\phi}, \bar{x}) = \mathcal{B}(\phi) \times \mathcal{A}$ , e mostre que o modelo condicional  $(\Omega, \mathcal{F}, \mu_{\mathcal{F}(\bar{\phi})})$  é regular, provando que

$$\mu[B \times A | \mathcal{F}(\bar{\phi})] = \int_A \nu_{\phi,x} P_{\phi}(dx)$$

d) Mostre que  $\nu_{\phi,x}(\cdot)$  coincide  $\mu$ -essencialmente com a medida de probabilidade a posteriori condicional em  $\phi, \nu_x(\cdot | \mathcal{B}(\phi))$ .

### Nota 9.3

Este procedimento de definição das funções de transição  $P_{\phi}, P_t, P_{\theta,t}$  e  $\nu_{\phi}, \nu_t$  e  $\nu_{\phi,x}$  pode ser aplicado a subálgebras  $\mathcal{C} \subset \mathcal{B}$  e  $\mathcal{D} \subset \mathcal{A}$ , sem explicitar se são ou não induzidas por funções mensuráveis não triviais. Neste quadro mais geral os símbolos  $P_{\mathcal{C}}, P_{\mathcal{D}}$  e  $P_{\mathcal{B} \times \mathcal{D}}$ , na linha do significado de  $P_{\mathcal{B}}$ , designarão as medidas no espaço amostral que se identificam com a restrição a  $\mathcal{A}$  ( $\equiv \mathcal{F}(\bar{x})$ ) das probabilidades condicionais em  $\mathcal{C}_+ = \{c \times \mathcal{X} : c \in \mathcal{C}\}$ ,  $\mathcal{D}_+ = \{\Theta \times D : D \in \mathcal{D}\}$  e  $\mathcal{B} \times \mathcal{D}$  ( $\equiv \mathcal{F}(\bar{\theta}) \vee \mathcal{D}_+$ ). Analogamente, os símbolos  $\nu_{\mathcal{C}}, \nu_{\mathcal{D}}$  e  $\nu_{\mathcal{A} \times \mathcal{C}}$  (recorde-se o significado de  $\nu_{\mathcal{A}}$ ) denotarão as medidas no espaço paramétrico que se identificam com a restrição a  $\mathcal{B}$  ( $\equiv \mathcal{F}(\bar{\theta})$ ) das probabilidades condicionais em  $\mathcal{C}_+, \mathcal{D}_+$  e  $\mathcal{C} \times \mathcal{A}$  ( $\equiv \mathcal{C}_+ \vee \mathcal{F}(\bar{x})$ ).

A separabilidade de  $\mathcal{D}$  (resp.  $\mathcal{C}$ ) garante a interpretação essencial de  $P_{\mathcal{B} \times \mathcal{D}}$  (resp.  $\nu_{\mathcal{A} \times \mathcal{C}}$ ) como uma medida amostral (resp. a posteriori) condicional em  $\mathcal{D}$  (resp.  $\mathcal{C}$ ). Nesse caso,  $P_{\mathcal{D}}$  (resp.  $\nu_{\mathcal{C}}$ ) traduz a medida preditiva (a priori) condicional em  $\mathcal{D}$  (resp. a medida a priori condicional em  $\mathcal{C}$ ). Também  $P_{\mathcal{C}}$  e  $\nu_{\mathcal{D}}$  representarão medidas amostral e a posteriori condicionais em  $\mathcal{C}$  e  $\mathcal{D}$ , respectivamente.

## 9.4 DOMINAÇÃO DE MODELOS MARGINAIS E CONDICIONAIS

Consideremos um modelo bayesiano  $(\Omega, \mathcal{F}, \mu)$  regular e dominado e seja  $h$  uma versão de  $d\mu/d\Pi$ , onde  $\Pi = \nu \times P$ . Por definição da restrição a  $\mathcal{G} \subset \mathcal{F}$  de  $\mu$  e  $\Pi$ , o modelo



marginal  $(\Omega, \mathcal{G}, \mu^{\mathcal{G}})$  é dominado por  $\Pi^{\mathcal{G}}$ . Como  $\forall f \in \mathcal{G}$

$$E_{\mu}[f] = \int_{\Omega} f d\mu = \int_{\Omega} fh d\Pi = \int_{\Omega} f E_{\Pi}(h|\mathcal{G}) d\Pi^{\mathcal{G}},$$

segue-se que  $E_{\Pi}(h|\mathcal{G})$  é uma versão de  $d\mu^{\mathcal{G}}/d\Pi^{\mathcal{G}}$ .

Como  $\mu$  e  $\Pi$  têm as mesmas medidas marginais, tem-se, por exemplo,  $E_{\Pi}(h|\mathcal{G}) = 1$ ,  $[\Pi]$  quando  $\mathcal{G} = \mathcal{C}_+ \equiv \{C \times \mathcal{X} : C \in \mathcal{C} \subset \mathcal{B}\}$  e  $\mathcal{G} = \mathcal{D}_+ \equiv \{\Theta \times D : D \in \mathcal{D} \subset \mathcal{A}\}$ . Consideremos agora  $\mathcal{G} = \mathcal{C} \times \mathcal{D} \equiv \sigma(\{C \times D : C \in \mathcal{C}, D \in \mathcal{D}\})$ . Observe-se que, por definição de  $\Pi$  e de  $\mathcal{G}$ ,

$$(9.23) \quad \Pi^{\mathcal{G}} = \nu^{\mathcal{C}} \times P^{\mathcal{D}}.$$

Por outro lado,  $\forall C \times D \in \mathcal{G}$

$$(9.24) \quad \mu^{\mathcal{G}}(C \times D) = \int_{\mathcal{C}} P_{\theta}(D) \nu(d\theta) = \int_{\mathcal{C}} E[P_{\theta}(D)|\mathcal{C}] \nu^{\mathcal{C}}(d\theta)$$

$$(9.25) \quad = \int_{\mathcal{D}} \nu_x(C) P(dx) = \int_{\mathcal{D}} E[\nu_x(C)|\mathcal{D}] P^{\mathcal{D}}(dx)$$

onde

$$\begin{aligned} E[\nu_x(C)|\mathcal{D}] &= E\{E[I_{\bar{C}}(\bar{\theta})|\mathcal{F}(\bar{x})|\mathcal{D}_+]\} \\ &= E[I_{\bar{C}}(\bar{\theta})|\mathcal{D}_+] = \nu_{\mathcal{D}}(C) \quad [\mu] \end{aligned}$$

e analogamente  $E[P_{\theta}(D)|\mathcal{C}] = P_{\mathcal{C}}(D) \quad [\mu]$ .

Assim,  $\mu^{\mathcal{G}}$  admite a decomposição

$$(9.26) \quad \mu^{\mathcal{G}} = \nu^{\mathcal{C}} \times P_{\mathcal{C}}^{\mathcal{D}} = P^{\mathcal{D}} \times \nu_{\mathcal{D}}^{\mathcal{C}}$$

onde  $P_{\mathcal{C}}^{\mathcal{D}}$  (resp.  $\nu_{\mathcal{D}}^{\mathcal{C}}$ ) indica a restrição a  $\mathcal{D}$  ( $\mathcal{C}$ ) da probabilidade amostral (a posteriori) condicional a  $\mathcal{C}$  (a  $\mathcal{D}$ ), que, sob a regularidade assumida, podem ser calculadas por

$$(9.27) \quad P_{\mathcal{C}}^{\mathcal{D}}(D) = \int_{\Theta} P_{\theta}(D) \nu_{\mathcal{C}}(d\theta), \quad \forall D \in \mathcal{D}$$

$$(9.28) \quad \nu_{\mathcal{D}}^{\mathcal{C}}(C) = \int_{\mathcal{X}} \nu_x(C) P_{\mathcal{D}}(dx), \quad \forall C \in \mathcal{C}$$

Daí a designação usual de probabilidade amostral marginalizada (restringida a  $\mathcal{D}$ ) para  $P_{\mathcal{C}}^{\mathcal{D}}$  e de probabilidade a posteriori marginalizada (restringida a  $\mathcal{C}$ ) para  $\nu_{\mathcal{D}}^{\mathcal{C}}$ .

Para a determinação de  $E_{\Pi}(h|\mathcal{C} \times \mathcal{D})$ , analisemos a estrutura da medida  $\Pi$  condicional a  $\mathcal{C} \times \mathcal{D}$ . Para todo  $B \in \mathcal{B}$ ,  $A \in \mathcal{A}$ , tem-se  $\forall C \times D \in \mathcal{C} \times \mathcal{D}$ , por um lado,

$$\Pi[(B \cap C) \times (A \cap D)] = \int_{\mathcal{C} \times \mathcal{D}} \Pi(B \times A|\mathcal{C} \times \mathcal{D})(\omega) \nu^{\mathcal{C}}(d\theta) P^{\mathcal{D}}(dx)$$

e por outro,

$$\begin{aligned} \Pi[(B \cap C) \times (A \cap D)] &= \nu(B \cap C) P(A \cap D) \\ &= \int_{\mathcal{C}} \nu_{\mathcal{C}}(B) \nu^{\mathcal{C}}(d\theta) \times \int_{\mathcal{D}} P_{\mathcal{D}}(A) P^{\mathcal{D}}(dx) \\ &= \int_{\mathcal{C} \times \mathcal{D}} \nu_{\mathcal{C}}(B) P_{\mathcal{D}}(A) \nu^{\mathcal{C}}(d\theta) P^{\mathcal{D}}(dx) . \end{aligned}$$

Conclui-se assim que a medida condicional  $\Pi_{\mathcal{G}}$ , com  $\mathcal{G} = \mathcal{C} \times \mathcal{D}$ , admite a decomposição

$$(9.29) \quad \Pi_{\mathcal{G}} = \nu_{\mathcal{C}} \times P_{\mathcal{D}} \quad [\Pi^{\mathcal{G}}]$$

pelo que, dada a regularidade das probabilidades condicionais  $\nu_{\mathcal{C}}$  e  $P_{\mathcal{D}}$ , se tem

$$(9.30) \quad E_{\Pi}(h|\mathcal{C} \times \mathcal{D}) = \int_{\Theta \times \mathcal{X}} h(\theta, x) \nu_{\mathcal{C}}(d\theta) P_{\mathcal{D}}(dx)$$

Nos casos particulares em que  $\mathcal{G} = \mathcal{B} \times \mathcal{D}$  e  $\mathcal{G} = \mathcal{C} \times \mathcal{A}$  tem-se obviamente

$$(9.31) \quad E_{\Pi}[h|\mathcal{B} \times \mathcal{D}] = \int_{\mathcal{X}} h(\theta, x) P_{\mathcal{D}}(dx)$$

$$(9.32) \quad E_{\Pi}[h|\mathcal{C} \times \mathcal{A}] = \int_{\Theta} h(\theta, x) \nu_{\mathcal{C}}(d\theta)$$

A estrutura da medida  $\mu$  condicional em  $\mathcal{G} = \mathcal{C} \times \mathcal{D}$  segue-se da aplicação de um argumento análogo àquele usado em 9.3 para a derivação da decomposição de  $\mu_{\mathcal{F}(\bar{t})}$  e  $\mu_{\mathcal{F}(\bar{\phi})}$ . Assim, sob a regularidade assumida de  $\nu_{\mathcal{C}}$  e  $P_{\mathcal{D}}$ , o modelo  $(\Omega, \mathcal{G}, \mu_{\mathcal{G}})$  é regular, sendo

$$(9.33) \quad \mu_{\mathcal{G}} = \nu_{\mathcal{G}} \times P_{\mathcal{B} \times \mathcal{D}} = P_{\mathcal{G}} \times \nu_{\mathcal{A} \times \mathcal{C}} .$$

Para a definição precisa da dominação de modelos condicionais, consideremos momentaneamente  $\mathcal{F}_1$  e  $\mathcal{F}_2$  duas subálgebras de  $\mathcal{F}$  e seja  $F_i \in \mathcal{F}_i$ ,  $i = 1, 2$ . Então, por um lado, de  $d\mu/d\Pi = h$ , vem

$$\begin{aligned} \mu(F_1 \cap F_2) &= \int_{F_2} I_{F_1}(\omega) \mu(d\omega) = \int_{F_2} I_{F_1}(\omega) E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2)(\omega) \Pi(d\omega) \\ (9.34) \quad &= \int_{F_2} E_{\Pi}[I_{F_1}(\omega) E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2)|\mathcal{F}_2](\omega) \Pi(d\omega) \end{aligned}$$

onde nas integrais não se explicitam, por comodidade, as restrições inerentes de  $\Pi$ .

Por outro lado,

$$\mu(F_1 \cap F_2) = \int_{F_2} \mu(F_1|\mathcal{F}_2)(\omega)\mu(d\omega) = \int_{F_2} \mu(F_1|\mathcal{F}_2)(\omega)E_{\Pi}[h|\mathcal{F}_2](\omega)\Pi(d\omega)$$

que, por comparação com (9.34), dada a arbitrariedade de  $F_2 \in \mathcal{F}_2$ , implica

$$(9.35) \quad E_{\Pi}[I_{F_1}E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2)|\mathcal{F}_2] = \mu(F_1|\mathcal{F}_2)E_{\Pi}(h|\mathcal{F}_2) \quad [\Pi^{\mathcal{F}_2}] .$$

Observe-se que  $g_2 \equiv E_{\Pi}(h|\mathcal{F}_2) > 0$   $[\mu]$  e  $\{g_2 = 0\} \subset \{E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2) = 0\}$   $[\Pi]$ , já que

$$\mu(\{g_2 = 0\}) = \int_{\{g_2=0\}} E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2) d\Pi = \int_{\{g_2=0\}} g_2 d\Pi = 0 .$$

Definamos então a função  $\mathcal{F}_1 \vee \mathcal{F}_2$ -mensurável

$$(9.36) \quad h_{\mathcal{F}_2}^{\mathcal{F}_1} = \begin{cases} E_{\Pi}(h|\mathcal{F}_1 \vee \mathcal{F}_2)/g_2 & , \text{ se } g_2 > 0 \\ 1 & , \text{ se } g_2 = 0 \end{cases}$$

De (9.35) obtém-se então,  $\forall F_1 \in \mathcal{F}_1$

$$(9.37) \quad E_{\Pi}[I_{F_1}h_{\mathcal{F}_2}^{\mathcal{F}_1}|\mathcal{F}_2] = \mu(F_1|\mathcal{F}_2) .$$

Esta relação diz-nos que a restrição a  $\mathcal{F}_1$  de  $\mu_{\mathcal{F}_2}$  é dominada pela correspondente restrição de  $\Pi_{\mathcal{F}_2}$ , com derivada de Radon-Nikodym definida por  $h_{\mathcal{F}_2}^{\mathcal{F}_1}$ .

Consideremos agora a particularização para  $\mathcal{F}_2 = \mathcal{C} \times \mathcal{D}$  e tomemos inicialmente  $\mathcal{F}_1 = \mathcal{F}(\bar{\theta})$ . Então,  $\forall B \subset \mathcal{B}$ ,

$$(9.38) \quad \begin{aligned} \nu_{\mathcal{C} \times \mathcal{D}}(B) &= E_{\mu}[I_{\bar{B}}|\mathcal{C} \times \mathcal{D}] = E_{\Pi}[I_{\bar{B}}h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}}] \\ &= \int_B \int_{\mathcal{X}} h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}} \nu_{\mathcal{C}}(d\theta) P_{\mathcal{D}}(dx) = \int_B h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}} \nu_{\mathcal{C}}(d\theta) \end{aligned}$$

onde as duas últimas igualdades se justificam por (9.29) e pelo fato de  $h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}}$  ser  $\mathcal{B} \times \mathcal{D}$ -mensurável. A expressão (9.38) revela que  $\nu_{\mathcal{C} \times \mathcal{D}}$  (a restrição a  $\mathcal{F}(\bar{\theta})$  de  $\mu_{\mathcal{C} \times \mathcal{D}}$ ) é dominada por  $\nu_{\mathcal{C}}$  com derivada de Radon-Nikodym dada por

$$(9.39) \quad h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}} = \frac{E_{\Pi}(h|\mathcal{B} \times \mathcal{D})}{E_{\Pi}(h|\mathcal{C} \times \mathcal{D})} = \frac{\int_{\mathcal{X}} h(\theta, x) P_{\mathcal{D}}(dx)}{\int_{\Theta \times \mathcal{X}} h(\theta, x) \nu_{\mathcal{C}}(d\theta) P_{\mathcal{D}}(dx)}$$

Tomando agora  $\mathcal{F}_1 = \mathcal{F}(\bar{x})$  prova-se que a restrição a  $\mathcal{F}(\bar{x})$  de  $\mu_{\mathcal{C} \times \mathcal{D}}$ ,  $P_{\mathcal{C} \times \mathcal{D}}$ , é dominada por  $P_{\mathcal{D}}$  com derivada de Radon-Nikodym

$$(9.40) \quad h_{\mathcal{C} \times \mathcal{D}}^{\bar{x}} = \frac{E_{\Pi}(h|\mathcal{A} \times \mathcal{C})}{E_{\Pi}(h|\mathcal{C} \times \mathcal{D})} = \frac{\int_{\Theta} h(\theta, x) \nu_{\mathcal{C}}(d\theta)}{\int_{\Theta \times \mathcal{X}} h(\theta, x) \nu_{\mathcal{C}}(d\theta) P_{\mathcal{D}}(dx)} .$$

Consequentemente, a medida produto  $\nu_{\mathcal{C} \times \mathcal{D}} \times P_{\mathcal{C} \times \mathcal{D}}$  é dominada por  $\nu_{\mathcal{C}} \times P_{\mathcal{D}}$  com derivada de Radon-Nikodym definida por  $h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}} \times h_{\mathcal{C} \times \mathcal{D}}^{\bar{x}}$ . Observe-se que, em particular,

$$(9.41) \quad h_{\mathcal{C} \times \mathcal{A}}^{\bar{\theta}} = \frac{h(\theta, x)}{\int_{\Theta} h(\theta, x) \nu_{\mathcal{C}}(d\theta)}$$

$$(9.42) \quad h_{\mathcal{B} \times \mathcal{D}}^{\bar{x}} = \frac{h(\theta, x)}{\int_{\mathcal{X}} h(\theta, x) P_{\mathcal{D}}(dx)}$$

Basta fazer  $\mathcal{D} = \mathcal{A}$  em (9.39) e  $\mathcal{C} = \mathcal{B}$  em (9.40), respectivamente.

A decomposição (9.33) permite constatar que  $\mu_{\mathcal{C} \times \mathcal{D}} \ll \nu_{\mathcal{C} \times \mathcal{D}} \times P_{\mathcal{C} \times \mathcal{D}}$  com derivada de Radon-Nikodym definida por

$$h_{\mathcal{C} \times \mathcal{D}} = \frac{d\mu_{\mathcal{C} \times \mathcal{D}}/d\Pi_{\mathcal{C} \times \mathcal{D}}}{h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}} \times h_{\mathcal{C} \times \mathcal{D}}^{\bar{x}}} .$$

Como por (9.37) ( $\mathcal{F}_1 = \mathcal{F}, \mathcal{F}_2 = \mathcal{C} \times \mathcal{D}$ )

$$(9.43) \quad \frac{d\mu_{\mathcal{C} \times \mathcal{D}}}{d\Pi_{\mathcal{C} \times \mathcal{D}}} = h_{\mathcal{C} \times \mathcal{D}}^{\bar{\theta}, \bar{x}} = \frac{E_{\Pi}(h|\mathcal{F})}{E_{\Pi}(h|\mathcal{C} \times \mathcal{D})} = \frac{h(\theta, x)}{\int_{\Theta \times \mathcal{X}} h(\theta, x) \nu_{\mathcal{C}}(d\theta) P_{\mathcal{D}}(dx)}$$

resulta que

$$(9.44) \quad h_{\mathcal{C} \times \mathcal{D}} = \frac{h E_{\Pi}(h|\mathcal{C} \times \mathcal{D})}{E_{\Pi}(h|\mathcal{B} \times \mathcal{D}) E_{\Pi}(h|\mathcal{C} \times \mathcal{A})} ,$$

no conjunto onde os dois fatores do denominador são positivos, podendo ser tomada igual a 1, no caso contrário. Observe-se que

$$\{h > 0\} \subset \{E_{\Pi}(h|\mathcal{B} \times \mathcal{D}) > 0\} \cap \{E_{\Pi}(h|\mathcal{C} \times \mathcal{A}) > 0\} \quad [\Pi] .$$

Pela decomposição regular de  $\mu_{\mathcal{C} \times \mathcal{D}}$ , tem-se

$$h_{\mathcal{C} \times \mathcal{D}} = \frac{dP_{\mathcal{B} \times \mathcal{D}}}{dP_{\mathcal{C} \times \mathcal{D}}} = \frac{d\nu_{\mathcal{A} \times \mathcal{C}}}{d\nu_{\mathcal{C} \times \mathcal{D}}}$$

pelo que as probabilidades amostral e a posteriori condicionais,  $P_{\mathcal{B} \times \mathcal{D}}$  e  $\nu_{\mathcal{A} \times \mathcal{C}}$ , podem ser definidas por

$$(9.45) \quad \forall A \in \mathcal{A}, P_{\mathcal{B} \times \mathcal{D}}(A) = \int_A h_{\mathcal{C} \times \mathcal{D}} dP_{\mathcal{C} \times \mathcal{D}} = \int_A h_{\mathcal{B} \times \mathcal{D}}^{\bar{x}} dP_{\mathcal{D}}$$

$$(9.46) \quad \forall B \in \mathcal{B}, \nu_{\mathcal{A} \times \mathcal{C}}(B) = \int_B h_{\mathcal{C} \times \mathcal{D}} d\nu_{\mathcal{C} \times \mathcal{D}} = \int_B h_{\mathcal{C} \times \mathcal{A}}^{\bar{\theta}} d\nu_{\mathcal{C}}$$

onde  $h_{\mathcal{B} \times \mathcal{D}}^{\bar{x}}$  e  $h_{\mathcal{C} \times \mathcal{A}}^{\bar{\theta}}$  são expressas por (9.41) e (9.42), respectivamente.

### Exemplo 9.2

Considere o modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  onde  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A}$  é a  $\sigma$ -álgebra de Borel de  $\mathbb{R}^n$  e  $\mathcal{P} = \{P_\theta : \theta = (\alpha, \delta) \in \mathbb{R} \times \mathbb{R}_+\}$ , onde  $P_\theta(dx) = f(x; \theta)m_n(dx)$ , com  $f(x; \theta) \equiv f(\theta, x)$ , para  $\theta$  fixo, a função densidade da distribuição Normal  $n$ -variada,  $N_n(\alpha 1_n, \delta I_n)$ , em relação à medida de Lebesgue  $n$ -dimensional  $m_n$  (os símbolos  $1_n$  e  $I_n$  representam o vetor  $n \times 1$  de 1's e a matriz identidade de ordem  $n$ , respectivamente).

Considere ainda a medida de probabilidade a priori definida em  $(\Theta, \mathcal{B})$ , onde  $\Theta = \mathbb{R} \times \mathbb{R}_+$  e  $\mathcal{B}$  a  $\sigma$ -álgebra dos borelianos de  $\Theta$ , pelo membro da família conjugada de  $\mathcal{P}$ , i.e.,

$$\nu(d\alpha \times d\delta) = q(\alpha, \delta)m_2(d\alpha \times d\delta)$$

onde  $q(\alpha, \delta)$  é a função densidade Normal-Gama Inversa,  $\text{NGI}(a, b, c, d)$ , relativamente à medida de Lebesgue bidimensional,  $m_2$ , em  $\Theta$ . Mais concretamente,

$$\nu(d\alpha \times d\delta) = \nu_\delta(d\alpha)\nu^\delta(d\delta)$$

com

$$\begin{aligned}\nu_\delta(d\alpha) &= q_1(\alpha; \delta)m_1(d\alpha) \\ \nu^\delta(d\delta) &= q_2(\delta)m_1(d\delta)\end{aligned}$$

onde

$$\begin{aligned}q_1(\alpha; \delta) &= (2\Pi)^{-1/2}\delta^{-1/2} \exp\left[-\frac{b}{2\delta}(\alpha - a)^2\right], \quad b > 0 \\ q_2(\delta) &= \frac{d^c}{\Gamma(c)} \left(\frac{1}{\delta}\right)^{c+1} \exp\left[-\frac{d}{\delta}\right] I_{(0,+\infty)}(\delta), \quad c, d > 0.\end{aligned}$$

Isto é,  $\alpha|\delta \sim N(a, \delta|b)$  e  $\delta \sim G_a I(c, d)$ .

Equivalentemente,

$$\nu(d\alpha \times d\delta) = \nu^\alpha(d\alpha)\nu_\alpha(d\delta)$$

com

$$\begin{aligned}\nu_\alpha(d\delta) &= q_3(\delta; \alpha)m_1(d\delta) \\ \nu^\alpha(d\alpha) &= q_4(\alpha)m_1(d\alpha)\end{aligned}$$

onde  $q_3(\delta; \alpha)$  é a densidade  $G_a I(c + \frac{1}{2}, d(\alpha))$ , com  $d(\alpha) = d + \frac{b}{2}(\alpha - a)^2$ , relativamente à medida de Lebesgue,  $m_1$ , em  $\mathbb{R}_+$ , e  $q_4(\alpha)$  é a densidade,  $t(2c, a, \frac{d}{bc})$ , de uma  $t$ -student com  $2c$  graus de liberdade e parâmetros de localização,  $a$ , e de escala,  $\sqrt{d/bc}$ , relativamente à mesma medida em  $\mathbb{R}$ . Isto é,

$$q_4(\alpha) = \frac{\Gamma\left(\frac{2c+1}{2}\right)}{\Gamma(c)\Gamma(1/2)} \left(\frac{2d}{b}\right)^{-1/2} \left[1 + \frac{(\alpha - a)^2}{2\frac{d}{b}}\right]^{-\frac{2c+1}{2}}.$$

A medida preditiva é definida por  $P(dx) = g(x)m_n(dx)$ , onde  $g(x)$  é a densidade  $t$ -Student  $n$ -variada,  $t_n(2c, a1_n, \frac{d}{c}V)$ ,

$$g(x) = \frac{\Gamma\left(\frac{2c+n}{2}\right)}{\Gamma(c)[\Gamma(1/2)]^n} |V|^{-1/2} (2d)^{-n/2} \left[1 + \frac{(x - a1_n)'V^{-1}(x - a1_n)}{2d}\right]^{-\frac{2c+n}{2}}$$

com  $V = I_n + \frac{1_n 1_n'}{b}$  e  $|V| = (n + b)/b$ . As componentes de  $x$  são assim marginalmente equicorrelacionadas, permutáveis mas dependentes, com distribuições marginais  $t(2c, a, \frac{d}{c}(1 + 1/b))$ .

A “densidade” de  $\mu = \nu \times P_\theta$  com respeito a  $\Pi = \nu \times P$  é então dada por  $h(\theta, x) = f(\theta, x)/g(x)$ .

Definamos o modelo marginal para  $(\alpha, x)$ ,  $(\Omega, \mathcal{G}, \mu^{\mathcal{G}})$ , onde  $\mathcal{G} = \mathcal{B}(\alpha) \times \mathcal{A}$ .

A medida de probabilidade amostral marginalizada,  $P_\alpha$  ( $\equiv P_{\mathcal{B}(\alpha)}$ ), é definida por  $P_\alpha(dx) = f(x; \alpha)m_n(dx)$ , onde, por (9.27),  $f(x; \alpha) = \int_0^\infty f(\theta, x)\nu_\alpha(d\delta)$ , traduzindo, para  $\alpha$  fixo, a densidade  $t_n(2c + 1, \alpha 1_n, \frac{d(\alpha)}{c + \frac{1}{2}}I_n)$ . De novo, as componentes de  $x$  são, dado  $\alpha$ , permutáveis e dependentes, mas agora não correlacionadas.

Por (9.32),  $E_\Pi[h(\theta, x)|\mathcal{B}(\alpha) \times \mathcal{A}] = f(x; \alpha)/g(x)$ , donde

$$\begin{aligned} \mu^{\mathcal{G}}(d\alpha \times dx) &= \frac{f(x; \alpha)}{g(x)} \nu^\alpha(d\alpha) P(dx) \\ &= f(x; \alpha) q_4(\alpha) m_1(d\alpha) m_n(dx) \end{aligned}$$

i.e.,  $d\mu^{\mathcal{G}}/d(m_1 \times m_n)$  é o produto de duas densidades  $t$ -Student.

Definamos agora o modelo marginal para  $(\theta, t)$ ,  $(\Omega, \mathcal{G}, \mu^{\mathcal{G}})$ , onde  $\mathcal{G} = \mathcal{B} \times \mathcal{A}(t)$  e  $t = \sum x_i/n$ .

A probabilidade amostral restringida a  $\mathcal{A}(t)$ ,  $P_\theta^t$ , é definida por  $P_\theta^t(dt) = f(t; \theta)m_1(dt)$ , onde  $f(t; \theta)$  é, para  $\theta$  fixo, a densidade  $N(\alpha, \delta/n)$ . Consequentemente, a probabilidade preditiva restringida a  $\mathcal{A}(t)$ ,  $P^t$ , é expressa por  $P^t(dt) = g(t)m_1(dt)$ , com  $g(t) = \int_{\Theta} f(\theta, t)\nu(d\theta)$  traduzindo a densidade  $t(2c, a, \frac{n+b}{nb} \frac{d}{c})$ .

Observe-se que a restrição a  $\mathcal{F}(\bar{x})$  de  $\mu$  condicional a  $\mathcal{F}(\bar{t})$ ,  $P_t$ , e a  $\mathcal{G} = \mathcal{B} \times \mathcal{A}(t)$ ,  $P_{\theta,t}$ , são definidas por

$$P_t(dx) = \frac{g(x)}{g(t)} m_n(dx)$$

e

$$P_{\theta,t}(dx) = \frac{f(x; \theta)}{f(t; \theta)} m_n(dx)$$

com  $x$  tal que  $t(x) = t$  (recorde, em particular, (9.42) e (9.45)).

A medida de probabilidade a posteriori,  $\nu_x$ , é definida por

$$\nu_x(d\theta) = h(\theta, x) \nu(d\theta) = q(\theta; x) m_2(d\theta)$$

onde  $q(\theta; x) = h(\theta, x) q(\theta)$  é a densidade NGI  $(A, B, C, D)$ , com  $A = \frac{ba+nt}{b+n}$ ,  $B = b + n$ ,  $C = c + n/2$  e  $D = d + \sum(x_i - t)^2/2$ .

Por outras palavras,

$$\nu_x(d\alpha \times d\delta) = \nu_{x,\delta}(d\alpha) \times \nu_x^\delta(d\delta) = \nu_x^\alpha(d\alpha) \times \nu_{x,\alpha}(d\delta)$$

onde

$$\nu_{x,\delta}(d\alpha) = q_1(\alpha; \delta, x) m_1(d\alpha)$$

$$\nu_x^\delta(d\delta) = q_2(\delta; x) m_1(d\delta)$$

$$\nu_{x,\alpha}(d\delta) = q_3(\delta; \alpha, x) m_1(d\delta)$$

$$\nu_x^\alpha(d\alpha) = q_4(\alpha; x) m_1(d\alpha)$$

com  $q_i$ ,  $i = 1, 2, 3, 4$  traduzindo, respectivamente, as densidades  $N(A, \delta/B)$ ,  $G_a I(C, D)$ ,  $G_a I(C + 1/2, D(\alpha))$ , com  $D(\alpha) = D + \frac{B}{2}(\alpha - A)^2$ , e  $t(2C, A, D/BC)$ . Note, em particular, a concretização de (9.41) e (9.46).

A probabilidade a posteriori marginalizada em  $t$ ,  $\nu_t$ , é definida, atendendo a (9.28), por

$$\nu_t(d\theta) = q(\theta; t) m_2(d\theta)$$

onde  $q(\theta, t) = q(\theta) f(\theta; t) / g(t)$  traduz a densidade NGI  $(A, B, C_t, D_t)$ , com  $C_t = c + 1/2$  e  $D_t = d + \frac{nb}{2B}(t - a)^2$ . Em particular, tem-se assim que, em relação à medida de Lebesgue unidimensional,  $\nu_t^\delta$  é definida pela densidade  $G_a I(C_t, D_t)$  e  $\nu_{t,\delta}$  coincide com  $\nu_{x,\delta}$  para  $t = t(x)$ .

Atendendo a (9.31),  $E_{\Pi}[h(\theta, x)|\mathcal{G}] = f(\theta, t)/g(t)$ , para  $\mathcal{G} = \mathcal{B} \times \mathcal{A}(t)$ , pelo que

$$\begin{aligned}\mu^{\mathcal{G}}(d\theta \times dt) &= \frac{f(\theta, t)}{g(t)} \nu(d\theta) P^t(dt) \\ &= f(\theta, t) q(\theta) m_2(d\theta) m_1(dt)\end{aligned}$$

define a medida produto em  $(\Omega, \mathcal{G})$ .

### Exercício 9.5

Relativamente ao Exemplo 9.2:

- Verifique os resultados distribucionais referidos.
- Mostre que o modelo marginal para  $(\sigma^2, x)$  é definido por uma densidade, relativamente à medida de Lebesgue em  $\mathbb{R}^n \times \mathbb{R}_+$ , que é o produto das densidades  $N_n(a1_n, \sigma^2 V)$  e  $G_a I(c, d)$ .
- Mostre que o modelo marginal para  $(\alpha, t)$  é definido por uma densidade, relativamente à medida de Lebesgue em  $\mathbb{R}^2$ , que é o produto das densidades  $t(c + 1, \alpha, d(\alpha)/[n(c + 1)])$  e  $t(2c, a, d/(bc))$ .
- Defina os modelos associados com:
  - a restrição a  $\mathcal{F}(\bar{x})$  e a  $\mathcal{F}(\bar{\theta})$  da medida produto  $\mu$  condicional a  $\mathcal{B}(\alpha) \times \mathcal{A}(t)$ ;
  - a restrição a  $\mathcal{F}(\bar{t})$  de  $\mu_{\mathcal{F}(\bar{\alpha})}$  e a restrição a  $\mathcal{F}(\bar{\alpha})$  de  $\mu_{\mathcal{F}(\bar{t})}$ .
- Sendo  $s = \sum (x_i - t)^2 / (n - 1)$ , defina a medida marginal  $P^{t,s}$ , mostrando em particular que  $s / [\frac{2d}{n-1} + s] \sim \mathcal{B}_e(\frac{n-1}{2}, c)$  e que  $s$  e  $t$  são marginalmente dependentes.
- Prove que, sendo  $\mathcal{A}(t, s)$  a  $\sigma$ -álgebra gerada por  $(t, s)$ , se verificam as identidades:

$$\begin{aligned}P_{\mathcal{B} \times \mathcal{A}(t,s)} &= P_{\mathcal{A}(t,s)} \\ \nu_{\mathcal{A}} &= \nu_{\mathcal{A}(t,s)}\end{aligned}$$



## CAPÍTULO 10

# INDEPENDÊNCIA CONDICIONAL E SUFICIÊNCIA BAYESIANA AMOSTRAL

### 10.1 CONCEITO E PROPRIEDADES DA INDEPENDÊNCIA CONDICIONAL

Antes de nos referirmos propriamente à definição e às propriedades relevantes da independência condicional no quadro do modelo bayesiano  $(\Omega, \mathcal{F}, \mu)$ , faremos uma breve digressão sobre o completamento de subálgebras de  $\mathcal{F}$  na linha do que foi feito no Capítulo 1 sobre as subálgebras de  $\mathcal{A}$  no contexto do modelo clássico.

Para o efeito, seja  $\overline{\mathcal{F}}_0 = \{F \in \mathcal{F} : \mu^2(F) = \mu(F)\}$  o completamento da  $\sigma$ -álgebra trivial  $\mathcal{F}_0 = \{\phi, \Omega\}$ , que é assim gerada pelo  $\sigma$ -anel,  $\mathcal{N}$ , dos conjuntos  $\mu$ -nulos de  $\mathcal{F}$ . O completamento de qualquer subálgebra  $\mathcal{F}_1 \subset \mathcal{F}$  é definido pela subálgebra  $\overline{\mathcal{F}}_1 = \mathcal{F}_1 \vee \overline{\mathcal{F}}_0$ .

As propriedades dos completamentos de subálgebras de  $\mathcal{A}$  com os conjuntos  $\mathcal{P}$ -nulos referidas no Capítulo 1 aplicam-se na sua essência a este contexto. Observe-se, no entanto, que os dois completamentos são naturalmente distintos já que  $\overline{\mathcal{F}}_0$  (e mesmo, em geral,  $\overline{\mathcal{F}}_0 \cap \mathcal{F}(\overline{x})$ ) contém estritamente os conjuntos  $\mathcal{P}$ -nulos de  $\mathcal{F}(\overline{x})$ . Consequentemente,  $\forall \mathcal{D} \subset \mathcal{A}$ ,  $\overline{\mathcal{D}}_+$  não está incluída necessariamente em  $\mathcal{F}(\overline{x})$ .

A este respeito convém notar que  $\forall \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$  o fato de  $\overline{\mathcal{F}}_1 = \mathcal{F}_1 \vee \mathcal{N}$  (constituindo

a classe de todos os conjuntos de  $\mathcal{F}$   $\mu$ -essencialmente iguais aos elementos de  $\mathcal{F}_1$ ), permite verificar que  $\mathcal{F}_1 \vee (\mathcal{F}_2 \cap \overline{\mathcal{F}}_0) = \overline{\mathcal{F}}_1 \cap \mathcal{F}_2$ . Ou seja, o completamento de  $\mathcal{F}_1$  com os conjuntos  $\mu$ -nulos de  $\mathcal{F}_2$  coincide com a  $\sigma$ -álgebra dos conjuntos de  $\mathcal{F}_2$  que são  $\mu$ -iguais aos conjuntos de  $\mathcal{F}_1$ . Deste modo, e voltando à ilustração acima,  $\overline{\mathcal{D}}_+ \cap \mathcal{F}(\overline{x})$  representa a menor subálgebra amostral contendo  $\mathcal{D}_+$  e todos os conjuntos  $\mu$ -nulos de  $\mathcal{F}(\overline{x})$ .

Para terminar este parêntesis, relembramos que  $\mathcal{F}_1$  diz-se completada (resp.  $\mathcal{F}_2$ -completada com  $\mathcal{F}_2 \supset \mathcal{F}_1$ ) se  $\overline{\mathcal{F}}_1 = \mathcal{F}_1$  (resp.  $\overline{\mathcal{F}}_1 \cap \mathcal{F}_2 = \mathcal{F}_1$ ).

Consideremos então que  $\mathcal{F}_i$ ,  $i = 1, 2, 3$  são subálgebras de  $\mathcal{F}$ .

### Definição 10.1

Diz-se que  $\mathcal{F}_1$  e  $\mathcal{F}_2$  são independentes condicionalmente a  $\mathcal{F}_3$  – e escreve-se  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$ , se  $\forall f_i \in \mathcal{L}(\mathcal{F}_i)$ ,  $i = 1, 2$

$$E(f_1 f_2 | \mathcal{F}_3) = E(f_1 | \mathcal{F}_3) E(f_2 | \mathcal{F}_3) \quad [\mu]$$

### Exercício 10.1

Prove que para a definição de  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  é suficiente restringirmo-nos a funções indicadoras dos conjuntos de  $\mathcal{F}_i$ ,  $i = 1, 2$ .

O resultado seguinte constitui uma formulação alternativa da Definição 10.1.

### Teorema 10.1

$\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  se, e somente se,  $\forall f_2 \in \mathcal{L}(\mathcal{F}_2)$

$$E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) = E(f_2 | \mathcal{F}_3) \quad [\mu]$$

### Prova

(Parte suficiente): Sejam  $f_i \in \mathcal{L}(\mathcal{F}_i)$ ,  $i = 1, 2$ . Pelas propriedades de “smoothing” e de idempotência (i. e.,  $E(f | \mathcal{F}_1) = f$   $[\mu]$  se  $f \in \mathcal{L}(\mathcal{F}_1)$ ), tem-se

$$E(f_1 f_2 | \mathcal{F}_3) = E\{E(f_1 f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) | \mathcal{F}_3\} = E\{f_1 E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) | \mathcal{F}_3\} \quad [\mu]$$

donde

$$E(f_1 f_2 | \mathcal{F}_3) = E\{f_1 E(f_2 | \mathcal{F}_3) | \mathcal{F}_3\} = E(f_1 | \mathcal{F}_1) E(f_2 | \mathcal{F}_3) \quad [\mu]$$

como pretendíamos.

(Parte necessária): Sejam  $F_i \in \mathcal{F}_i$ ,  $i = 1, 3$ . Como  $F_1 \cap F_3 \in \mathcal{F}_1 \vee \mathcal{F}_3$  e atendendo à definição de esperança condicional dado  $\mathcal{F}_1 \vee \mathcal{F}_3$  e dado  $\mathcal{F}_3$ , resulta

$$\int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) d\mu = \int_{F_1 \cap F_3} f_2 d\mu = \int_{F_3} f_2 I_{F_1} d\mu = \int_{F_3} E(f_2 I_{F_1} | \mathcal{F}_3) d\mu$$

onde, por comodidade, se omitiu a explicitação das restrições apropriadas de  $\mu$ .

Atendendo à hipótese e à definição de  $E(I_{F_1} | \mathcal{F}_3)$ , vem

$$\int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) d\mu = \int_{F_3} E(f_2 | \mathcal{F}_3) E(I_{F_1} | \mathcal{F}_3) d\mu = \int_{F_1 \cap F_3} E(f_2 | \mathcal{F}_3) d\mu .$$

Definamos agora

$$\begin{aligned} \mathcal{J}_0 &= \{F_1 \cap F_3 : F_i \in \mathcal{F}_i, i = 1, 3\} \quad \text{e} \\ \mathcal{M} &= \{F \in \mathcal{F}_1 \vee \mathcal{F}_3 : \int_F f_2 d\mu = \int_F E(f_2 | \mathcal{F}_3) d\mu\} . \end{aligned}$$

É fácil constatar que  $\mathcal{J}_0$  é uma classe-II geradora de  $\mathcal{F}_1 \vee \mathcal{F}_3$  que contém  $\mathcal{M}$  que é um sistema de Dynkin. O teorema de Dynkin assegura-nos então que  $\mathcal{M} = \mathcal{F}_1 \vee \mathcal{F}_3$  e, como tal,  $E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) = E(f_2 | \mathcal{F}_3) \quad [\mu]$ , Q.E.D.  $\square$

Dado que toda versão de  $E(f_2 | \mathcal{F}_3)$  é  $\overline{\mathcal{F}}_3$ -mensurável e atendendo ao significado de mensurabilidade  $\mu$ -essencial para variáveis aleatórias, tem-se

### Corolário 10.2

$\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  se, e somente se,  $\forall f_2 \in \mathcal{L}(\mathcal{F}_2)$ ,  $E(f_2 | \mathcal{F}_1 \vee \mathcal{F}_3) \in \overline{\mathcal{F}}_3$ .

No caso particular em que  $\mathcal{F}_3 \subset \mathcal{F}_1$ ,  $\mathcal{F}_1 \vee \mathcal{F}_3 = \mathcal{F}_1$  e assim

### Corolário 10.3

Se  $\mathcal{F}_3 \subset \mathcal{F}_1$ , então  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  se, e somente se,  $E(f_2 | \mathcal{F}_1) \in \overline{\mathcal{F}}_3$ ,  $\forall f_2 \in \mathcal{L}(\mathcal{F}_2)$ .

O resultado seguinte ilustra a propriedade de monotonicidade da independência condicional relativamente a subálgebras de  $\mathcal{F}_1$ .

**Teorema 10.4**

Se  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  e  $\mathcal{F}_3 \subset \mathcal{F}_4 \subset \mathcal{F}_1$ , então  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_4$ .

**Prova**

Pelo Corolário 10.3, tem-se  $E(f_2 | \mathcal{F}_1) = E(f_2 | \mathcal{F}_3) \ [\mu] \ \forall f_2 \in \mathcal{L}(\mathcal{F}_2)$ . Mas, pelo fato de  $\mathcal{F}_4 \subset \mathcal{F}_1$ , a definição de  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  implica que  $\mathcal{F}_4 \amalg \mathcal{F}_2 | \mathcal{F}_3$ , ou seja, que  $E(f_2 | \mathcal{F}_4) = E(f_2 | \mathcal{F}_3) \ [\mu], \ \forall f_2 \in \mathcal{L}(\mathcal{F}_2)$ . Conseqüentemente,  $\forall f_2 \in \mathcal{L}(\mathcal{F}_2), E(f_2 | \mathcal{F}_1) = E(f_2 | \mathcal{F}_4) \ [\mu]$ . Q.E.D.  $\square$

Um resultado de importância crucial sobre o conceito de independência condicional é indicado no teorema seguinte.

**Teorema 10.5**

As seguintes condições são equivalentes:

- i)  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  e  $\mathcal{F}_1 \amalg \mathcal{F}_4 | \mathcal{F}_2 \vee \mathcal{F}_3$ ;
- ii)  $\mathcal{F}_1 \amalg (\mathcal{F}_2 \vee \mathcal{F}_4) | \mathcal{F}_3$ ;
- iii)  $\mathcal{F}_1 \amalg \mathcal{F}_4 | \mathcal{F}_3$  e  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_4 \vee \mathcal{F}_3$ .

**Prova**

Pela simetria de  $\mathcal{F}_2$  e  $\mathcal{F}_4$ , basta provar a equivalência entre i) e ii).

i)  $\Rightarrow$  ii):  $\forall f_1 \in \mathcal{L}(\mathcal{F}_1)$ , a relação

$$E(f_1 | \mathcal{F}_2 \vee \mathcal{F}_4 \vee \mathcal{F}_3) = E(f_1 | \mathcal{F}_2 \vee \mathcal{F}_3) = E(f_1 | \mathcal{F}_3)$$

decorre de i), a qual prova ii).

ii)  $\Rightarrow$  i):  $\forall f_1 \in \mathcal{L}(\mathcal{F}_1)$ , ii) garante

$$E(f_1 | \mathcal{F}_2 \vee \mathcal{F}_4 \vee \mathcal{F}_3) = E(f_1 | \mathcal{F}_3) \ [\mu]$$

e o fato de  $\mathcal{F}_3 \subset \mathcal{F}_2 \vee \mathcal{F}_3 \subset \mathcal{F}_2 \vee \mathcal{F}_4 \vee \mathcal{F}_3$

$$E(f_1 | \mathcal{F}_2 \vee \mathcal{F}_4 \vee \mathcal{F}_3) = E(f_1 | \mathcal{F}_2 \vee \mathcal{F}_3) \ [\mu] .$$

Conseqüentemente, obtém-se i).  $\square$

Uma implicação importante deste teorema mostra que a relação  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$  permanece válida se  $\mathcal{F}_1$  (ou  $\mathcal{F}_2$ ) são alargadas na direção de  $\mathcal{F}_3$ , ou se  $\mathcal{F}_3$  é alargada na direção de  $\mathcal{F}_1$  e/ou de  $\mathcal{F}_2$ .

**Corolário 10.6**

Se  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$ ,  $\mathcal{F}_4 \subset \mathcal{F}_2 \vee \mathcal{F}_3$  e  $\mathcal{F}_5 \subset \mathcal{F}_1 \vee \mathcal{F}_3$ , então:

- i)  $\mathcal{F}_1 \vee \mathcal{F}_5 \amalg \mathcal{F}_2 \vee \mathcal{F}_4 | \mathcal{F}_3$ ;
- ii)  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3 \vee \mathcal{F}_4 \vee \mathcal{F}_5$ .

**Prova**

Obviamente que o fato de  $\mathcal{F}_4 \subset \mathcal{F}_2 \vee \mathcal{F}_3$  implica que  $\mathcal{F}_1 \amalg \mathcal{F}_4 | \mathcal{F}_2 \vee \mathcal{F}_3$ . O Teorema 10.5 garante então que  $\mathcal{F}_1 \amalg \mathcal{F}_2 \vee \mathcal{F}_4 | \mathcal{F}_3$ . Esta relação juntamente com  $\mathcal{F}_5 \amalg \mathcal{F}_2 \vee \mathcal{F}_4 | \mathcal{F}_1 \vee \mathcal{F}_3$ , implicada por  $\mathcal{F}_5 \subset \mathcal{F}_1 \vee \mathcal{F}_3$ , conduzem pelo mesmo teorema a  $\mathcal{F}_1 \vee \mathcal{F}_5 \amalg \mathcal{F}_2 \vee \mathcal{F}_4 | \mathcal{F}_3$ . Esta relação implica (faça-se  $\mathcal{F}_1^* = \mathcal{F}_1 \vee \mathcal{F}_5$  no Teorema 10.5)  $\mathcal{F}_1 \vee \mathcal{F}_5 \amalg \mathcal{F}_2 | \mathcal{F}_3 \vee \mathcal{F}_4$  que, por sua vez, implica a relação ii), Q.E.D. □

A caracterização de independência condicional em termos de densidades é proporcionada pelo seguinte resultado:

**Teorema 10.7**

Seja  $(\Omega, \mathcal{F}, \mu)$  dominado pelo modelo  $(\Omega, \mathcal{F}, \Pi)$  e  $h = d\mu/d\Pi$ . Se as subálgebras  $\mathcal{F}_1$  e  $\mathcal{F}_2$  são independentes condicionalmente a  $\mathcal{F}_3$  relativamente à medida  $\Pi$ , então em termos de  $\mu$

$$\begin{aligned} \mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3 &\Leftrightarrow h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2} = h_{\mathcal{F}_3}^{\mathcal{F}_2} [\mu] \text{ em } \{E_{\Pi}(h | \mathcal{F}_3) > 0\} \\ &\Leftrightarrow h_{\mathcal{F}_2 \vee \mathcal{F}_3}^{\mathcal{F}_1} = h_{\mathcal{F}_3}^{\mathcal{F}_1} [\mu] \text{ em } \{E_{\Pi}(h | \mathcal{F}_2 \vee \mathcal{F}_3) > 0\} \end{aligned}$$

**Prova**

Por hipótese,  $\forall f_i \in \mathcal{L}(\mathcal{F}_i)$ ,  $i = 1, 2$

$$\begin{aligned} E_{\mu}(f_1 f_2 | \mathcal{F}_3) &= E_{\Pi}(f_1 f_2 h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2} | \mathcal{F}_3) [\mu] \\ E_{\mu}(f_1 | \mathcal{F}_3) &= E_{\Pi}(f_1 h_{\mathcal{F}_3}^{\mathcal{F}_1} | \mathcal{F}_3) [\mu] \\ E_{\mu}(f_2 | \mathcal{F}_3) &= E_{\Pi}(f_2 h_{\mathcal{F}_3}^{\mathcal{F}_2} | \mathcal{F}_3) [\mu] \end{aligned}$$

Pela independência de  $\mathcal{F}_1$  e  $\mathcal{F}_2$  condicional a  $\mathcal{F}_3$  com respeito a  $\Pi$  e usando o Teorema 10.5,

$$E_{\Pi}(f_1 h_{\mathcal{F}_3}^{\mathcal{F}_1} | \mathcal{F}_3) E_{\Pi}(f_2 h_{\mathcal{F}_3}^{\mathcal{F}_2} | \mathcal{F}_3) = E_{\Pi}(f_1 f_2 h_{\mathcal{F}_3}^{\mathcal{F}_1} h_{\mathcal{F}_3}^{\mathcal{F}_2} | \mathcal{F}_3) \quad [\mu] .$$

Por conseguinte,

$$E_{\mu}(f_1 f_2 | \mathcal{F}_3) = E_{\mu}(f_1 | \mathcal{F}_3) E_{\mu}(f_2 | \mathcal{F}_3) \quad [\mu]$$

se, e somente se,

$$E_{\Pi}(f_1 f_2 h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2} | \mathcal{F}_3) = E_{\Pi}(f_1 f_2 h_{\mathcal{F}_3}^{\mathcal{F}_1} h_{\mathcal{F}_3}^{\mathcal{F}_2} | \mathcal{F}_3) \quad [\mu] .$$

Isto equivale a

$$h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2} = h_{\mathcal{F}_3}^{\mathcal{F}_1} h_{\mathcal{F}_3}^{\mathcal{F}_2} \quad [\mu^{\mathcal{F}_3}] \Leftrightarrow E_{\Pi}(h | \mathcal{F}_3) h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2} = E_{\Pi}(h | \mathcal{F}_3) h_{\mathcal{F}_3}^{\mathcal{F}_1} h_{\mathcal{F}_3}^{\mathcal{F}_2} \quad [\mu] ,$$

ou seja, ao resultado pretendido.

Como, por definição

$$h_{\mathcal{F}_2 \vee \mathcal{F}_3}^{\mathcal{F}_1} = \frac{E_{\Pi}(h | \mathcal{F}_1 \vee \mathcal{F}_2 \vee \mathcal{F}_3)}{E_{\Pi}(h | \mathcal{F}_2 \vee \mathcal{F}_3)} = \frac{h_{\mathcal{F}_3}^{\mathcal{F}_1 \vee \mathcal{F}_2}}{h_{\mathcal{F}_3}^{\mathcal{F}_2}}$$

sempre que  $E_{\Pi}(h | \mathcal{F}_2 \vee \mathcal{F}_3) > 0$ , o teorema fica cabalmente demonstrado.  $\square$

Consideremos agora um conceito da Teoria das Probabilidades enquadrado no espaço  $(\Omega, \mathcal{F}, \mu)$ .

### Definição 10.2

Sendo  $\mathcal{F}_1$  e  $\mathcal{F}_2$  duas subálgebras de  $\mathcal{F}$ , chama-se projeção de  $\mathcal{F}_2$  em  $\mathcal{F}_1$  – e denota-se por  $\mathcal{F}_1[\mathcal{F}_2]$  – a  $\sigma$ -álgebra gerada por qualquer versão de  $E(f_2 | \mathcal{F}_1)$  para toda a função  $f_2 \in \mathcal{L}(\mathcal{F}_2)$ , i.e.,

$$\mathcal{F}_1[\mathcal{F}_2] = \sigma(\{E(f_2 | \mathcal{F}_1) : f_2 \in \mathcal{L}(\mathcal{F}_2)\}) .$$

O resultado seguinte esclarece o papel que este conceito pode desempenhar: o de constituir essencialmente a menor das subálgebras de  $\mathcal{F}_1$  condicionalmente às quais  $\mathcal{F}_1$  se torna independente de  $\mathcal{F}_2$ .

### Teorema 10.8

Sejam  $\mathcal{F}_i$ ,  $i = 1, 2, 3$  subálgebras de  $\mathcal{F}$  tal que  $\mathcal{F}_3 \subset \mathcal{F}_1$ . Então

- i)  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_1[\mathcal{F}_2]$ ;
- ii) Se  $\mathcal{F}_1 \amalg \mathcal{F}_2 | \mathcal{F}_3$ , então  $\mathcal{F}_1[\mathcal{F}_2] \subset \mathcal{F}_3$   $[\mu]$ .

**Prova**

Como  $\mathcal{F}_1[\mathcal{F}_2] \subset \mathcal{F}_1$ , o Corolário 10.3 assegura i). Por outro lado, este corolário, ao afirmar que na hipótese de ii),  $E(f_2|\mathcal{F}_1) \in \overline{\mathcal{F}}_3$ ,  $\forall f_2 \in \mathcal{L}(\mathcal{F}_2)$ , implica que  $\mathcal{F}_1[\mathcal{F}_2] \subset \overline{\mathcal{F}}_3$ , pela definição de  $\mathcal{F}_1[\mathcal{F}_2]$ .  $\square$

**Exercício 10.2**

Mostre que

- i)  $\mathcal{F}_1 \cap \overline{\mathcal{F}}_0 \subset \mathcal{F}_1 \cap \overline{\mathcal{F}}_2 \subset \mathcal{F}_1[\mathcal{F}_2] \subset \mathcal{F}_1$ ;
- ii)  $\overline{\mathcal{F}_1[\mathcal{F}_2]} \cap \mathcal{F}_1 = \mathcal{F}_1[\mathcal{F}_2]$ .

**Nota 10.1**

Os resultados acima mostram que  $\mathcal{F}_1[\mathcal{F}_2]$  é a intersecção de todas as subálgebras de  $\mathcal{F}_1$  que contêm os conjuntos nulos de  $\mathcal{F}_1$  e condicionalmente às quais  $\mathcal{F}_1$  e  $\mathcal{F}_2$  são independentes.

O Corolário 10.3 estabelece uma condição de mensurabilidade que é necessária e suficiente para a independência condicional entre  $\mathcal{F}_1$  e  $\mathcal{F}_2$  dado uma subálgebra de  $\mathcal{F}_1$ . Nestas condições, uma outra condição equivalente com relevância estatística, como referiremos na Seção 10.3, é fornecida pelo seguinte teorema:

**Teorema 10.9**

Se  $\mathcal{F}_3 \subset \mathcal{F}_1$ ,  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2|\mathcal{F}_3$  se, e somente se,

$$\forall f \in \mathcal{L}(\mathcal{F}_1), \quad E\{E(f|\mathcal{F}_3)|\mathcal{F}_2\} = E(f|\mathcal{F}_2) \quad [\mu]$$

**Prova**

(Parte necessária): Por definição,  $\forall f \in \mathcal{L}(\mathcal{F}_1)$

$$E(f|\mathcal{F}_2 \vee \mathcal{F}_3) = E(f|\mathcal{F}_3) \quad [\mu]$$

o que implica que, pela propriedade de “smoothing”,

$$E[E(f|\mathcal{F}_2 \vee \mathcal{F}_3)|\mathcal{F}_2] = E(f|\mathcal{F}_2) = E[E(f|\mathcal{F}_3)|\mathcal{F}_2] \quad [\mu].$$

(Parte suficiente): Seja  $f_i \in \mathcal{L}(\mathcal{F}_i)$ ,  $i = 1, 2, 3$ . Como  $f_1 f_3 \in \mathcal{L}(\mathcal{F}_1)$ ,

$$\begin{aligned} E(f_1 f_2 f_3|\mathcal{F}_2) &= f_2 E(f_1 f_3|\mathcal{F}_2) = f_2 E[E(f_1 f_3|\mathcal{F}_3)|\mathcal{F}_2] \\ &= f_2 E[f_3 E(f_1|\mathcal{F}_3)|\mathcal{F}_2] = E[f_2 f_3 E(f_1|\mathcal{F}_3)|\mathcal{F}_2] \quad [\mu]. \end{aligned}$$



Conseqüentemente,

$$E(f_1 f_2 f_3) = E[E(f_1 f_2 f_3 | \mathcal{F}_2)] = E[E(f_2 f_3 E(f_1 | \mathcal{F}_3))] \quad [\mu] .$$

Por definição de  $E(\cdot | \mathcal{F}_3)$ , tem-se

$$\begin{aligned} E(f_1 f_2 f_3) &= E[f_3 E(f_1 f_2 | \mathcal{F}_3)] \\ E[f_2 f_3 E(f_1 | \mathcal{F}_3)] &= E\{f_3 E[f_2 E(f_1 | \mathcal{F}_3) | \mathcal{F}_3]\} \\ &= E\{f_3 E(f_1 | \mathcal{F}_3) E(f_2 | \mathcal{F}_3)\} \end{aligned}$$

donde, pela arbitrariedade de  $f_3 \in \mathcal{L}(\mathcal{F}_3)$ ,

$$E(f_1 f_2 | \mathcal{F}_3) = E(f_1 | \mathcal{F}_3) E(f_2 | \mathcal{F}_3) \quad [\mu]$$

traduzindo  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$ , Q.E.D. □

### Exercício 10.3

Prove que  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$  pode ser caracterizado pela seguinte condição:

$$\forall f \in \mathcal{L}(\mathcal{F}_1 \vee \mathcal{F}_3) , \quad E[E(f | \mathcal{F}_3) | \mathcal{F}_2] = E(f | \mathcal{F}_2) \quad [\mu]$$

(Sugestão: use a relação  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3 \Leftrightarrow (\mathcal{F}_1 \vee \mathcal{F}_3) \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3$  e o Teorema 10.9.)

## 10.2 SUFICIÊNCIA BAYESIANA AMOSTRAL

A idéia intuitiva de suficiência bayesiana de uma estatística  $t \in \mathcal{A}$  está ligada à característica de a medida a posteriori depender dos dados apenas através de  $t$ , i.e., à relação  $\nu_x(B) = \nu_t(B)$ ,  $\forall B \in \mathcal{B}$ . É o que sucede no Exemplo 9.2, com a estatística  $(\sum x_i/n, \sum(x_i - \sum x_i/n)^2/(n-1))$  – reveja o Exercício 9.5, alínea  $f$ .

Esta noção pode ser aplicada a qualquer subálgebra  $\mathcal{D} \subset \mathcal{A}$  de acordo com a seguinte definição:

### Definição 10.3

A subálgebra  $\mathcal{D} \subset \mathcal{A}$  diz-se suficiente bayesiana (abreviadamente,  $B$ -suficiente) se no modelo  $(\Omega, \mathcal{F}, \mu)$  se tem

$$\forall B \in \mathcal{B}, \quad \mu[\overline{B} | \mathcal{F}(\overline{x})] = \mu[\overline{B} | \mathcal{D}_+] \quad [P]$$

onde  $\mathcal{D}_+ = \{\Theta \times D : D \in \mathcal{D}\}$ .

Em termos da notação definida em 9.3, esta relação traduz a identidade essencial das funções  $\nu_{\mathcal{A}}$  e  $\nu_{\mathcal{D}}$ , significando pois que para todo  $B \in \mathcal{B}$ , a v.a.  $\nu_{\mathcal{A}}(B)$  tem uma versão  $\mathcal{D}$ -mensurável. Isto implica que tal v.a. seja constante nos átomos de  $\mathcal{D}$  (elementos da partição induzida por  $\mathcal{D}$ ). Sob a natureza euclideana de  $(\mathcal{X}, \mathcal{A})$  que temos vindo a admitir, a Definição 10.3 para  $\mathcal{D}$  separável equivale mesmo a afirmar que  $\nu_{\mathcal{A}}(B)$  é constante nos átomos de  $\mathcal{D}$ ,  $\forall B \in \mathcal{B}$ .

A definição do conceito de independência condicional em  $(\Omega, \mathcal{F}, \mu)$  implica, considerando  $\mathcal{A}_+ \equiv \mathcal{F}(\bar{x})$  e  $\mathcal{B}_+ \equiv \mathcal{F}(\bar{\theta})$ , que:

**Teorema 10.10**

$\mathcal{D} \subset \mathcal{A}$  é  $B$ -suficiente se, e somente se, verifica uma das seguintes condições equivalentes:

- i)  $\forall B \in \mathcal{B}, \nu_{\mathcal{A}}(B) = \nu_{\mathcal{D}}(B) \quad [P]$ ;
- ii)  $\mathcal{A}_+ \perp\!\!\!\perp \mathcal{B}_+ | \mathcal{D}_+$ ;
- iii)  $\forall A \in \mathcal{A}, P_{\mathcal{B} \times \mathcal{D}}(A) = P_{\mathcal{D}}(A) \quad [P]$ .

**Prova**

Basta atender ao significado das funções de transição mencionadas no espaço  $(\Omega, \mathcal{F}, \mu)$  e à simetria do conceito de independência condicional. □

A inspeção da condição iii) revela a sua semelhança com a definição de suficiência clássica (abreviadamente,  $C$ -suficiência) com respeito ao modelo clássico, dada no Capítulo 3, nas condições de identificação de  $P_{\mathcal{B} \times \mathcal{D}}$  com uma probabilidade amostral condicional. Mais concretamente, tem-se

**Teorema 10.11**

Se a subálgebra separável  $\mathcal{D}$  é  $C$ -suficiente, então  $\mathcal{D}$  é  $B$ -suficiente para toda a medida de probabilidade a priori  $\nu$ .

**Prova**

Pela  $C$ -suficiência de  $\mathcal{D}$ , seja  $g_A$  uma versão comum de  $P_{\theta}(A|\mathcal{D})$  para todo  $\theta \in \Theta$ . A separabilidade de  $\mathcal{D}$  garante que existe uma versão de  $P_{\theta}(A|\mathcal{D})$  que é  $\mathcal{B} \times \mathcal{D}$ -mensurável

no espaço produto, coincidindo então com  $P_{\mathcal{B} \times \mathcal{D}}(A)$   $\mu$ -essencialmente. Assim,

$$\forall A \in \mathcal{A}, \quad \mu(\overline{A} | \mathcal{F}(\overline{\theta}) \vee \mathcal{D}_+) = \overline{g}_A \equiv g_A \circ \overline{x} \quad [\mu]$$

donde  $\forall D \in \mathcal{D}$

$$P(A \cap D) = \mu(\overline{A} \cap \overline{D}) = \int_{\overline{D}} \mu(\overline{A} | \mathcal{B} \times \mathcal{D}) d\mu = \int_{\overline{D}} \overline{g}_A d\mu = \int_D g_A dP .$$

Por conseguinte,  $g_A$  é uma versão de  $P(A|D) = P_{\mathcal{D}}(A)$  para todo  $A \in \mathcal{A}$  e toda a medida  $\nu$ . □

É este resultado que justifica que toda a estatística real  $C$ -suficiente seja igualmente  $B$ -suficiente para toda a probabilidade a priori. Contudo, nem toda a estatística real  $B$ -suficiente é necessariamente  $C$ -suficiente. Basta pensar no fato de que a  $B$ -suficiência de  $\mathcal{D}$  implica a  $B$ -suficiência para qualquer subálgebra que a contenha (recorde-se a monotonicidade do conceito de independência condicional) quando tal propriedade não se verifica para a  $C$ -suficiência, como se frisou no Capítulo 3 (relembre-se o exemplo de Burkholder, 1961).

A colocação de certas suposições sobre a probabilidade a priori  $\nu$  já permite obter a proposição recíproca do teorema anterior. Para o efeito, e usando a terminologia de Florens et al. (1990, op. cit.), consideremos

#### Definição 10.4

A medida de probabilidade a priori  $\nu$  diz-se regular para  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  se

$$\forall t \in \mathcal{L}(\mathcal{A}), E_{P_\theta}(t) = 0 \quad [\nu] \Rightarrow E_{P_\theta}(t) = 0, \quad \forall \theta \in \Theta . \quad \square$$

A particularização desta definição a funções indicadoras de conjuntos  $\mathcal{A}$ -mensuráveis mostra que o uso de uma medida  $\nu$  regular na construção da medida produto assegura a continuidade absoluta mútua entre  $\mathcal{P}$  e  $P$  e, em particular, a dominação do modelo clássico.

A condição de regularidade de  $\nu$  é verificada em inúmeras situações usuais sob um espaço paramétrico euclideano:

- a) Quando  $\Theta$  é numerável, a regularidade de  $\nu$  para  $\mathcal{P}$  é garantida se o conjunto vazio for o único conjunto  $\nu$ -nulo;
- b) Quando  $\Theta$  é não numerável e as probabilidades amostrais são tais que  $E_{P_\theta}(t) = \psi_t(\theta)$  é uma função contínua em  $\theta$ ,  $\forall t \in \mathcal{L}(\mathcal{A})$ ,  $\nu$  é regular para  $\mathcal{P}$  se atribuir probabilidade positiva a qualquer aberto de  $\mathcal{B}$ .

Note-se que a suposição de continuidade de  $\psi_t(\theta)$  no caso de dominação de  $\mathcal{P}$  pela medida de Lebesgue  $m$  é assegurada quando a função  $\mathcal{F}$ -mensurável  $dP_\theta(x)/dm(x)$  é contínua em  $\theta$  para cada  $x$  e dominada para todo  $\theta$  por uma função de  $x$  Lebesgue-integrável.

#### Exercício 10.4

Prova as afirmações feitas nas alíneas a) e b) acima referidas.

Posto isto, pode afirmar-se que:

#### Teorema 10.12

Se  $\mathcal{D} \subset \mathcal{A}$  é  $B$ -suficiente e  $\nu$  é regular para  $\mathcal{P}$ , então  $\mathcal{D}$  é  $C$ -suficiente.

#### Prova

Da relação (9.22) aplicada a  $\mathcal{D}$  (em vez de  $\mathcal{A}(t)$ ) tem-se, sob a hipótese, para todo  $A \in \mathcal{A}$

$$\forall A_1 \in \mathcal{D}, \quad E_{P_\theta}[I_{A \cap A_1} - I_{A_1} P_{\mathcal{D}}(A)] = 0 \quad [\nu]$$

o que implica, pela regularidade de  $\nu$  para  $\mathcal{P}$ , que

$$\forall A_1 \in \mathcal{D}, \quad P_\theta(A \cap A_1) = \int_{A_1} P_{\mathcal{D}}(A) P_\theta^{\mathcal{D}}(dx), \quad \forall \theta \in \Theta.$$

Isto revela que a probabilidade preditiva condicional a  $\mathcal{D}$  de  $A$  é uma versão comum  $\mathcal{D}$ -mensurável de  $P_\theta(A|\mathcal{D})$ , para todo  $\theta$ ,  $\forall A \in \mathcal{A}$ , ou seja, que  $\mathcal{D}$  é suficiente com respeito a  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ .  $\square$

A definição de  $B$ -suficiência através do conceito de independência condicional permite caracterizá-la num modelo bayesiano dominado através de propriedades de densidades adequadas, pelo uso do Teorema 10.7. Com efeito, tendo em conta que em termos de

$\Pi = \nu \times P$ ,  $\mathcal{B}_+$  e  $\mathcal{A}_+$  são independentes não condicionalmente, o que implica a sua independência condicional a  $\mathcal{D}_+$ , o Teorema 10.7 garante-nos que  $\mathcal{D}$  é  $B$ -suficiente quando  $\mu \ll \Pi$  se, e somente se

$$h_{\mathcal{D}_+}^{\bar{\theta}, \bar{x}} = h_{\mathcal{D}_+}^{\bar{\theta}} h_{\mathcal{D}_+}^{\bar{x}} \quad [\Pi] \Leftrightarrow h_{\mathcal{A}_+}^{\bar{\theta}} = h_{\mathcal{D}_+}^{\bar{\theta}} \quad [\Pi] .$$

Recordando o significado destas funções (veja-se Seção 9.4), tais relações correspondem a

$$E_{\Pi}[h|\mathcal{B} \times \mathcal{A}] \equiv h = E_{\Pi}[h|\mathcal{B} \times \mathcal{D}] = \int_{\mathcal{X}} h P_{\mathcal{D}}(dx) \quad [\Pi]$$

onde  $h = d\mu/d\Pi$ . Deste modo,  $\mathcal{D}$  é  $B$ -suficiente se, e somente se,  $h$  é  $\overline{\mathcal{B} \times \mathcal{D}}^{\Pi}$ -mensurável, onde  $\overline{\mathcal{B} \times \mathcal{D}}^{\Pi}$  representa o complemento de  $\mathcal{B} \times \mathcal{D}$  com os conjuntos  $\Pi$ -nulos. Quando  $\mathcal{P}$  é dominada pela medida de referência  $m$  em  $(\mathcal{X}, \mathcal{A})$ , obtém-se

$$\frac{d\mu(\omega)}{d(\nu \times m)(\omega)} = h(\theta, x)g(x) \quad [\nu \times m]$$

que pode ser vista como o análogo bayesiano do teorema de fatorização de Halmos-Savage para a caracterização da suficiência de uma subálgebra  $\mathcal{D}$ .

### Exercício 10.5

Do ponto de vista clássico,  $\mathcal{C} \subset \mathcal{A}$  diz-se ancilar (abreviadamente,  $C$ -ancilar) se  $\forall A \in \mathcal{D}$ ,  $P_{\theta}(A) = K_A$ ,  $\forall \theta \in \Theta$ . A definição bayesiana natural de ancilaridade é a seguinte:  $\mathcal{D}$  diz-se  $B$ -ancilar se  $\forall A \in \mathcal{D}$ ,  $P_B^{\mathcal{D}}(A) = P^{\mathcal{D}}(A)$   $[\mu]$ .

- a) Mostre que a definição bayesiana de ancilaridade equivale a qualquer das condições:
  - i)  $\mathcal{B}_+ \amalg \mathcal{D}_+$ ;
  - ii)  $\forall B \in \mathcal{B}$ ,  $\nu_{\mathcal{D}}(B) = \nu(B)$   $[\mu]$ .
- b) Prove que se  $\mathcal{D}$  é  $B$ -ancilar, então qualquer subálgebra de  $\mathcal{D}$  também é  $B$ -ancilar.
- c) Mostre que os conceitos clássico e bayesiano de ancilaridade são equivalentes se  $\nu$  é regular para  $\mathcal{P}$ .
- d) Num modelo bayesiano dominado, mostre que  $\mathcal{D}$  é  $B$ -ancilar se, e somente se,  $E_{\Pi}[h|\mathcal{B} \times \mathcal{D}] = 1$   $[\Pi]$ .

As definições de  $B$ -suficiência e  $B$ -ancilaridade em termos de independência condicional permitem obter algumas relações entre elas com base em propriedades referidas na Seção 10.1. Uma delas dispõe que:

**Teorema 10.13**

Sejam  $\mathcal{E}$  e  $\mathcal{D}$  subálgebras de  $\mathcal{A}$  preditivamente independentes. Se  $\mathcal{D}$  é  $B$ -suficiente, então  $\mathcal{E}$  é  $B$ -ancilar e independente de  $\mathcal{D}$ , para todo  $\theta \in \Theta$ .

**Prova**

O fato de  $\mathcal{E}_+ \amalg \mathcal{D}_+ \equiv \mathcal{E}_+ \amalg \mathcal{D}_+ | \mathcal{F}_0$  e de  $\mathcal{B}_+ \amalg \mathcal{E}_+ | \mathcal{D}_+$  pela  $B$ -suficiência de  $\mathcal{D}$  e  $\mathcal{E} \subset \mathcal{A}$  conduzem pelo Teorema 10.5 a  $\mathcal{B}_+ \amalg \mathcal{E}_+$  e  $\mathcal{E}_+ \amalg \mathcal{D}_+ | \mathcal{B}_+$ , Q.E.D.  $\square$

A introdução de condições adicionais sobre  $\mathcal{B}$  e  $\mathcal{D} \subset \mathcal{A}$  (e, eventualmente,  $\mathcal{E} \subset \mathcal{A}$ ) permite obter novas relações entre os conceitos de  $B$ -suficiência,  $B$ -ancilaridade e independência clássica (i.e., dado  $\mathcal{B}$ ), constituindo as versões bayesianas dos famosos teoremas de Basu.

Restringir-nos-emos por agora ao teorema que estabelece a condição em que  $\mathcal{E}_+ \amalg \mathcal{D}_+ | \mathcal{B}_+$  e  $\mathcal{E}_+ \amalg \mathcal{B}_+$  implicam  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$ .

**Teorema 10.14**

Sejam  $\mathcal{E}$  e  $\mathcal{D}$  subálgebras de  $\mathcal{A}$  classicamente (i.e., dado  $\mathcal{B}_+$ ) independentes e  $\mathcal{E}$   $B$ -ancilar. Se  $\mathcal{E}_+ \vee \mathcal{D}_+$  é  $B$ -suficiente, então  $\mathcal{D}$  é igualmente  $B$ -suficiente.

**Prova**

Por hipótese,  $\mathcal{E}_+ \amalg \mathcal{D}_+ | \mathcal{B}_+$  e  $\mathcal{E}_+ \amalg \mathcal{B}_+$  que, pelo Teorema 10.5 equivalem a  $\mathcal{E}_+ \amalg \mathcal{D}_+$  e  $\mathcal{E}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$ . Pelo mesmo teorema, esta última condição conjugada com a hipótese adicional  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{E}_+ \vee \mathcal{D}_+$  equivale a  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$  (note-se que  $\mathcal{E}_+ \vee \mathcal{A}_+ = \mathcal{A}_+$ ), Q.E.D.  $\square$

**Exercício 10.6**

Considere que  $(\mathcal{X}, \mathcal{A})$  é o espaço euclidiano  $n$ -dimensional no qual se considera a família produto de  $n$  distribuições  $N(\theta, \sigma^2)$ , com  $\sigma^2$  conhecido. Seja  $\nu$  uma medida de probabilidade a priori equivalente à medida de Lebesgue em  $\mathbb{R}$ . Mostre através do resultado anterior (versão bayesiana do 3º teorema de Basu) que a estatística  $t = \frac{1}{n} \sum_{i=1}^n x_i$  é

$B$ -suficiente.

### 10.3 COMPLETUDE E SUFICIÊNCIA MÍNIMA BAYESIANAS

Vimos na seção anterior que a condição de independência clássica entre duas subálgebras, das quais uma é  $B$ -ancilar, implica a  $B$ -suficiência da outra desde que a menor subálgebra que as contenha seja  $B$ -suficiente. Trataremos agora da versão bayesiana do 1<sup>o</sup> teorema de Basu, a qual estabelece a condição que assegura a independência clássica entre uma subálgebra  $B$ -suficiente e uma subálgebra  $B$ -ancilar.

Para o efeito, recordemos que a definição clássica de uma subálgebra completa (que rotularemos de  $C$ -completa) é a seguinte:

$\mathcal{D} \subset \mathcal{A}$  é  $C$ -completa se para toda função  $f \in \mathcal{D}$  e integrável

$$E_{P_\theta}(f) = 0, \quad \forall \theta \in \Theta \Rightarrow f = 0 \quad [P_\theta], \quad \forall \theta.$$

A restrição desta relação a funções  $f \in \mathcal{L}(\mathcal{D})$  define o conceito de  $C$ -completude limitada.

A versão bayesiana deste conceito radica numa relação probabilística entre  $\mathcal{D}_+$  e  $\mathcal{B}_+$  (vulgarmente denominada de identificação forte de  $\mathcal{D}_+$  por  $\mathcal{B}_+$ ), descrita por:

#### Definição 10.5

$\mathcal{D} \subset \mathcal{A}$  diz-se  $B$ -completa (resp.  $B$ -completa limitada) se para toda a função  $f \in \mathcal{D}$  e integrável (resp.  $f \in \mathcal{L}(\mathcal{D})$ ), se tem

$$E_\mu(\bar{f}|\mathcal{B}_+) = 0 \Rightarrow \bar{f} = 0 \quad [\mu], \quad \text{com } \bar{f} = f \circ \bar{x} \in \mathcal{D}_+.$$

Recordando que, por construção de um modelo bayesiano regular,  $E_\mu(\bar{f}|\mathcal{B}_+) = E_{P_\theta}(f) \quad [\nu]$ , a definição de  $B$ -completude pode ser, nesse quadro, traduzida pela relação

$$\forall f \in \mathcal{D} \text{ e integrável, } E_{P_\theta}(f) = 0 \quad [\nu] \Rightarrow f = 0 \quad [P].$$

#### Exercício 10.7

Mostre que

- a) Os conceitos de  $C$ -completude e de  $B$ -completude num modelo bayesiano regular são equivalentes se  $\nu$  é regular para  $\mathcal{P}$ .

b) Qualquer subálgebra  $B$ -ancilar contida numa subálgebra  $B$ -completa é trivial (no sentido de estar incluída em  $\overline{\mathcal{F}}_0$ ).

Posto isto, tem-se

**Teorema 10.15**

Se  $\mathcal{D} \subset \mathcal{A}$  é  $B$ -suficiente completa, então  $\mathcal{D}$  é independente classicamente (e preditivamente) de qualquer subálgebra  $B$ -ancilar.

**Prova**

Pela  $B$ -suficiência de  $\mathcal{D}$ ,  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$ , o que implica  $\mathcal{E}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$ , para qualquer  $\mathcal{E} \subset \mathcal{A}$ . Assim,  $\forall f \in \mathcal{L}(\mathcal{E}_+)$

$$E(f | \mathcal{B}_+ \vee \mathcal{D}_+) = E(f | \mathcal{D}_+) \quad [\mu]$$

e, por conseguinte, dado que  $\mathcal{B}_+ \subset \mathcal{B}_+ \vee \mathcal{D}_+$ ,

$$E(f | \mathcal{B}_+) = E[E(f | \mathcal{B}_+ \vee \mathcal{D}_+) | \mathcal{B}_+] = E[E(f | \mathcal{D}_+) | \mathcal{B}_+] \quad [\mu].$$

Sendo  $\mathcal{E}$   $B$ -ancilar,  $\mathcal{B}_+ \amalg \mathcal{E}_+$ , tem-se  $\forall f \in \mathcal{L}(\mathcal{E}_+)$

$$E(f | \mathcal{B}_+) = E(f) \quad [\mu]$$

pelo que

$$E[E(f | \mathcal{D}_+) - E(f) | \mathcal{B}_+] = 0 \quad [\mu].$$

Pela  $B$ -completude de  $\mathcal{D}$  tem-se então

$$E(f | \mathcal{D}_+) = E(f) \quad [\mu],$$

ou seja,  $\mathcal{E}_+ \amalg \mathcal{D}_+$ .

Como  $\mathcal{E}_+ \amalg \mathcal{D}_+$  e  $\mathcal{E}_+ \amalg \mathcal{B}_+ | \mathcal{D}_+$  equivalem a  $\mathcal{E}_+ \amalg (\mathcal{B}_+ \vee \mathcal{D}_+)$  que, por sua vez, implica  $\mathcal{E}_+ \amalg \mathcal{D}_+ | \mathcal{B}_+$  (vide Teorema 10.5), fica demonstrada a versão bayesiana do 1<sup>o</sup> teorema de Basu. □

**Exercício 10.8**

Seja  $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$  tal que  $x_i | \theta \underset{\text{iid}}{\sim} G_a(\alpha, \delta)$ ,  $\theta = (\alpha, \delta) \in \mathbb{R}_+^2$  e  $\nu$  definida nos borelianos de  $\mathbb{R}_+^2$  e equivalente à respectiva medida de Lebesgue. Denote por  $t_1$  e  $t_2$  a média aritmética e a média geométrica, respectivamente, dos  $\{x_i\}$ . Mostre que:



- a) a estatística  $(t_1, t_2)$  é  $B$ -suficiente completa;
- b) no modelo condicional a  $\alpha$  ( $\equiv \mathcal{F}(\bar{\alpha})$ ),  $t_1$  é  $B$ -suficiente completa e  $u = t_2/t_1$  é  $B$ -ancilar;
- c)  $\mathcal{F}(\bar{u}) \amalg (\mathcal{F}(\bar{t}_1) \vee \mathcal{F}(\bar{\delta})) | \mathcal{F}(\bar{\alpha})$  e  $\mathcal{F}(\bar{u}) \amalg \mathcal{F}(\bar{t}_1) | \mathcal{F}(\bar{\alpha}) \vee \mathcal{F}(\bar{\delta})$ .

Uma outra implicação do conceito de completude para a estimação clássica foi já referida no Capítulo 6. A este respeito, é interessante referir que a caracterização de independência condicional estabelecida no Teorema 10.9 tem a seguinte tradução estatística:

$\mathcal{D} \subset \mathcal{A}$  é  $B$ -suficiente se, e somente se, para toda a estatística  $f \in \mathcal{A}_+$  (integrável)

$$E[E(f|\mathcal{D}_+) | \mathcal{B}_+] = E(f|\mathcal{B}_+) \quad [\mu].$$

Com base nesta caracterização pode então derivar-se a formulação bayesiana dos resultados clássicos a que acima aludimos. É esse o propósito do seguinte exercício.

### Exercício 10.9

Seja  $\mathcal{U} = \{f \in \mathcal{A}_+ : E(f|\mathcal{B}_+) = \psi(\theta) \text{ e } E[(f - \psi(\theta))^2 | \mathcal{B}_+] < \infty\}$ .

- a) Mostre que  $\forall f \in \mathcal{U}$ ,  $E(f|\mathcal{D}_+) \in \mathcal{U}$  se  $\mathcal{D}$  é  $B$ -suficiente.
- b) Se  $\mathcal{D}$  é adicionalmente  $B$ -completa, prove que  $E(f|\mathcal{D}_+)$  é o  $\mu$ -essencialmente único elemento de  $\mathcal{U}$  com menor variância amostral.

Como já referimos anteriormente, a suficiência bayesiana goza da propriedade de monotonicidade, independentemente da dominação do modelo bayesiano. Para a determinação do elemento mínimo da classe de  $\sigma$ -álgebras  $B$ -suficientes devemos ter em conta que o completamento de qualquer subálgebra  $\mathcal{D}_+$  de  $\mathcal{A}_+$  não está incluído necessariamente em  $\mathcal{A}_+$  (recorde-se que  $\overline{\mathcal{D}_+} = \mathcal{D}_+ \vee \overline{\mathcal{F}_0}$ ) e que  $\overline{\mathcal{D}_+} \cap \mathcal{A}_+$  representa a menor subálgebra de  $\mathcal{A}_+$  contendo  $\mathcal{D}_+$  e os conjuntos nulos de  $\mathcal{A}_+$ . Daí que seja natural definir a  $B$ -suficiência mínima do seguinte modo:

### Definição 10.6

$\mathcal{D}^* \subset \mathcal{A}$  é  $B$ -suficiente mínima se  $\mathcal{B}_+ \amalg \mathcal{A}_+ | \mathcal{D}_+^*$  e  $\mathcal{D}_+^* \subset \overline{\mathcal{D}_+^*} \cap \mathcal{A}_+$ ,  $\forall \mathcal{D} \subset \mathcal{A}$  tal que  $\mathcal{B}_+ \amalg \mathcal{A}_+ | \mathcal{D}_+$ .

Um pouco de reflexão deixa perceber que com esta definição é possível existirem duas subálgebras de  $\mathcal{A}$   $B$ -suficientes mínimas com uma estritamente contida na outra.

Por outro lado, o Teorema 10.8 garante-nos que a projeção de  $\mathcal{B}_+$  sobre  $\mathcal{A}_+$ ,  $\mathcal{A}_+[\mathcal{B}_+] = \sigma(\{E(\phi|\mathcal{A}_+) : \phi \in \mathcal{L}(\mathcal{B}_+)\})$  é  $B$ -suficiente mínima. Além disso, tem a particularidade de ser  $\mathcal{A}_+$ -completada, i.e., de conter os conjuntos nulos de  $\mathcal{A}_+$  (recorde-se o Exercício 10.2) e de ser a única subálgebra  $B$ -suficiente mínima dentro da classe de subálgebras  $\mathcal{A}_+$ -completadas.

Assim, e tendo em conta que a  $B$ -suficiência de  $\mathcal{D} \subset \mathcal{A}$  equivale à  $B$ -suficiência do seu complemento, pode afirmar-se que:

**Teorema 10.16**

$\mathcal{D}^* \subset \mathcal{A}$  é  $B$ -suficiente mínima se, e somente se,  $\overline{\mathcal{D}_+^*} \cap \mathcal{A}_+ = \mathcal{A}_+[\mathcal{B}_+]$ .

**Prova**

Deixada para exercício.

No contexto clássico, a suficiência mínima(1) é definida em termos de inclusão  $\mathcal{P}$ -essencial na classe das subálgebras suficientes. No Capítulo 3 viu-se que o elemento mínimo desta classe não existe em geral – reveja-se o exemplo de Pitcher –, contrariamente ao que sucede com o elemento  $B$ -suficiente mínimo, como vimos.

No caso dominado, a subálgebra suficiente mínima pode ser definida em termos de uma probabilidade dominadora específica,  $\tilde{P}$ , que é uma combinação linear convexa de elementos de  $\mathcal{P}$ . Com efeito, segundo o Teorema 4.3,  $\mathcal{D}^* = \sigma(\{g(\theta, x) : \theta \in \Theta\})$ , onde  $g(\theta, x)$  é uma versão de  $dP_\theta(x)/d\tilde{P}(x)$ , é suficiente mínima no sentido em que  $\mathcal{D}^* \subset \mathcal{D} [P]$  (ou equivalentemente,  $\mathcal{D}^* \subset \mathcal{D} [\tilde{P}]$ ),  $\forall \mathcal{D}$   $C$ -suficiente.

Observe-se agora que a  $C$ -suficiência de  $\mathcal{D}^*$  implica sua  $B$ -suficiência e, sendo assim,  $\mathcal{A}_+[\mathcal{B}_+] \subset \overline{\mathcal{D}_+^*} \cap \mathcal{A}_+$ , onde  $\overline{\mathcal{D}_+^*} \cap \mathcal{A}_+$  designa o complemento de  $\mathcal{D}_+^*$  com os conjuntos de  $\mathcal{A}_+$   $P$ -nulos. Por outro lado, se  $\nu$  é regular para  $\mathcal{P}$ ,  $\mathcal{A}_+[\mathcal{B}_+]$  também é  $C$ -suficiente e a medida preditiva  $P$  é equivalente à medida  $\tilde{P}$ . Assim, pelo fato de  $\mathcal{D}^*$  ser  $C$ -suficiente mínima e de  $\mathcal{A}_+[\mathcal{B}_+]$  ser  $\mathcal{A}_+$ -completada, tem-se

$$\mathcal{D}_+^* \subset \mathcal{A}_+[\mathcal{B}_+] [P] \Leftrightarrow \mathcal{D}_+^* \subset \mathcal{A}_+[\mathcal{B}_+] .$$

Consequentemente,  $\mathcal{A}_+[\mathcal{B}_+] = \mathcal{D}_+^* [P]$ , i.e.,  $\mathcal{A}_+[\mathcal{B}_+]$  coincide com o complemento de  $\mathcal{D}_+^*$

com os conjuntos  $P$ -nulos (ou  $\tilde{P}$ -nulos) de  $\mathcal{A}_+$ , o que assegura a  $B$ -suficiência mínima de  $\mathcal{D}^*$ , pelo Teorema 10.16. Fica pois demonstrado que:

**Teorema 10.17**

Se  $\nu$  é regular para  $\mathcal{P}$ , a subálgebra  $C$ -suficiente mínima  $\mathcal{D}^* = \sigma(\{dP_\theta/d\tilde{P} : \theta \in \Theta\})$  é  $B$ -suficiente mínima. O seu completamento com os conjuntos  $\tilde{P}$ -nulos de  $\mathcal{A}$  define então a subálgebra  $B$ -suficiente mínima  $\mathcal{A}_+$ -completada.

**Exercício 10.10**

Num modelo bayesiano regular, seja  $\xi^*$  a menor subálgebra de  $\mathcal{A}$  que torna as probabilidades a posteriori mensuráveis, i.e.,

$$\xi^* = \sigma(\{\nu_x(B) : B \in \mathcal{B}\})$$

- a) Mostre que o completamento de  $\xi^*$  com os conjuntos  $P$ -nulos de  $\mathcal{A}$  define a subálgebra  $B$ -suficiente mínima  $\mathcal{A}_+$ -completada.
- b) Sob a dominação de  $\mu$  por  $\Pi$  com  $d\mu/d\Pi = h$ , mostre que o completamento de  $\xi^*$  referido é definido por  $\sigma(\{h(\theta, x) : \theta \in \Theta - N\})$ , onde  $\nu(N) = 0$ .

Uma vez estabelecidos os conceitos bayesianos de completude e suficiência mínima, estamos em condições de derivar a versão bayesiana do conhecido teorema de Lehmann-Scheffé.

**Teorema 10.18**

Sendo  $\mathcal{D} \subset \mathcal{A}$  uma subálgebra  $B$ -suficiente e  $B$ -completa limitada, então  $\mathcal{D}$  é  $B$ -suficiente mínima.

**Prova**

Denotando  $\mathcal{A}_+[\mathcal{B}_+]$  por  $\xi_+$  e sendo  $f \in \mathcal{L}(\mathcal{D}_+)$

$$E(f|\mathcal{B}_+) = E[E(f|\mathcal{B}_+ \vee \xi_+)|\mathcal{B}_+] \quad [\mu] = E[E(f|\xi_+)|\mathcal{B}_+] \quad [\mu].$$

A 1ª igualdade resulta da propriedade de “smoothing” da esperança condicional e a 2ª da relação  $\mathcal{D}_+ \perp\!\!\!\perp \mathcal{B}_+|\xi_+$ , que decorre da  $B$ -suficiência de  $\xi$  e do fato de  $\mathcal{D} \subset \mathcal{A}$ .

Assim, fazendo  $g = f - E(f|\xi_+)$ , que é  $\overline{\mathcal{D}}_+$ -mensurável por  $\xi_+ \subset \overline{\mathcal{D}}_+$  (Teorema 10.8), tem-se  $E(g|\mathcal{B}_+) = 0$   $[\mu]$ . Dado que  $\mathcal{D}$  é  $B$ -completa, conclui-se então que  $f$  é  $\mu$ -essencialmente igual a  $E(f|\xi_+)$  ou, por outras palavras, que  $\mathcal{D}_+ \subset \overline{\xi}_+$ , o que implica que

$$\overline{\mathcal{D}}_+ \cap \mathcal{A}_+ \subset \overline{\xi}_+ \cap \mathcal{A}_+ = \xi_+$$

pelo fato de  $\xi_+$  ser  $\mathcal{A}_+$ -completada.

Em suma,  $\overline{\mathcal{D}}_+ \cap \mathcal{A}_+ = \xi_+$ , o que prova a  $B$ -suficiência mínima de  $\mathcal{D}$  pelo Teorema 10.16.  $\square$

Este resultado indica-nos assim que a classe de subálgebras  $B$ -suficientes completas está contida na classe das subálgebras  $B$ -suficientes mínimas. O fato de as duas classes serem distintas pode ser comprovado pelas subálgebras  $B$ -suficientes mínimas que contêm uma subálgebra  $B$ -ancilar não trivial (recorde-se o Exercício 10.7 b), como se ilustra em seguida.

### Exemplo 10.1

Seja  $P_\theta$  uma medida produto das distribuições  $x_1|\theta \sim \exp(\theta)$  e  $x_2|\theta \sim \exp(1/\theta)$ ,  $\theta > 0$ , e  $\nu$  uma medida equivalente à medida de Lebesgue em  $\mathbb{R}^+$ . A estatística  $x = (x_1, x_2)$  é claramente  $B$ -suficiente mínima (Teorema 10.17) e equivalente à estatística  $(x_1, u)$ , onde  $u = x_1x_2$ .

Como a distribuição amostral de  $u$  é o produto de duas distribuições  $\text{Exp}(1)$ ,  $U$  é  $B$ -ancilar e, portanto,  $x$  não é  $B$ -completa.

### Exercício 10.11

Seja  $P_\theta$  definida pela distribuição Multinomial  $(N_1, N_2, N_3)|\theta \sim M_2(n, (\theta/2, (1-\theta)/2, 1/2))$ ,  $\theta \in (0, 1)$  e  $\nu$  uma medida atribuindo probabilidade positiva a qualquer aberto de  $(0, 1)$ . Mostre que  $(N_1, N_2)$  é  $B$ -suficiente mínima mas não é  $B$ -completa.

# CAPÍTULO 11

## IDENTIFICABILIDADE PARAMÉTRICA

### 11.1 CONCEITOS E RESULTADOS BÁSICOS NO CONTEXTO CLÁSSICO

Consideremos o modelo estatístico clássico  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  e definamos em  $\Theta$  a relação binária  $\sim$  tal que

$$\theta_1 \sim \theta_0 \Leftrightarrow P_{\theta_1}(A) = P_{\theta_0}(A), \quad \forall A \in \mathcal{A}.$$

#### Definição 11.1

Dois pontos de  $\Theta$ ,  $\theta_1$  e  $\theta_0$ , dizem-se observacionalmente equivalentes se  $\theta_1 \sim \theta_0$ .

A relação  $\sim$ , dita de equivalência observacional, é uma relação de equivalência em  $\Theta$  induzindo neste a partição  $\Theta / \sim$  (o chamado conjunto quociente de  $\Theta$  segundo  $\sim$ ) nas classes  $[\theta_0] = \{\theta_1 \in \Theta : P_{\theta_1} = P_{\theta_0}\}$ .

#### Definição 11.2

O modelo (ou  $\Theta$ ) diz-se identificável se  $\Theta / \sim = \{\{\theta\}, \theta \in \Theta\}$ .

Deste modo, a identificabilidade de  $\Theta$  é definida em termos da injetividade da aplicação  $\gamma_A(\theta) = P_\theta(A) : \Theta \rightarrow [0, 1]$  para pelo menos um  $A \in \mathcal{A}$ . Se  $\gamma_A(\theta), \forall A \in \mathcal{A}$ , não for injetiva,

o modelo diz-se inidentificável. No caso dominado,  $dP_\theta(x)/dm(x) = f(\theta, x)$  é constante em  $[\theta]$  para quase todo  $x$  e  $f(\theta_1, x) \neq f(\theta_0, x)$  para  $\theta_1 \notin [\theta_0]$  para pelo menos um  $x$ .

### Exercício 11.1

- a) Mostre que num modelo inidentificável não há estimadores consistentes para o parâmetro e a sua estimação é puramente arbitrária.
- b) Mostre que o modelo é necessariamente identificável se o respectivo parâmetro é interpretável como  $\theta = E_{P_\theta}(t)$  para alguma função  $t \in \mathcal{A}$  e integrável.

Como conseqüência da definição de um modelo inidentificável,  $\Theta / \sim$  é a partição mais fina de  $\Theta$  (i.e., com mais partes) que consegue ser identificada (diferenciada probabilisticamente) pelos dados. Obviamente que toda a partição  $\Pi$  mais grossa do que  $\Theta / \sim$  também é identificada pelos dados.

Introduzindo na classe de partições de  $\Theta$  a relação binária “ $\Pi_1$  é uma redução de  $\Pi_2$ ” (denotada por  $\Pi_1 < \Pi_2$ ), entendida no sentido de  $\Pi_1$  ser mais grossa do que  $\Pi_2$ , tem-se

### Definição 11.3

- i) A partição  $\Pi$  diz-se identificável se  $\Pi < \Theta / \sim$ .
- ii) A função  $\phi(\theta)$  diz-se identificável se a partição por ela induzida é identificável, i.e., se  $\phi(\theta)$  é constante em cada parte de  $\Theta / \sim$ .

Como conseqüência, uma partição não identificável não pode ser redução de uma partição identificável, o que não significa que seja necessariamente mais fina (note-se que  $<$  não goza de dicotomia). Uma partição que seja mais fina do que a partição identificável  $\Theta / \sim$  é “suficiente” para identificar medidas idênticas em  $\mathcal{P}$  já que os elementos de cada uma das suas partes são observacionalmente equivalentes. Assim,

### Definição 11.4

- i) A partição  $\Pi$  diz-se suficiente se  $\Theta / \sim < \Pi$ .
- ii) A função  $\phi(\theta)$  diz-se suficiente se a partição por ela induzida é suficiente, i.e., se assume valores diferentes em classes de equivalência distintas.

Como conseqüência, toda a partição (resp. função) identificável é redução (função) de qualquer partição (função) suficiente.

As funções suficientes podem ser caracterizadas através de uma fatorização das medidas  $P_\theta$ , permitindo compreender melhor a razão de ser do epíteto suficiente.

**Teorema 11.1**

A função  $\phi(\theta)$  é suficiente se, e somente se, existe uma função  $P_\phi^*(\cdot)$  definida em  $\phi(\Theta) \times \mathcal{A}$  tal que  $P_\theta(A) = P_{\phi(\theta)}^*(A)$ ,  $\forall A \in \mathcal{A}$ , para todo  $\theta$ .

**Prova**

(Parte suficiente): Seja  $\theta_1$  e  $\theta_0$  tal que  $\phi(\theta_1) = \phi(\theta_0)$ . Pela fatorização referida,  $P_{\theta_1} = P_{\theta_0}$ , o que traduz a suficiência de  $\phi$ .

(Parte necessária): Pela suficiência de  $\phi(\theta)$ ,  $P_{\theta_1} = P_{\theta_0}$  sempre que  $\phi(\theta_1) = \phi(\theta_0)$ , o que significa que  $P_\theta$  depende de  $\theta$  apenas através de  $\phi(\theta)$ . Definindo para cada  $A \in \mathcal{A}$ , a função  $P_\phi^*(A) = P_\theta(A)$  para  $\theta$  tal que  $\phi(\theta) = \phi$ , fica demonstrada a fatorização referida. □

Num contexto bayesiano, a suficiência de uma estatística  $t \in \mathcal{A}$  é essencialmente definida pela característica de a medida de probabilidade à posteriori,  $\nu_x$ , depender dos dados só através de  $t$ . O resultado do Teorema 11.1 evidencia assim que o conceito de suficiência paramétrica pode ser visto como o dual da suficiência amostral, num quadro bayesiano.

Como  $\Theta/\sim$  é a única partição simultaneamente suficiente e identificável, qualquer função que a induza é, por um lado, a “menor” função suficiente e, por outro, a “maior” função identificável.

Os qualificativos de “menor” e “maior” reportam-se aqui ao menor e maior número de partes da respectiva partição induzida na classe das partições respectivamente suficientes e identificáveis. Daí as designações frequentes de suficiente minimal e identificável maximal para uma função desse tipo. Esta função tem a particularidade de caracterizar tudo o que é ou não é identificável, já que das definições anteriores, uma função é identificável se, e somente se, é função de toda aquela que induza a partição  $\Theta/\sim$ . Esta é a razão do termo identificante também usado para designar tal tipo de função.

A partição  $\Theta/\sim$  é claramente induzida pela função de conjunto  $\theta \in \Theta \rightarrow [\theta] \in \Theta/\sim$ , ou

por qualquer transformação real bijetiva sua. Em modelos dominados regulares usuais, tal partição pode ser definida à custa de funções reais associadas com medidas de informação amostral, como mostra o seguinte exercício.

### Exercício 11.2

Seja  $\theta_0$  um valor determinado do parâmetro  $\theta \in \Theta$  indexante do modelo  $(\mathcal{X}, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$ . A função de informação de Kullback-Leibler para discriminar  $P_{\theta_0}$  contra  $P_\theta$  por observação de  $P_{\theta_0}$  é a função  $K(\theta_0, \theta) : \Theta \times \Theta \rightarrow [0, \infty]$  definida por

$$K(\theta_0, \theta) = E_{P_{\theta_0}} \left[ \ln \frac{P_{\theta_0}(dx)}{P_\theta(dx)} \right] .$$

a) Mostre que se  $P_{\theta_1}$  for equivalente (no sentido da continuidade absoluta mútua) a  $P_{\theta_0}$

$$\theta_1 \in [\theta_0] \Leftrightarrow K(\theta_0, \theta_1) = 0 .$$

b) Conclua de a) que se a família  $\{P_\theta\}$  for autodominada por  $P_{\theta_0}$ ,  $\forall \theta_0 \in \Theta$ , o modelo é identificável se, e somente se, para todo  $\theta_0$ ,  $K(\theta_0, \theta) > 0$ ,  $\forall \theta \neq \theta_0$ .

c) Para uma família satisfazendo as usuais condições de regularidade (sob as quais se derivam as conhecidas propriedades da medida de informação de Fisher,  $I(\theta)$ )  $K(\theta_0, \theta)$  é contínua em  $\theta$  para cada  $\theta_0$ . Mostre que nestas condições o modelo é identificável localmente se  $I(\theta_0)$  é não singular,  $\forall \theta_0$ .

Para ilustração dos conceitos e resultados descritos, consideremos o seguinte exemplo:

### Exemplo 11.1

Consideremos o modelo ANOVA II balanceado sem interação descrito por

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \varepsilon_{ijk} , \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, c$$

$$\{\varepsilon_{ijk}\} \text{ i.i. d. (dado } \sigma^2) \sim N(0, \sigma^2)$$

e tomemos  $\beta = (\mu \ \alpha_1 \ \dots \ \alpha_a \ \gamma_1 \ \dots \ \gamma_b)'$  e  $\theta = (\beta' \ \sigma^2)'$ .

Observe-se que o vetor  $\mu$  de todas as médias distintas,  $\mu_{ij} = \mu + \alpha_i + \gamma_j$ , das observações é definido condensadamente por  $\mu = X\beta$ , onde a matriz  $X$  de 0's e 1's, relacionando as componentes de  $\beta$  e  $\mu$ , tem dimensão  $ab \times (a + b + 1)$  e posto incompleto  $r = a +$



$b - 1$ . Conseqüentemente, há uma infinidade de valores de  $\beta$  compatíveis com a mesma distribuição Normal multivariada para o vetor das observações, o que evidencia a falta de identificabilidade do modelo. A partição identificante  $\Theta / \sim$  pode ser então definida por

$$[\theta_0] = \{(\beta_1' \sigma_1^2) : X\beta_1 = X\beta_0, \sigma_1^2 = \sigma_0^2\}, \quad \theta_0 = (\beta_0' \theta_0^2).$$

A classe das funções identificáveis inclui  $\sigma^2$ ,  $X\beta$ ,  $X'X\beta$ ,  $(X\beta \sigma^2)$ ,  $\alpha_i - \alpha_{i'}$  ( $i \neq i'$ ),  $\gamma_j - \gamma_{j'}$  ( $j \neq j'$ ) e  $(\mu + \alpha_1 + \gamma_1)^2$ , enquanto  $\theta$ ,  $\sum_i \alpha_i$  e  $\sum \gamma_j$  são exemplos de funções não identificáveis. As funções  $\theta$  e  $(X\beta \sigma^2)$  exemplificam funções suficientes enquanto  $(\mu + \alpha_1 + \gamma_1)^2$  e  $(\sum \alpha_i \sigma^2)$  já não satisfazem a definição de suficiência. As funções identificantes são exemplificadas por  $(X\beta \sigma^2)$  ou por  $(X'X\beta \sigma^2)$ .

Observe-se que os exemplos dados de funções lineares de  $\beta$  identificáveis correspondem ao que na literatura de Modelos Lineares se chama de funções linearmente estimáveis (i.e., que admitem um estimador linear não viesado). Esta correspondência seria previsível se tivéssemos em conta que os conceitos de identificabilidade e de estimabilidade linear coincidem no quadro de funções lineares. De fato, a caracterização de funções identificáveis como transformações de qualquer função identificante (em particular,  $(X\beta \sigma^2)$ ) permite concluir que toda a função linear de  $\beta$  é identificável se, e somente se, é linearmente estimável.

Como conseqüência da estimabilidade (existência de um estimador não viesado, não necessariamente linear) de qualquer função implicar identificabilidade – reveja-se o argumento do Exercício 11.1b) –, concluiu-se ainda que o conceito de identificabilidade é mesmo equivalente ao de estimabilidade, quando aplicados a funções lineares.

Este resultado, ao indicar que não pode haver nenhuma função linear que seja estimável sem ser linearmente estimável, permite compreender o porquê da restrição do conceito de estimabilidade de funções lineares ao de estimabilidade linear na análise clássica de modelos lineares. A consideração de funções não lineares (como  $(\mu + \alpha_1 + \gamma_1)^2$ , neste exemplo) ou de modelos identificáveis onde a dimensionalidade de  $\Theta$  é superior à de  $\mathcal{X}$  comprova que o conceito de identificabilidade é mais geral que o de estimabilidade, encarado quer no sentido restrito dado, quer no sentido mais lato que o termo encerra.

### Exercício 11.3

- a) Mostre que, num modelo geral, a estimabilidade da função  $\phi(\theta)$  implica a sua identificabilidade.
- b) Mostre que, no modelo linear geral  $y|\beta, \sigma^2 \sim N_n(X\beta, \sigma^2 I_n)$  onde  $\beta \in \mathbb{R}^p$  e  $X$  tem posto  $r \leq p$ , a função linear  $A\beta$  é identificável se, e somente se, é linearmente estimável.

(Sugestão: Use a condição necessária e suficiente usual de estimabilidade linear.)

- c) Conclua de a) e b) que  $A\beta$  é identificável se, e somente se, é estimável.

A arbitrariedade das inferências sobre parâmetros inidentificáveis justifica a concentração do trabalho inferencial em funções identificáveis. Para o efeito, a estatística clássica segue frequentemente uma via de prévia identificação do modelo baseada na imposição de restrições exatas e consistentes com as classes de equivalência observacional que tornem estas singulares.

Pela própria definição de funções identificáveis, fica claro que tal objetivo só poderá ser atingido por restrições consistentes definidas por funções inidentificáveis e em número determinado pela dimensionalidade dos elementos de  $\Theta / \sim$ . Voltando ao exemplo precedente, isso é assegurado, nomeadamente, pelas duas restrições  $\sum_i \alpha_i = \sum_j \gamma_j = 0$ , tão nossas conhecidas da literatura, e que têm a particularidade de conferir uma interpretação intuitiva aos parâmetros ANOVA do modelo identificado. No contexto do modelo linear geral, esta via tem ainda a particularidade de manter imutáveis as inferências sobre as funções identificáveis no modelo original. Deste modo, ela pode ser vista como um meio apenas de desencadear as inferências possíveis através da obtenção de estimativas dos parâmetros inidentificáveis originais.

Contudo, em geral, esta via de desbloqueamento inferencial tem sérias implicações devido à forte arbitrariedade das restrições impostas, ainda que estas pretendam ser uma tentativa de materialização de algum tipo de informação a priori. Com efeito, a natureza geralmente pouco precisa deste tipo de informação não se coaduna com tal forma de concretização, afigurando-se mais justificável uma especificação estocástica dessa (ou de parte dessa) informação, mas que o edifício estatístico clássico não permite.

Deste modo, parece-nos que a abordagem bayesiana desfruta de uma maior maleabilidade e naturalidade na viabilização inferencial sob um modelo inidentificável. Além disso, como a distribuição a posteriori será própria sempre que a distribuição a priori o for, as implicações da falta de identificabilidade são menos graves, embora não possam ser subestimadas (pelo menos, em certos casos) devido aos problemas decorrentes da especificação de uma distribuição a priori apropriada. Para mais detalhes sobre esta questão, veja-se Paulino (1992); As estratégias clássica e bayesiana na análise de modelos não identificáveis; Actas da 3<sup>a</sup> Conferência sobre Aplicações da Matemática à Economia e Gestão, CEMAPRE, Lisboa.

## 11.2 SUFICIÊNCIA BAYESIANA PARAMÉTRICA

Num contexto bayesiano têm aparecido na literatura várias concepções, nomeadamente, do termo identificabilidade/identificação que não são equivalentes em geral. Adotaremos aqui o ponto de vista de que os conceitos paramétricos de identificabilidade e suficiência são uma propriedade “essencial” do modelo  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , intervindo o espaço de probabilidade a priori  $(\Theta, \mathcal{B}, \nu)$  na especificação do significado preciso do termo “essencial”.

Mais concretamente, a visão que tomaremos como bayesiana desses conceitos é traduzida pelas definições clássicas mas encaradas quase em toda parte relativamente ao espaço  $(\Theta, \mathcal{B}, \nu)$ . Esta concepção permite encarar tais conceitos paramétricos como duais de correspondentes conceitos amostrais, como evidenciaremos ao longo deste capítulo.

Naturalmente que a constituição do modelo bayesiano justifica que a atenção se dirija a funções  $\mathcal{B}$ -mensuráveis. Observe-se que a uma função paramétrica  $\phi$  arbitrária se pode fazer corresponder a transformação mensurável  $(\Theta, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_1)$ , onde  $\Phi = \phi(\Theta)$  e  $\mathcal{B}_1 = \{E \subset \Phi : \phi^{-1}(E) \in \mathcal{B}\}$ . A respectiva subálgebra induzida é assim a classe das imagens inversas dos subconjuntos de  $\Phi$  que são elementos de  $\mathcal{B}$ . Frise-se, contudo, que nem toda a subálgebra de  $\mathcal{B}$  é induzida por uma função neste sentido, como é exemplificado pela classe da  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  formada pelos subconjuntos que são enumeráveis ou cujos complementos são enumeráveis. Outra coisa não seria de esperar dada a correspondência entre partições e funções e o que foi referido no Capítulo 2.

#### Exercício 11.4

Prove que se  $\mathcal{C}$  é uma subálgebra própria de  $\mathcal{B}$  contendo todos os conjuntos singulares  $\{\theta\}$ ,  $\mathcal{C}$  não pode ser induzida por uma função (Lema de Blackwell).

Nesta linha de raciocínio e tendo em vista a definição de suficiência bayesiana paramétrica, o Teorema 11.1 deve ser reformulado de modo a tomar em consideração a natureza de  $P_\theta(A)$  como função de transição em  $\Theta \times \mathcal{A}$ . A visualização da função  $\phi(\theta)$  como a transformação  $(\Theta, \mathcal{B}) \rightarrow (\Phi, \mathcal{B}_1)$  permite estender a função de transição  $P_\theta^*(\cdot)$  em  $\phi(\Theta) \times \mathcal{A}$  a  $\Phi \times \mathcal{A}$ , desde que  $\phi(\Theta) \in \mathcal{B}_1$ , fixando-a igual a uma medida de probabilidade arbitrária em  $(\mathcal{X}, \mathcal{A})$  para todo  $\phi \in \Phi - \phi(\Theta)$ , que é um conjunto de probabilidade (a priori) nula.

Deste modo, diremos que a função  $\mathcal{B}$ -mensurável  $\phi$  é  $B$ -suficiente se  $\forall A \in \mathcal{A}$ ,  $P_\theta(A) = P_{\phi(\theta)}^*(A) [\nu]$ . Em consonância com o conteúdo do capítulo anterior, estenderemos esta definição a subálgebras paramétricas. A sua especialização a funções  $\mathcal{B}$ -mensuráveis, partições e funções arbitrárias é definida através das respectivas subálgebras induzidas. Assim, e recordando a notação do Capítulo 9, teremos:

#### Definição 11.4

$\mathcal{C} \subset \mathcal{B}$  diz-se  $B$ -suficiente se

$$\forall A \in \mathcal{A}, \quad P_B(A) = P_C(A) [\nu].$$

Observe-se que no espaço produto esta definição corresponde a afirmar que  $\forall A \in \mathcal{A}$ ,  $\mu(\bar{A}|\mathcal{F}(\bar{\theta})) = \mu(\bar{A}|\mathcal{C}_+) [\nu]$  e, por conseguinte, é imediato que (recorde-se que  $\mathcal{B}_+ = \mathcal{F}(\bar{\theta})$ ,  $\mathcal{A}_+ = \mathcal{F}(\bar{x})$  e  $\mathcal{C}_+ = (C \times \mathcal{X} : C \in \mathcal{C})$ ):

#### Teorema 11.2

$\mathcal{C} \subset \mathcal{B}$  é  $B$ -suficiente se, e somente se, se verifica uma das seguintes condições equivalentes:

- i)  $\forall A \in \mathcal{A}$ ,  $P_B(A) = P_C(A) [\nu]$ ;
- ii)  $\mathcal{B}_+ \amalg \mathcal{A}_+ | \mathcal{C}_+$ ;
- iii)  $\forall B \in \mathcal{B}$ ,  $\nu_{\mathcal{A} \times \mathcal{C}}(B) = \nu_C(B) [\nu]$ .

## Prova

Exercício. □

A comparação deste teorema com o Teorema 10.10 patenteia o estreito paralelismo dual entre a  $B$ -suficiência paramétrica e a  $B$ -suficiência amostral, e deixa antever a larga gama de resultados que se poderão derivar por dualidade com aqueles demonstrados na Seção 10.2.

Antes porém, convém determo-nos um pouco sobre o significado deste conceito menos familiar. A condição i) do Teorema 11.2, ao indicar que cada elemento da família de funções  $\{\gamma_A(\theta) = P_\theta(A) : \forall A \in \mathcal{A}\}$  tem uma versão  $\mathcal{C}$ -mensurável, implica que cada  $\gamma_A$  é essencialmente constante nas partes da partição  $\Pi(\mathcal{C})$  (a equivalência entre estas duas asserções é garantida para  $\mathcal{C}$  separável sob a natureza euclideana de  $(\Theta, \mathcal{B})$ ). Por outras palavras, essencialmente todos os elementos de cada átomo de  $\mathcal{C}$  são observacionalmente equivalentes pelo que o processo amostral é suficientemente descrito por  $\mathcal{C}$ .

Em termos inferenciais, torna-se então impossível que o processo observacional opere qualquer refinamento dos átomos de  $\mathcal{C}$  na direção dos átomos mais finos de  $\mathcal{B}$ . Esta asserção é inequivocamente evidenciada pela condição iii) ao traduzir que a amostra observada não contribui em essência para a atualização do conhecimento a priori sobre  $\theta$  dentro de cada átomo de  $\mathcal{C}$ .

Em virtude da dualidade entre os conceitos bayesianos de suficiência paramétrica e suficiência amostral, é de esperar que o conceito de  $B$ -suficiência paramétrica num modelo regular possa ser associado ao dual do conceito clássico de suficiência amostral, traduzido por:

### Definição 11.5

$\mathcal{C} \subset \mathcal{B}$  diz-se suficiente com respeito ao modelo  $(\Theta, \mathcal{B}, Q)$ , onde  $Q = \{\nu_x : x \in \mathcal{X}\}$  – e escreve-se  $\mathcal{C} \text{ suf } (\mathcal{B}, Q)$  –, se para todo  $B \in \mathcal{B}$  existe uma função real  $\psi_B$   $\mathcal{C}$ -mensurável tal que

$$\nu_x(B \cap C) = \int_C \psi_B \nu_x(d\theta) , \quad \forall C \in \mathcal{C}, \forall x \in \mathcal{X} .$$

Para a ligação entre estes dois conceitos interessa considerar o dual da Definição 10.4:

**Definição 11.6**

A medida preditiva  $P$  diz-se regular para  $Q$  se

$$\forall \phi \in \mathcal{L}(\mathcal{B}), \quad E_{\nu_x}(\phi) = 0 \ [P] \Rightarrow E_{\nu_x}(\phi) = 0, \quad \forall x \in \mathcal{X}.$$

Observe-se que a regularidade de  $P$  para  $Q$  é verificada, nomeadamente, no caso dos modelos amostrais discretos em que o conjunto vazio é o único conjunto  $P$ -nulo e no caso em que a média a posteriori de qualquer função  $\mathcal{L}(\mathcal{B})$  é contínua em  $\mathcal{X}$  com  $P$  atribuindo probabilidade positiva a qualquer aberto de  $\mathcal{A}$ .

Note-se ainda que esta condição implica que  $Q$  seja equivalente a  $\nu$  no sentido da continuidade absoluta mútua. Deste modo:

**Teorema 11.3**

Se  $\mathcal{C} \subset \mathcal{B}$  é separável e  $P$  é regular para  $Q = \{\nu_x : x \in \mathcal{X}\}$ ,  $\mathcal{C}$  é  $B$ -suficiente se, e somente se, para todo  $B \in \mathcal{B}$  existe uma função  $\psi_B$   $\mathcal{C}$ -mensurável tal que  $\psi_B = \nu_x(B|\mathcal{C}) \ [\nu_x], \forall x \in \mathcal{X}$ .

**Prova**

Exercício. (Sugestão: aplique-se o argumento da demonstração dos Teoremas 10.11 e 10.12.) □

Nestas condições, a repetição do procedimento de Halmos-Savage-Bahadur para o modelo  $(\Theta, \mathcal{B}, Q)$  quando dominado por uma medida  $\sigma$ -finita  $\lambda$ , conduz-nos ao dual do teorema de fatorização clássico:

**Teorema 11.4**

Seja  $Q \ll \lambda$  e  $q(\theta; x)$  uma versão de  $d\nu_x/d\lambda$ . Então, a subálgebra separável  $\mathcal{C} \subset \mathcal{B}$  é  $B$ -suficiente se, e somente se, existe uma função não negativa  $\mathcal{B}$ -mensurável,  $q^*$ , e uma família de funções não negativas  $\mathcal{C}$ -mensuráveis  $\{g_x : x \in \mathcal{X}\}$ , tal que para  $P$ -essencialmente todo o  $x$

$$q(\theta; x) = g_x(\theta)q^*(\theta) \ [\lambda].$$

**Nota**

Sob a regularidade de  $P$  para  $Q$ ,  $\nu$  é um candidato para  $\lambda$ , e a fatorização de  $d\nu_x/d\lambda$  (em particular, de  $h(\theta; x)$ ) verifica-se para todo  $x \in \mathcal{X}$ .

### Prova

Exercício. (Sugestão: recorde o Capítulo 4 e tenha em conta o Teorema 11.3.)  $\square$

A dualidade com correspondentes conceitos bayesianos amostrais dados em 10.2 é expressa, nomeadamente, nas questões postas no seguinte exercício:

### Exercício 11.5

Por dualidade com conceito amostral de  $B$ -ancilaridade, Florens et al. (1990, op. cit.) definem a  $B$ -ancilaridade de  $\mathcal{C} \subset \mathcal{B}$  através da relação

$$\forall B \in \mathcal{C}, \quad \nu_{\mathcal{A}}^{\mathcal{C}}(B) = \nu^{\mathcal{C}}(B) \quad [\mu].$$

- a) Mostre que  $\mathcal{C}$  é  $B$ -ancilar se, e somente se, verifica qualquer das condições:
  - i)  $\mathcal{C}_+ \amalg \mathcal{A}_+$ ;
  - ii)  $\forall A \in \mathcal{A}, P_{\mathcal{C}}(A) = P(A) \quad [\mu]$ .
- b) No quadro de um modelo bayesiano dominado (por  $\Pi = \nu \times P$ , com  $d\mu/d\Pi = h$ ), prove que:
  - i)  $\mathcal{C}$  é  $B$ -suficiente se, e somente se,  $h = E_{\Pi}[h|\mathcal{C} \times \mathcal{A}] \quad [\Pi]$  se, e somente se  $h \in \overline{\mathcal{C} \times \mathcal{A}^{\Pi}}$ ;
  - ii)  $\mathcal{C}$  é  $B$ -ancilar se, e somente se  $E_{\Pi}[h|\mathcal{C} \times \mathcal{A}] = 1 \quad [\Pi]$ .
- c) Sejam  $\mathcal{C}_1$  e  $\mathcal{C}_2$  subálgebras de  $\mathcal{B}$  independentes a priori. Se  $\mathcal{C}_2$  é  $B$ -suficiente, então  $\mathcal{C}_1$  é  $B$ -ancilar e independente a posteriori de  $\mathcal{C}_2$ .
- d) Se  $\mathcal{D} \subset \mathcal{A}$  é independente de  $\mathcal{C} \subset \mathcal{B}$  e  $\mathcal{C}$  é  $B$ -suficiente, então  $\mathcal{D}$  é  $B$ -ancilar.
- e) Se  $\mathcal{C} \subset \mathcal{B}$  é independente de  $\mathcal{D} \subset \mathcal{A}$  e  $\mathcal{D}$  é  $B$ -suficiente, então  $\mathcal{C}$  é  $B$ -ancilar.

### Exemplo 11.2

Tomemos inicialmente o produto de duas distribuições de Poisson independentes de parâmetros  $\mu_1$  e  $\mu_2$ , positivos, e formemos o modelo amostral condicional, no valor  $N$  da

soma dessas duas variáveis, i.e.,  $\mathcal{X} = \{0, 1, \dots, N\}$ ,  $\mathcal{A}$  é o conjunto das partes de  $\mathcal{X}$  e  $\mathcal{P} = \{P_\theta : \theta = (\mu_1, \mu_2) \in \mathbb{R}_+^2\}$  é definida relativamente à medida de contagem  $m$  em  $\mathcal{X}$  pela densidade

$$f(\theta, x) = \binom{N}{x} \left( \frac{\mu_1}{\mu_1 + \mu_2} \right)^x \left( 1 - \frac{\mu_1}{\mu_1 + \mu_2} \right)^{n-x}.$$

Consideremos que a medida de probabilidade a priori é definida relativamente à medida de Lebesgue em  $\mathbb{R}_+^2$  pelo produto das densidades  $G_a(b_i, 1)$ ,  $i = 1, 2$ . Observe-se que a reparametrização para  $(\psi, \phi)$ , onde  $\psi = \mu_1/(\mu_1 + \mu_2)$  e  $\phi = \mu_1 + \mu_2$  conduz a que a medida de probabilidade a priori seja equivalentemente traduzida pelo produto das densidades  $B_e(b_1, b_2)$  e  $G_a(b, 1)$  com  $b = b_1 + b_2$ , com respeito à medida de Lebesgue em  $[0, 1] \times \mathbb{R}_+$ .

A medida preditiva é também dominada por  $m$  com densidade beta-binomial definida por

$$g(x) = \binom{n}{x} \frac{B(a_1, a_2)}{B(b_1, b_2)}$$

onde  $a_1 = b_1 + x$ ,  $a_2 = b_2 + n - x$ , sendo pois regular para  $Q$ .

O fato de  $P_\theta(A)$ ,  $\forall A \subset \mathcal{X}$  só depender de  $\theta$  através de  $\psi$  ilustra a suficiência (quer clássica, quer bayesiana) de  $\mathcal{B}(\psi)$ , i.e., de  $\psi$ . Por outro lado, a independência a priori entre  $\psi$  e  $\phi$  conduz a que a medida amostral condicional em  $\phi$  seja traduzida pela densidade preditiva  $g(x)$ , o que traduz a  $B$ -ancilaridade de  $\mathcal{B}(\phi)$  (ou de  $\phi$ ) – condição ii) de a) do Exercício 11.5.

Estas conclusões obtém-se igualmente da análise da medida de probabilidade a posteriori que, em termos de  $(\psi, \phi)$ , é traduzida pelo produto das densidades  $\mathcal{B}(a_1, a_2)$  e  $G_a(b, 1)$  relativamente à medida de Lebesgue em  $[0, 1] \times \mathbb{R}_+$ . A  $B$ -ancilaridade de  $\phi$  segue-se da identidade das suas medidas de probabilidade a priori e a posteriori.

As distribuições a priori e a posteriori de  $\theta$  condicionais a  $\psi$  são degeneradas ao estarem concentradas na parte da reta  $\mu_1/\mu_2 = \psi/(1 - \psi)$  contida em  $\Theta$ . Atendendo a que  $\phi$  e  $\mu_1$ , dado  $\psi$ , estão bijetivamente relacionados, as medidas a posteriori e a priori condicionais em  $\psi$  coincidem, refletindo a  $B$ -suficiência de  $\psi$ , e são caracterizadas em  $\mathbb{R}_+$  pela distribuição  $G_a(b, 1/\psi)$  para  $\mu_1$  dado  $\psi$  e por  $\mu_2 = \mu_1(1 - \psi)/\psi$  com probabilidade 1. Note-se que a  $B$ -suficiência de  $\psi$  e a sua independência a priori de  $\phi$  justifica não só a sua independência a posteriori mas também a  $B$ -ancilaridade de  $\phi$  (vide c) do Exercício 11.5). Observe-se ainda a ilustração do resultado e) do mesmo exercício, dada a independência entre  $x$ , que é  $B$ -suficiente, e  $\phi$ .



O fato de o modelo bayesiano ser dominado, com

$$h(\theta, x) = f(\theta, x)/g(x) = d\mu(\theta, x)/d\Pi(\theta, x) ,$$

pode igualmente ser aproveitado para a inspeção da  $B$ -ancilaridade e/ou  $B$ -suficiência. Por exemplo,

$$E_{\Pi}[h|\mathcal{B}(\phi) \times \mathcal{A}] \equiv \int_{\Theta} h(\theta, x)\nu_{\phi}(d\theta) = \int_0^1 h(\psi, x)q(\psi) d\psi = 1$$

onde  $q(\psi)$  é a densidade  $B_l(b_1, b_2)$  relativamente à medida de Lebesgue em  $[0, 1]$ .

Consideremos agora  $\gamma = \mu_1/\mu_2$ . Relativamente às medidas de Lebesgue apropriadas, a densidade a priori de  $\gamma$  é a de uma distribuição Beta de 2<sup>a</sup> espécie de parâmetros  $b_1$  e  $b_2$ , enquanto a densidade a priori de  $\mu_2$  condicional a  $\gamma$  é a de uma  $G_a(b, \gamma + 1)$ . Como  $h(\theta, x)$  só depende de  $\theta$  através de  $\gamma$ , segue-se que

$$E_{\Pi}[h|\mathcal{B}(\gamma) \times \mathcal{A}] = \int_0^{\infty} h(\theta, x)q(\mu_2|\gamma) d\mu_2 = h(\theta, x)$$

revelando a  $B$ -suficiência de  $\gamma$ . Observe-se ainda que a densidade a posteriori de  $\theta$  pode ser posta na forma

$$q(\theta; x) = \frac{(\mu_1/\mu_2)^{a_1-1}}{B(a_1, a_2)(1 + \mu_1/\mu_2)^{b+n}} \times \frac{\left(1 + \frac{\mu_1}{\mu_2}\right)^b \mu_2^{b-2} e^{-(1+\mu_1/\mu_2)\mu_2}}{\Gamma(b)}$$

que, de acordo com o Teorema 11.4, evidencia a  $B$ -suficiência de  $\gamma = \mu_1/\mu_2$ . Como não é possível obter uma fatorização desse tipo para  $\mu_2$  ou para  $\phi = \mu_1 + \mu_2$ , estas duas funções não são  $B$ -suficientes.  $\square$

Quando o modelo bayesiano é dominado, é possível pesquisar simultaneamente parâmetros e estatísticas  $B$ -suficientes através de uma propriedade de mensurabilidade da densidade  $h = d\mu/d\Pi$ , como estabelece o teorema seguinte.

### Teorema 11.5

Num modelo bayesiano dominado com densidade  $h = d\mu/d\Pi$ ,  $\mathcal{C} \subset \mathcal{B}$  e  $\mathcal{D} \subset \mathcal{A}$  são  $B$ -suficientes se, e somente se,  $h \in \overline{\mathcal{C} \times \mathcal{D}}^{\Pi}$ .

### Prova

(Parte suficiente): Se  $h$  é  $\Pi$ -essencialmente  $\mathcal{C} \times \mathcal{D}$ -mensurável, então é simultaneamente  $\mathcal{C} \times \mathcal{A}$ -mensurável e  $\mathcal{B} \times \mathcal{D}$ -mensurável a menos de um conjunto  $\Pi$ -nulo. Ou seja,  $h \in \overline{\mathcal{C} \times \mathcal{A}}^\Pi$  e  $h \in \overline{\mathcal{B} \times \mathcal{D}}^\Pi$ , o que caracteriza a  $B$ -suficiência de  $\mathcal{C}$  e de  $\mathcal{D}$ .

(Parte necessária): A  $B$ -suficiência de  $\mathcal{C}$  e de  $\mathcal{D}$  significam que  $E_\Pi(h|\mathcal{C} \times \mathcal{A}) = h = E_\Pi(h|\mathcal{B} \times \mathcal{D})$  [II]. Por outro lado, a independência entre  $\mathcal{B}_+$  e  $\mathcal{A}_+$  em termos de  $\Pi$  implica pelo Corolário 10.6 ii) e i) que  $\mathcal{B}_+ \perp\!\!\!\perp \mathcal{A}_+|\mathcal{C} \times \mathcal{D}$ ;  $\Pi$  e  $\mathcal{B} \times \mathcal{D} \perp\!\!\!\perp \mathcal{C} \times \mathcal{A}|\mathcal{C} \times \mathcal{D}$ ;  $\Pi$ , respectivamente.

Assim,

$$\begin{aligned} E_\Pi(h|\mathcal{C} \times \mathcal{D}) &= E_\Pi[E_\Pi(h|\mathcal{B} \times \mathcal{D})|\mathcal{C} \times \mathcal{D}] = E_\Pi[E_\Pi(h|\mathcal{B} \times \mathcal{D})|\mathcal{C} \times \mathcal{A}] \\ &= E_\Pi[E_\Pi(h|\mathcal{C} \times \mathcal{A})|\mathcal{C} \times \mathcal{A}] = E_\Pi(h|\mathcal{C} \times \mathcal{A}) = h \quad [\text{II}] \end{aligned}$$

refletindo que  $h \in \overline{\mathcal{C} \times \mathcal{D}}^\Pi$ , Q.E.D. □

O dual do conceito amostral de  $B$ -completude, frequentemente denominado de identificação forte por  $\mathcal{A}_+$ , é naturalmente traduzido por:

### Definição 11.7

$\mathcal{C} \subset \mathcal{B}$  diz-se  $B$ -completa (resp.  $B$ -completa limitada) se para toda função  $\phi \in \mathcal{C}$  e integrável (resp.  $\phi \in \mathcal{L}(\mathcal{C})$ ), se tem

$$E_\mu(\bar{\phi}|\mathcal{A}_+) = 0 \Rightarrow \bar{\phi} = 0 \quad [\mu]$$

com  $\bar{\phi} = \phi \circ \bar{\theta} \in \mathcal{C}_+$ . Desta forma:

### Exercício 11.6

1. Mostre que num modelo bayesiano regular com  $P$  regular para  $Q$ ,  $\mathcal{C} \subset \mathcal{B}$  é  $B$ -completa se, e somente se,  $\forall \phi \in \mathcal{C}$  e integrável

$$E_{\nu_x}(\phi) = 0, \quad \forall x \in \mathcal{X} \Rightarrow \phi = 0 \quad [\nu_x] \quad \forall x.$$

2. Prove que se  $\mathcal{C}$  é  $B$ -completa então  $\mathcal{C}$  não contém nenhuma subálgebra  $B$ -ancilar não trivial.
3. Justifique as afirmações:

- a)  $\mathcal{C} \subset \mathcal{B}$  é  $B$ -suficiente se, e somente se, a esperança a posteriori de qualquer função  $\phi$   $\mathcal{B}$ -mensurável não muda  $\mu$ -essencialmente com a aplicação a  $\phi$  do operador esperança condicional em  $\mathcal{C}$ ;
- b) Se  $\mathcal{C}$  é  $B$ -suficiente completa, então  $E(\phi|\mathcal{C})$  é a  $\mu$ -essencialmente única função com menor variância a posteriori dentro daquelas com a esperança a posteriori de  $\phi$  e com variância a posteriori finita.

Com base nos conceitos paramétricos de  $B$ -suficiência,  $B$ -ancilaridade e  $B$ -completude facilmente se derivam os duais das versões bayesianas dos teoremas de Basu referidos no Capítulo 10.

### Exercício 11.7

Mostre que:

1. Sejam  $\mathcal{H}$  e  $\mathcal{C}$  subálgebras de  $\mathcal{B}$  independentes a posteriori e  $\mathcal{H}$   $B$ -ancilar. Se  $\mathcal{H}_+ \vee \mathcal{C}_+$  é  $B$ -suficiente, então  $\mathcal{C}$  é igualmente  $B$ -suficiente.
2. Se  $\mathcal{C} \subset \mathcal{B}$  é  $B$ -suficiente completa, então  $\mathcal{C}$  é independente a priori e a posteriori de qualquer subálgebra paramétrica  $B$ -ancilar.
3. Verifique os resultados 1. e 2. no Exemplo 11.2 com  $\mathcal{H} = \mathcal{B}(\phi)$  e  $\mathcal{C} = \mathcal{B}(\psi)$ .

Em consonância com as considerações feitas em 10.3 sobre o conceito amostral de  $B$ -suficiência mínima diremos que:

### Definição 11.8

$\mathcal{C}^* \subset \mathcal{B}$  é  $B$ -suficiente mínima se  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{C}_+^*$  e  $\mathcal{C}_+^* \subset \bar{\mathcal{C}}_+ \cap \mathcal{B}_+$ ,  $\forall \mathcal{C} \subset \mathcal{B}$  tal que  $\mathcal{A}_+ \amalg \mathcal{B}_+ | \mathcal{C}_+$ .

Consideremos agora a projeção de  $\mathcal{A}_+$  sobre  $\mathcal{B}_+$ , i.e.,

$$\mathcal{B}_+[\mathcal{A}_+] = \sigma(\{E(t|\mathcal{B}_+) : t \in \mathcal{L}(\mathcal{A}_+)\}) .$$

De novo se pode provar que o conceito da Definição 11.8 gira em torno desta subálgebra de  $\mathcal{B}$ , como o atestam as afirmações seguintes.

### Exercício 11.8

Justifique as afirmações:

- a)  $\mathcal{B}_+[\mathcal{A}_+]$  é a única subálgebra  $B$ -suficiente mínima dentro da classe de subálgebras  $\mathcal{B}_+$ -completadas (i.e., que incluem os conjuntos  $\nu$ -nulos de  $\mathcal{B}$ ).
- b)  $\mathcal{C}^*$  é  $B$ -suficiente mínima se, e somente se,  $\overline{\mathcal{C}_+^*} \cap \mathcal{B}_+ = \mathcal{B}_+[\mathcal{A}_+]$ .
- c) Sendo  $\mathcal{H}_0$  a menor subálgebra de  $\mathcal{B}$  que torna as probabilidades amostrais mensuráveis, i.e.,  $\mathcal{H}_0 = \sigma(\{P_\theta(A) : A \in \mathcal{A}\})$ ,  $\mathcal{H}_0$  é  $B$ -suficiente mínima.
- d) Sendo  $P$  regular para  $Q$ , o completamento de  $\sigma(\{h(\theta, x) : x \in \mathcal{X}\})$  com os conjuntos  $\nu$ -nulos de  $\mathcal{B}$  coincide com  $\mathcal{B}_+[\mathcal{A}_+]$ .
- e) Toda a subálgebra de  $\mathcal{B}$   $B$ -suficiente completa limitada é  $B$ -suficiente mínima.

## 11.3 IDENTIFICABILIDADE BAYESIANA PARAMÉTRICA

A definição e existência de uma subálgebra paramétrica  $B$ -suficiente mínima justifica, dada a monotonicidade do conceito de independência condicional, que possamos definir a  $B$ -suficiência do seguinte modo:

### Definição 11.9

$\mathcal{C} \subset \mathcal{B}$  é  $B$ -suficiente se, e somente se,  $\mathcal{H}_0 \subset \mathcal{C}$  onde  $\mathcal{H}_0 = \sigma(\{P_\theta(A) : A \in \mathcal{A}\})$ .

### Nota

Obtém-se uma definição equivalente se substituirmos  $\mathcal{H}_0$  por  $\mathcal{B}_+[\mathcal{A}_+]$ , dado que esta projeção coincide com o completamento de  $\mathcal{H}_0$  com os conjuntos  $\nu$ -nulos de  $\mathcal{B}$  (vide Exercício 11.8 c)).

Com o intuito de ligarmos a especialização desta definição a partições (ou funções) com a Definição 11.4, devemos notar que a partição induzida por  $\mathcal{H}_0$  coincide com  $\Theta / \sim = \{[\theta], \theta \in \Theta\}$ .

Com efeito, sendo  $\gamma_A(\theta) = P_\theta(A) : (\Theta, \mathcal{B}) \rightarrow ([0, 1], G_0)$  onde  $G_0$  é a classe dos borelianos de  $[0, 1]$ , indutora da subálgebra  $\mathcal{B}_A \equiv \gamma_A^{-1}(G_0)$ ,  $\mathcal{H}_0$  fica definida por

$\mathcal{H}_0 = \bigvee_{A \in \mathcal{A}} \mathcal{B}_A$ . Por definição da partição induzida por uma  $\sigma$ -álgebra,  $\Pi(\mathcal{H}_0) = \{\Pi(\theta_0) : \theta_0 \in \Theta\}$ , onde  $\Pi(\theta_0)$  (os chamados átomos de  $\mathcal{H}_0$ ) são definidos por

$$\begin{aligned} \Pi(\theta_0) &= \bigcap \{B \in \mathcal{H}_0 : \theta_0 \in B\} \\ &= \bigcap_{B \in \mathcal{H}_0} \{\theta_1 : I_B(\theta_1) = I_B(\theta_0)\} \end{aligned}$$

Por outras palavras,  $\Pi(\mathcal{H}_0)$  não é mais do que o conjunto quociente de  $\Theta$  segundo a relação de equivalência definida por  $I_B(\theta_1) = I_B(\theta_0), \forall B \in \mathcal{H}_0$ . Atendendo à definição de  $\mathcal{H}_0$ , tem-se então

$$\begin{aligned} \Pi(\theta_0) &= \{\theta_1 : I_B(\theta_1) = I_B(\theta_0), \forall B \in \mathcal{B}_A, \forall A \in \mathcal{A}\} \\ &= \{\theta_1 : \gamma_A(\theta_1) = \gamma_A(\theta_0) : \forall A \in \mathcal{A}\} = [\theta_0] \end{aligned}$$

ou seja, os átomos de  $\mathcal{H}_0$  são as classes de equivalência observacional.

Observe-se ainda que  $\mathcal{H}_0$  é necessariamente separável e, conseqüentemente,  $\Theta / \sim$  é  $\mathcal{H}_0$ -mensurável quando  $(\mathcal{X}, \mathcal{A})$  é um espaço euclideano.

### Exercício 11.9

Mostre que a separabilidade de  $\mathcal{A}$  garante a separabilidade de  $\mathcal{H}_0$  e a mensurabilidade dos seus átomos:

- definindo uma classe-  $\Pi \mathcal{I}$  de conjuntos  $A_n$  gerando  $\mathcal{A}$  e provando a separabilidade de  $\mathcal{B}_{A_n}$ ;
- mostrando que  $\mathcal{M} = \{A \in \mathcal{A} : \gamma_A \in \bigvee_{n=1}^{\infty} \mathcal{B}_{A_n}\}$  é um sistema de Dynkin contendo  $\mathcal{I}$ ;
- concluindo que  $\mathcal{H}_0 = \bigvee_{n=1}^{\infty} \mathcal{B}_{A_n}$  e que  $\Pi(\mathcal{H}_0)$  é  $\mathcal{H}_0$ -mensurável.

Por outro lado, deve notar-se que no contexto de  $(\Theta, \mathcal{B})$  euclideano, a relação de redução  $\nu$ -essencial entre duas partições equivale à relação de inclusão  $\nu$ -essencial entre as respectivas subálgebras induzidas, se as partições forem induzidas por estatísticas reais.

### Exercício 11.10

Prove a afirmação anterior.

Deste modo, a especialização da Definição 11.9 à subálgebra induzida por uma partição  $\Pi$  conduz-nos a dizer, no quadro em que  $(\mathcal{X}, \mathcal{A})$  e  $(\Theta, \mathcal{B})$  são ambos euclidianos, que  $\Pi$  é  $B$ -suficiente se, e somente se,  $\Theta / \sim < \Pi \ [\nu]$ . Trata-se assim da extensão da Definição 11.4 a que aludimos no início de 11.2.

Este argumento de definição da  $B$ -suficiência em termos da  $B$ -suficiência mínima e do paralelismo referido com os respectivos conceitos clássicos motiva-nos a definir o conceito bayesiano de identificabilidade da seguinte forma:

**Definição 11.10**

$\mathcal{C} \subset \mathcal{B}$  diz-se  $B$ -identificável se  $\mathcal{C} \subset \mathcal{H}_0 \ [\nu]$ , ou equivalentemente, se  $\mathcal{C} \subset \mathcal{B}_+[\mathcal{A}_+]$ . Em particular, o modelo diz-se  $B$ -identificável se  $\mathcal{B} = \mathcal{H}_0 \ [\nu]$ , ou seja,  $\mathcal{B} = \mathcal{B}_+[\mathcal{A}_+]$ .

A especialização desta definição à subálgebra induzida por uma partição  $\Pi$  traduz-se no contexto mencionado pela afirmação de que  $\Pi$  é  $B$ -identificável se, e somente se,  $\Pi < \Theta / N \ [\nu]$ , ou seja, pela extensão da definição clássica (reveja a Definição 11.3).

Este conceito de identificabilidade do modelo corresponde efetivamente ao conceito clássico de injetividade da aplicação  $P_\theta : \Theta \rightarrow [0, 1]$  a menos de um conjunto  $\nu$ -nulo, como o atesta o seguinte teorema:

**Teorema 11.6**

O modelo é  $B$ -identificável quando  $(\mathcal{X}, \mathcal{A})$  e  $(\Theta, \mathcal{B})$  são euclidianos se, e somente se, a função  $P_\theta : \Theta \rightarrow [0, 1]$  é injetiva  $\nu$ -quase sempre.

**Prova**

(Parte necessária): Seja  $\mathcal{I} = \{E_n\}$  uma álgebra enumerável geradora de  $\mathcal{B}$  suposto separável. Como, por hipótese,  $\mathcal{B} = \overline{\mathcal{H}_0}^\nu$ ,  $\exists F_n \in \mathcal{H}_0$  tal que  $E_n = F_n \ [\nu]$ , para cada  $E_n \in \mathcal{I}$ . Sendo  $N = \bigcup_{n=1}^\infty (E_n \Delta F_n)$ , então  $\nu(N) = 0$ .

Seja agora  $\theta_1$  e  $\theta_0$  tal que  $I_{E_n}(\theta_1) = I_{E_n}(\theta_0)$ ,  $\forall n \geq 1$ . Então  $\theta_1$  e  $\theta_0$  fazem parte do mesmo átomo de  $\mathcal{B}$ . Como  $\mathcal{B}$  é uma  $\sigma$ -álgebra de Borel, os seus átomos (que são mensuráveis pela separabilidade de  $\mathcal{B}$ ) são os conjuntos singulares e, portanto,  $\theta_1 = \theta_0$ . De outra forma,  $\forall \theta_1 \neq \theta_0$ ,  $\exists E_n : \theta_1 \in E_n$  e  $\theta_0 \in E_n^c$ .

Suponha agora  $\theta_1 \in \Theta - N$  e equivalente observacionalmente a  $\theta_0 \neq \theta_1$ , i.e.,  $\theta_1$  faz parte do átomo de  $\mathcal{H}_0$  que contém  $\theta_0$ . Então  $\theta_1 \in E_n \cap F_n$  e  $\theta_0 \in E_n^c \cap F_n \subset E_n \Delta F_n$

necessariamente. Assim,  $\Theta - N$  só contém um ponto de cada um dos átomos de  $\mathcal{H}_0$  (note-se que  $\theta_0 \in N$ ) e, portanto, a aplicação  $P_\theta$  é injetiva em  $\Theta - N$ .

(Parte suficiente): Seja  $\mathcal{B}^* = \mathcal{B} \cap (\Theta - N)$  e  $\mathcal{H}_0^* = \mathcal{H}_0 \cap (\Theta - N)$ . Por hipótese, os átomos de  $\mathcal{H}_0^*$  são os conjuntos singulares de  $\Theta - N$ . Como a separabilidade de  $\mathcal{A}$  garante a separabilidade de  $\mathcal{H}_0^* \equiv \sigma(\{P_\theta : \theta \in \Theta - N\})$ , cada átomo de  $\mathcal{H}_0^*$  é  $\mathcal{B}^*$ -mensurável e portanto os conjuntos singulares de  $\Theta - N$  são necessariamente os átomos de  $\mathcal{B}^*$ . Como  $\mathcal{B}^*$  é a  $\sigma$ -álgebra de Borel de  $\Theta - N$ , isto significa que os átomos de  $\mathcal{H}_0^*$  e de  $\mathcal{B}^*$  coincidem, ou seja,  $\mathcal{H}_0^* = \mathcal{B}^*$ . Pela definição de  $\mathcal{H}_0^*$  e  $\mathcal{B}^*$ , esta relação significa que  $\mathcal{H}_0$  e  $\mathcal{B}$  coincidem num conjunto de probabilidade  $\nu = 1$ , i.e.,  $\mathcal{B} = \overline{\mathcal{H}_0}^\nu$ , Q.E.D.  $\square$

É importante realçar o papel crucial da medida  $\nu$  na definição destes conceitos bayesianos de suficiência e identificabilidade paramétricas. O exemplo seguinte retrata-o bem.

### Exemplo 11.3

Seja  $P_\theta$  definida no espaço euclideano  $n$ -dimensional pela medida produto de  $n$  distribuições  $N(|\theta|, 1)$  independentes e seja  $(\Theta, \mathcal{B})$  o espaço euclideano unidimensional.

Observe-se que  $\mathcal{H}_0 = \mathcal{B}(|\theta|) = \{D \in \mathcal{B} : D = -D\}$ , pelo que  $\mathcal{B}_+[\mathcal{A}_+] = \overline{\mathcal{H}_0}^\nu$ . Sendo  $\nu$  equivalente à medida de Lebesgue, o complemento de  $\mathcal{H}_0$  com os borelianos  $\nu$ -nulos de  $\mathcal{B}$  está incluído estritamente em  $\mathcal{B}$ , pelo que, o modelo não é  $B$ -identificável, como também não o é no sentido clássico – note-se que  $[\theta] = \{\theta, -\theta\}$ .

Contudo, se  $\nu$  for equivalente à medida de Lebesgue em  $\mathbb{R}^+$  e  $\nu(\mathbb{R}^-) = 0$ , o complemento referido de  $\mathcal{H}_0$ , ao incluir os borelianos de  $\mathbb{R}^-$  e, em decorrência, os borelianos de  $\mathbb{R}^+$ , coincide com  $\mathcal{B}$ , pelo que o modelo passa a ser  $B$ -identificável. Este caso ilustra uma situação em que o modelo não é identificável classicamente mas é o do ponto de vista bayesiano, ainda que  $\mathcal{H}_0$  esteja contido estritamente em  $\mathcal{B}$ .

### Exercício 11.11

Mostre que a  $B$ -identificabilidade é implicada para toda medida  $\nu$ , quer pela identificabilidade clássica, quer pela condição  $\mathcal{H}_0 = \mathcal{B}$ .

Com esta definição de  $B$ -identificabilidade, todo o elemento minimal da classe de subálgebras  $B$ -suficientes é um elemento maximal da classe de subálgebras  $B$ -identificá-

veis. Daí que o conceito que denominamos de  $B$ -suficiência mínima possa ser rotulado de  $B$ -identificabilidade máxima. Os átomos de uma subálgebra como  $\mathcal{H}_0$ , que permitem caracterizar tudo o que é ou não identificável, são em termos essenciais, os átomos mais grossos que são deixados invariantes pelos dados no processo da operação bayesiana e, simultaneamente, os átomos mais finos que são discriminados probabilisticamente pelos dados amostrais.

Observe-se ainda que as Definições 11.9 e 11.10 não são mais do que o dual dos conceitos clássicos de subálgebras amostrais suficientes e necessárias (na acepção de Bahadur), quando  $\nu$  é equivalente a  $Q$  (como acontece se  $P$  for regular para  $Q$ ).

**Exercício 11.12**

Seja  $\psi$  uma função paramétrica mensurável real identificante, i.e., indutora da partição  $\Theta / \sim$ . Mostre que a estatística real  $\phi$  é  $B$ -identificável se, e somente se,  $\phi$  é  $\nu$ -essencialmente  $\mathcal{B}(\psi)$ -mensurável.



## CAPÍTULO 12

### $F$ -SUFICIÊNCIA E $K$ -SUFICIÊNCIA

#### 12.1 INTRODUÇÃO

Seja  $\theta \in \Theta$  o parâmetro que define a distribuição de probabilidade  $P_\theta$  de um elemento aleatório observável  $x$  e seja  $T = T(x)$  uma estatística. Segundo Fisher (1922), a estatística  $T$  é chamada suficiente se a distribuição condicional de  $X$ , dado  $T$ , é a mesma para todo  $\theta \in \Theta$ . Chamaremos a tal estatística  $F$ -suficiente, isto é, suficiente no sentido de Fischer.

A noção de  $F$ -suficiência está orientada ao espaço amostral. Esta noção de suficiência tem sido estudada extensivamente na literatura.

Em 1942, Kolmogorov propôs uma definição de suficiência orientada ao espaço paramétrico nos seguintes termos bayesianos. Para uma distribuição de probabilidade a priori  $v$  sobre o espaço paramétrico  $\Theta$ , seja  $v * (\cdot|x)$  a correspondente distribuição a posteriori dado  $x$ . Se é achado que a medida a posteriori  $v * (\cdot|x)$  depende de  $x$  só através de  $T = T(x)$ , então  $T$  pode ser considerada suficiente com respeito a  $v$ . Se  $T$  é suficiente no sentido acima com respeito a toda a priori  $v$ , então  $T$  será considerada suficiente no sentido de Kolmogorov e diremos que é  $K$ -suficiente.

A noção de  $K$ -suficiência não tem recebido muita atenção.

Neste capítulo faremos um estudo comparativo das noções de  $F$ -suficiência e  $K$ -suficiência numa formulação abstrata, utilizando a teoria da Medida. Mostraremos que a noção de  $K$ -suficiência é levemente mais geral que a noção de  $F$ -suficiência e que a noção de  $K$ -suficiência está livre das conseqüências paradoxais da definição de Fisher como Burkholder pontualizou em 1961. Também mostraremos que, no caso dominado, as noções de  $F$ -suficiência e  $K$ -suficiência coincidem.

## 12.2 PRELIMINARES

Seja  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , onde  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ , o **modelo estatístico** para o elemento aleatório observável  $x$  e seja  $(\Theta, \mathcal{F})$  a estrutura mensurável do espaço paramétrico. Então o espaço mensurável produto

$$(\Omega, \mathcal{G}) = (\mathcal{X} \times \Theta, \mathcal{A} \times \mathcal{F})$$

corresponde a  $\omega = (x, \theta)$ .

### Definição 12.1

Uma **função de transição** de  $(\Theta, \mathcal{F})$  a  $(\mathcal{X}, \mathcal{A})$  é uma aplicação  $K(\cdot, \cdot) : \mathcal{A} \times \Theta \rightarrow [0, 1]$  tal que  $\forall A \in \mathcal{A}, K(A, \cdot) \in \mathcal{F}$ , isto é,  $K(A, \cdot)$  é  $\mathcal{F}$ -mensurável, e  $\forall \theta \in \Theta, K(\cdot, \theta)$  é uma medida de probabilidade sobre  $(\mathcal{X}, \mathcal{A})$ . Escreveremos  $K(A, \theta)$  como  $K(A|\theta)$  para cada  $(A, \theta) \in \mathcal{A} \times \Theta$ .

Assumiremos que a aplicação  $K(\cdot, \cdot) : \mathcal{A} \times \Theta \rightarrow [0, 1]$ , definida por  $K(A, \theta) = P_\theta(A) \forall (A, \theta) \in \mathcal{A} \times \Theta$ , é uma função de transição e frequentemente escreveremos  $P_\theta(A)$  como  $P(A|\theta)$ .

Estabelecemos os seguintes fatos sobre funções de transição:

### Lema 12.1

Se  $f \in \mathcal{A}$  é  $\mathcal{P}$ -integrável, então a função  $g$  definida por  $g(\theta) = \int_{\mathcal{X}} f dP_\theta$  é  $\mathcal{F}$ -mensurável.

### Lema 12.2

Correspondendo a cada medida de probabilidade a priori  $\nu$  sobre  $(\Theta, \mathcal{F})$ , existe uma

medida de probabilidade marginal  $\mu_v$  sobre  $(\mathcal{X}, \mathcal{A})$  definida por

$$\mu_v(A) = \int_{\Theta} P_{\theta}(A)v(d\theta), \quad \forall A \in \mathcal{A}$$

**Lema 12.3**

Dada a medida de probabilidade  $v$  sobre  $(\Theta, \mathcal{F})$ , a função

$$\Pi_v(A \times F) = \int_{\mathcal{F}} P_{\theta}(A)v(d\theta)$$

definida sobre a classe  $\mathcal{R} = \{A \times F; A \in \mathcal{A}, F \in \mathcal{F}\}$  dos retângulos mensuráveis, pode ser estendida a uma medida de probabilidade sobre o espaço produto  $(\Omega, G) = (\mathcal{X} \times \Theta, \mathcal{A} \times \mathcal{F})$ . Esta medida de probabilidade sobre  $(\Omega, G)$  também é denotada por  $\Pi_v$ .

A  $\sigma$ -álgebra  $\mathcal{A}^0 = \{A \times \Theta; A \in \mathcal{A}\} \subseteq \mathcal{A} \times \mathcal{F}$  é chamada a subálgebra marginal associada a  $\mathcal{A}$ . A cada subálgebra  $C$  de  $\mathcal{A}$  está associada uma subálgebra  $C^0$  de  $\mathcal{A} \times \mathcal{F}$ , definida por  $C^0 = \{C \times \Theta; C \in \mathcal{C}\}$  e chamada a **subálgebra marginal** associada a  $\mathcal{C}$ . A subálgebra marginal  $\mathcal{F}^0$  de  $\mathcal{A} \times \mathcal{F}$  associada a  $\mathcal{F}$  é definida por  $\mathcal{F}^0 = \{\mathcal{X} \times F; F \in \mathcal{F}\}$ .

Observamos que a medida de probabilidade sobre o espaço produto  $(\Omega, G) = (\mathcal{X} \times \Theta, \mathcal{A} \times \mathcal{F})$

$$\Pi_v : \mathcal{A} \times \mathcal{F} \rightarrow [0, 1]$$

do Lema 12.3, satisfaz:

- (i)  $\Pi_v(A \times \Theta) = \mu_v(A) \quad \forall A \in \mathcal{A}$ ,
- (ii)  $\Pi_v(\mathcal{X} \times F) = v(F) \quad \forall F \in \mathcal{F}$ .

Portanto  $\mathcal{A}^0$  pode ser identificado com  $\mathcal{A}$  e  $\Pi_v|_{\mathcal{A}^0}$  (a restrição de  $\Pi_v$  a  $\mathcal{A}^0$ ) pode ser identificada com  $\mu_v$  (a distribuição de probabilidade marginal associada a  $v$ , do Lema 12.2);  $\mathcal{F}^0$  e  $\Pi_v|_{\mathcal{F}^0}$  podem ser identificadas com  $\mathcal{F}$  e  $v$ , respectivamente.

Se  $f \in \mathcal{A}$ , isto é,  $f : \mathcal{X} \rightarrow \mathbb{R}$  é uma função  $\mathcal{A}$ -mensurável, então a função  $(x, \theta) \rightarrow f(x)$  definida em  $\mathcal{X} \times \Theta = \Omega$  é  $\mathcal{A} \times \mathcal{F}$ -mensurável e é denotada também por  $f$ .

**Lema 12.4**

Seja  $v$  uma medida de probabilidade a priori sobre  $(\Theta, \mathcal{F})$ .

(a) Se  $h \in \mathcal{A} \times \mathcal{F}$  é  $\Pi_v$ -integrável, então

$$\int_{\mathcal{A} \times \mathcal{F}} h d\Pi_v = \int_{\mathcal{F}} \left[ \int_{\mathcal{A}} h(x, \theta) P_{\theta}(dx) \right] v(d\theta) \quad \forall A \in \mathcal{A}, \forall F \in \mathcal{F}$$

(b) Se  $f \in \mathcal{A}$  é  $\mathcal{P}$ -integrável, isto é,  $f$  é  $P_\theta$ -integrável  $\forall \theta \in \Theta$ , então

$$\int_{\mathcal{A} \times \Theta} f d\Pi_v = \int_{\mathcal{A}} f(x) \mu_v(dx) \quad \forall A \in \mathcal{A}.$$

### Notação

Se  $\mathcal{H}$  é uma subálgebra de  $\mathcal{A} \times \mathcal{F}$  e  $h \in \mathcal{A} \times \mathcal{F}$  é  $\Pi_v$ -integrável, então denotaremos  $E_{\Pi_v}(h|\mathcal{H})$  por  $E_v(h|\mathcal{H})$ .

### Lema 12.5

Suponhamos que o espaço paramétrico  $(\Theta, \mathcal{F})$  é euclidiano, isto é,  $\Theta$  tem estrutura de espaço vetorial e está munido de um produto interno. Então, correspondendo a cada medida de probabilidade a priori  $v$  sobre  $(\Theta, \mathcal{F})$ , existe uma função de transição

$$v * (\cdot|\cdot) : \mathcal{F} \times \mathcal{X} \rightarrow [0, 1]$$

de  $(\mathcal{X}, \mathcal{A})$  a  $(\Theta, \mathcal{F})$  tal que

- (a) para cada  $F \in \mathcal{F}$ ,  $v * (F|\cdot) \in \mathcal{A}$  é uma versão de  $E_v(I_{\mathcal{X} \times \mathcal{F}}|\mathcal{A}^0)$ ;
- (b) para cada  $x \in \mathcal{X}$ ,  $v * (\cdot|x)$  é uma medida de probabilidade sobre  $(\Theta, \mathcal{F})$ .

Além disto, a função de transição  $v * (\cdot|\cdot)$  é  $\mu_v$ -essencialmente única no seguinte sentido: se  $v_1(\cdot|\cdot)$  é outra tal função de transição, então existe  $N \in \mathcal{A}$  tal que  $\mu_v(N) = 0$  e  $v_1(\cdot|x) = v * (\cdot|x) \quad \forall x \in \mathcal{X} - N$ .

### Observação

Assumiremos daqui para a frente que o espaço paramétrico  $(\Theta, \mathcal{F})$  é euclidiano.

Nos termos abstratos da Teoria da Medida, a noção de  $F$ -suficiência pode ser caracterizada como segue.

### Definição 12.2

Uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é chamada  $F$ -suficiente para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  se  $\forall f \in \mathcal{A}$   $\mathcal{P}$ -integrável, existe uma função  $f^* \in \mathcal{C}$   $\mathcal{P}$ -integrável tal que  $f^* = E_\theta(f|\mathcal{C})[P_\theta]$ ,  $\forall \theta \in \Theta$ , isto é,  $\forall \theta \in \Theta$ ,  $\forall C \in \mathcal{C}$ ,  $\int_C f^* dP_\theta = \int_C f dP_\theta$ .

### Definição 12.3

Uma sub- $\sigma$ -álgebra  $\mathcal{C}$  de  $\mathcal{A}$  é chamada  $F$ -suficiente por par para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  se  $\forall \theta_1, \theta_2 \in \Theta$ ,  $\forall f \in \mathcal{A}$   $\mathcal{P}$ -integrável, existe uma função  $f^* \in \mathcal{C}$   $\mathcal{P}$ -integrável tal que

$$\forall \theta \in \{\theta_1, \theta_2\}, \forall C \in \mathcal{C}, \int_C f dP_\theta = \int_C f^* dP_\theta$$

**Exemplo 12.1**

Seja  $\mathcal{X} = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}_1$ , a  $\sigma$ -álgebra dos borelianos de  $\mathbb{R}$  e seja  $\mathcal{P}$  a família de todas as medidas de probabilidade discretas sobre  $(\mathcal{X}, \mathcal{A})$ . É fácil verificar que a  $\sigma$ -álgebra  $\mathcal{C} = \{C \in \mathcal{A}; C \text{ é contável ou } C^c \text{ é contável}\}$  é  $F$ -suficiente por par para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  e que a única subálgebra  $F$ -suficiente para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  é  $\mathcal{A}$ .

Este exemplo demonstra que a noção de  $F$ -suficiência por par é mais fraca do que a noção de  $F$ -suficiência. Porém, no caso dominado, isto é, quando existe uma medida  $\sigma$ -finita  $\lambda$  sobre  $(\mathcal{X}, \mathcal{A})$  tal que cada  $P_\theta \in \mathcal{P}$  é absolutamente contínua com respeito a  $\lambda$ , temos o seguinte resultado

**Lema 12.6**

Suponhamos que  $\mathcal{P}$  é dominada. Então uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é  $F$ -suficiente para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  se e só se  $\mathcal{C}$  é  $F$ -suficiente por par para  $\mathcal{A}$  com respeito a  $\mathcal{P}$ .

A seguir apresentamos a noção de  $K$ -suficiência, orientada ao espaço paramétrico, nos termos abstratos da teoria da Medida.

**Definição 12.4**

Uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é chamada  $K$ -suficiente para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  se para cada a priori  $v$  sobre  $(\Theta, \mathcal{F})$ , a função de transição a posteriori  $v * (\cdot|\cdot)$  do Lema 12.5 satisfaz a seguinte condição:

$$\forall F \in \mathcal{F}, v * (F|\cdot) \text{ tem uma versão } \mathcal{C}\text{-mensurável, isto é, } E_v(I_{\mathcal{X} \times F} | \mathcal{A}^0) \in \mathcal{C}^0[\mathcal{P}].$$

**Definição 12.5**

Uma sub- $\sigma$ -álgebra  $\mathcal{C}$  de  $\mathcal{A}$  é chamada  $K$ -suficiente por par para  $\mathcal{A}$  com respeito a  $\mathcal{P}$  se para cada a priori  $v$  uniforme sobre dois pontos de  $\Theta$ ,  $v * (F|\cdot) \in \mathcal{C}[\mathcal{P}] \forall F \in \mathcal{F}$ .

Agora fixemos uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  e consideremos uma medida de probabilidade  $P$  sobre  $(\mathcal{X}, \mathcal{A})$ .

**Definição 12.6**

Dois eventos  $B$  e  $C$  em  $\mathcal{A}$  são ditos condicionalmente independentes dada  $\mathcal{D}$  (com respeito a  $\mathcal{P}$ ) se

$$P(B \cap C/\mathcal{D}) = P(B/\mathcal{D})P(C/\mathcal{D}) [P]$$

### Notação

Escreveremos  $B \perp C/P$  para indicar que  $B$  e  $C$  são  $P$ -independentes, e escreveremos  $B \perp C \mid \mathcal{D}, P$  para indicar que  $B$  e  $C$  são condicionalmente independentes da  $\mathcal{D}$  com respeito a  $P$ . No que segue omitiremos  $P$ .

Desde a definição é fácil ver que  $B \perp D/\mathcal{D} \forall D \in \mathcal{D}$ , e que  $B \perp C/\mathcal{D}$  implica  $B \perp (C \cap D)/\mathcal{D} \forall D \in \mathcal{D}$ .

Seja  $B \in \mathcal{A}$  fixo. É facilmente visto que a classe

$$\mathcal{K}_B = \{K \in \mathcal{A}; B \perp K/\mathcal{D}\}$$

é um sistema de Dynkin.

### Definição 12.7

O evento  $B$  e a subálgebra  $\mathcal{C}$  são ditos condicionalmente independentes dada  $\mathcal{D}$  (com respeito a  $P$ ) se  $\mathcal{C} \subseteq \mathcal{K}_B$ , isto é,  $B \perp C/\mathcal{D} \forall C \in \mathcal{C}$ , e neste caso escreveremos  $B \perp \mathcal{C}/\mathcal{D}$ .

Dada uma subálgebra  $\mathcal{B}$  de  $\mathcal{A}$ , a classe

$$\mathcal{K}_{\mathcal{B}} = \{K \in \mathcal{A}; K \perp \mathcal{B}/\mathcal{D}\} = \bigcap_{B \in \mathcal{B}} \mathcal{K}_B$$

é também um sistema de Dynkin

### Definição 12.8

Duas subálgebras  $\mathcal{B}$  e  $\mathcal{C}$  de  $\mathcal{A}$  são ditas condicionalmente independentes dada  $\mathcal{D}$  (com respeito a  $P$ ) se  $\mathcal{C} \subseteq \mathcal{K}_{\mathcal{B}}$ , isto é,  $B \perp C/\mathcal{D} \forall B \in \mathcal{B}, \forall C \in \mathcal{C}$ .

Estabelecemos os seguintes fatos sobre independência condicional. Sejam  $\mathcal{B}_0, \mathcal{C}, \mathcal{B}$  e  $\mathcal{D}_1$ , subálgebras de  $\mathcal{A}$ . Temos:

### Lema 12.7

Se  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}$  e  $\mathcal{B}_0 \subseteq \mathcal{B}$ , então  $\mathcal{B}_0 \perp \mathcal{C}/\mathcal{D}$ .

**Lema 12.8**

As seguintes afirmações são equivalentes:

- (a)  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}$ ;
- (b)  $\mathcal{B} \perp (\mathcal{C} \vee \mathcal{D})/\mathcal{D}$ ;
- (b)  $(\mathcal{B} \vee \mathcal{D}) \perp (\mathcal{C} \vee \mathcal{D})/\mathcal{D}$ .

**Lema 12.9**

Se  $\mathcal{B} \perp \mathcal{C}$  e  $\mathcal{B}_0 \subseteq \mathcal{B}$ , então  $\mathcal{B} \perp \mathcal{C}/\mathcal{B}_0$ .

**Lema 12.10**

$\mathcal{B} \perp \mathcal{C}/\mathcal{D}$  se e só se  $E_P(I_C/\mathcal{B} \vee \mathcal{D}) = E_P(I_C/\mathcal{D}) [P] \quad \forall C \in \mathcal{C}$ .

**Corolário 12.1**

$\mathcal{B} \perp \mathcal{C}/\mathcal{D}$  se e só se para cada  $C \in \mathcal{C}$ ,  $E_P(I_C/\mathcal{B} \vee \mathcal{D})$  tem uma versão  $\mathcal{D}$ -mensurável.

**Corolário 12.2**

Suponhamos que  $\mathcal{D} \subseteq \mathcal{D}_1 \subseteq \mathcal{B}$ .  $\mathcal{D}_1 \perp \mathcal{C}/\mathcal{D}$  e  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}_1$  implica  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}$ .

**Corolário 12.3**

Se  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}$  e  $\mathcal{D} \subseteq \mathcal{D}_1 \subseteq \mathcal{B}$ , então  $\mathcal{B} \perp \mathcal{C}/\mathcal{D}_1$ .

## 12.3 RELAÇÃO ENTRE AS NOÇÕES DE $F$ -SUFICIÊNCIA E $K$ -SUFICIÊNCIA

Faremos um estudo comparativo das noções de  $F$ -suficiência e  $K$ -suficiência. Em 1961 Burkholder demonstrou que existe uma subálgebra  $\mathcal{D}$  não- $F$ -suficiente que contém uma subálgebra  $\mathcal{C}$   $F$ -suficiente. Aqui mostraremos que toda subálgebra  $K$ -suficiente é também  $K$ -suficiente. Assim, a definição de Kolmogorov está livre das conseqüências paradoxais da definição de Fisher.

Também mostraremos que toda subálgebra  $F$ -suficiente é também  $K$ -suficiente, mas a recíproca não é verdadeira. Porém, no caso dominado, as noções de  $F$ -suficiência e  $K$ -suficiência são equivalentes.

Temos a seguinte caracterização de  $K$ -suficiência em termos de independência condicional, a qual fornece uma interpretação mais intuitiva da noção de  $K$ -suficiência.

**Teorema 12.1**

Uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é  $K$ -suficiente (para  $\mathcal{A}$  com respeito a  $\mathcal{P}$ ) se e só se para toda a priori  $\nu$  sobre  $(\Theta, \mathcal{F})$ , temos  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}^0, \Pi_\nu$ , isto é,  $\mathcal{A}^0$  e  $\mathcal{F}^0$  são condicionalmente independentes dada  $\mathcal{C}^0$  com respeito a  $\Pi_\nu$ .

**Demonstração**

Pela definição 2.4,  $\mathcal{C}$  é  $K$ -suficiente se e só se para toda a priori  $\nu$  sobre  $(\Theta, \mathcal{F})$ ,  $\Pi_\nu(\mathcal{X} \times F^0 | \mathcal{A}^0) \in \mathcal{C}^0 [\mathcal{P}] \forall F \in \mathcal{F}$ . Portanto  $\mathcal{C}$  é  $K$ -suficiente se e só se para toda a priori  $\nu$  sobre  $(\Theta, \mathcal{F})$ ,  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}^0, \Pi_\nu$ . □

O correspondente resultado para estatísticas é: “Uma estatística  $T = T(x)$  é  $K$ -suficiente para o observável  $x$  (para propósitos de fazer inferências sobre o parâmetro desconhecido  $\theta$ ) se e só se, para toda a priori  $\nu$ ,  $x \perp \theta/T, \Pi_\nu$ ”.

**Corolário 12.4**

$\mathcal{C}$  é  $K$ -suficiente por par (para  $\mathcal{A}$  com respeito a  $\mathcal{P}$ ) se e só se para toda a priori  $\nu$  uniforme sobre dois pontos de  $\Theta$ ,  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}^0, \Pi_\nu$ .

Estas caracterizações de  $K$ -suficiência e  $K$ -suficiência por par serão as chaves para o nosso estudo.

**Teorema 12.2**

Sejam  $\mathcal{C} \subseteq \mathcal{D}$  duas subálgebras de  $\mathcal{A}$ . Se  $\mathcal{C}$  é  $K$ -suficiente, então  $\mathcal{D}$  também é  $K$ -suficiente.

**Demonstração**

Seja  $\nu$  uma priori qualquer sobre  $(\theta, \mathcal{F})$ . Como  $\mathcal{C}$  é  $K$ -suficiente, então  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}^0, \Pi_\nu$ . Pelo Corolário 12.3, temos que  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{D}^0, \Pi_\nu$ , pois  $\mathcal{C}^0 \subseteq \mathcal{D}^0 \subseteq \mathcal{A}^0$ . Consequentemente,  $\mathcal{D}$  é  $K$ -suficiente. □

**Teorema 12.3**



Se  $\mathcal{C}$  é uma subálgebra de  $\mathcal{A}$   $F$ -suficiente, então  $\mathcal{C}$  é  $K$ -suficiente.

### Demonstração

Seja  $v$  uma priori qualquer sobre  $(\Theta, \mathcal{F})$ . Seja  $A \in \mathcal{A}$  qualquer. Como  $\mathcal{C}$  é  $F$ -suficiente, então existe uma função  $I_A^* \in \mathcal{C}$  tal que  $\forall \theta \in \Theta, \forall C \in \mathcal{C}$ ,

$$P_\theta(C \cap A) = \int_C I_A^* dP_\theta .$$

Para demonstrar que  $\mathcal{C}$  é  $K$ -suficiente, isto é,  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}$ ,  $\Pi_\nu$  basta mostrar que  $E_v(I_{\mathcal{A} \times \Theta}/\mathcal{C} \vee \mathcal{F}^0)$  tem uma versão  $\mathcal{C}^0$ -mensurável.

Vejamos que  $I_A^*$  é uma versão de  $E_v(I_{\mathcal{A} \times \Theta}/\mathcal{C}^0 \vee \mathcal{F}^0)$ . Seja

$$\mathcal{K} = \left\{ K \in \mathcal{C}^0 \vee \mathcal{F}^0; \int_K I_{\mathcal{A} \times \Theta} d\Psi_v = \int_K I_A^* d\Pi_\nu \right\}$$

É fácil verificar que  $\mathcal{K}$  é um sistema de Dynkin. Sejam  $C \in \mathcal{C}$ ,  $F \in \mathcal{T}$  quaisquer. Então

$$\begin{aligned} \int_{C \times F} I_A^* d\Pi_\nu &= \int_F \left[ \int_C I_A^*(x) P_\theta(dx) \right] v(d\theta) \\ &= \int_F P_\theta(C \cap A) v(d\theta) \\ &= \Pi_\nu((C \cap A) \times F) \\ &= \Pi_\nu((C \times F) \cap (A \times \Theta)) \\ &= \int_{\mathcal{X} \times \Theta} I_{C \times F} I_{A \times \Theta} d\Pi_\nu \\ &= \int_{C \times F} I_{A \times \Theta} d\Pi_\nu . \end{aligned}$$

Portanto  $C \times F \in \mathcal{K}$ . Logo o sistema de Dynkin  $\mathcal{K}$  contém o  $\pi$ -sistema (classe fechada por intersecção finita)  $\{C \times F; C \in \mathcal{C}, F \in \mathcal{F}\}$ . Segue-se que  $\mathcal{C}^0 \vee \mathcal{F}^0 = \sigma(\{C \times F; C \in \mathcal{C}, F \in \mathcal{F}\}) \subseteq \mathcal{K}$ , pelo teorema de Dynkin. Então  $\forall D \in \mathcal{C}^0 \vee \mathcal{F}^0$ ,

$$\int_D I_{A \times \Theta} d\Pi_\nu = \int_D I_A^* d\Pi_\nu$$

e consequentemente  $I_A^*$  é uma versão de  $E_v(I_{\mathcal{A} \times \Theta}/\mathcal{C}^0 \vee \mathcal{F}^0)$ , como queríamos mostrar.  $\square$

Foi mostrado por Burkholder que existe uma subálgebra  $\mathcal{D}$  de  $\mathcal{A}$  não- $F$ -suficiente que contém uma subálgebra  $\mathcal{C}$   $F$ -suficiente. Pelo Teorema 12.3  $\mathcal{C}$  é também  $K$ -suficiente. Logo,  $\mathcal{D}$  é  $K$ -suficiente, pelo Teorema 12.2. Portanto, a recíproca do Teorema 12.3 não é verdadeira.  $\square$

**Teorema 12.4**

Uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é  $K$ -suficiente por par se e só se é  $F$ -suficiente por par.

**Demonstração**

Suponhamos que  $\mathcal{C}$  é  $F$ -suficiente por par. Seja  $\nu$  uma priori uniforme sobre dois pontos  $\theta_1, \theta_2 \in \Theta$  quaisquer. Como  $\mathcal{C}$  é  $F$ -suficiente para  $\mathcal{A}$  com respeito a  $\{P_{\theta_1}, P_{\theta_2}\}$ , então para todo  $F \in \mathcal{F}$ ,  $\nu^*(F|\cdot)$  tem uma versão  $\mathcal{C}$ -mensurável. Portanto  $\mathcal{C}$  é  $K$ -suficiente por par.

Agora suponhamos que  $\mathcal{C}$  é  $K$ -suficiente por par. Sejam  $P_{\theta_1}, P_{\theta_2} \in \mathcal{P}$  quaisquer. Consideremos uma a priori  $\nu$  uniforme sobre  $\{\theta_1, \theta_2\} \subseteq \Theta$ . Pelo Corolário 12.4 tem-se que  $\mathcal{A}^0 \perp \mathcal{F}^0/\mathcal{C}^0, \Pi_\nu$ . Logo para cada  $A \in \mathcal{A}$ , existe uma função  $I_A^* \in \mathcal{C}$  tal que  $E_\nu(I_{A \times \Theta}/\mathcal{C}^0 \vee \mathcal{F}^0) = I_A^* [\Pi_\nu]$ , isto é,  $E_\nu(I_{A \times \Theta}/\mathcal{C}^0 \vee \mathcal{F}^0)$  tem uma versão  $\mathcal{C}^0$ -mensurável.

Seja  $A \in \mathcal{A}$  qualquer. Então existe uma função  $I_A^* \in \mathcal{C}$  tal que  $\forall C \in \mathcal{C}, \forall i \in \{1, 2\}$ ,

$$\int_{C \times \{\theta_i\}} I_{A \times \Theta} d\Pi_\nu = \int_{C \times \{\theta_i\}} I_A^* d\Pi_\nu$$

(lembrar que  $\mathcal{C}^0 \vee \mathcal{F}^0 = \sigma(\{C \times F; C \in \mathcal{C}, F \in \mathcal{F}\})$ ).

Mas  $\forall C \in \mathcal{C}, \forall i \in \{1, 2\}$ ,

$$\begin{aligned} \int_{C \times \{\theta_i\}} I_{A \times \Theta} d\Pi_\nu &= \int_{\{\theta_i\}} \left[ \int_C I_{A \times \Theta}(x, \theta) P_\theta(dx) \right] \nu(d\theta) \\ &= \int_{\{\theta_i\}} \left[ \int_C I_A(x) I_\Theta(\theta) P_\theta(dx) \right] \nu(d\theta) \\ &= \int_{\{\theta_i\}} \left[ \int_C I_A(x) P_\theta(dx) \right] \nu(d\theta) \\ &= \nu(\{\theta_i\}) \int_C I_A(x) P_{\theta_i}(dx) \\ &= \frac{1}{2} \int_C I_A(x) P_{\theta_i}(dx) \end{aligned}$$

e similarmente mostra-se que  $\forall C \in \mathcal{C}, \forall i \in \{1, 2\}$ ,

$$\int_{C \times \{\theta_i\}} I_A^* d\Pi_\nu = \frac{1}{2} \int_C I_A^*(x) P_{\theta_i}(dx)$$

Portanto  $\forall C \in \mathcal{C}, \forall \theta \in \{\theta_1, \theta_2\}$ ,

$$\int_C I_A(x) P_\theta(dx) = \int_C I_A^*(x) P_\theta(dx),$$

e consequentemente  $\mathcal{C}$  é  $F$ -suficiente por par. □

**Teorema 12.5**

Suponhamos que  $\mathcal{P}$  é dominada. Então uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é  $K$ -suficiente se e só se é  $F$ -suficiente.

**Demonstração**

Suponhamos que  $\mathcal{C}$  é  $F$ -suficiente. Então  $\mathcal{C}$  é  $K$ -suficiente, pelo Teorema 12.3.

Reciprocamente, se  $\mathcal{C}$  é  $K$ -suficiente, então  $\mathcal{C}$  é  $K$ -suficiente por par. Logo  $\mathcal{C}$  é  $F$ -suficiente por par, pelo Teorema 12.4. Portanto,  $\mathcal{C}$  é  $F$ -suficiente, pelo Lema 12.6.  $\square$

**Corolário 12.5**

Suponhamos que  $\mathcal{P}$  seja dominada. Então uma subálgebra  $\mathcal{C}$  de  $\mathcal{A}$  é  $K$ -suficiente se, e só se, é  $K$ -suficiente por par.

**Demonstração**

Se  $\mathcal{C}$  é  $K$ -suficiente, então  $\mathcal{C}$  é  $K$ -suficiente por par, pela definição.

Reciprocamente, se  $\mathcal{C}$  é  $K$ -suficiente por par, então  $\mathcal{C}$  é  $F$ -suficiente por par, pelo Teorema 12.4. Logo  $\mathcal{C}$  é  $F$ -suficiente, pelo Lema 12.6. Portanto  $\mathcal{C}$  é  $K$ -suficiente, pelo Teorema 12.3.  $\square$

## REFERÊNCIAS

- [1] ASH, R.B. (1972). *Real analysis and probability*. New York: Academic Press.
- [2] BAHADUR, R.R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.*, **25**, 423–462.
- [3] — (1955). A characterization of sufficiency. *Ann. Math. Statist.*, **26**, 286–293.
- [4] — (1955). Statistics and subfields. *Ann. Math. Statist.*, **26**, 490–497.
- [5] — AND LEHMANN, E.L. (1955). Two comments on sufficiency and statistical decision functions. *Ann. Math. Statist.*, **26**, 139–141.
- [6] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā Ser. A*, **15**, 377–380.
- [7] — (1958). On statistics independent of a sufficient statistic. *Sankhyā Ser. A*, **20**, 223–226.
- [8] — (1959). The family of ancillary statistics. *Sankhyā Ser. A*, **21**, 247–256.
- [9] — (1964). Recovery of ancillary information. *Sankhyā Ser. A*, **26**, 3–16.
- [10] — (1975). Statistical information and likelihood. *Sankhyā Ser. A*, **37**, 1–71.
- [11] — (1982). Basu theorems. *Encyclopedia of Statistical Sciences*, **1**, 193–196.
- [12] — AND CHENG, S.C. (1981). A note on sufficiency in coherent models. *Int. J. Math. & Math. Sci.*, **4** (3), 571–582.
- [13] — AND GHOSH, J.K. (1967). Sufficient statistic in sampling from a finite universe. *Proc. 36th Session Internat. Statist. Inst.* (in ISI Bulletin), 850–858.

- [14] — AND PEREIRA, C.A. DE B. (1983). Conditional independence in statistics. *Sankhyã Ser. A*, **45**, 324–337.
- [15] — AND TIWARI, R.C. (1980). *Note on Dirichlet process*. Tallahassee, FSU, 29p. (FSU Statistics Report, M.536)
- [16] BLACKWELL, D. (1947). Copnditional expectation and unbiased sequential estimation. *Ann. Math. Statist.*, **18**, 105–110.
- [17] BREIMAN, L. (1968). *Probability*. Addison Wesley, Massachusetts.
- [18] BURKHOLDER, D.L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.*, **32**, 1191–1200.
- [19] — AND CHOW, Y.S. (1961). Iterates of conditional expectation operators. *Proc. Amer. Math. Soc.*, **12**, 490–495.
- [20] CHENG, S.C. (1979). *A mathematical study of sufficiency and adequacy in statistical theory*. PhD dissertation. The Florida State University.
- [21] CHUNG, K.L. (1974). *A course in probability theory*. 2.ed. Academic Press, New York.
- [22] DAWID, A.P. (1979a). Conditional independence in statistical theory. *J. Royal Statist. Soc.*, **B**, **41**, 1–31.
- [23] — (1979b). Some misleading arguments involving conditional independence. *J. Royal Statist. Soc.*, **B**, **41**, 249–252.
- [24] BRANCO, M. D’ELIA (1991). *Um estudo sobre suficiência, ancilaridade, independência estatística e os teoremas de Basu*. Dissertação de Mestrado. Universidade de São Paulo.
- [25] DOOB, J.L. (1953). *Stochastic processes*. John Wiley and Sons, New York.
- [26] DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika*, **48**, 41–55.
- [27] HALMOS, P. (1970). *Measure theory - graduate texts in mathematics*. New York: Springer.
- [28] — AND SAVAGE, L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.*, **20**, 225–241.
- [29] HASEGAWA, M. AND PERLMAN, M.D. (1974). On the existence of a minimal sufficient subfield. *Ann. Statist.*, **2**, 1049–1055.
- [30] — AND — (1975). Correction to “on the existence of a minimum sufficient subfield”. *Ann. Statist.*, **3**, 1371–1372.

- [31] KOEHN, U. AND THOMAS, D.L. (1975). On statistics independent of sufficient statistic: Basu's Lemma. *The Amer. Statist.*, **29**, 40–42.
- [32] KUSAMA, T. AND YAMADA, S. (1972). On compactness of the statistical structure and sufficiency. *Osaka J. Math.*, **9**, 11–18.
- [33] LEHMANN, E.L. (1959). *Testing statistical hypotheses*. John Wiley and Sons, New York.
- [34] — AND SCHEFFÉ, H. (1950). Completeness, similar regions, and unbiased estimation. Part I. *Sankhyā Ser. A*, **10**, 305–340.
- [35] LOÉVE, M. (1963). *Probability theory*. 3.ed. Van Nostrand, Princeton, New Jersey.
- [36] PATHAK, P.K. (1975). *Note on Basu's Lemma*. New Mexico, UNM, 5p. (Technical Report, 308)
- [37] PAULINO, C.D.M. (1988). *Incomplete categorical data analysis: foundations, methods, and applications*. Doctoral Thesis, University of São Paulo.
- [38] PEREIRA, C.A. DE B. (1980). *Bayesian solutions to some classical problems of statistics*. PhD. Thesis, Florida State University.
- [39] PICCI, G. (1974). *Theory of sufficient statistics and structure analysis of transition probabilities with applications to identifiability*. Padova, LASEB-CNR, 50p. (CCNRLASEB, 06/74)
- [40] IGLESIAS ZUAZOLA, P.L. (1993). *Formas finitas do teorema de de Finetti. A visão preditivista das inferências em populações finitas*. Tese de Doutorado, Universidade de São Paulo.
- [41] PITCHER, T.S. (1959). Sets of measures not admitting necessary and sufficient statistics or subfields. *Ann. Math. Statist.*, **28**, 267–268.
- [42] — (1965). A more general property than domination for sets of probability measures. *Pacif J. Math.*, **15**, 597–611.
- [43] RAO, C.R. (1949). Sufficient statistics and minimum variance unbiased estimates. *Cambridge Phil. Soc.*, **45**, 213–218.