Um estudo sobre suficiência, ancilaridade,independência estatística e os teoremas de Basu

Márcia D'Elia Branco

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Carlos Alberto de Bragança Pereira**

-São Paulo, abril de 1991-

Dedico este trabalho a,

Carlota e Alcides

que possibilitaram minha formação .

E as pessoas que pelo apoio e acolhida, tornaram mais leve minha adaptação a esta cidade

Iara, Chico, Jana, Letícia, Leandro

em especial

a Társis

Agradeço ao meu orientador **Carlos Alberto de Bragança Pereira** e aos colegas **Victor Salinas e Carlos Daniel Paulino** pelo conjunto de idéias que possibilitaram a realização deste trabalho.

Agradeço aos meus amigos

Adilson Simonis

Fábio Machado

Giovani da Silva

Maria Paula Ferreira

pelo apoio em inúmeros momentos, antes e durante o trabalho.

NOTAÇÕES

$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$

$\{0,1\}^2 = \{0,1\} \times \{0,1\} = \{(0,0),(0,1),(1,0),(1,1)\}$

$[a,b]^n = [a,b] \times [a,b] \times \ldots\ldots \times [a,b]$

$\mathbb{P}(\mathcal{X})$: conjunto das partes de $\mathcal{X}$

$\mathcal{B}$: sigma álgebra de Borel do $\mathbb{R}$

$\mathcal{B}_n$: sigma álgebra de Borel do $\mathbb{R}^n$

$\mathcal{B}_{[0,1]}$: sigma álgebra de Borel do $[0,1]$

$$I_A(x) = \begin{cases} 1, & \text{se } x \in A; \\ 0, & \text{caso contrário.} \end{cases} \qquad \text{(função indicadora)}$$

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \quad , \ a > 0 \qquad \text{(função gama)}$$

$$Be(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx \quad , \ a > 0, b > 0 \qquad \text{(função beta)}$$

ÍNDICE

INTRODUÇÃO

Os conceitos de Suficiência, Ancilaridade e Independência Estatística são de grande importância na Teoria Estatística. Nossa intenção, neste trabalho, é apresenta-los numa abordagem matemática da Teoria da Medida, explorando cada um deles de maneira diferente daquela normalmente apresentada nos textos básicos de Teoria Estatística. Posteriormente, pretendemos relaciona-los através de três teoremas, apresentados por D.Basu. Usualmente, somente o primeiro teorema, conhecido como Teorema de Basu, é estudado.

Os conceitos básicos aqui trabalhados foram, a principio, abordados em artigos segundo uma perspectiva Clássica e, nesta visão, construimos o Capítulo 1. Posteriormente, o próprio Basu reescreveu seus teoremas sob um ponto de vista bayesiano, deixando de considerar o parâmetro como um ente fixo mas tratando-o como um objeto aleatório. Nesta perspectiva, a independência estatística é uma forma de independência condicional e os demais conceitos podem ser reconstruídos de modo análogo, utilizando-se a esperança e a independência condicional. Além disso, os Teoremas de Basu têm suas provas simplificadas e a relação dos conceitos de ancilaridade e suficiência com a idéia de informação torna-se mais clara. Nesta abordagem, apresentada no Capítulo 2, trabalha-se num universo matemático mais complexo, o que é possível perceber na unidade 2.1 onde construimos o modelo bayesiano.

Ao longo do trabalho procuramos apresentar, sempre que possível, exemplos. Nosso objetivo foi tornar mais claras as definições e proposições ou colocar em xeque algum senso comum. D.Basu parece gostar muito de contra-exemplos, procuramos apresentar aqui alguns de seus favoritos nesta área. No último capítulo destacamos dois exemplos. O primeiro, mais conhecido, devido sua importância na literatura e o segundo, devido sua utilização em problemas mais aplicados. Além disso, a respeito deste segundo exemplo, destacamos o modo como foi refeita a demonstração do Teorema de Durbin utilizando-se os argumentos estatísticos apresentados neste trabalho.

CAPITULO 1: ABORDAGEM CLÁSSICA

1.1. PRELIMINARES

Sejam $\mathcal{X}$ um conjunto qualquer observável, $\mathcal{A}$ uma sigma álgebra de subconjuntos de $\mathcal{X}$ e $\mathcal{P}$ uma família de medidas de probabilidade em $(\mathcal{X}, \mathcal{A})$. A trinca $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ será denominada neste trabalho de espaço estatístico. Considerando uma estrutura paramétrica, a família $\mathcal{P}$ é indexada por um parâmetro θ, $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Como na maioria das situações esta representação pode ser adotada, vamos utilizá-la.

Definição 1.1 : O modelo $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ é dominado se existe uma medida λ, σ-finita, tal que, se $N \in \mathcal{A}$ e $\lambda(N) = 0$, então $P_\theta(N) = 0$, $\forall P_\theta \in \mathcal{P}$. A notação usada será $\mathcal{P} \ll \lambda$.

Seja $\mathcal{P}^*$ uma outra família de medidas de probabilidade em $(\mathcal{X}, \mathcal{A})$. Se $\forall N \in \mathcal{A}$ tal que $P_\theta(N) = 0$, $\forall P_\theta \in \mathcal{P}$, tivermos que $P_\theta^*(N) = 0$, $\forall P_\theta^* \in \mathcal{P}^*$, dizemos que $\mathcal{P}^* \ll \mathcal{P}$. As duas famílias são equivalentes se $\mathcal{P}^* \ll \mathcal{P}$ e $\mathcal{P} \ll \mathcal{P}^*$. Neste caso a notação é $\mathcal{P} \equiv \mathcal{P}^*$.

Teorema de Radon-Nikodym : Sejam λ e μ duas medidas sigma finitas no espaço mensurável $(\mathcal{X}, \mathcal{A})$, tais que $\mu \ll \lambda$. Nestas condições, existe uma função Borel mensurável $g : \mathcal{X} \to \overline{\mathbb{R}}$, tal que

$$\mu(A) = \int_A g d\lambda \qquad \forall A \in \mathcal{A}.$$

Além disso, g é λ−essencialmente única. A função g é denominada derivada de Radon-Nikodym de μ em relação a λ e denotada por $\frac{d\mu}{d\lambda}$.

Se $\mathcal{P} \ll \lambda$, pelo teorema de Radon-Nikodym , existe $\frac{dP_\theta}{d\lambda}$, também denominada função densidade (f_θ). Usualmente, apenas são denominadas densidades as derivadas obtidas em relação a medida de Lebesgue.

Definição 1.2 : Considere dois espaços mensuráveis $(\mathcal{X}, \mathcal{A})$ e $(\mathcal{Y}, \mathcal{T})$. Uma estatística T é uma função mensurável de $(\mathcal{X}, \mathcal{A})$ em $(\mathcal{Y}, \mathcal{T})$. Se $\mathcal{Y}$ é o $\mathbb{R}^n$, $n \in \mathbb{N}$, e $\mathcal{T}$ a sigma álgebra de Borel, então T é uma variável (ou vetor) aleatória (o).

A toda estatística T está associada uma sigma álgebra no espaço de partida $(\mathcal{X}, \mathcal{A})$ que é gerada pelo conjunto $\{T^{-1}(B) : B \in \mathcal{T}\}$. Vamos denominá-la de sigma álgebra gerada por T e denotá-la por $\sigma(T)$.

Teorema 1.1: Seja $(\mathcal{X}, \mathcal{A}, P_\theta)$ um espaço de probabilidade. Seja f uma função $\mathcal{A}$- mensurável tal que, $\int_{\mathcal{X}} f dP_\theta < \infty$ e $\mathcal{D}$ uma sub-sigma álgebra de $\mathcal{A}$. Nestas condições, existe uma função $f^*, \mathcal{D}$-mensurável, tal que

$$\int_D f dP_\theta = \int_D f^* dP_\theta \qquad \forall D \in \mathcal{D}.$$

A notação usada será $f^* = E[f/\mathcal{D}]$, a esperança condicional de f dado $\mathcal{D}$.

No caso de $\mathcal{D}$ ser gerada por uma estatística T $\left(\mathcal{D} = \sigma(T)\right)$ usaremos, alternativamente, a notação $E[f/T]$. Além disso, se $f = I_A$ (função indicadora),temos

$$E[I_A/T] = P_\theta(A/T), \ \forall A \in \mathcal{A},$$

a probabilidade condicional de A dado T.

Em muitas situações no decorrer deste trabalho, utilizaremos a probabilidade e a esperança condicionais, alternativamente, com o objetivo de simplificar . Entretanto, em todas elas poderiamos ter usado qualquer um dos dois operadores.

1.2.CONCEITOS FUNDAMENTAIS

Independência Estatística

A independência estatística difere da probabilística pelo fato de sua referência ser um espaço estatístico. Neste caso, trabalha-se com uma família de medidas de probabilidade e devemos garantir a independência para todo $P_\theta \in \mathcal{P}$.

Definição 1.3: Um evento $A \in \mathcal{A}$ é estatisticamente independente de uma estatística T se $P_\theta(A/T) = P_\theta(A) \quad \mathcal{P} - q.c.$. Uma estatística Y é estatisticamente independente de uma outra estatística T, se todo $A \in \sigma(Y)$ é estatisticamente independente de T.

Proposição 1.1: Duas estatísticas Y e T são independentes estatisticamente se,e somente se, para todo $A \in \sigma(Y)$ e todo $B \in \sigma(T)$,

$$P_\theta(A \cap B) = P_\theta(A)P_\theta(B) \qquad \mathcal{P} - q.c..$$

Prova: $(\Rightarrow)$ Seja $B \in \sigma(T)$ e $A \in \sigma(Y)$ então,

$$P_\theta(A \cap B) = \int_B I_A dP_\theta = \int_B P_\theta(A/T) dP_\theta = \int_B P_\theta(A) dP_\theta = P_\theta(A)P_\theta(B) \ , \ \forall \theta \in \Theta.$$

$(\Leftarrow)$ Seja $A \in \sigma(Y)$ então,

$$\int_B I_A dP_\theta = P_\theta(A \cap B) = P_\theta(A)P_\theta(B) = \int_B P_\theta(A) dP_\theta \ , \ \forall B \in \sigma(T).$$

Logo $P_\theta(A)$ é uma versão da probabilidade condicional de A dado T . $\Diamond$

Exemplo 1.1: Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um espaço estatístico onde $\mathcal{P} = \{P_1, P_2\}$ e existe um $A \in \mathcal{A}$ tal que $P_1(A) = P_2(A^c) = 1$. Considere $Y = I_A$. Observe que $A = \{x \in \mathcal{X} : Y = 1\}$ e $A^c = \{x \in \mathcal{X} : Y = 0\}$.

Seja $B \in \mathcal{A}$, temos que,

$$P_\theta(B \cap A) = P_\theta(B) I_{\{1\}}(\theta) = P_\theta(B) P_\theta(A) \ , \quad \forall \theta \in \{1, 2\}.$$

Similarmente, temos que

$$P_\theta(B \cap A^c) = P_\theta(B) P_\theta(A^c) \quad \forall \theta \in \{1, 2\}.$$

Como A e A^c geram $\sigma(Y)$, $\forall D \in \sigma(Y)$ e $\forall B \in \mathcal{A}$, temos

$$P_\theta(B \cap D) = P_\theta(B) P_\theta(D) \quad \forall \theta \in \{1, 2\}.$$

Portanto, Y é estatisticamente independente de X .

Exemplo 1.2: Seja $(I\!R, \mathcal{B}, \mathcal{P})$ um espaço estatístico onde $\mathcal{P} = \{P_\theta : \theta \in \mathbb{Z}\}$, tal que $P_\theta(x) = I_{[\theta, \theta+1[}(x)$. Considere $Y(x) = [x]$, onde $[x]$ é o maior inteiro de x.

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

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

Como $\{D_y : y \in \mathbb{Z}\}$ gera $\sigma(Y)$, a igualdade acima é válida $\forall D \in \sigma(Y)$ e $\forall \theta \in \mathbb{Z}$. Portanto, Y e X são estatisticamente independentes.

Suponha agora que observamos $X = 3,2$. Podemos afirmar algo sobre Y?

Claramente $Y = 3$. Mas, X e Y não são independentes ?

A idéia de independência, quando trabalha-se num espaço probabilístico, é de que se X e Y são variáveis independentes, o conhecimento da variável Y não altera a distribuição de probabilidade da variável X e vice versa. O que ocorre no caso da independência estatística, é que o conhecimento de X não altera o de Y dada a medida P_θ, isto é ,supondo θ fixo.

No contexto Bayesiano fica mais fácil entender este conceito . Como naquele contexto θ tem uma distribuição de probabilidade, podemos pensar na probabilidade condicional (em relação a θ) , isto é, X e Y são condicionalmente independentes dado θ .

Exemplo 1.3: Considere $(I\!R, \mathcal{B}, \mathcal{P})$ onde $\mathcal{P} = \{P_\theta : \theta \in \mathbb{Z}\}$ e P_θ é uniforme em $[\theta-1, \theta+1]$.

Sejam X_1 e X_2 duas observações estatisticamente independentes. A princípio, X_1 e X_2 podem assumir quaisquer valores em $I\!R$. Entretanto, se observamos $X_1 = 3$ o que poderiamos dizer sobre X_2 ?

Evidentemente, após o conhecimento do valor de X_1 o conjunto de valores possíveis para X_2 fica reduzido, isto é,

$$\theta - 1 \leq X_1 \leq \theta + 1 \quad \Leftrightarrow \quad X_1 - 1 \leq \theta \leq X_1 + 1 \quad \Leftrightarrow \quad 2 \leq \theta \leq 4.$$

Logo, $1 \leq X_2 \leq 5$ com probabilidade um .

No entanto, se estivessemos trabalhando com o modelo de probabilidade $(I\!R, \mathcal{B}, P_1)$, onde P_1 é uniforme em $[0, 2]$, teriamos que $P_1(X_2 = x \mid X_1) = P_1(X_2 = x)$, $x \in [0, 2]$, isto é, o conhecimento sobre o valor de X_1 não altera a distribuição de X_2.

Estatística Suficiente

O conceito de estatística suficiente é crucial quando discute-se a redução do que se observa, sem prejuízo da informação sobre o parâmetro. Os teoremas de Basu abordam justamente este tipo de redução.

Definição 1.5: T é uma estatística suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ se para toda f, $\mathcal{A}$-mensurável, onde $\int_{\mathcal{X}} f dP_\theta < \infty$, $\forall \theta \in \Theta$, existe uma função f^* (livre de θ) , $\sigma(T)$-mensuravel, tal que,

$$f^*(T) = E_\theta[f \mid T] \ , \quad \forall \theta \in \Theta.$$

Em particular se $f = I_A$, onde $A \in \mathcal{A}$, então

$$f^*(T) = P_\theta(A \mid T) \qquad \forall \theta \in \Theta.$$

Exemplo 1.4: Seja $(I\!R^n, \mathcal{B}_n, \mathcal{P})$ um espaço estatístico onde P_θ é uma medida de probabilidade Uniforme em $[0,\theta]^n$, $\forall \theta \in \Theta$. A função densidade neste caso é

$$f_\theta(x_1, x_2, ..., x_n) = \theta^{-n} I_{[0,\theta]^n}(x_1, x_2, ..., x_n).$$

A estatística $T(x_1, ..., x_n) = \max(x_1, ..., x_n)$ é suficiente para o modelo. De fato, definindo

$$A_t = \{\omega \in \mathcal{X} : T(\omega) \leq t\},$$

$$B_u = \{\omega \in \mathcal{X} : X_1(\omega) > u_1, X_2(\omega) > u_2, ..., X_n(\omega) > u_n\}$$

e trabalhando com a função indicadora temos

$$\int_{A_t} I_{B_u} dP_\theta = P_\theta(B_u \cap A_t) = P_\theta(T \leq t, X_1 > u_1, ..., X_n > u_n) =$$

$$= P_\theta(u_1 \leq X_1 \leq t, u_2 \leq X_2 \leq t, ..., u_n \leq X_n \leq t) = \prod_{i=1}^{n} (t - u_i)\theta^{-n} =$$

$$= \prod_{i=1}^{n} (t - u_i) t^{-n} P_\theta(T \leq t) = \int_{A_t} h dP_\theta,$$

onde $h = \prod_{i=1}^{n} (t - u_i) t^{-n}$ é uma função livre de θ.

Como $\{A_t : t \in I\!R\}$ e $\{B_u : u \in I\!R^n\}$ são geradores de $\sigma(T)$ e $I\!R^n$ respectivamente, temos que, para todo $A \in \sigma(T)$ e $B \in I\!R^n$,

$$\int_A I_B dP_\theta = \int_A h dP_\theta \qquad \forall \theta \in \Theta.$$

Como h é uma função $\sigma(T)$-mensuravel e livre de θ , T é suficiente.

Em geral não é facil verificar a suficiência pela definição. A seguir apresentamos um teorema, conhecido como Teorema da Fatoração (ou Halmos-Savage), que resolve este problema em muitos casos.

Teorema da Fatoração : Se o modelo estatístico $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ é dominado pela medida λ, uma condição necessária e suficiente para T ser uma estatística suficiente é que a derivada de Radon-Nikodym de P_θ com respeito a λ , $dP_\theta/d\lambda = f_\theta$, tenha a seguinte versão

$$f_\theta = g_\theta h,$$

onde g_θ é uma função $\sigma(T)$-mensurável e h é uma função livre de θ.

Prova: $(\Rightarrow)$ Como o modelo é dominado, existe um conjunto enumerável $\mathcal{P}_0 = \{P_1, P_2, P_3, ...\}$ tal que $\mathcal{P} \equiv \mathcal{P}_0$ (apêndice C) .

Seja λ_0 uma probabilidade em $(\mathcal{X}, \mathcal{A})$, tal que

$$\lambda_0(A) = \sum_{i=1}^{\infty} a_i P_i(A) \quad \forall A \in \mathcal{A},$$

onde $a_i = \frac{1}{2^i}$, $\quad i = 1, 2, 3, ...$ Então, $\lambda_0 \equiv \mathcal{P}_0 \ \Rightarrow \ \lambda_0 \equiv \mathcal{P}$.

Além disso, usando o fato de T ser suficiente,$\forall A \in \mathcal{A}$ e $\forall B \in \sigma(T)$, temos

$$\lambda_0(A \cap B) = \sum_{i=1}^{\infty} a_i P_i(A \cap B) = \sum_{i=1}^{\infty} a_i \int_B f(T) dP_i = \int_B f(T) d\lambda_0.$$

Portanto, $f(T)$ é também uma versão da probabilidade condicional para λ_0.

Vamos mostrar agora que $\frac{dP_\theta}{dP_{\lambda_0}} = g_\theta$ é uma função $\sigma(T)$-mensurável , $\forall P_\theta \in \mathcal{P}$.

Seja $A \in \mathcal{A}$ e $P_\theta \in \mathcal{P}$, então

$$\int_A g_\theta d\lambda_0 = P_\theta(A) = \int_{\mathcal{X}} f(T) dP_\theta = \int_{\mathcal{X}} f(T) g_\theta d\lambda_0 =$$

$$= \int_{\mathcal{X}} f(T) E_{\lambda_0}(g_\theta \mid T) d\lambda_0 = \int_{\mathcal{X}} E_{\lambda_0}(I_A g_\theta \mid T) d\lambda_0 =$$

$$= \int_A E_{\lambda_0}(g_\theta \mid T) d\lambda_0 \qquad \forall A \in \mathcal{A}.$$

Logo, $g_\theta = E_{\lambda_0}(g_\theta / T) \ \ \lambda_0 - q.c.$. Portanto, g_θ é $\sigma(T)$- mensurável.

Podemos escrever $f_\theta = \frac{dP_\theta}{d\lambda} = \frac{dP_\theta}{d\lambda_0} \frac{d\lambda_0}{d\lambda} = g_\theta h$, com h livre de θ.

$(\Leftarrow)$ Por hipótese $f_\theta = g_\theta h$.Usando a medida λ_0 definida anteriormente, podemos escrever $g_\theta = \frac{dP_\theta}{d\lambda_0}$ que por hipótese é $\sigma(T)$-mensurável.

Fixo $E \in \mathcal{A}$ e $P_\theta \in \mathcal{P}$, seja ν uma medida de probabilidade tal que

$$\nu(A) = P_\theta(A \cap E) \quad \forall A \in \mathcal{A}. \tag{1.1}$$

Então,

$$\int_A d\nu = \int_A I_E dP_\theta = \int_A P_\theta(E \mid T) dP_\theta =$$

$$= \int_A P_\theta(E \mid T) g_\theta(T) d\lambda \quad \forall A \in \mathcal{A}\,. \tag{1.2}$$

Por outro lado, de (1.1) temos que

$$\int_A d\nu = \int_A I_E dP_\theta = \int_A I_E g_\theta(T) d\lambda_0 =$$

$$= \int_A \lambda_0(E \mid T) g_\theta(T) d\lambda_0 \quad \forall A \in \mathcal{A}\,. \tag{1.3}$$

De (1.2) e (1.3) temos que

$$P_\theta(E \mid T) g_\theta(T) = \lambda_0(E \mid T) g_\theta(T) \quad \lambda_0 - q.c..$$

Como $g_\theta(T) = \frac{dP_\theta}{\lambda_0}$, temos que $P_\theta(\{x : g_\theta o T(x) = 0\}) = 0$, isto é, $g_\theta(T) \neq 0 \quad P_\theta - q.c..$ Logo, $P_\theta(T) = \lambda_0(E \mid T) \quad P_\theta - q.c..$

Como P_θ é arbitrário, a igualdade acima vale $\forall P_\theta \in \mathcal{P}$. Portanto, $P_\theta(E \mid T)$ é constante em relação a θ. Logo, T é suficiente. $\diamondsuit$

Um teorema análogo foi apresentado por Basu e Ghosh (1967), substituindo a condição de modelo dominado pela de modelo discreto, entendendo como modelo discreto aquele que satisfaz as seguintes condições:

(a) todo $P_\theta \in \mathcal{P}$ é uma medida de probabilidade discreta;

(b) $\mathcal{A}$ é o conjunto das partes de $\mathcal{X}$;

(c) o único conjunto P_θ-nulo , $\forall \theta \in \Theta$, é o conjunto vazio .

No Exemplo 1.4 o modelo é dominado e podemos escrever

$$f_\theta(x_1, x_2, ..., x_n) = \theta^{-n} I_{(-\infty,\theta)}(T) I_{(0,\infty)}(S),$$

onde $S(x_1, x_2, ..., x_n) = \min(x_1, x_2, ..., x_n)$.

Identificando $g_\theta = \theta^{-n} I_{(-\infty,\theta)}(T)$ e $h = I_{(0,\infty)}(S)$, concluimos pela suficiência de T.

Proposição 1.2 : Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um modelo dominado. Sejam T_1 e T_2 duas estatísticas tal que $\sigma(T_1) \subset \sigma(T_2)$ $\mathcal{P}-q.c.$, isto é , T_2 é essencialmente função de T_1. Se T_1 é suficiente, então T_2 também é suficiente para o modelo.

Utilizando o Teorema da Fatoração, a prova deste resultado é imediata.

Exemplo 1.5: Seja $(I\!R, \mathcal{B}, \mathcal{P})$ um modelo estatístico onde

$$\mathcal{P} = \{P : P(A) = P(-A), \forall A \in \mathcal{B}\}.$$

Se $\mathcal{B}_0$ é a σ-álgebra gerada pela estatística $Y(x) = x^2$ então,

$$\mathcal{B}_0 = \{A \in \mathcal{B} : A = -A\},$$

que é a σ-álgebra dos borelianos simétricos.

Se f é uma função $\mathcal{B}$-mensurável, então $g(x) = \frac{[f(x)+f(-x)]}{2}$ é $\mathcal{B}_0$-mensurável, e

$$\int_A f dP = \int_A g dP \ , \quad \forall A \in \mathcal{B}_0 \ , \quad \forall P \in \mathcal{P}.$$

Logo, $\mathcal{B}_0$ é suficiente para o modelo.

Seja $S \subset I\!R, S \notin \mathcal{B}$ e

$$\mathcal{B}_1 = \{A \cup B_0 : A \subset S, A \in \mathcal{B} \text{ e } B_0 \in \mathcal{B}_0\},$$

então $\mathcal{B}_1$ é uma σ-álgebra tal que, $\mathcal{B}_0 \subset \mathcal{B}_1 \subset \mathcal{B}$.

Vamos supor que $\mathcal{B}_1$ é suficiente. Então, $\forall f$ $\mathcal{B}$-mensurável, existe uma g ,$\mathcal{B}_1$- mensurável, tal que, $\forall B \in \mathcal{B}_1$

$$\int_B f dP = \int_B g dP \qquad \forall P \in \mathcal{P}.$$

Seja $x \in S$ e $B = \{x\}$, então

$$\int_{\{x\}} f dP = \int_{\{x\}} g dP \qquad \Leftrightarrow$$

$$\Leftrightarrow \quad f(x)P(x) = g(x)P(x) \quad \forall P \in \mathcal{P}$$

Agora, seja $x \in S^c$ e considerando $B = \{-x, x\}$, temos que $B \in \mathcal{B}_0 \Rightarrow B \in \mathcal{B}_1$. Portanto,

$$\int_{\{-x,x\}} f dP = \int_{\{-x,x\}} g dP \quad \Leftrightarrow$$
$$\Leftrightarrow \quad [f(x) + f(-x)]P(x) = 2g(x)P(x).$$

Como existe $P \in \mathcal{P}$ tal que $P(x) > 0, \ \forall x$, temos que

$$g(x) = \frac{f(x) + f(-x)}{2} I_{S^c}(x) + f(x) I_S(x).$$

Usando $f(x) = I_{(0.\infty)}(x) - I_{(-\infty,0)}(x)$, temos que $g(x) = 0$ se $x \in S^c$ e $g(x) \neq 0$ se $x \in S$. Logo, $g^{-1}(0) = S^c$. Mas $S^c \notin \mathcal{B}$ e portanto $S^c \notin \mathcal{B}_1$. Isto contradiz a hipótese de que g é $\mathcal{B}_1$-mensurável. Portanto, $\mathcal{B}_1$ não é suficiente.

Este exemplo mostra que é possível ter uma sigma álgebra que não é suficiente mas contenha uma suficiente . Este fato pode ocorrer quando a condição de dominação não esta satisfeita. A seguir vamos discutir a dominação do modelo.

Suponha que $\mathcal{P}$ seja dominado. Então, existe uma medida μ sigma finita tal que $\mathcal{P} \ll \mu$. Em particular, considere as medidas em $\mathcal{P}$ do tipo $P_j(j) = P_j(-j) = 1/2$ onde $j \in I\!R^+$. Temos que $\mu(j) > 0, \ \forall j \in I\!R$. Isto implica que μ não é sigma finita.

Com efeito, seja I um conjunto enumerável de indices e $(A_i)_{i \in I}$ uma partição de $I\!R$ com $I\!R = \bigcup_{i \in I} A_i$, então existe $k \in I$ tal que A_k é não enumerável. Mas,

$$\mu(A_k) = \sum_{j \in A_k} \mu(j) = \infty.$$

Como a partição de $I\!R$ é arbitrária, concluimos que μ não é σ-finita e consequentemente o modelo não é dominado. $\diamondsuit$

Proposição 1.3: Sejam T_1 e T_2 duas estatísticas suficientes tal que $\sigma(T_1) \subset \sigma(T_2)$ $\mathcal{P} - q.c.$. Se T_1 é suficiente para $(\mathcal{X}, \sigma(T_2), \mathcal{P})$ e T_2 é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, então T_1 é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$.

Prova : Seja f uma função $\mathcal{A}$-mensurável. Pela suficiência de T_2 existe uma função f^0 , $\sigma(T_2)$-mensurável, tal que,

$$f^0 = E_\theta[f \mid T_2] \quad \forall \theta \in \Theta.$$

Da suficiência de T_1 segue que existe uma f^* , $\sigma(T_1)$- mensurável, tal que

$$f^* = E_\theta[f^0 \mid T_1] \quad \forall \theta \in \Theta.$$

Mas, como $\sigma(T_1) \subset \sigma(T_2)$,

$$E_\theta[f \mid T_1] = E_\theta[E_\theta[f \mid T_2] \mid T_1].$$

Utilizando as igualdades anteriores temos que,

$$E_\theta[f \mid T_1] = E_\theta[f^0 \mid T_1] = f^*.$$

Logo, T_1 é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. $\diamondsuit$

Como já foi dito, o objetivo do estudo da estatística suficiente é a redução das observações sem perda de informação. Assim, poderiamos nos interessar pela redução máxima. Dois conceitos relacionados a este tipo de redução são os de suficiência minima e minimal. Prova-se que os dois conceitos são equivalentes nos modelos dominados e discreto. Contudo, nem sempre a equivalência é satisfeita. A estatística suficiente mínima sempre existe e não é necessariamente única. Em relação a estatística suficiente minimal, caso ela exista, ela é essencialmente única e também é mínima.

Definição 1.6: Uma estatística suficiente T_0 é suficiente minimal, se para toda outra estatística suficiente T, temos que $\sigma(T_0) \subset \sigma(T)$ $\mathcal{P} - q.c.$.

Definição 1.7: Uma estatística suficiente T_0 é suficiente mínima, se não existe outra suficiente T, tal que $\sigma(T) \subset \sigma(T_0)$ $\mathcal{P} - q.c.$.

Para o modelo dominado a existência de uma estatística suficiente minimal foi estudada por Lehmann e Scheffé (1950) e Bahadur (1955). Para o modelo discreto , definido anteriormente, foram feitos estudos por Basu e Ghosh (1967) . Posteriormente, Hasegawa e Perlman (1974) introduziram o conceito de modelos coerentes, um modelo estatístico mais geral que inclui os casos dominado e discreto. Entretanto, o modelo coerente não se limita a estes dois casos. Basu e Cheng (1979) apresentam um estudo organizado sobre tais modelos, sua relação com a estatística suficiente e, um exemplo de um modelo coerente que não é discreto nem dominado. Para o caso dos modelos coerentes a existência da estatística suficiente minimal está provada. Contudo, a condição do modelo ser coerente é uma condição suficiente mas não necessária para a existência da suficiente minimal.

Estes conceitos não serão explorados ao longo deste trabalho. Aqui trabalharemos , na maioria dos exemplos, com modelos dominados, onde nada patológico ocorre. Outro conceito de interesse, que pode facilitar o trabalho no caso dominado, é o de suficiência pareada.

Definição 1.8: Uma estatística T é suficiente pareada para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ se, para todo $A \in \mathcal{A}$ e todo par $\theta_1, \theta_2 \in \Theta$, existe uma função f (livre de θ) , $\sigma(T)$-mensurável, tal que

$$f(T) = P_{\theta_i}(A \mid T) \qquad i = 1, 2.$$

Ao invés de nos referirmos a uma estatística como sendo suficiente (ou suficiente pareada) poderiamos nos referir , com esses termos, a sua sigma álgebra. Deste modo, poderiamos ter definido sigmas álgebras suficientes (minimas, minimal e pareada). No próximo exemplo trabalharemos com sigmas álgebras, sem nos referirmos a alguma estatística.

Exemplo 1.6 : Seja $(I\!R, \mathcal{A}, \mathcal{P})$ um modelo estatístico onde $\mathcal{A}$ é a σ-álgebra de conjuntos simétricos que contêm pelo menos um conjunto não enumerável e,$\mathcal{P}$ é a família de probabilidade discreta simétrica, isto é, aquela que concentra a massa de probabilidade em apenas dois pontos $-a$ e a , $a \in I\!R$.

Seja $\mathcal{C} \subset \mathcal{A}$ a $\sigma-$álgebra dos conjuntos simétricos enumeráveis e coenumeráveis (cujo complementar é enumerável). Vamos mostrar que $\mathcal{C}$ é suficiente pareada mas, não é suficiente.

Seja $A \in \mathcal{A}$, P_a, $P_b \in \mathcal{P}$ onde $a \neq b$ com, $P_a(\{-a, a\}) = 1$ e $P_b(\{-b, b\}) = 1$.

Defina $C = A \cap \{-a, -b, a, b, \}$. Como A é simétrico, C também é simétrico e $C \in \mathcal{C}$. Além disso, $A = C \;\; P_i - q.c. \;\; i = a, b$.

Se $D \in \mathcal{C}$, então

$$\int_D I_A dP_i = \int_D I_C dP_i, \qquad i = a, b.$$

Como a e b são valores arbitrários, a igualdade acima é válida para todo $P_a, P_b \in \mathcal{P}$. Logo, $\mathcal{C}$ é suficiente pareada para o modelo.

Vamos mostrar agora que $\mathcal{C}$ não pode ser suficiente. Para isto observe que, $\mathcal{C} \neq \mathcal{A}$ $\mathcal{P} - q.c.$. Mostraremos que nenhuma outra σ-álgebra menor que $\mathcal{A}$ pode ser suficiente para o modelo.

Suponha que exista uma σ-álgebra $\mathcal{D} \subset \mathcal{A}$ que seja suficiente. Como $\mathcal{D} \neq \mathcal{A}$, então $\mathcal{A} - \mathcal{D} \neq \emptyset$. Seja $A \in \mathcal{A} - \mathcal{D}$ e $P_a \in \mathcal{P}$. Pela suficiência de $\mathcal{D}$, existe h , $\mathcal{D}$ -mensurável, tal que

$$\int_{\mathbb{R}} I_A dP_a = \int_{\mathbb{R}} h dP_a.$$

Mas,

$$\int_{\mathbb{R}} I_A dP_a = \frac{I_A(-a) + I_A(a)}{2} = I_A(a).$$

E por outro lado,

$$\int_{\mathbb{R}} h dP_a = \frac{h(-a) + h(a)}{2} = h(a).$$

Portanto,

$$I_A(a) = h(a) \qquad P_a - q.c. \quad \forall a \in \mathbb{R}.$$

Como h é $\mathcal{D}$-mensurável, então $A \in \mathcal{D}$, o que é uma contradição.

Lema 1.1: Uma condição necessária e suficiente para T ser uma estatística suficiente pareada para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ é que, para todo par $\theta_1, \theta_2 \in \Theta$, $\frac{dP_{\theta_1}}{d(P_{\theta_1} + P_{\theta_2})}$ seja $\sigma(T)-$mensurável.

Prova: Fixados $\theta_1, \theta_2 \in \Theta$, vamos considerar a seguinte família de probabilidades em $(\mathcal{X}, \mathcal{A})$,

$$\mathcal{P}^* = \{P_{\theta_1}, P_{\theta_2}\}.$$

É facil verificar que $\mathcal{P}^*$ é dominada pela medida $\lambda_0 = P_{\theta_1} + P_{\theta_2}$.

Usando os mesmos argumentos da prova do Teorema da Fatoração temos que, $\frac{dP_{\theta_1}}{d\lambda_0}$ é $\sigma(T)$-mensurável, se e somente se, T é suficiente para $\mathcal{P}^*$. Como isto é válido para todo par θ_1, θ_2, o lema esta provado. $\Diamond$

Lema 1.2: Se T é suficiente pareada para a família finita $\mathcal{P} = \{P_{\theta_0}, ..., P_{\theta_n}\}$, então

$$\frac{dP_{\theta_0}}{d\Big(\sum\limits_{i=0}^{n} P_{\theta_i}\Big)} \quad \text{é } \sigma(T) - \text{ mensurável .}$$

Prova: Quando $n = 1$ estamos no caso do Lema 1.1 .

Parte I - Vamos mostrar que a afirmação é válida para $n = 2$.

Seja $\mathcal{P} = \{P_{\theta_0}, P_{\theta_1}, P_{\theta_2}\}$. Como T é suficiente pareado, pelo lema 1.1 ,

$$f_1 = \frac{dP_{\theta_0}}{d(P_{\theta_0} + P_{\theta_1})} \quad \text{e} \quad f_2 = \frac{dP_{\theta_0}}{d(P_{\theta_0} + P_{\theta_2})}$$

são funções $\sigma(T)$-mensuráveis. Temos então que,

$$(f_1 + f_2 - f_1 f_2)dP_{\theta_0} = f_1 f_2 d(P_{\theta_0} + P_{\theta_1} + P_{\theta_2}).$$

Como $0 \leq f_1 \leq 1, 0 \leq f_2 \leq 1$ então $f_1 + f_2 - f_1 f_2 = 0 \;\Leftrightarrow\; f_1 = f_2 = 0$. Mas, $P_{\theta_0}\Big(\{x \in \mathcal{X} : f_1(x) = f_2(x) = 0\}\Big) = 0$. Então definimos

$$f = \frac{f_1 f_2}{f_1 + f_2 - f_1 f_2} \quad P_{\theta_0} - q.c.,$$

e concluimos que f é $\sigma(T)$-mensurável.

Por outro lado,

$$f = \frac{dP_{\theta_0}}{d(P_{\theta_0} + P_{\theta_1} + P_{\theta_2})}) \quad P_{\theta_0} - q.c.$$

Parte II - Vamos mostrar, por indução, que a propriedade

$$\frac{dP_{\theta_0}}{d(\sum_{i=0}^{n} P_{\theta_i})} \quad \text{é } \sigma(T)\text{-mensurável}$$

é válida , $\forall n \in I\!N$.

Considerando que

$$\frac{dP_{\theta_0}}{d(\sum_{i=0}^{n+1} P_{\theta_i})} = \frac{dP_{\theta_0}}{d(P_{\theta_0} + \sum_{i=1}^{n} P_{\theta_i} + P_{\theta_{n+1}})} = \frac{dP_{\theta_0}}{d(P_{\theta_0} + \mu + P_{\theta_{n+1}})}.$$

Pela suficiência pareada de T, temos que $\frac{dP_{\theta_0}}{d(P_{\theta_0}+P_{\theta_{n+1}})}$ é $\sigma(T)$−mensurável e, pela hipótese de indução, $\frac{dP_{\theta_0}}{d(P_{\theta_0}+\mu)}$ é $\sigma(T)$- mensurável . Portanto, T é suficiente pareada para $\{P_{\theta_0}, \mu, P_{\theta_{n+1}}\}$ e, pela parte I, concluimos a prova. $\diamond$

Lema 1.3: Seja T suficiente pareada para a família infinita e enumerável $\mathcal{P} = \{P_{\theta_1}, P_{\theta_1}, ...\}$. Se $\mu(A) = \sum_{i=1}^{\infty} a_i P_{\theta_i}(A)$, $\forall A \in \mathcal{A}$, onde $\sum_{i=1}^{\infty} a_i = 1$, então,

(i) $\lim_{n\to\infty} \frac{dP_{\theta_0}}{d(\sum_{i=0}^{n} a_i P_{\theta_i})} = \frac{dP_{\theta_0}}{d\mu} \quad \mu - q.c.$

(ii) $\frac{dP_{\theta_0}}{d\mu}$ é $\sigma(T)$-mensurável.

Prova: Temos que, $\forall n$,

$$\frac{dP_{\theta_0}}{d\mu} = \frac{dP_{\theta_0}}{d(\sum_{i=0}^{n} a_i P_{\theta_i})} \frac{d(\sum_{i=0}^{n} a_i P_{\theta_i})}{d\mu}.$$

Fazendo $n \to \infty$, temos que $\frac{d(\sum_{i=0}^{n} a_i P_{\theta_i})}{d\mu} \to 1$. Portanto vale (i).

Para a validade de (ii) basta observar que $\frac{dP_{\theta_0}}{d(\sum_{i=0}^{n} a_i P_{\theta_i})}$ é $\sigma(T)$- mensurável, então $\forall n$, $\frac{dP_{\theta_0}}{d\mu}$ é $\sigma(T)$-mensurável. $\diamond$

Proposição 1.4: Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um modelo dominado então, a estatística T é suficiente se, e somente se, ela é suficiente pareada.

Prova:

$(\Rightarrow)$ Imediata

$(\Leftarrow)$ Suponha que T é suficiente pareada. Como o modelo é dominado existe um conjunto $\mathcal{P}_0 = \{P_1, P_2, ...\}$ enumerável contido em $\mathcal{P}$, tal que $\mathcal{P}_0 \equiv \mathcal{P}$ (apêndice C).

Seja P_θ uma medida qualquer em $\mathcal{P}$. Defina a medida μ, onde $\sum_{i=1}^{\infty} a_i = 1$, como

$$\mu(E) = \sum_{i=1}^{\infty} a_i P_i(E), \quad \forall E \in \mathcal{A}.$$

Pela suficiência pareada de T e pelo Lema 1.3, $\frac{dP_0}{d(P_0+\mu)}$ é $\sigma(T)$ -mensurável. Além disso,

$$\frac{dP_0}{d\mu} = \frac{dP_0}{d(P_0+\mu)} \frac{d(P_0+\mu)}{d\mu} =$$

$$= \frac{dP_0}{d(P_0+\mu)} \left[\frac{d\mu}{d(P_0+\mu)} \right]^{-1} = \frac{dP_0}{d(P_0+\mu)} \left[1 - \frac{dP_0}{d(P_0+\mu)} \right]^{-1}.$$

Logo, $\frac{dP_0}{d\mu}$ é $\sigma(T)$-mensurável.

Utilizando o teorema da Fatoração, concluimos que T é suficiente. $\diamond$

Este último resultado simplifica bastante o trabalho no caso do modelo dominado. Neste caso, basta mostrar a suficiência pareada para garantir a suficiência.

Estatística Ancilar

Na seção anterior discutimos um conceito relacionado com informação contida na amostra a respeito do parâmetro. Nesta seção apresentamos um conceito que, a princípio, foi considerado relacionado a " não-informação " .

Definição 1.9: Um evento $A \in \mathcal{A}$ é ancilar em relação ao modelo estatístico se, $P_\theta(A) = P(A)$,$\forall \theta \in \Theta$, onde P é uma função $\mathcal{A}$-mensurável. Uma estatística Y é ancilar, para o modelo estatístico, se todo $A \in \sigma(Y)$ é ancilar.

Observe que se a familia $\mathcal{P}$ é dominada, então a estatística é ancilar se, e somente se, sua função densidade é livre de θ. Este fato é facilmente verificado utilizando-se o Teorema de Radon-Nikodym .

Exemplo 1.7 : Seja $(\mathbb{R}^2, \mathcal{B}_2, \mathcal{P})$ um modelo estatístico onde $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}^+\}$ é dominado pela medida de Lebesgue e sua densidade é

$$f_\theta(x,y) = e^{-\theta x - y/\theta} \qquad x, y > 0.$$

Seja Z a estatística $Z(x,y) = x.y$. Considere $A_z \in \sigma(Z)$ tal que

$$A_z = \{(x,y) \in \mathbb{R}^2 : xy \leq z\} = \{(x,y) \in \mathbb{R}^2 : y \leq z/x\}.$$

Temos que,

$$P_\theta(A_z) = \int_{A_z} f_\theta dP_\theta = \int_0^\infty \int_0^{z/x} f_\theta(x,y)dydx.$$

Fazendo substituição de variáveis, obtemos

$$P_\theta(A_z) = \int_0^z \int_0^\infty \frac{1}{x} f_\theta(x, v/x)dxdv.$$

A integral interna é a função densidade de Z a qual denotaremos por g_Z. Então,

$$g_Z(v) = \int_0^\infty \frac{1}{x} e^{-\theta x - v/\theta x} dx = \int_0^\infty \frac{1}{w} e^{-w - v/w} dw.$$

Portanto, Z é ancilar.

O estimador de máxima verossimilhança para θ é a estatística $T(x,y) = (y/x)^{1/2}$. Utilizando-se o Teorema da Fatoração verifica-se que T não é suficiente para o modelo, mas (T, Z) é suficiente.

Embora nosso objetivo não seja aprofundar questões sobre o conceito de informação, abriremos aqui um parêntese para entender melhor o exemplo em termos estatísticos.

Informação, segundo Fisher, é a função

$$I(\theta) = E_\theta[\frac{-\delta^2}{\delta\theta^2} log\ f_\theta(X)]$$

onde $f_\theta(X)$ é a função de verossimilhança.

A informação de Fisher para uma estatística T é definida pela expressão

$$J(\theta) = E_\theta[\frac{-\delta^2}{\delta\theta^2} log\ h_\theta(T)]$$

onde $h_\theta(T)$ é a função densidade de T.

Uma estatística tem a mesma informação de Fisher da amostra $[I(\theta) = J(\theta)]$ se, e somente se, ela é suficiente (Basu, 1975) . Isto caracteriza, em termos de Informação de Fisher, a suficiência. Por outro lado, se Y é uma estatística ancilar sua informação de Fisher é zero.

Em termos de σ-álgebra temos, no Exemplo 1.7, que $\sigma(T)$ não é suficiente e portanto, sua informação $J(\theta)$ é menor que $I(\theta)$ (a informação máxima de Fisher). Entretanto, quando acrescentamos elementos ancilares (com informação de Fisher zero), resulta que $\sigma(T) \vee \sigma(Z)$ tem informação de Fisher igual a $I(\theta)$. Este fato nos faz questionar a idéia de que uma sigma álgebra ancilar não contêm informação relevante para o parâmetro. Portanto, nem sempre podemos dispensar as estatísticas ancilares quando estamos

interessados em estimar determinado parâmetro. Muitas vezes as estatísticas ancilares são de grande utilidade.

A seguir apresentamos alguns conceitos relacionados com ancilaridade.

Definição 1.10: A estatística ancilar Y é ancilar maximal, se não existe outra estatística ancilar Y^*, tal que $\sigma(Y) \subset \sigma(Y^*)$.

Definição 1.11: Uma estatística S é informativa, se existe pelo menos um conjunto $A \in \sigma(S)$, tal que A não é ancilar.

Definição 1.12: Dois conjuntos ancilares A e B se adaptam, se $A \cap B$ também é ancilar e escrevemos $A \sim B$. Caso contrário, escrevemos $A \not\sim B$. Note que $A \sim B \Rightarrow A \sim B^c$ pois,

$$P_\theta(A \cap B) + P_\theta(A \cap B^c) = P_\theta(A).$$

Exemplo 1.8: Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ onde $\mathcal{X} = \{0,1\}^2, \mathcal{A} = I\!P(\mathcal{X})$ e $\mathcal{P} = \{P_\theta : \theta \in [0,1]\}$, tal que $P_\theta(x,y)$ é descrita na tabela abaixo.

TABELA 1

	Y		
X	0	1	Total
0	$\theta/2$	$(1-\theta)/2$	1/2
1	$(1-\theta)/2$	$\theta/2$	1/2
Total	1/2	1/2	1

Se $A = \{(0.0);(0,1)\}$ e $B = \{(0,1);(1,1)\}$, então $P_\theta(A) = P_\theta(B) = 1/2$, $\forall \theta \in [0,1]$. Portanto, A e B são ancilares. Entretanto, $P_\theta(A \cap B) = \theta/2, \quad \forall \theta \in [0,1]$. Logo, A e B não se adaptam.

Proposição 1.5: Se A é um evento ancilar que não pertence a $\sigma(Y)$, onde Y é uma ancilar maximal, então existe pelo menos um elemento $B \in \sigma(Y)$ tal que $A \not\sim B$.

Prova : Seja $\mathcal{M}$ uma σ-álgebra ancilar maximal e A um conjunto ancilar tal que $A \notin \mathcal{M}$. Vamos supor que $A \sim M$ para todo $M \in \mathcal{M}$. Considere

$$\mathcal{M}^* = \{(A \cap X) \cup (A^c \cap Y) : X, Y \in \mathcal{M}\}.$$

Observe que $\mathcal{M} \subset \mathcal{M}^*$ e $\mathcal{M} \neq \mathcal{M}^*$. Além disso,

$$P_\theta[(A \cap X) \cup (A^c \cap Y)] = P_\theta(A \cap X) + P_\theta(A^c \cap Y).$$

Por hipótese os dois termos à direita da equação são livres de θ, e portanto como isto é válido para todo elemento de $\mathcal{M}^*$, concluimos que $\mathcal{M}^*$ é uma σ-álgebra ancilar. Mas $\mathcal{M} \subset \mathcal{M}^*$, o que contradiz a maximalidade de $\mathcal{M}$. $\diamond$

Corolário 1.1: Seja Y uma estatística ancilar maximal e S uma estatística ancilar tal que $\sigma(S) \not\subset \sigma(Y)$, então $\sigma(S) \vee \sigma(Y)$ é informativa. Isto é, (Y, S) é informativa.

Exemplo 1.9: Seja $(\mathbb{R}^2, \mathcal{B}_2, \mathcal{P})$ um espaço estatístico onde $\mathcal{P}$ é a família da Normal bivariada com vetor de médias $\binom{0}{0}$ e matriz de covariância $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.Sejam X, Y duas estatísticas, tais que $X(x, y) = x$ e $Y(x, y) = y$. Cada uma delas tem medida de probabilidade $N(0, 1)$ e portanto são ancilares. No entanto, (X, Y) é suficiente para o modelo e portanto informativa.

Família Completa

O conceito de família completa, bem como os conceitos de família conectada e conjunto separador que serão abordados a seguir, são conceitos auxiliares para estabelecermos os Teoremas de Basu. A condição de que a família seja completa é a condição exigida para que o espaço estatístico possa ser particionado , de forma independente, numa parte informativa (suficiente) e outra não informativa (ancilar) .

Optamos por falar em família Completa e não estatística Completa, pelo fato deste conceito ser uma propriedade da família de medidas de Probabilidade.

Definição 1.13: Uma família de medidas de probabilidade $\mathcal{P}$ é (limitadamente) completa se, para toda função real g (essencialmente limitada) ,

$$\int_{\mathcal{X}} g dP_\theta = 0 \ , \quad \forall \theta \in \Theta, \quad \Rightarrow \quad g = 0 \ \ \mathcal{P} - q.c.$$

Uma estatística T é (limitadamente) completa se a correspondente família de medidas induzidas por T é (limitadamente) completa.

O fato da família ser completa implica que ela é limitadamente completa. Podemos entretanto, ter uma família limitadamente completa que não é completa. Vejamos o próximo exemplo.

Exemplo 1.10: Seja $(I\!R, \mathcal{B}, \mathcal{P})$ um espaço estatístico onde $\mathcal{P} = \{P_\theta : \theta \in (0,1)\}$ e cada P_θ tem a seguinte função densidade,

$$f_\theta(x) = (1-\theta)^2 \theta^x I_{I\!N}(x) + \theta I_{\{-1\}}(x).$$

Seja g uma função Borel mensurável tal que $\int_{I\!R} g P_\theta = 0, \quad \forall \theta \in (0,1)$. Então,

$$g(-1)\theta + \sum_{j=0}^{\infty} g(j)(1-\theta)^2\theta^j = 0 \quad \Rightarrow$$

$$\Rightarrow \quad (1-\theta)^2 \sum_{j=0}^{\infty} g(j)\theta^j = -\theta g(-1) \quad \Rightarrow$$

$$\Rightarrow \quad \sum_{j=0}^{\infty} g(j)\theta^j = -g(-1)\theta(1-\theta)^{-2}.$$

Como $0 < \theta < 1$, temos que $\theta(1-\theta)^{-2} = \sum_{j=0}^{\infty} j\theta^j$. Portanto,

$$\sum_{j=0}^{\infty} g(j)\theta^j = \sum_{j=0}^{\infty} -g(-1)j\theta^j.$$

Então,

$$g(j) = -jg(-1) \qquad j = 0, 1, 2, \ldots\ldots \tag{1.4}$$

Caso I: Suponha g limitada, então $g(j) = g(-1) = 0$ para $j = 0, 1, 2, \ldots..$

Caso II: Supoha que a condição de g ser limitada não é imposta. Considere g_0, tal que

$$g_0(-1) = 1, \quad g_0(0) = 0 \quad e \quad g_0(j) = -j \quad j = 1, 2, 3, \ldots\ldots$$

Então, g_0 satisfaz (1.4) . Logo, $\mathcal{P}$ é limitadamente completa, mas não é completa.

Exemplo 1.11 : Seja $(I\!N, I\!P(I\!N), \mathcal{P})$ um espaço estatístico onde $\mathcal{P} = \left\{P_\theta : \theta \in \{\frac{1}{4}, \frac{1}{2}\}\right\}$ e P_θ é Binomial $(2, \theta)$.

Vamos mostrar que a família $\mathcal{P}$ não é limitadamente completa.

Seja g uma função essencialmente limitada e $\mathcal{A}$-mensurável, tal que $\int_{\mathcal{X}} g dP_\theta = 0$ $\forall \theta \in \{\frac{1}{4}, \frac{1}{2}\}$.Então ,

$$g(0)(1-\theta)^2 + 2g(1)\theta(1-\theta) + g(2)\theta^2 = 0 \quad \Rightarrow$$

$$\Rightarrow \quad g(0)(1-2\theta+\theta^2) + 2g(1)(\theta-\theta^2) + g(2)\theta^2 = 0 \quad \Rightarrow$$

$$\Rightarrow \quad [g(0) - 2g(1) + g(2)]\theta^2 + [g(1) - g(0)]2\theta + g(0) = 0.$$

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

$$8\theta^2 - 6\theta + 1 = 0 \ , \quad \forall \theta \in \{\frac{1}{4}, \frac{1}{2}\} \tag{1.5}$$

Logo, a família não é limitadamente completa e portanto não é completa.

Entretanto, se completarmos o espaço paramétrico de modo que $\Theta = \{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}\}$, não podemos mais garantir que vale (1.5), $\forall \theta \in \Theta$. Neste caso, não existirá uma $g \neq 0$ tal que a equação de segundo grau se anule, $\forall \theta$, pois ela possui apenas duas raízes. Então, $\mathcal{P}^* = \{P_\theta : \theta \in \{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}\}\}$ é limitadamente completa (também completa pois, a condição de g ser limitada não é relevante no caso) e é claro que incluindo mais elementos no espaço paramétrico ela continua sendo completa.

Proposição 1.6: (Lehmann e Scheffé, 1950) Se T é uma estatística suficiente e limitadamente completa, então T é suficiente minimal.

Prova: Seja S uma estatística suficiente. Vamos mostrar que $\sigma(T) \subset \sigma(S)$ $\mathcal{P} - q.c.$.

Seja $A \in \sigma(T)$. Considere g uma versão da probabilidade condicional de A dado S, livre de θ, e $h = E_\theta[g/T]$, também livre de θ. Temos que

$$\int_{\mathcal{X}} I_A dP_\theta = \int_{\mathcal{X}} g dP_\theta = \int_{\mathcal{X}} h dP_\theta \qquad \forall \theta \in \Theta,$$

então

$$\int_{\mathcal{X}} (I_A - h) dP_\theta = 0 \qquad \forall \theta \in \Theta.$$

Como $(I_A - h)$ é $\sigma(T)$-mensurável e limitada, pois $I_A \leq 1$ e $h = P_\theta(A/T) \leq 1$. Utilizando o fato de T ser limitadamente completa, concluimos que

$$I_A = h \qquad \mathcal{P} - q.c..$$

Mas,

$$h = E_\theta[g/T] \leq g = E_\theta[I_A/S] \leq I_A.$$

Portanto, $I_A = g \quad \mathcal{P} - qc$. Então, $A \in \sigma(S)$ $\diamond$

Proposição 1.7: Seja T uma estatística suficiente e completa, então para qualquer estatística S ancilar, tal que $\sigma(T) \vee \sigma(S)$ é essencialmente igual a σ-álgebra de Borel, S é essencialmente ancilar maximal.

Vamos omitir a prova deste resultado pois com o Teorema de Basu 1, que será enunciado e demonstrado posteriormente, ela torna-se simples.

Familia Conectada e Conjuntos Separadores

Definição 1.14: Duas medidas de probabilidade P_1 e P_2 em $(\mathcal{X}, \mathcal{A})$ são sobrepostas se, para cada $A \in \mathcal{A}$, tal que $P_1(A) = 1$, temos que $P_2(A) > 0$. Observe que a propriedade é simétrica.

Outro modo de definir medidas sobrepostas é dizer que seus suportes não são disjuntos.

Definição 1.15: Uma família de medidas de probabilidade é conectada, se para todo par P_θ e P_α em $\mathcal{P}$, existem medidas de probabilidade $P_{\theta_1}, P_{\theta_2}, \ldots.., P_{\theta_n}$ em $\mathcal{P}$, onde $\theta_1 = \theta$ e $\theta_n = \alpha$ e $n \in I\!N$, tais que P_{θ_i} e $P_{\theta_{i+1}}$ são soprepostas para $i = 1, 2, 3, \ldots.., n-1$.

Exemplo 1.12: Seja $(\mathbb{R}, \mathcal{B}, \mathcal{P})$ um espaço estatístico onde $\mathcal{P} = \{P_\theta : \theta \in \mathbb{Z}\}$ e P_θ é uniforme em $[\theta, \theta+1)$. Considerando $A = [0,1]$ e $\theta = 0$, temos que $P_0(A) = 1$ e para todo outro $\theta \in \mathbb{Z}$ $P_\theta(A) = 0$. Logo, não existe $\theta \in \mathbb{Z} - \{0\}$ tal que $P_\theta(A) > 0$. Portanto, P_0 não se sobrepõe a nenhuma outra medida em $\mathcal{P}$, então a família não é conectada.

Exemplo 1.13: Utilizando o mesmo modelo do exemplo anterior e alterando o espaço paramétrico para $\Theta = \{k/2 : k \in \mathbb{Z}\}$, temos que a família é conectada. Pois, se $\theta \in \Theta$ e $A = [\theta, \theta+1)$, então $P_\theta(A) = 1$ e $P_{\theta+1/2}(A) > 0, \quad \forall \theta \in \Theta$.

Definição 1.16: $A \in \mathcal{A}$ 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 $P_{\theta_1}(A) = P_{\theta_2}(A^c) = 1$.

Em alguns casos a condição de $\mathcal{P}$ ser conectada é equivalente a condição da não existência de um conjunto separador. A equivalência entre as duas condições foi apresentada por Pathak (1975) nos casos do modelo dominado e do modelo discreto. Posteriormente, Basu e Cheng (1979) generalizaram este resultado para os modelos coerentes. A seguir apresentaremos os dois lemas introduzidos por Pathak.

Lema 1.4: Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um modelo discreto. A família $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ 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.

$(\Rightarrow)$ Seja A um conjunto separador. Defina

$$\mathcal{P}_1 = \{P_\theta \in \mathcal{P} : P_\theta(A) = 1\} \quad \text{e}$$

$$\mathcal{P}_2 = \{P_\theta \in \mathcal{P} : P_\theta(A) = 0\}.$$

Como A é separador $\mathcal{P}_1$ e $\mathcal{P}_2$ são conjuntos não vazios. Além disso, todos elementos de $\mathcal{P}_1$ são sobrepostos e o mesmo é válido para $\mathcal{P}_2$. Entretanto, para nenhum par $(P_{\theta_1}, P_{\theta_2})$, tal que $P_{\theta_1} \in \mathcal{P}_1$ e $P_{\theta_2} \in \mathcal{P}_2$, as medidas são sobrepostas.

Logo, o modelo não é conectado.

($\Leftarrow$) Suponha que $\mathcal{P}$ não é conectado.

Para cada θ_1, θ_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 $P_1 \in \mathcal{P}_1$. Considere $\mathcal{P}_1 = \{P_\theta \in \mathcal{P} : P_\theta \sim P_1\}$. Como $\mathcal{P}$ não é conectado, $\mathcal{P}_2 = \mathcal{P} - \mathcal{P}_1 \neq \emptyset$. Além disso, sejam

$$\Gamma_1 = \{\theta : P_\theta \in \mathcal{P}_1\} \quad \text{e} \quad \Gamma_2 = \{\theta : P_\theta \in \mathcal{P}_2\}.$$

Se $\mathcal{X}_\theta$ é o suporte da medida de probabilidade P_θ, isto é, $P_\theta(\mathcal{X}_\theta) = 1$ e $\forall B \subset \mathcal{X}_\theta, P_\theta(B) > 0$, temos que

$$\mathcal{X}_{\theta_1} \cap \mathcal{X}_{\theta_2} = \emptyset, \quad \forall \theta_1 \in \Gamma_1, \;\; \forall \theta_2 \in \Gamma_2.$$

Defina

$$A = \bigcup_{\theta_1 \in \Gamma_1} \mathcal{X}_{\theta_1}.$$

Como o modelo é discreto a σ-álgebra associada é o conjunto das partes de $\mathcal{X}$. Logo, A é mensurável . Além disso,

$$P_{\theta_1}(A) = 1 \qquad \forall \theta_1 \in \mathcal{P}_1 \quad \text{e}$$

$$P_{\theta_2}(A) = 0 \qquad \forall \theta_2 \in \mathcal{P}_2.$$

Portanto, A é um conjunto separador. $\Diamond$

Lema 1.5: 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: ($\Rightarrow$) A mesma prova do Lema 1.4 .

($\Leftarrow$) Suponha $\mathcal{P}$ conectada e considere $\mathcal{P}_1, \mathcal{P}_2$ definidos anteriormente no Lema 1.4 . Como $\mathcal{P}_1$ é dominado, segue que, existe uma família enumerável $\mathcal{P}_0 \subset \mathcal{P}_1$, tal que $\mathcal{P}_1 \equiv \mathcal{P}_0$ (apêndice C) .

Sejam

$$\Gamma_0 = \{\theta : P_\theta \in \mathcal{P}_0\}, \quad \Gamma_1 = \{\theta : P_\theta \in \mathcal{P}_1\} \quad \text{e} \quad \Gamma_2 = \{\theta : P_\theta \in \mathcal{P}_2\}.$$

Defina

$$A = \bigcup_{\theta \in \Gamma_0} \mathcal{X}_\theta.$$

Como Γ_0 é enumerável, A é mensurável. Então,

$$P_\theta(A^c) = 0, \quad \forall \theta \in \Gamma_0 \;\Rightarrow\; P_\theta(A^c) = 0, \quad \forall \theta \in \Gamma_1 \;\Rightarrow\; P_\theta(A) = 1, \quad \forall \theta \in \Gamma_1.$$

Como nenhuma medida em $\mathcal{P}_2$ é conectada com alguma medida em $\mathcal{P}_1$, temos que $A \cap \mathcal{X}_\theta = \emptyset, \quad \forall \theta \in \Gamma_2$. Portanto,

$$P_\theta(A) = 0 \quad \forall \theta \in \Gamma_2.$$

Logo, A é um conjunto separador. $\diamond$

Exemplo 1.14: Seja $\mathcal{P} = \{P : P(x) = P(y) = 1/2, 0 \leq x < y\}$ uma família de medidas de probabilidade em $(I\!R, \mathcal{B})$ e Q a medida Normal Padrão . A família $\mathcal{P} \cup Q$ não é conectada, pois Q e P não são sobrepostas, $\forall P \in \mathcal{P}$.

Os candidatos a conjunto separador são:

i) $I\!R$, pois $Q(I\!R) = 1$. Contudo, $P(I\!R) = 1, \ \forall P \in \mathcal{P}$. Logo, $I\!R$ não é um conjunto separador.

ii) $I\!R^+$, pois $P(I\!R^+) = 1, \quad \forall P \in \mathcal{P}$. Contudo, $Q(I\!R^+) = 1/2$. Logo, $I\!R^+$ também não é separador.

Portanto, o modelo não é conectado e não possui um conjunto separador.

1.3.OS TEOREMAS DE BASU SOB A PERSPECTIVA CLÁSSICA

Teorema de Basu 1 (Basu, 1955)

Seja T uma estatística limitadamente completa e suficiente para o modelo $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. Se S é uma estatística ancilar, então S e T são estatisticamente independentes.

Prova : Considere $A \in \sigma(S)$. Como T é suficiente existe uma versão da probabilidade condicional de A dado T (livre de θ) ,$f(T)$, tal que

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

Pela ancilaridade de S, temos que $P_\theta(A) = k$ (constante em relação a θ). Portanto,

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

onde $(f(T) - k)$ é $\sigma(T)$- mensurável. Usando o fato de T ser limitadamente completa, temos que

$$f(T) = P_\theta(A) \quad \mathcal{P} - q.c..$$

Logo, A é estatisticamente independente de T e portanto, S e T são estatisticamente independentes . $\diamond$

Exemplo 1.15: Seja $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ um espaço estatístico 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 θ e variância 1 .

Considere as estatísticas conhecidas como média e variância amostrais, respectivamente,

$\overline{X}(x_1, ..., x_n) = \frac{\sum x_i}{n}$,
$S^2(x_1, ..., x_n) = \frac{\sum (x_i - \overline{x})^2}{n-1}$.

Vamos mostrar que $\overline{X}$ é suficiente e completa para este modelo estatístico.

O modelo é dominado pela medida de Lebesgue e sua função densidade é dada por,

$$f_\theta(x_1, ..., x_n) = (2\pi)^{-n/2} e^{\frac{-\sum (x_i-\theta)^2}{2}} =$$

$$= (2\pi)^{-n/2} e^{\frac{-1}{2}\left[\sum x_i^2 - 2\theta n\overline{X} - n\theta^2\right]} =$$

$$= (2\pi)^{-n/2} e^{\frac{-\sum x_i^2}{2}} e^{\frac{\theta n\overline{X} - n\theta^2}{2}} = h.g(\theta, \overline{X}).$$

Portanto, pelo Teorema da Fatoração, $\overline{X}$ é suficiente.

Seja $\mathcal{P}^{\overline{X}}$ a família de medidas induzida por $\overline{X}$. Se $Q_\theta \in \mathcal{P}^{\overline{X}}$, então Q_θ tem distribuição $N(\theta, \frac{1}{n})$. O modelo estatístico induzido por $\overline{X}$ é $(I\!R, \mathcal{B}, \mathcal{P}^{\overline{X}})$.

Seja f uma função Borel mensurável, tal que

$$\int_{I\!R} f dQ_\theta = 0 \quad \Rightarrow \quad \int_{-\infty}^{\infty} f(y)[2\pi n]^{-2} e^{\frac{-(y-\theta)^2}{2n}} dy = 0.$$

Fazendo transformação de variáveis e usando a unicidade da Transformada de Laplace, temos que $f = 0$ (essencialmente) , $\forall Q_\theta \in \mathcal{P}^{\overline{X}}$. Portanto, $\overline{X}$ é completa.

Vamos mostrar agora que S^2 é uma estatística ancilar. Temos que,

$$(n-1)S^2 = \sum_{i=1}^{n} (x_i - \overline{X})^2 = \sum_{i=1}^{n} (x_i - \theta)^2 - n(\overline{X} - \theta)^2.$$

Mas, $(x_i - \theta), \forall i$, e $(\overline{X} - \theta)$ tem distribuição $N(0,1)$. Então, $\frac{\sum_{i=1}^{n} (x_i-\theta)^2}{n}$ tem distribuição Qui-quadrado com n graus de liberdade, e $(\overline{X} - \theta)^2$ tem distribuição Qui-quadrado com 1 grau de liberdade. Portanto, $\frac{(n-1)}{n} S^2$ tem distribuição Qui-quadrado com $(n-1) g.l.$, então S^2 tem distribuição livre de θ.

Utilizando o Teorema de Basu 1, concluimos que $\overline{X}$ e S^2 são estatisticamente independentes.

Exemplo 1.16 Seja $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ um espaço estatístico onde

$$\mathcal{P} = \{P_\theta : \theta = (\mu, \sigma), t.q., \mu \in \mathbb{R} \; e \; \sigma \in \mathbb{R}^+\}$$

é a família da probabilidade Normal n-variada, com covariância zero.

Utilizando-se o Teorema de Basu 1 é possível mostrar, também neste caso, que $\overline{X}$ e S são estatisticamente independentes. Entretanto, Basu acreditou que utilizando-se a mesma técnica seria possível mostrar a independência estatística entre $\frac{\overline{X}}{S}$ e S.

Observe que fixado $\mu = \mu_0$, o parâmetro de interesse σ é um parâmetro escala . É facil verificar que $\frac{\overline{X}}{S}$ é invariante escala e, portanto é ancilar para σ (apêndice D). Por outro lado, sabe-se que $(\overline{X}, S)$ é uma estatística suficiente e completa. Logo, pelo Teorema de Basu 1, $\frac{\overline{X}}{S}$ e $(\overline{X}, S)$ são estatisticamente independentes. Deste fato resulta a independência entre $\frac{\overline{X}}{S}$ e S.

Analisando melhor a solução anterior, veja que a afirmação de que $\frac{\overline{X}}{S}$ e $(\overline{X}, S)$ são estatisticamente independentes é absurda, pois $\frac{\overline{X}}{S}$ fica completamente determinada com o conhecimento de $(\overline{X}, S)$. Então, qual o problema? Será algum erro do Teorema ? Não, o que ocorre é que ao analisarmos a ancilaridade de $\frac{\overline{X}}{S}$ haviamos fixado $\mu = \mu_0$ e neste caso, a estatistica $(\overline{X}, S)$ não é completa para o parâmetro de interesse σ. Pois, considerando a estatistica

$$h(x) = \frac{\overline{X} - \mu_0}{S},$$

temos que

$$E[h(x)] = 0.$$

Este exemplo é interessante para salientar a importância da condição de que a família da estatística suficiente seja completa. Outra maneira de verificar que $(\overline{X}, S)$ não é completa, no caso de μ_0 conhecido, é observar que a estatística $\sum_{i=1}^{n}(x_i - \mu_0)^2$ é suficiente e menor que $(\overline{X}, S)$. Portanto, esta última não é suficiente minima e não poderia ser completa, pois seria uma contradição a Proposição 1.6 .

Antes de enunciarmos o próximo teorema lembramos que Basu, num primeiro momento, acreditou que através da independência de uma estatística suficiente seria possível garantir a ancilaridade. Ele inclusive apresentou, incorretamente, uma prova desta afirmação (Basu, 1955) . O contra-exemplo a seguir (Basu, 1958) nega a idéia de que a independência de uma estatística suficiente garante a ancilaridade.

Exemplo 1.17: Uma urna contêm 10 bolas identicas numeradas de $\theta+1, \theta+2, ..., \theta+10$. Retira-se duas bolas aleatoriamente da urna, uma após a outra, com reposição. Seja $\Theta = \{0, 10, 20, 30, ...\}$ o espaço paramétrico e T_i o numero da i-ésima bola retirada, $i = 1, 2$. Observe que T_1 e T_2 são identicamente distribuídas e estatisticamente independentes e cada uma é suficiente para θ. Isso contradiz a idéia de que uma estatística independente de uma suficiente é ancilar. Para que tal afirmação seja verdadeira, deve-se impor uma restrição à família de medidas de probabilidade.

Teorema de Basu 2 (Basu, 1958)

Seja $\mathcal{P}$ uma família de medidas de probabilidade conectada. Sejam T e S estatísticas, tais que T é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ e estatisticamente independente de S então, S é ancilar .

Prova: Considere $A \in \sigma(S)$. Pela suficiência de T existe uma versão da probabilidade condicional de A dado T (livre de θ), $f(T)$, tal que

$$P_\theta(A \cap B) = \int_B f(T) dP_\theta, \quad \forall \theta \in \Theta \quad \forall B \in \sigma(T).$$

Por outro lado, pela independência de S e T temos que

$$P_\theta(A \cap B) = P_\theta(A) P_\theta(B) = \int_B P_\theta(A) dP_\theta, \quad \forall \theta \in \Theta.$$

Das duas igualdades anteriores, concluimos que

$$P_\theta(A) = f(T), \quad \mathcal{P} - q.c. \tag{1.6}$$

Considere $P_{\theta_1}, P_{\theta_2} \in \mathcal{P}$ medidas sobrepostas. Então, existe $x_0 \in \mathcal{X}_{\theta_1} \cap \mathcal{X}_{\theta_2}$, onde $\mathcal{X}_{\theta_i}$ é o suporte de P_{θ_i}, $i = 1, 2$. Logo, para pelo menos um ponto x_0 a igualdade (1.6) é válida para P_{θ_1} e P_{θ_2}. Portanto,

$$P_{\theta_1}(A) = P_{\theta_2}(A).$$

Considere agora que as medidas são apenas conectadas. Então, existem $P_{\theta_1}, P_{\theta_2}, ..., P_{\theta_n}$ em $\mathcal{P}$, tais que $P_{\theta_i}, P_{\theta_{i+1}}$ são sobrepostas, para $i = 1, 2, ..., n-1$. Neste caso, $P_{\theta_i}(A) = P_{\theta_{i+1}}(A)$, $i = 1, 2, ..., n-1$, então $P_{\theta_1}(A) = P_{\theta_n}(A)$, para todo par θ_1 e $\theta_n \in \Theta$. Logo, A é ancilar e portanto, S é ancilar. $\diamond$

A condição da família ser conectada é uma condição suficiente para garantir a ancilaridade de S. Thomas e Koehn (1975) estabeleceram uma condição necessária e suficiente para a ancilaridade de S.

Teorema de Basu 2.A (Thomas e Koehn, 1975)

Seja T uma estatística suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. Toda S, estatisticamente independente de T, é ancilar se, e somente se, não existe um conjunto $A \in \mathcal{A}$, tal que A é um conjunto separador.

Prova : Vamos mostrar que existe uma S não ancilar se, e somente se, existe um A separador.

($\Leftarrow$) Suponha que existe A separador.

Seja $\Phi = \{\theta \in \Theta : P_\theta(A) = 1\}$.

Defina a estatística $S(x) = I_A(x)$, para $x \in \mathcal{X}$. Temos que,

$$P_\theta(S(x) = 1) = 1, \quad \theta \in \Phi \quad \text{e}$$

$$P_\theta(S(x) = 0) = 1, \quad \theta \in \Phi^c.$$

Então, para θ fixo, S é essencialmente constante o que implica que S é estatisticamente independente de qualquer outra estatística. Em particular, S é independente de T. Contudo, S não é ancilar.

($\Rightarrow$) Suponha que existe S não ancilar. Neste caso, existe $B \in \sigma(S)$ tal que , para algum par $\theta_1, \theta_2 \in \Theta$,

$$P_{\theta_1}(B) \neq P_{\theta_2}(B). \tag{1.7}$$

Pela suficiência de T e independencia de S [lembre da prova do Teorema de Basu 2 (1.6)], temos

$$P_\theta(B) = f(T) \qquad P_\theta - q.c.$$

Considere

$$A_\theta = \{x \in \mathcal{X} : foT(x) = P_\theta(B)\}.$$

Então, $P_\theta(A_\theta) = 1$.

De (1.7) temos que existem $\theta_1, \theta_2 \in \Theta$, tal que $A_{\theta_1} \cap A_{\theta_2} = \emptyset$.

Seja $A = A_{\theta_1}$ e $\Phi = \{\theta \in \Theta : A_\theta = A\}$. Φ e Φ^c são não vazios pois, $\theta_1 \in \Phi$ e $\theta_2 \in \Phi^c$. Temos que

$$\theta \in \Phi \;\Rightarrow\; P_\theta(A) = P_\theta(A_\theta) = 1 \qquad \text{e}$$

$$\theta \in \Phi^c \;\Rightarrow\; P_\theta(A^c) \geq P_\theta(A_\theta) = 1 \;\Rightarrow P_\theta(A^c) = 1.$$

Logo, A é um conjunto separador. $\Diamond$

Pelos lemas 1.4 e 1.5 obtemos a equivalência entre as condições do Teorema de Basu 2 e Teorema de Basu 2.A . Portanto, nos casos dos modelos discreto e dominado a condição dada por Basu é uma condição necessária e suficiente para a ancilaridade de S.

Considere os três conceitos básicos apresentados,

(a) Suficiência,

(b) Independência Estatística e

(c) Ancilaridade.

O Teorema de Basu 1 estabelece que sob certas condições adicionais, (a) e (c) implicam em (b) e, o Teorema de Basu 2 estabelece que, também sob certas condições, (a) e (b) implicam em (c) . Para completar, apresentamos o Teorema de Basu 3 que estabelece condições, para que (b) e (c) impliquem em (a) .

Teorema de Basu 3

Sejam S e T estatísticas, tais que S é ancilar e estatisticamente independente de T. Se (S, T) é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, então T é suficiente para o modelo estatístico.

Prova : Considere o espaço estatístico $(\mathcal{X}, \sigma(T) \vee \sigma(S), \mathcal{P})$. Vamos mostrar que T é suficiente para este espaço.

Seja

$$\mathcal{D} = \{D \in \mathcal{A} : D = A \cup B \text{ ou } D = A \cap B,\ A \in \sigma(S), B \in \sigma(T)\}.$$

Então, $\mathcal{D} = \sigma(T) \vee \sigma(S)$ pois, $\mathcal{D}$ é um D-sistema que contém $\sigma(S) \cup \sigma(T)$.

Para concluir, necessitamos mostrar que para todo $D \in \mathcal{D}$ existe uma versão da probabilidade condicional de D dado T, livre de θ.

Caso 1 : Considere $D \in \mathcal{D}$ tal que $D = A \cap B$ e, $E \in \sigma(T)$, então

$$P_\theta(D \cap E) = P_\theta(A \cap B \cap E) = P_\theta(A \cap E^*),$$

onde $E^* = B \cap E \quad \in \sigma(T)$. Utilizando a independência de S e T e a ancilaridade de A temos

$$P_\theta(D \cap E) = P_\theta(A) P_\theta(E^*) =$$

$$= \int_{E^*} P_\theta(A) dP_\theta = \int_E k I_B dP_\theta \quad \forall \theta \in \Theta,$$

onde $k.I_B$ é $\sigma(T)$-mensurável e livre de θ. Portanto, existe uma versão da probabilidade condicional de D dado T livre de θ.

Caso 2: Seja $D \in \mathcal{D}$, tal que $D = A \cup B$. Considere os seguintes conjuntos,

$$D_1 = A^c \cap B \quad , \quad D_2 = A \cap B^c \;\text{ e }\; D_3 = A \cap B.$$

Então, $D = D_1 \cup D_2 \cup D_3$ onde os D_i são disjuntos , $i = 1, 2, 3$

Se $E \in \sigma(T)$, utilizando o Caso 1, temos

$$P_\theta(D \cap E) = P_\theta(D_1 \cap E) + P_\theta(D_2 \cap E) + P_\theta(D_3 \cap E) =$$

$$\int_E f_1(T)dP_\theta + \int_E f_2(T)dP_\theta + \int_E f_3(T)dP_\theta = \int_E (f_1(T) + f_2(T) + f_3(T))dP_\theta.$$

Logo, também existe uma versão da probabilidade condicional de D dado T, livre de θ. Portanto, T é suficiente para $(\mathcal{X}, \sigma(S) \vee \sigma(T), \mathcal{P})$.

Por hipótese, $\sigma(S) \vee \sigma(T)$ é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. Mas $\sigma(T) \subset \sigma(S) \vee \sigma(T) \subset \mathcal{A}$. Utilizando a Proposição 1.3 concluimos que T é suficiente para $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. ◊

No Exemplo 1.9 , observamos que (X, Y) é suficiente para o modelo e $X \sim N(0, 1)$ é ancilar. Mas Y não é suficiente, ao contrário, também é ancilar. O que ocorre neste caso é que a condição de independência entre X e Y não esta satisfeita . Note entretanto, que a condição de independência não é uma condição necessária. Observe o Exemplo 1.8 , onde $(X, X+Y)$ é suficiente e X é ancilar, pois $P_\theta(X = 0) = P_\theta(X = 1) = 1/2, \;\; \forall\theta \in [0, 1]$.Além disso, a distribuição de probabilidade era dada por

$$f_\theta(x, y) = \frac{\theta}{2} I_{\{x=y\}}(x, y) + \frac{1-\theta}{2} I_{\{x \neq y\}} = \frac{\theta}{2} I_{\{x+y \neq 1\}}(x, y) + \frac{1-\theta}{2} I_{\{x+y=1\}}.$$

Pelo Teorema da Fatoração, $X + Y$ é suficiente. Entretanto, X e $X + Y$ não são independentes.

CAPITULO 2 : ABORDAGEM BAYESIANA

2.1. CONSTRUÇÃO DO MODELO BAYESIANO

Na visão bayesiana além do espaço estatístico $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, admite-se um espaço probabilistico $(\Theta, \mathcal{T}, \nu)$, associado ao parâmetro θ, conhecido como espaço de probabilidade a priori. O modelo Bayesiano é construído no espaço produto $(\mathcal{X} \text{x} \Theta, \mathcal{A} \otimes \mathcal{T})$ onde define-se uma medida de probabilidade . Entretanto, para a existência desta medida é necessário que o espaço de probabilidade associado a θ satisfaça certas condições. A seguir apresentamos uma série de resultados relacionados com esta medida.

Considere um espaço mensurável $(\mathcal{X}, \mathcal{A})$ e um espaço de probabilidade $(\Theta, \mathcal{T}, \nu)$. A função $P_\theta(A)$ é uma *função de transição* de $(\Theta, \mathcal{T})$ em $(\mathcal{X}, \mathcal{A})$ se as condições abaixo estão satisfeitas.

(a) Para cada $\theta \in \Theta, P_\theta(.)$ é uma medida de probabilidade em $(\mathcal{X}, \mathcal{A})$ e,

(b) para cada $A \in \mathcal{A}, P_.(A)$ é uma função de θ, $\mathcal{T}$-mensurável.

Teorema 2.1: Existe uma única probabilidade no espaço produto $(\Omega, \mathcal{F})$ onde, $\Omega = \mathcal{X} \text{x } \Theta$ e $\mathcal{F} = \mathcal{A} \otimes \mathcal{T}$ (sigma álgebra gerada pelos retângulos de Ω) tal que,

$$\pi(A \text{ x } B) = \int_B P_\theta(A) d\nu(\theta) \quad , \forall A \in \mathcal{A} \ , \ \forall B \in \mathcal{T} \ ,$$

onde $P_\theta(A)$ é uma função de transição.

De um modo geral, esta medida é definida como

$$\pi(F) = \int_\Theta P_\theta(F_\theta) d\nu(\theta) \qquad \forall F \in \mathcal{F},$$

onde $F_\theta = \{x \in \mathcal{X} : (x, \theta) \in F\}$.

Prova:

Parte I - Vamos mostrar que se $F \in \mathcal{F}$ então, $F_\theta \in \mathcal{A}$, $\forall \theta \in \Theta$.

Seja $\mathcal{L} = \{F \in \mathcal{F} : F_\theta \in \mathcal{A}\}$. Não é difícil verificar que $\mathcal{L}$ é uma sigma álgebra.

Sejam $A \in \mathcal{A}$ e $B \in \mathcal{T}$, então

$(A\text{x}B)_\theta = B$, se $\theta \in A$ e

$(A\text{x}B)_\theta = \emptyset$, se $\theta \notin A$.

Portanto, $(A \text{ x } B)_\theta \in \mathcal{A} \quad \Rightarrow \quad A\text{x}B \in \mathcal{L}$. Então, $\mathcal{L}$ é uma sigma álgebra que contêm todos os retângulos de Ω. Logo, $\mathcal{F} \subset \mathcal{L} \quad \Rightarrow \quad \mathcal{F} = \mathcal{L}$.

Parte II - Vamos mostrar que se $F \in \mathcal{F}$, então $P_\theta(F_\theta)$ é $\mathcal{T}$-mensurável.

Seja $\mathcal{L} = \{F \in \mathcal{F} : P_\theta(F_\theta)$ é $\mathcal{T}$-mensurável $\}$. Temos que $\mathcal{L}$ é um D-sistema, pois

i) $\Omega_\theta = \mathcal{X} \quad \Rightarrow \quad P_\theta(\Omega_\theta) = 1$, $\forall \theta \in \Theta$ e portanto, $\Omega \in \mathcal{L}$.

ii)Se $F^1, F^2 \in \mathcal{L}$ tal que $F^1 \subset F^2$, então

$$P_\theta[(F^2 - F^1)_\theta] = P_\theta(F^2_\theta - F^1_\theta) = P_\theta(F^2_\theta) - P_\theta(F^1_\theta),$$

que é $\mathcal{T}$-mensurável.

iii) Se $(F^n)_{n \geq 1} \in \mathcal{L}$ é tal que $\lim_{n \to \infty} F^n = F$, então

$$P_\theta(F_\theta) = P_\theta(\lim_{n \to \infty} F_\theta^n) = \lim_{n \to \infty} P_\theta(F_\theta^n),$$

onde cada $P_\theta(F_\theta^n)$ é $\mathcal{T}$-mensurável. Portanto, $P_\theta(F_\theta)$ é $\mathcal{T}$-mensurável.

Considerando $F = A$ x B temos que,

$$P_\theta[(A\text{x}B)_\theta] = P_\theta(A) I_B(\theta).$$

Mas, $P_\theta(A)$ é por hipótese $\mathcal{T}$-mensurável e como $B \in \mathcal{T}, I_B(\theta)$ também é $\mathcal{T}$-mensurável. Logo, $P_\theta[(A$ x $B)_\theta]$ é $\mathcal{T}$-mensurável e portanto, $\mathcal{L}$ contêm os retângulos. Utilizando o teorema sobre sistemas Dynkin (apêndice B) concluímos que,

$$\mathcal{F} \subset \mathcal{L} \quad \Rightarrow \quad \mathcal{F} = \mathcal{L}.$$

Parte III - Seja

$$\pi(F) = \int_\Theta P_\theta(F_\theta) d\nu(\theta) \qquad \forall F \in \mathcal{F}.$$

A integral existe pois, $P_\theta(F_\theta)$ é $\mathcal{T}$-mensurável. Basta mostrar que π é uma medida de probabilidade.

Seja $F = \bigcup_{n=1}^{\infty} F^n$ onde $F^n \in \mathcal{F}$ são disjuntos , $\forall n$, então

$$\pi(F) = \int_\Theta P_\theta(F_\theta) d\nu(\theta) = \int_\Theta P_\theta(\bigcup_{n=1}^{\infty} F_\theta^n) d\nu(\theta) =$$

$$= \int_\Theta \sum_{n=1}^{\infty} P_\theta(F_\theta^n) d\nu(\theta) = \sum_{n=1}^{\infty} \int_\Theta P_\theta(F_\theta^n) d\nu(\theta) = \sum_{n=1}^{\infty} \pi(F_\theta^n).$$

Além disso,

$$\pi(\Omega) = \int_\Theta P_\theta(\mathcal{A}) d\nu(\theta) = \nu(\Theta) = 1.$$

Em particular se $F = A \times B$ temos

$$\pi(AxB) = \int_A P_\theta(A)\nu(\theta).$$

Para verificar a unicidade utiliza-se o teorema de extensão de Carathéodory. $\diamond$

O Teorema 2.1 continua válido se as medidas não forem de probabilidade mas apenas, medidas sigmas finitas e, é conhecido como Teorema da Medida Produto Generalizado (Ash,1972).

Nas condições do teorema trabalharemos, a partir de agora, no espaço de medida produto $(\Omega, \mathcal{F}, \pi)$ chamado modelo bayesiano.

A restrição da probabilidade π à $(\Theta, \mathcal{T})$, que é $\pi(\mathcal{X} \times B), \forall B \in \mathcal{T}$, será identificada pela medida ν, pois

$$\pi(\mathcal{X} \times B) = \int_B P_\theta(\mathcal{X}) d\nu(\theta) = \nu(B) \quad \forall B \in \mathcal{T}.$$

A medida ν será denominada *probabilidade a priori* .

Por outro lado, a restrição de π à $(\mathcal{X}, \mathcal{A})$, que é $\pi(A \times \Theta), \forall A \in \mathcal{A}$, será identificada por uma medida de probabilidade P em $(\mathcal{X}, \mathcal{A})$, dada por

$$P(A) = \pi(A \times \Theta) = \int_\Theta P_\theta(A) d\nu(\theta) \qquad \forall A \in \mathcal{A}.$$

A medida P será denominada *Probabilidade marginal ou preditiva* .

Exemplo 2.1 : Seja $(I\!R, \mathcal{B}, \mathcal{P})$ um espaço estatístico onde P_θ é $N(\theta, 1), \forall P_\theta \in \mathcal{P}$ e $\Theta = \{0, 1, 2, ..., 10\}$. Considere a probabilidade a priori ν, Uniforme em Θ,isto é, $\nu(i) = 1/11$, $\forall i \in \Theta$. Associamos a Θ a sigma álgebra das partes , $I\!P(\Theta)$.

O objetivo aqui é saber se o modelo está bem definido.

Por definição, a condição (a) do Teorema 2.1 é satisfeita. Vamos assim, estudar a condição (b). Tal condição é facilmente verificada pois, fixado $A \in \mathcal{A}$, $P_.(A) \equiv f : \Theta \to [0,1]$ é certamente $\mathbb{P}(\Theta)$ mensurável. Logo, o modelo bayesiano está bem definido.

Exemplo 2.2 : Utilizando o mesmo espaço estatístico do exemplo anterior. Considere agora, $\Theta = [0,10]$ com a seguinte medida de probabilidade a priori,

$\nu([0,2)) = \nu((8,10]) = 1/4$ e,

$\nu([2,8]) = 1/2.$

Associando a Θ a única sigma álgebra em que é possivel determinar a medida ν de seus elementos, temos

$$\mathcal{T} = \sigma\Big(\{\emptyset; [0,10]; [0,2); [2,8]; (8,10]\}\Big).$$

Seja $A = [0,\infty) \in \mathcal{B}$, $B = \{1/2\} \in \mathcal{B}_{[0,1]}$. Então,

$$f^{-1}(B) = \{j \in \Theta : P_j(A) = 1/2\} = \{0\} \notin \mathcal{T}.$$

Portanto, $f = P_.(A)$ não é $\mathcal{T}$-mensurável e não podemos definir um modelo bayesiano para essa priori.

Exemplo 2.3 : Utilizando, ainda, o mesmo espaço estatístico do exemplo anterior, seja $\Theta = [0,10]$ e ν a probabilidade triangular com função densidade dada por,

$$g(\theta) = \frac{\theta}{25} I_{[0,5)}(\theta) + \frac{(1-\theta)}{25} I_{[5,10]}(\theta).$$

Associando a sigma álgebra de Borel do $[0,10]$, temos o seguinte espaço de probabilidade a priori $([0,10]; \mathcal{B}_{[0,10]}; \nu)$. Claramente, $P_.(A)$ é Borel mensurável, $\forall A \in \mathcal{A}$.

Definição 2.1: Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ dois espaços mensuráveis. Um objeto aleatório é uma função $f : \Omega_1 \rightarrow \Omega_2$ onde, $\forall F_2 \in \mathcal{F}_2$ temos que, $f^{-1}(F_2) \in \mathcal{F}_1$. Em particular se $\Omega_2 = I\!R^n$ e $\mathcal{F}_2$ é a sigma álgebra de Borel, $\forall n \in I\!N$ ou $n = \infty$, então f é uma variável (vetor) aleatória (o).

Definição 2.2: Seja f um objeto aleatório e $\mathcal{D}$ uma sigma álgebra contida em $\mathcal{F}_1$. A função $g : \Omega_1 \text{ x } \mathcal{F}_2 \rightarrow [0,1]$ é uma probabilidade condicional regular de f dado $\mathcal{D}$ se, e somente se,

(i) para cada $\omega \in \Omega_1$ fixado , $g(\omega, .)$ é uma probabilidade em $(\Omega_2, \mathcal{F}_2)$ e,

(ii) para cada $F_2 \in \mathcal{F}_2$ fixado, $g(., F_2)$ é uma versão da probabilidade condicional de F_2 dado $\mathcal{D}$.

Teorema 2.2: Seja Ω_2 um espaço métrico, completo e separável (em particular o $I\!R^n$) e $\mathcal{F}_2$ a sigma álgebra de Borel, então existe uma $g : \Omega_1 \text{ x } \mathcal{F}_2 \rightarrow [0,1]$ que é uma versão regular da probabilidade condicional de f dado $\mathcal{D}$.

A prova deste Teorema pode ser encontrada em Ash, 1972 .

Regularidade Bayesiana

Seja $A \text{ x } B$ um retângulo de $\mathcal{F}$. Pela definição da medida π, temos

$$\pi(A \text{ x } B) = \int_B P_\theta(A) d\nu(\theta) \quad \forall B \in \mathcal{T}. \tag{2.1}$$

Por outro lado, definindo $\overline{A} \equiv A \text{ x } \Theta$ e $\overline{B} \equiv \mathcal{X} \text{ x } B$. Temos que, $A \text{ x } B = \overline{A} \cap \overline{B}$ e, usando a definição de probabilidade condicional temos,

$$\pi(A \text{ x } B) = \int_{(\mathcal{X} \text{ x } B)} I_{(A\text{x } \Theta)} d\pi = \int_{\overline{B}} \pi(\overline{A}/\mathcal{T}) d\pi(\theta) \quad \forall B \in \mathcal{T} \ . \tag{2.2}$$

Lembrando que a restrição de π à $(\Theta, \mathcal{T})$ é a probabilidade ν, segue de (2.1) e (2.2) que

$$\pi(\overline{A}/\mathcal{T}) = P_\theta(A) \qquad \nu - q.c..$$

Como $P_\theta(A)$ satisfaz a condição (i) da definição 2.2, temos que ela é uma versão regular da probabilidade condicional de A dado $\mathcal{T}$, $\forall A \in \mathcal{A}$.

A medida P_θ é chamada *probabilidade amostral* e, alternativamente, podemos usar a notação $P_\theta(A) = E[I_A/\mathcal{T}]$. De um modo geral, $\forall f$, $\mathcal{A}$-mensurável , $E[f/\mathcal{T}]$ é chamada *esperança amostral* .

De forma similar, pela definição de probabilidade condicional, fazendo as identificações devidas e lembrando que a probabilidade marginal P é a restrição de π à $(\mathcal{X}, \mathcal{A})$ temos,

$$\pi(A \text{ x } B) = \int_{\overline{A}} \pi(\overline{B}/\mathcal{A})dP(x) \qquad \forall A \in \mathcal{A}.$$

Sob a hipótese de Θ ser um espaço métrico, completo e separável e $\mathcal{T}$ ser a sigma álgebra de Borel, existe (Teorema 2.2) uma versão regular da probabilidade condicional de B dado $\mathcal{A}$. A seguinte notação será adotada,

$$\nu_x(B) = \pi(\overline{B}/\mathcal{A})(x) \qquad \forall x \in \mathcal{X}.$$

Como para todo $x \in \mathcal{X}$, ν_x é uma medida regular, então possui as seguintes propriedades:

(a) $\forall B \in \mathcal{T}$ fixado, $\nu_.(B)$ é uma função $\mathcal{A}$-mensurável e,

(b) $\forall x \in \mathcal{X}$ fixado, $\nu_x(.)$ é uma probabilidade em $(\Theta, \mathcal{T})$.

Portanto, $\nu_x(B)$ é uma função de transição de $(\mathcal{X}, \mathcal{A})$ em $(\Theta, \mathcal{T})$. E é tal que para todo retângulo A x B,

$$\pi(A \text{ x } B) = \int_A \nu_x(B)dP(x).$$

Pelo Teorema 2.1, podemos escrever,

$$\pi(F) = \int_{\mathcal{X}} \nu_x(F^x)dP(x) \qquad \forall F \in \mathcal{F},$$

onde $F^x = \{\theta \in \Theta : (x, \theta) \in F\}$.

A esta família de probabilidades $\{\nu_x : x \in \mathcal{X}\}$ em $(\Theta, \mathcal{T})$ denominamos de família de *probabilidades a posteriori* .

Por construção $\pi = P_\theta$ x ν e agora podemos escrever $\pi = P$ x ν_x. Esta propriedade é chamada regularidade do modelo Bayesiano, e a condição exigida é que Θ seja métrico, completo e separável e $\mathcal{T}$ a sigma álbegra de Borel .

Ainda, podemos escrever a probabilidade a priori como,

$$\nu(B) = \int_{\mathcal{X}} \nu_x(B) dP(x) \quad \forall B \in \mathcal{T}.$$

Dominação do modelo Bayesiano

Considere os seguintes espaços de probabilidade $(\mathcal{X}, \mathcal{A}, P)$ e $(\Theta, \mathcal{T}, \nu)$. Pelo Teorema da Medida Produto (Ash,1972), existe uma única medida produto P x ν em $(\Omega, \mathcal{F})$, tal que

$$P \text{ x } \nu(A \text{ x } B) = P(A)\nu(B) \quad \forall A \in \mathcal{A},\ \forall B \in \mathcal{T}.$$

Suponha que A x B é tal que Px$\nu(A$ x $B) = 0$, então $P(A) = 0$ ou $\nu(B) = 0$. No primeiro caso, pela definição de P, temos que $P_\theta(A) = 0\ \forall \theta \in \Theta$. Mas,

$$\pi(A \text{ x } B) = \int_B P_\theta(A) d\nu(\theta)\ . \tag{2.3}$$

Logo $\pi(A$ x $B) = 0$. No caso de $\nu(B) = 0$ temos, também de (2.3) , que $\pi(A$ x $B) = 0$.

Portanto, restrito aos retângulos de $\mathcal{F}$, temos a dominação de π pela medida produto P x ν. É possível estender esta dominação para toda $\mathcal{F}$ sob a condição de Ω ser um espaço métrico, completo e separável (Picci, 1974). Neste caso,

$$\pi \ll P \text{ x } \nu$$

Além disso, sob a hipótese do modelo ser dominado, sabemos que existe uma função $\mathcal{F}$-mensurável, $g(x,\theta)$, tal que

$$\pi(F) = \int_F g(x,\theta)d(P \text{ x } \nu) \quad \forall F \in \mathcal{F}.$$

Se F é um retângulo, pelo Teorema de Fubini (Ash,1972) temos,

$$\pi(A \text{ x } B) = \int_B \Big[\int_A g(x,\theta)dP(x)\Big] d\nu(\theta) = \int_A \Big[\int_B g(x,\theta)d\nu(\theta)\Big] dP(x).$$

Lembrando que

$$\pi(A \text{ x } B) = \int_B P_\theta(A)d\nu(\theta) \qquad \text{e,}$$

$$\pi(A \text{ x } B) = \int_A \nu_x(B)dP(x).$$

Concluimos que

$$P_\theta(A) = \int_A g(x,\theta)dP(x) \qquad \nu - q.c. \ (2.4) \qquad \text{e}$$

$$\nu_x(B) = \int_B g(x,\theta)d\nu(\theta) \qquad P - q.c. \ (2.5).$$

É possível mostrar (Teorema 3.6, Picci, 1974) que existe $N_1 \in \mathcal{T}$ ($N_2 \in \mathcal{A}$), que não depende de A (B), tal que $\nu(N_1) = 0$ [$P(N_2) = 0$] e a relação (2.4) [(2.5)] é válida para todo $\theta \in N_1^c$ ($\forall x \in N_2^c$). Daí resulta que $P_\theta \ll P$ e $\nu_x \ll \nu$. Mas, sabemos que $P \ll P_\theta$ e $\nu \ll \nu_x$e portanto, $P \equiv P_\theta$ e $\nu \equiv \nu_x$. Assim,

$$\frac{dP_\theta}{dP}(x) = \frac{d\nu_x}{d\nu}(\theta) = g(x,\theta).$$

Suponha agora que os modelos estatístico e a priori são dominados, respectivamente, pelas medidas sigma finitas μ e λ. É claro que também é válido $P \ll \mu$ e $\nu_x \ll \lambda, \forall x$. Vamos achar uma expressão para derivada de Radon-Nykodim , $\frac{d\nu_x}{d\lambda}$, da probabilidade a posteriori. Temos que,

$$\frac{d\nu_x}{d\lambda}(\theta) = \frac{d\nu_x}{d\nu}(\theta).\frac{d\nu}{d\lambda}(\theta) = \frac{dP_\theta}{dP}(x).\frac{d\nu}{d\lambda}(\theta) = \frac{\frac{dP_\theta}{d\mu}(x)}{\frac{dP}{d\mu}(x)}.\frac{d\nu}{d\lambda}(\theta).$$

Mas,

$$\frac{dP}{d\mu}(x) = \frac{dP}{d\mu}(x).\nu_x(\Theta) = \int_\Theta \frac{dP}{d\mu}(x).\frac{d\nu_x}{d\nu}(\theta)d\nu(\theta) = \int_\Theta \frac{dP}{d\mu}(x).\frac{dP_\theta}{dP}(x)d\nu(\theta) =$$

$$\int_\Theta \frac{dP_\theta}{d\mu}(x)d\nu(\theta) = \int_\Theta \frac{dP_\theta}{d\mu}(x).\frac{d\nu}{d\lambda}(\theta).d\lambda(\theta) = \int_\Theta f(x,\theta).g(\theta)d\lambda(\theta).$$

Considerando $g(\theta) = \frac{d\nu}{d\lambda}(\theta)$ (densidade a priori), $f(x,\theta) = \frac{dP_\theta}{d\mu}(x)$ (densidade amostral, também conhecida como verossimilhança) e $h(x,\theta) = \frac{d\nu_x}{d\lambda}(\theta)$ (densidade a posteriori), temos

$$h(x,\theta) = \frac{f(x,\theta).g(\theta)}{\int_\Theta f(x,\theta).g(\theta)d\lambda(\theta)}.$$

Que é a conhecida fórmula de Bayes.

Exemplo 2.4 : Seja $\mathcal{X} = \{0,1,2,...,n\}$, $\mathcal{A} = \mathbb{P}(\mathcal{X})$ e $\mathcal{P} = \{P_\theta : \theta \in [0,1]\}$ onde P_θ tem distribuição binomial (n,θ). Considerando que μ é a medida de contagem, temos

$$\frac{dP_\theta}{d\mu}(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}.$$

Além disso, seja $\Theta = [0,1], \mathcal{T} = \mathcal{B}_{[0,1]}$ e ν a medida Beta(a , b) , onde a , b $\in \mathbb{N}$. Se λ é a medida de lebesgue, temos

$$\frac{d\nu}{d\lambda}(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{Be(a,b)}.$$

A medida marginal P em $(\mathcal{X}, I\!P(\mathcal{X}))$ é dada por,

$$P(A) = \int_\Theta P_\theta(A) d\nu(\theta) = \int_\Theta \Big[\int_A \binom{n}{x} \theta^\theta (1-\theta)^{n-x} d\mu \Big] d\nu(\theta) =$$

$$\int_A \Big[\int_\Theta \binom{n}{x} \theta^x (1-\theta)^{n-x} d\nu(\theta) \Big] d\mu(x) = \int_A \Big[\int_0^1 \frac{\binom{n}{x} \theta^{x+a-1} (1-\theta)^{n+b-x-1}}{Be(a,b)} d\theta \Big] d\mu(x).$$

Então,

$$P(A) = \int_A \binom{n}{x} \frac{Be(a+x, b+a-x)}{Be(a,b)} d\mu(x)$$

e, pela fórmula de Bayes, temos que,

$$\frac{d\nu_x}{d\lambda}(\theta) = \frac{\theta^{x+a-1}(1-\theta)^{n+b-x-1}}{Be(a+x, b+a-x)}.$$

Portanto, a medida a posteriori, $\forall x \in \mathcal{X}$, é

$$\nu_x(B) = \int_B \frac{\theta^{x+a-1}(1-\theta)^{n+b-x-1}}{Be(a+x, b+a-x)} d\lambda(\theta).$$

Note que a função do integrando é uma densidade Beta$(a+x, b+n-x)$.

O estatístico bayesiano usualmente tem como interesse principal determinar a probabilidade a posteriori, com objetivo de fazer inferências sobre determinado parâmetro baseando-se nas informações obtidas da amostra. De um modo geral, considerando o objeto aleatório X assumindo valores em $\mathcal{X}$ como a amostra, o interesse está em determinar $E[f \mid \sigma(X)]$, onde f é uma função $\mathcal{T}-$mensurável.

Utilizaremos a seguinte notação

$E[f \mid X] = E[f \mid \sigma(X)]$

$E[f \mid (Y, Z)] = E[f \mid \sigma(Y) \vee \sigma(Z)]$

2.2 CONCEITOS FUNDAMENTAIS

Independência Condicional

Proposição 2.1: Sejam Y e Z objetos aleatórios, tais que $Y \subset Z$, então $\forall f$ $\mathcal{F}-$ mensurável e integrável, temos que

$(i) E[E(f \mid Y) \mid Z] = E[f \mid Y]$

$(ii) E[E(f \mid Z) \mid Y] = E[f \mid Y]$

Prova:

(i) Pelo Teorema 1.1, $E[f \mid Y]$ é $\sigma(Y)-$mensurável e integrável $\Big(E[f \mid Y] \in L_\infty(Y)\Big)$. Mas $Y \subset Z$, então $E[f \mid Y]$ é $\sigma(Z)-$mensurável $\Big(E[f \mid Y] \in L_\infty(Z)\Big)$. Logo, ela é a própria versão de sua esperança condicional dado Z. ◇

(ii) Utilizando o Teorema 1.1 temos que,

$$\int_A f d\pi = \int_A f^* d\pi \qquad \forall A \in \sigma(Z), \tag{2.6}$$

onde $f^* = E[f \mid Z] \ \in L_\infty(Z)$. Mas, como $Y \subset Z$, em particular (2.6) é válido $\forall A \in \sigma(Y)$.

Por outro lado,

$$\int_A f d\pi = \int_A E[f \mid Y] d\pi \qquad \forall A \in \sigma(Y)$$

e portanto, $E[f^* \mid Y] = E[f \mid Y]$. ◇

Definição 2.3.: Sejam T, Y e Z três objetos aleatórios. T e Y são condicionalmente independentes dado Z se,

$$E[f \mid (Y, Z)] = E[f \mid Z] \qquad \forall f \in L_\infty(T).$$

Notação : $T \amalg Y \mid Z$.

Em particular se Z é essencialmente constante, isto é, Z gera a σ-álgebra trivial $[\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\pi - q.c.]$, X e Y são simplesmente independentes e a notação é $X \amalg Y$.

Proposição 2.2 : Sejam T, Y e Z três objetos aleatórios. T e Y são condicionalmente independentes dado Z se, e somente se, $\forall f \in L_\infty(T)$ e $\forall g \in L_\infty(Y)$ a igualdade

$$E[f.g \mid Z] = E[f \mid Z]E[g \mid Z].$$

é válida.

Através desta proposição fica clara a simetria da independência condicional, isto é, $T \amalg Y \mid Z \Leftrightarrow Y \amalg T \mid Z$.

Prova: $(\Rightarrow)$

Seja $f \in L_\infty(T)$ e $g \in L_\infty(Y)$. Usando a Proposição 2.1 e a Definição 2.3 temos,

$$E[f.g \mid Z] = E\Big[E[f.g \mid (Y,Z)] \mid Z\Big] = E\Big[gE[f \mid (Y,Z)] \mid Z\Big] =$$
$$E\Big[gE[f \mid Z] \mid Z\Big] = E[f \mid Z]E[g \mid Z].$$

$(\Leftarrow)$ Seja,

$$\mathcal{D} = \Big\{D : D \in \sigma(Y) \vee \sigma(Z), \int_D f d\pi = \int_D E(f \mid Z) d\pi, \forall f \in L_\infty(X)\Big\}$$

Vamos mostrar que $\mathcal{D}$ é um D-sistema (apêndice B) .

i) Claramente $\Omega \in \mathcal{D}$, pois $E[f] = E[E[f \mid Z]]$

ii) Se $A, B \in \mathcal{D}$ com $B \subset A$, então $A - B = A \cap B^c \quad \in (Y, Z)$.

Além disso,

$$\int_{A \cap B^c} f d\pi = \int_A f I_{B^c} d\pi = \int_A E[f I_{B^c} \mid Z] d\pi =$$
$$\int_A I_{B^c} E[f \mid Z] d\pi = \int_{A \cap B^c} E[f \mid Z] d\pi.$$

Logo $A - B \in \mathcal{D}$.

iii) Seja $(A_n)_{n \geq 1} \in \mathcal{D}$ tal que, $\lim_{n \to \infty} A_n = A$. Pelo teorema de convergência dominada temos

$$\int_A f d\pi = \int_\Omega f I_A d\pi = \int_\Omega \lim_{n\to\infty} f I_{A_n} d\pi =$$

$$= \lim_{n\to\infty} \int_\Omega f I_{A_n} d\pi = \lim_{n\to\infty} \int_\Omega I_{A_n} E[f \mid Z] d\pi =$$

$$= \int_\Omega I_A E[f \mid Z] d\pi = \int_A E[f \mid Z] d\pi.$$

Seja $\mathcal{E} = \{C \cap D : C \in \sigma(Y) e D \in \sigma(Z)\}$. Vamos mostrar que $\mathcal{E} \subset \mathcal{D}$.

Seja $C \cap D \in \mathcal{E}$ e $f \in L_\infty(T)$, então utilizando a condição da proposição, temos

$$\int_{C\cap D} f d\pi = \int_D f I_C d\pi = \int_D E[f I_C \mid Z] d\pi =$$

$$= \int_D E[f \mid Z] E[I_C \mid Z] d\pi = \int_D E[E(f \mid Z) I_C \mid Z] d\pi =$$

$$= \int_D E[f \mid Z] I_C d\pi = \int_{D\cap C} E[f \mid Z] d\pi.$$

Logo $C \cap D \in \mathcal{D}$ e portanto, $\mathcal{E} \subset \mathcal{D}$.

Mas, claramente $\mathcal{E}$ é fechado por uniões finitas e $\sigma(\mathcal{E}) = \sigma(Y) \vee \sigma(Z)$.

Utilizando o teorema sobre sistemas Dinkyn (apêndice B), temos que $\sigma(Y) \vee \sigma(Z) \subset \mathcal{D}$ logo, $\mathcal{D} = \sigma(Y) \vee \sigma(Z)$. ◊

Sejam X e θ dois objetos aleatórios com domínio em $(\Omega, \mathcal{F})$. Dizemos que X representa a amostra e θ o parâmetro se

$$\sigma(X) = \{A \text{ x } \Theta : A \in \mathcal{A}\}$$

$$\sigma(\theta) = \{\mathcal{X} \text{ x } B : B \in \mathcal{T}\}$$

Um objeto aleatório T é uma estatística se $T \subset X$.

T e Y são estatisticamente independentes se $\forall f \in L_\infty(T)$ e $\forall g \in L_\infty(Y)$,

$$E[f.g \mid \theta] = E[f \mid \theta] E[g \mid \theta], \quad \forall \theta \in \Theta.$$

Portanto, a independência estatística é uma independência condicional, $T \amalg Y \mid \theta$.Este fato não implica necessariamente que $T \amalg Y$.

Exemplo 2.5 : Sejam X_1 e X_2 observações independentes de uma Uniforme em $\{\theta, \theta+1, ..., \theta+10\}$, onde θ tem distribuição Uniforme em $\{10, 20, 30, ...\}$. Isto é, $X_1 \amalg X_2 \mid \theta$.

Suponha que $X_1 = 13$, então $\theta = 10$ e portanto $10 \leq X_2 \leq 20$. Logo, X_1 não é independente de X_2 no sentido probabilístico ($X_1 \not\amalg X_2$).

Exemplo 2.6 : (Paradoxo de Simpson)

Com o interesse de verificar o efeito de determinado tratamento na cura de uma doença realizou-se um estudo através de uma amostra . Os pacientes foram classificados segundo uma variável X, assumindo dois valores T (tratamento) e T^c (controle), conforme o paciente tenha ou não se submetido ao tratamento em estudo. A variável resposta de interesse Y, assume também somente dois valores C (curado) e C^c (doente). Além disso, uma variável auxiliar Z foi considerada no problema, o sexo do paciente. Na tabela abaixo apresentamos os resultados.

TABELA 2

	Homem		Mulher	
X	C	C^c	C	C^c
T	20	40	100	50
T^c	20	40	20	10

Observe que para os dois valores de Z (homem e mulher) $X \amalg Y$. Entretanto, desconsiderando a variável auxiliar, temos o resultado na Tabela 3 .

TABELA 3

X	C	C^c
T	120	90
T^c	40	50

Observando a Tabela 3, concluimos que não é possível afirmar que X e Y são independentes. Este fato, bastante interessante para os estatísticos é conhecido como o Paradoxo de Simpson . Em termos de independência condicional podemos dizer que $X \amalg Y \mid Z$ não implica em $X \amalg Y$. Para uma análise mais detalhada do problema, inclusive sobre qual a conclusão estatística a respeito do efeito do tratamento na cura da doença, ver Dawid (1979a).

Proposição 2.3: Sejam T_1, T, Y, Z objetos aleatórios tais que $T \amalg Y/Z$ e $T_1 \subset T$, então

(i) $T_1 \amalg Y \mid Z$ e,

(ii) $T \amalg Y \mid (Z, T_1)$.

Prova:

$i)$ Por hipótese, $\forall f \in L_\infty(T)$, temos que $E[f \mid (Y,Z)] = E[f \mid Z]$. (2.7)

Como $T_1 \subset T$ em particular temos que, $\forall f \in L_\infty(T_1)$, (2.7) é válido. Portanto, $T_1 \amalg Y \mid Z$.

$ii)$ Por hipótese, $\forall g \in L_\infty(Y)$ temos que $E[g \mid (T,Z)] = E[g \mid Z]$. Mas $T_1 \subset T$, então $(T,Z) = (T,T_1,Z)$ e portanto $E[g \mid (T,T_1,Z)] = E[g \mid Z]$. (2.8)

Por outro lado, usando a proposição 2.1 temos

$$E[g \mid (T_1,Z)] = E\Big[E[g \mid (T,T_1,Z)] \mid (T_1,Z)\Big] =$$

$$= E\Big[E[g \mid Z] \mid (T_1,Z)\Big] = E[g \mid Z] \qquad (2.9)$$

De (2.8) e (2.9) temos

$$E[g \mid (T,T_1,Z)] = E[g \mid (T_1,Z)] \quad \forall g \in L_\infty(Y).$$

Logo, $T \amalg Y \mid (T_1, Z)$. $\diamond$

Proposição 2.4 : Sejam T, Y, Z, W objetos aleatórios . As seguintes propriedades são equivalentes,

(i) $T \amalg Y \mid Z$ e $T \amalg W \mid (Y, Z)$

(ii) $T \amalg (Y, W) \mid Z$

Prova: $(i) \Rightarrow (ii)$

$f \in L_\infty(T) \Rightarrow E[f \mid (Y, W, Z)] = E[f \mid (Y, Z)] = E[f \mid Z]$.

$(ii) \Rightarrow (i)$

Pela Proposição 2.3 temos

$a)$ $T \amalg Y \mid Z$

$b)$ $T \amalg (Y, W) \mid (Z, Y) \;\Rightarrow\; T \amalg W \mid (Y, Z)$ $\diamond$

Corolário 2.1 : Seja T_1 um objeto aleatório tal que $T_1 \subset (T, Z, W)$. Se $T \amalg Y \mid (Z, W)$ e $T_1 \amalg W \mid Z$, então $T_1 \amalg Y \mid Z$ e $T_1 \amalg W \mid (Y, Z)$.

Proposiçào 2.5: Sejam T_1 e T_2 objetos aleatórios tais que $(T_1, T_2) \subset X$. Se $X \amalg Y \mid T_1$ e $X \amalg Y \mid T_2$, então $X \amalg Y \mid T_1 \wedge T_2$.

Prova :

$\forall f \in L_\infty(Y)$ temos

$$E[f \mid (X, T_1)] = E[f \mid T_1] \quad \text{e,}$$

$$E[f \mid (X, T_2)] = E[f \mid T_2].$$

Como $(T_1, T_2) \subset X$, temos que $(X, T_1) = X$ e $(X, T_2) = X$.Logo, $E[f \mid X]$ é mensurável em relação a $T_1 \wedge T_2$ e

$$E[f \mid X] = E[f \mid (X, T_1 \wedge T_2)]. \tag{2.10}$$

Por outro lado,

$$\int_A f d\pi = \int_A E[f \mid X] d\pi \quad \forall A \in X.$$

Em particular, a igualdade acima é valida $\forall A \in T_1 \wedge T_2$. Utilizando (2.10) temos que,

$$E[f \mid (X, T_1 \wedge T_2)] = E[f \mid T_1 \wedge T_2].$$

Portanto, $X \amalg Y \mid T_1 \wedge T_2$ $\diamond$

Corolário 2.2 : Se T, Y, Z são objetos aleatórios tais que $T \amalg Y \mid Z$ e $T \amalg Z \mid Y$, então $T \amalg (Y, Z) \mid Y \wedge Z$.

Neste corolário é fundamental observar que concluímos por uma independência condicional à sigma álgebra da interseção de Y e Z. Em várias situações em estatística, há uma tendência em concluir pela simples independência entre T e (Y, Z). Alguns exemplos deste tipo de erro são apresentados por Dawid (1979b) como " falácias da independência condicional". Convêm lembrar a primeira versão (1955) do Teorema de Basu 2, cujo erro foi concluir por uma independência simples sem levar em conta o condicionamento. Este problema será abordado na unidade 2.3.

Suficiência

Definição 2.4 : T é uma estatística suficiente para X com respeito a θ, se $X \coprod \theta \mid T$.

O interesse do pesquisador bayesiano, como já foi dito anteriormente, está em determinar a esperança a posteriori $E[f \mid X]$, onde f é uma função $\sigma(\theta)$-mensurável. No caso de T ser uma estatística suficiente $E[f \mid X] = E[f \mid T]$ e portanto, o aprendizado relevante da amostra X a respeito de θ está concentrado na estatística suficiente T. Este aprendizado muitas vezes é chamado de informação. Vejamos a seguir a idéia de Basu (1975) sobre informação.

Segundo Basu, dois experimentos $\mathcal{E}_1$ e $\mathcal{E}_2$ geradores de dois pontos amostrais x_1 e x_2 são igualmente informativos para θ se obtemos a mesma inferência na previsão de θ.

Exemplo 2.7 : Uma urna contêm bolas numeradas de $\theta + 1, \theta + 2, ..., \theta + 100$, onde $\Theta = I\!N$. Realiza-se os dois experimentos caracterizados a seguir.

$\mathcal{E}_1$: retira-se 10 bolas da urna e observa-se os seguintes valores nas bolas retiradas,

$$x_1 = (15, 16, 27, 32, 57, 83, 92, 100, 102, 113).$$

$\mathcal{E}_2$: retira-se 2 bolas da urna e observa-se os seguintes valores,

$$x_2 = (15, 113).$$

Em ambos os casos ($\mathcal{E}_1$ e $\mathcal{E}_2$) o que podemos inferir a respeito do parâmetro θ é que

$$\theta + 1 \leq 15 \quad \text{e} \quad \theta + 100 \geq 113 \;\Rightarrow \theta = 13 \quad \text{ou} \quad 14.$$

Portanto, a informação a respeito de θ é a mesma nos dois casos, isto é,

$$Inf(\mathcal{E}_1, x_1) = Inf(\mathcal{E}_2, x_2).$$

Com seria de esperar, se T é uma estatística suficiente a informação contida em um determinado ponto amostral x é a mesma contida em t, onde $T(x) = t$ (Basu,1975).

Exemplo 2.8: Sejam X_1, X_2, X_3 variáveis aleatórias estatisticamente independentes com distribuição comum N(θ , 1).

É facil verificar que as estatísticas $T_1 = (X_1, X_2 + X_3)$ e $T_2 = (X_1 + X_2, X_3)$ são individualmente suficiente para $(X_1, X_2, X_3) = X$ com respeito a θ. Isto é, $X \amalg \theta \mid T_1$ e $X \amalg \theta \mid T_2$. O que implica que $\theta \amalg T_2 \mid T_1$ e $\theta \amalg T_1 \mid T_2$.

Se por um lado θ não depende de T_2, e por outro lado não depende de T_1, então podemos afirmar que θ não depende de (T_1, T_2)? A afirmação não é verdadeira pois se fosse resultaria em $\theta \amalg X$, o que é um absurdo. Entretanto, utilizando o corolário 2.2, é possível afirmar que $\theta \amalg (T_1, T_2) \mid T_1 \wedge T_2$.

Proposição 2.6: Sejam T e T_1 duas estatísticas, tais que T é suficiente para X com respeito a θ e $T \subset T_1$. Então, T_1 é suficiente para θ.

Prova: Pela Proposição 2.3 temos que

$$X \amalg \theta \mid T \;\; \text{e} \;\; T_1 \subset X \quad \Rightarrow \quad X \amalg \theta \mid (T, T_1).$$

Mas, $T \subset T_1 \Rightarrow (T, T_1) = T_1$. Logo, $X \amalg \theta \mid T_1$. $\diamond$

Através desta proposição observamos que de um modo geral a suficiência Bayesiana não é equivalente a Clássica, pois nesta última a Proposição 2.6 não é verdadeira. Basta lembrar o Exemplo 1.5 onde encontramos uma sigma álgebra que contêm uma outra suficiente e contudo, não é suficiente.

Ancilaridade

Definição 2.5 : S é uma estatística ancilar com respeito a θ se $S \amalg \theta$.

Neste caso a esperança a posteriori $E[f \mid S]$ é igual a priori, onde f é $\sigma(\theta)-$mensurável. Podemos dizer que S, isoladamente, não contêm informação a respeito de θ. No entanto, os exemplos e comentários do capítulo 1, a respeito da importância das estatísticas ancilares, quando em conjunto com outra estatística, continuam válidos.

Proposição 2.7: Sejam S e S_1 estatísticas tais que S é ancilar para θ e $S_1 \subset S$, então S_1 é ancilar para θ.

Prova: imediata da proposição 2.3 .

Vamos retomar o Exemplo 1.9 onde (X, Y) tem distribuição Normal bivariada e onde o parâmetro desconhecido θ é a covariância . Além disso, $X \sim N(0,1)$ e $Y \sim N(0,1)$. Portanto, X e Y são individualmente ancilares para θ. No entanto, (X, Y) é suficiente. Escrevendo em termos de esperança condicional resulta

$$X \amalg \theta, \quad Y \amalg \theta \quad \not\Rightarrow \quad (X, Y) \amalg \theta.$$

A condição que deve ser acrescentada para que a implicação acima seja válida é, $Y \amalg \theta / X$. Pois neste caso, ela é a própria Proposição 2.4 com Z constante.

Proposição 2.8: Sejam S_1 e S_2 duas estatísticas ancilares tal que $S_1 \amalg S_2 \mid \theta$, então (S_1, S_2) é ancilar.

Prova: Utilizando o Corolário 2.1, temos que

$$S_1 \amalg S_2 \mid \theta \text{ e } S_1 \amalg \theta \quad \Rightarrow \quad S_1 \amalg S_2 \text{ e } S_1 \amalg \theta \mid S_2.$$

Pela proposição 2.4 ,utilizando o fato de que $\theta \amalg S_2$ e $S_1 \amalg \theta \mid S_2$, temos que $\theta \amalg (S_1, S_2)$. $\diamond$

Fortemente Identificável e Conjunto Separador

Definição 2.5: Se $\forall f \in L_\infty(T)$ existe $f^* \in L_\infty(Y)$, tal que $f_1 \neq f_2 \quad \Rightarrow \quad f_1^* \neq f_2^*$ $\pi-$essencialmente, então T é fortemente identificável por Y.

Notação : $T \not< < Y$

Em particular, se $f^* = E[f/Y]$, temos $E[f \mid Y] = 0 \Rightarrow f = 0$ $\pi-$essencialmente. Considerando $Y = \theta$ temos a versão bayesiana de estatística completa.

Definição 2.6: $A \in \mathcal{F}$ é um conjunto separador se

(i) $0 < \pi(A) < 1$ e,

(ii) $E[I_A \mid \theta] = E^2[I_A \mid \theta]$.

Proposição 2.9: Não existe um conjunto A separador se, e somente se, $\theta \wedge X$ for essencialmente constante.

Prova: $(\Leftarrow)$ Seja A um conjunto separador, então

$$E\Big[[I_A - E(I_A \mid \theta)]^2 \mid \theta\Big] = E\Big[I_A^2 - 2I_A E[I_A \mid \theta] + E^2[I_A \mid \theta] \mid \theta\Big] =$$

$$= E[I_A \mid \theta] - 2E[I_A \mid \theta]E[I_A \mid \theta] + E^2[I_A \mid \theta] = E[I_A \mid \theta] - E^2[I_A \mid \theta] = 0.$$

Portanto,

$$E\Big[[I_A - E(I_A \mid \theta)]^2 \mid \theta\Big] = 0 \quad \Rightarrow \quad I_A = E[I_A \mid \theta] \quad \pi - q.c..$$

Logo, A é $\sigma(\theta)$-mensurável e portanto, A é $\sigma(\theta) \wedge \sigma(X)$- mensurável. Mas $0 < \pi(A) < 1$, o que implica que $\theta \wedge X$ não é essencialmente constante. $\diamond$

($\Rightarrow$) Suponha que $\theta \wedge X$ não é essencialmente constante. Então, existe $B \in \sigma(\theta) \wedge \sigma(X)$ tal que, $0 < \pi(B) < 1$. Como $B \in \sigma(\theta)$, temos que $E[I_B \mid \theta] = I_B$ e $E^2[I_B \mid \theta] = I_B^2 = I_B$.

Portanto B é um conjunto separador, o que contradiz a hipótese. $\diamond$

2.3 OS TEOREMAS DE BASU SOB A PERSPECTIVA BAYESIANA

Teorema de Basu 1

Sejam T, S, θ três objetos aleatórios tais que (T, S) é uma estatística e θ o parâmetro de interesse. Se $S \amalg \theta$, $X \amalg \theta \mid T$ e $T \not\ll \theta$, então $S \amalg T \mid \theta$.

Em palavras, se S é ancilar para θ e T é fortemente identificável por θ e suficiente para X, então S e T são condicionalmente independentes dado θ.

Prova : Utilizando a proposição 2.3 temos que

$$S \amalg \theta \mid T \qquad (2.11).$$

Utilizando (2.11) e a proposição 2.2 ,temos que, $\forall f \in L_\infty(S)$,

$$E[f \mid \theta] = E\Big[E[f \mid (\theta, T)] \mid \theta\Big] = E\Big[E[f \mid T] \mid \theta\Big].$$

Portanto, $E[E(f \mid \theta) - E(f \mid T) \mid \theta] = 0$ e pela ancilaridade de S, temos que

$$E[E(f) - E(f \mid T) \mid \theta] = 0.$$

Logo $E[f] = E[f \mid T]$ $\pi - q.c.$ pois $T \not\ll \theta$. Portanto,

$$S \amalg T \qquad (2.12).$$

Utilizando (2.11) , (2.12) e o corolário 2.1 concluimos que $S \amalg T \mid \theta$. $\diamond$

Antes de apresentarmos o segundo teorema de Basu vamos análisar o erro da sua primeira versão (Basu, 1955) dada a seguir.

Sejam T e S duas estatísticas onde T é suficiente para θ e S é estatisticamente independente de T, então S é ancilar para θ. Caracterizava-se assim a classe das estatísticas ancilares. Podemos escrever esta afirmação de uma outra maneira,

$$X \amalg \theta \mid T \quad \text{e} \quad S \amalg T \mid \theta \Rightarrow S \amalg \theta.$$

Fica facil para nos, depois de visto Corolário 2.2, perceber o erro da afirmação. Para que a sentença acima seja verdadeira é necessário que $\theta \wedge T$ seja essencialmente constante.

Teorema de Basu 2

Seja T uma estatística suficiente para X com respeito a θ. A estatística $\theta \wedge T$ é essencialmente constante se, e somente se, para toda S tal que $S \amalg T \mid \theta, S$ é ancilar.

Prova : ($\Rightarrow$) imediata do Corolário 2.2 .

($\Leftarrow$) Seja $S_1 = \theta \wedge T$.Como $S_1 \subset \theta$, temos que $S_1 \amalg T \mid \theta$, então por hipótese, $S_1 \coprod \theta$. Portanto, usando a Proposição 2.3, temos que $S_1 \coprod S_1$. Logo, S_1 é essencialmente constante. $\diamondsuit$

Teorema de Basu 3

Sejam S e T duas estatísticas, onde S é ancilar em relação a θ e $S \amalg T \mid \theta$. Se (S,T) é suficiente para X com respeito a θ, então T também é suficiente para X com respeito a θ.

Prova: Utilizando o corolário 2.1, temos que

$$S \amalg \theta \ \text{e} \ S \amalg T \mid \theta \Rightarrow S \amalg T \ \text{e} \ S \amalg \theta \mid T.$$

Aplicando a Proposição 2.4, resulta

$$S \amalg \theta \mid T \ \text{e} \ X \amalg \theta \mid (S,T) \Rightarrow X \amalg \theta \mid T. \qquad \diamondsuit$$

CAPITULO 3: APLICAÇÕES

3.1.CARACTERIZAÇÃO DA DISTRIBUIÇÃO DE DIRICHLET ATRAVÉS DE DISTRIBUIÇÕES GAMA

Nesta unidade definiremos a distribuição de Dirichlet através de variáveis aleatórias independentes Gamas com o objetivo de apresentar a Proposição 3.1 , cuja prova é uma aplicação do Teorema de Basu 1 .

Definição 3.1 : Sejam $X_1, X_2, ..., X_{n+1}$ variáveis aleatórias estatisticamente independentes com distribuição Gama $(\alpha_i, \beta), i = 1, 2, ..., n+1$, respectivamente. Isto significa que a medida de probabilidade induzida por $X_i, i = 1, 2, ..., n+1$, é dada por

$$P_{X_i}(B) = \int_B \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} e^{-\beta x_i} x_i^{\alpha_i - 1} d\lambda(x_i) \quad \forall B \in \mathcal{B},$$

onde λ é a medida de Lebesgue de $I\!R$ e $\mathcal{B}$ é a sigma álgebra de Borel.

Sejam $S = \sum_{i=1}^{n+1} X_i$ e $Y_k = \frac{X_k}{S}, k = 1, 2, ..., n+1$, então $Y^* = (Y_1, Y_2, ..., Y_{n+1})$ têm distribuição de Dirichlet com parâmetro $(\alpha_1, \alpha_2, ..., \alpha_{n+1})$.

Definição 3.2 : Sejam P e Q duas medidas de probabilidade num espaço mensurável $(\mathcal{X}, \mathcal{A})$. Se existe um conjunto $A \in \mathcal{A}$ tal que $P(A) = 0$ e $Q(A^c) = 0$, então P é singular em relação a Q.

Por construção Y^* tem a seguinte restrição sob suas componentes, $\sum_{i=1}^{n+1} Y_i = 1$. Portanto, se P_{Y^*} é a medida de probabilidade induzida por Y^* e

$$A = \left\{(y_1, ..., y_{n+1}) \in [0,1]^{n+1} : \sum_{i=1}^{n+1} y_i = 1\right\},$$

temos que $P_{Y^*}(A) = 1$.

Por outro lado, se λ^{n+1} é a medida de Lebesgue do $[0,1]^{n+1}$, temos que $\lambda^{n+1}(A) = 0$, pois A é de dimensão n . Portanto, P_{Y^*} é singular em relação a medida λ^{n+1} e então não existe a função densidade (em relação a medida de Lebesgue) de Y^* no $\mathbb{R}^{n+1}$.

Para solucionar este problema defina o vetor $Y = (Y_1, Y_2, ..., Y_n)$ de tal forma que $\sum_{i=1}^{n} Y_i < 1$ e $Y_{n+1} = 1 - \sum_{i=1}^{n} Y_i$. Neste caso, a medida $P_Y$induzida pelo vetor Y é dominada pela medida λ^n do $\mathbb{R}^n$ e portanto podemos definir a função densidade de Y.

Através de transformações de variáveis não é difícil verificar que sua densidade é dada por

$$f(y_1, ..., y_n) = K \prod_{i=1}^{n} y_i^{\alpha_i - 1}(1 - \sum_{i=1}^{n} y_i)^{\alpha_{n+1}-1}.$$

onde K é uma constante.

Para diferenciar os dois vetores diremos que Y^* tem uma distribuição singular de Dirichlet e Y tem distribuição de Dirichlet. Observe que no caso de n = 1 a variável aleatória Y tem distribuição Beta.

Proposição 3.1 : Se Y tem distribuição de Dirichlet com parâmetro $(\alpha_1, ..., \alpha_{n+1})$, então Y é estatisticamente independente de S.

Prova:

Fixado $\alpha_0 = (\alpha_1, \alpha_2, ..., \alpha_{n+1})$, vamos mostrar que S é uma estatística suficiente e completa para β.

Como $X_1, X_2, ..., X_{n+1}$ são variáveis aleatórias Gamas independentes, sua densidade conjunta é dada por

$$g(x_1, x_2, ..., x_{n+1}) = \beta^{\sum_{i=1}^{n+1} \alpha_i} \prod_{i=1}^{n+1} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} exp\{-\beta \sum_{i=1}^{n+1} x_i\}.$$

Utilizando o Teorema da Fatoração, temos que S é suficiente para β.

É sabido que soma de Gamas independentes com igual parâmetro β, é também uma Gama . Portanto, S tem distribuição Gama(α, β) onde $\alpha = \sum_{i=1}^{n+1} \alpha_i$. Denotamos por $Q_{\alpha,\beta}$ a medida induzida por S.

Seja h uma função $\sigma(S)$-mensurável tal que

$$\int_{\mathbb{R}^+} h dQ_{\alpha,\beta} = 0.$$

Temos que

$$\int_0^\infty h(u) \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta u} u^{\alpha-1} du = 0 \quad \Rightarrow$$

$$\Rightarrow \quad \int_0^\infty k h^*(u) e^{-\beta u} du = 0,$$

onde k é uma constante em relação a u e $h^*(u) = u^{\alpha-1} h(u)$. Pela unicidade da transformada de Laplace temos que

$$h^* = 0 \;\; Q_{\alpha,\beta} - q.c. \quad \Rightarrow \quad h = 0 \;\; Q_{\alpha,\beta} - q.c..$$

Portanto S é uma estatística completa.

Além disso, $(Y_1, Y_2, ..., Y_n) = (\frac{X_1}{S}, \frac{X_2}{S}, ..., \frac{X_n}{S})$ é invariante escala. Como β é um parâmetro escala (apêndice D), temos que $(Y_1, ..., Y_n)$ é ancilar para β. Finalmente, utilizando o Teorema de Basu 1 concluimos que S e $(Y_1, ..., Y_n)$ são estatisticamente independentes para todo α_0 .

Lembre que, pelo capítulo 2, a independência estatística é uma independência condicional. Logo, sob as hipóteses de que $X \amalg \beta \mid (S, \alpha)$, $Y \amalg \beta \mid \alpha$ e $S \not<< \beta \mid \alpha$, temos que $S \amalg (Y_1, ..., Y_n) \mid (\alpha, \beta)$. Assim, S e $(Y_1, ..., Y_n)$ são estatisticamente independentes.

3.2. O RESULTADO DE DURBIN

O resultado apresentado por Durbin em 1961 e utilizado pelos estatísticos como Teorema de Durbin, se refere a um teste de aderência à distribuição Normal . Na ocasião, Durbin provou que para testar a aderência de um conjunto de observações independentes à distribuição Normal(μ, σ^2) basta testar a aderência de determinada transformação das observações a uma Normal padrão [$N(0, 1)$]. Como o objetivo é testar a normalidade, a técnica de Durbin é eliminar através de transformações de variáveis o parâmetro pertubador (μ, σ^2). Durbin, utiliza-se de transformadas polares em sua prova, necessitando mostrar a independência entre cossenos e senos . Vamos refazer a prova do Teorema de Durbin utilizando os conceitos aqui trabalhados, em especial, o Teorema de Basu 1 .

Teorema de Durbin: Seja $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ um espaço estatístico onde, $\mathcal{P} = \{P_\theta : \theta = (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$ e P_θ tem medida de probabilidade Normal n-variada com covariância zero. Isto é, $X_1, X_2, ..., X_n$ são variáveis aleatórias estatisticamente independentes $N(\mu, \sigma^2)$.

Considere as seguintes estatísticas

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{e}$$

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{(n-1)}.$$

Considere agora as variáveis aleatórias $\overline{x}$, com medida de probabilidade $N(0, \frac{1}{n})$ e, $(n-1)s^2$ com medida de probabilidade qui-quadrado com $(n-1)g.l.$. Observe que podemos pensar em $\overline{x}$ e s^2 como sendo a média e a variância amostrais de n variáveis $N(0,1)$.

Defina as variáveis $x_1, x_2, ..., x_n$ pela tranformação

$$\frac{x_i - \overline{x}}{s} = \frac{X_i - \overline{X}}{S} \qquad i = 1, 2, ..., n.$$

Então, $x_1, x_2, ..., x_n$ são estatisticamente independentes com medida de probabilidade $N(0,1)$.

Prova:

Parte I: Seja $W_{k-1} = (Z_1, Z_2, ..., Z_{k-1})$, onde

$$Z_i = \frac{X_i - \overline{X}}{S} \qquad i = 1, 2, ..., n.$$

Claramente W_{k-1} é invariante em relação ao grupo

$$G = \{g_{a,b} : I\!R^n \rightarrow I\!R^n / g_{a,b}(x) = (\frac{x_1 - a}{b}, ..., \frac{x_n - a}{b}), a \in I\!R, b > 0\}, k = 2, 3, ..., n.$$

Logo, W_{k-1} é ancilar para (μ, σ^2) (Apêndice D) .

Parte II: Vamos mostrar que $(\overline{X}, S^2)$ é uma estatística suficiente e completa para (μ, σ^2).

A função densidade conjunta das v.a. $X_1, X_2, ..., X_n$ é dada por

$$f(x_1, ..., x_n) = (\sqrt{2\pi}\sigma)^{-n} exp\left\{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right\} =$$

$$= (\sqrt{2\pi}\sigma)^{-n} exp\Big\{-\Big[\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2 + n(\overline{X} - \mu)^2}{2\sigma^2}\Big]\Big\} =$$

$$= (\sqrt{2\pi}\sigma)^{-n} exp\Big\{-\Big[\frac{(n-1)S^2 + n(\overline{X} - \mu)^2}{2\sigma^2}\Big]\Big\}.$$

Utilizando o Teorema da Fatoração temos que $(\overline{X}, S^2)$ é suficiente para (μ, σ^2).

Para mostrar que a estatística é completa, vamos nos valer do conhecido fato que $\overline{X}$ tem distribuição $N(\mu, \frac{\sigma^2}{n})$ e $Y = \frac{(n-1)S^2}{\sigma^2}$ tem distribuição qui-quadrado com $(n-1)$ graus de liberdade . Então, é possível encontrar a distribuição de $S^2 = \frac{\sigma^2}{(n-1)}Y$ que é,

$$g(y) = \frac{(n-1)((n-1)y)^{\frac{n-1}{2}-1}}{(2\sigma^2)^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} exp\Big\{\frac{-(n-1)y}{2\sigma^2}\Big\}.$$

Além disso, não é dificil mostrar, usando o Teorema de Basu 1, que $\overline{X}$ e S^2 são estatisticamente independentes (ver exemplo 1.15). Portanto, a função densidade de $(\overline{X}, S^2)$ é o produto das densidades de $\overline{X}$ e S^2 .

Seja h uma função $(\overline{X}, S^2)$-mensurável, tal que $E[h] = 0$, então

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} h(x,y)\Big(\frac{n}{2\pi\sigma^2}\Big)^{1/2} exp\Big\{\frac{-n(x-\mu)^2}{2\sigma^2}\Big\} g(y)dydx = 0.$$

Portanto,

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} kh^*(x,y) exp\Big\{\frac{-n(x-\mu)^2 - (n-1)y}{2\sigma^2}\Big\} dydx = 0,$$

onde k é uma constante e $h^*(x,y) = h(x,y)[(n-1)y]^{\frac{n-1}{2}-1}$.

Considerando $u = (x-\mu)^2$ temos

$$\int_{-\infty}^{\infty} h^*(x,y) exp\Big\{\frac{-n(x-\mu)^2}{2\sigma^2}\Big\} dx = 2\int_{0}^{\infty} \frac{h^*(u,y) exp\{\frac{-nu}{2\sigma^2}\}}{2\sqrt{u}} du =$$

$$= \int_{0}^{\infty} h^{**}(u,y) exp\{\frac{-nu}{2\sigma^2}\} du,$$

onde $h^{**}(u,y) = \frac{h^*(u,y)}{\sqrt{u}}$, para $u \neq 0$, e $h^{**}(u,y) = 0$, caso contrário. Logo,

$$\int_0^\infty \int_0^\infty kh^{**}(x,y)exp\Big\{\frac{-nu-(n-1)y}{2\sigma^2}\Big\}dudy = 0.$$

Pela unicidade da transformada de Laplace $h^{**} = 0$ e portanto, $h = 0$.

Parte III: Dos resultados anteriores temos que W_{k-1} é ancilar, $k = 2, 3, ..., n$ e, $(\overline{X}, S^2)$ é suficiente e completa para (μ, σ^2). Utilizando o Teorema de Basu 1 resulta que $(\overline{X}, S^2)$ e W_{k-1} são estatisticamente independentes.

Além disso, W_{k-1} não depende de X_k e

$$Z_k = \frac{X_k - \overline{X}}{S}.$$

Segue que, W_{k-1} é estatisticamente independente de Z_k, para todo valor de k. Logo, $Z_1, Z_2, ..., Z_n$ são estatisticamente independentes.

Parte IV: Por definição

$$x_i = Z_i s + \overline{x}.$$

Observando-se s e $\overline{x}$, x_i é uma função linear de Z_i, $i = 1, ..., n$. Portanto, os $x_i's$ são também estatisticamente independentes entre si. Além disso, por construção, são identicamente distribuidas.

Parte V : Observe que

$$\sum_{i=1}^n a_i x_i = \overline{x} \sum_{i=1}^n a_i,$$

onde $\overline{x}$ tem distribuição $N(0, \frac{1}{n})$. Portanto, para toda sequência de reais $a_{i's}$ temos que $\sum_{i=1}^n a_i x_i$ tem distribuição Normal . Logo, cada x_i tem distribuição Normal.

Considere $a_i = 1$, $\forall i$. Temos que $\sum_{i=1}^n x_i$ é $N(0, n)$. Logo, cada x_i é $N(0, 1)$. $\diamond$

APÊNDICE A : CONCEITOS BÁSICOS DE TEORIA DA MEDIDA

Nesta seção apresentaremos algumas definições, usuais na Teoria da Medida, que foram utilizadas no decorrer do trabalho. É interessante que o leitor tenha certa familiaridade com tais conceitos, o que poderá ser facilmente adquirido num curso básico de Teoria da Medida ou, num livro da área.

Definição A.1 : Seja $\mathcal{X}$ um conjunto qualquer . Uma classe $\mathcal{A}$, de subconjuntos de $\mathcal{X}$, é uma sigma álgebra ($\sigma-$álgebra) se satisfaz,

$i)$ $\mathcal{X} \in \mathcal{A}$,

$ii)$ $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,

$iii)$ $(A_n)_{n \geq 1} \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

A dupla $(\mathcal{X}, \mathcal{A})$ é chamada espaço mensurável.

Será denotada por $\sigma(A)$ a sigma álgebra gerada pelo conjunto A $(A \subset \mathcal{X})$, que é a menor sigma álgebra que contêm A, isto é, se existe outra σ-álgebra $\mathcal{A}^*$ tal que $A \in \mathcal{A}^*$, então $\sigma(A) \subset \mathcal{A}^*$.

Definição A.2 : Se F é a classe de todos os abertos de $\mathcal{X}$, então $\sigma(F)$ é denominada sigma álgebra de Borel e seus elementos de borelianos. Em particular, se $\mathcal{X} = I\!R$, então F é o conjunto dos intervalos abertos e utilizamos a notação $\mathcal{B} = \sigma(F)$. Além disso, utilizaremos a notação $\mathcal{B}_n$ para os borelianos do $I\!R^n$.

Definição A.3 : Seja $(\mathcal{X}, \mathcal{A})$ um espaço mensurável. Uma função μ de $\mathcal{A}$ em $\overline{I\!R} = I\!R \cup \{-\infty, +\infty\}$ é uma medida em $(\mathcal{X}, \mathcal{A})$, se satisfaz,

i) $\mu(\emptyset) = 0$,

ii) se $(A_n)_{n \leq 1}$ é uma sequência disjunta de elementos de $\mathcal{A}$, então

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

Definição A.4 : Uma medida μ em $(\mathcal{X}, \mathcal{A})$ é $\sigma-$finita se existe uma sequência $(A_n)_{n \geq 1} \in \mathcal{A}$, tal que $\mathcal{X} = \bigcup\limits_{n=1}^{\infty} A_n$ e $\mu(A_n) < \infty$, $\forall n \in I\!N$. Se $\mu(\mathcal{X}) < \infty$, então μ é finita.

Se μ for $\sigma-$finita (finita) diremos que $(\mathcal{X}, \mathcal{A}, \mu)$ é um espaço σ-finito (finito) .

Definição A.5 : Considerando o espaço mensurável $(I\!R, \mathcal{B})$ a medida λ, neste espaço, que satisfaz $\lambda[(a, b)] = b - a$, $\forall a, b \in I\!R$, é chamada de medida de Lebesgue.

Definição A.6 Uma relação $\sim$ é válida essencialmente com respeito a uma medida μ se, $\mu(\{x \in \mathcal{X},\ talque,\ \sim \text{ não ocorre }\}) = 0$. Em particular, se a relação for de igualdade, diremos que A e B são iguais essencialmente se $\mu(\{x \in \mathcal{X} : A \neq B\}) = 0$. A notação adotada será $A = B \quad \mu - q.c.$.

Se $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\}$ é uma família de medidas, a relação $\sim$ é $\mathcal{M}$ essencialmente válida se $\mu_\theta(\{x \in \mathcal{X}\ talque\ \ \sim \text{ não ocorre }\}) = 0, \forall \mu_\theta \in \mathcal{M}$. No caso da relação ser de igualdade temos que se $\mu_\theta(\{x \in \mathcal{X} : A \neq B\}) = 0$, $\forall \mu_\theta \in \mathcal{M}$, então $A = B$ $\mathcal{M} - q.c.$ ($\mathcal{M}$ essencialmente) .

APÊNDICE B : SISTEMAS DYNKIN

Em várias provas, no decorrer deste trabalho, utilizamos o recurso de mostrar que uma classe de subconjuntos era um D - sistema, e após utilizar o Teorema sobre sistemas Dynkin. Nesta seção, vamos apresentar e provar este teorema.

Definição B.1: Uma classe $\mathcal{D}$ de subconjuntos de $\mathcal{X}$ é um D-sistema ou sistema Dynkin, se as seguintes condições são válidas,

(i) $\mathcal{X} \in \mathcal{D}$,

(ii) se $A, B \in \mathcal{D}$ com $B \subset A$, então $A - B \in \mathcal{D}$,

(iii) se $(A_n)_{n \geq 1} \in \mathcal{D}$ e $\lim\limits_{n \to \infty} A_n = A$, então $A \in \mathcal{D}$.

Teorema B.1: Seja $\mathcal{L}$ uma classe de subconjuntos de $\mathcal{X}$ fechada por intersecções finitas. Se $\mathcal{D}$ é um D-sistema e $\mathcal{L} \subset \mathcal{D}$, então $\sigma(\mathcal{L}) \subset \mathcal{D}$.

Prova: Seja $\mathcal{D}_0$ o menor D-sistema que contêm $\mathcal{L}$. Vamos mostrar que $\sigma(\mathcal{L}) = \mathcal{D}_0$. Assim, fica provado o teorema.

Como $\sigma(\mathcal{L})$ é uma sigma álgebra ela é um D-sistema. Portanto, $\mathcal{D}_0 \subset \sigma(\mathcal{L})$. (B.1)

Seja $\mathcal{A} = \{A \in \mathcal{D}_0 : A \cap B \in \mathcal{D}_0, \forall B \in \mathcal{L}\}$. Como, $\mathcal{L}$ é fechado por intersecções finitas temos que

$$A \in \mathcal{L} \Rightarrow A \cap B \in \mathcal{L}, \quad \forall B \in \mathcal{L}.$$

Mas,

$$\mathcal{L} \subset \mathcal{D}_0 \Rightarrow A \cap B \in \mathcal{D}_0 \Rightarrow A \in \mathcal{A}.$$

Logo $\mathcal{L} \subset \mathcal{A}$. Não é difícil verificar que $\mathcal{A}$ é um D-sistema. Portanto, $\mathcal{D}_0 \subset \mathcal{A} \Rightarrow \mathcal{D}_0 = \mathcal{A}$.

Seja $\mathcal{A}^* = \{A \in \mathcal{D}_0 : A \cap B \in \mathcal{D}_0, \forall B \in \mathcal{D}_0\}$, então

$$A \in \mathcal{L} \;\Rightarrow\; A \cap B \in \mathcal{D}_0, \;\; \forall B \in \mathcal{A} = \mathcal{D}_0 \;\Rightarrow\; A \in \mathcal{A}^* \;\Rightarrow\; \mathcal{L} \subset \mathcal{A}^*.$$

Como $\mathcal{A}^*$ é um D-sistema, temos que $\mathcal{D}_0 \subset \mathcal{A}^* \;\Rightarrow\; \mathcal{D}_0 = \mathcal{A}^*$. Portanto, $\mathcal{D}_0$ é uma sigma álgebra, pois é fechada por intersecções finitas. Logo, $\sigma(\mathcal{L}) \subset \mathcal{D}_0$ (B.2)

De (B.1) e (B.2) concluimos que $\sigma(\mathcal{L}) = \mathcal{D}_0$. $\Diamond$

APÊNDICE C : FAMÍLIAS DOMINADAS

A intenção desta unidade é provar o fato, muito usado no decorrer da dissertação, de que para toda família de medidas dominada $\mathcal{P}$, existe uma família $\mathcal{P}_0 \subset \mathcal{P}$ enumerável, tal que $\mathcal{P} \equiv \mathcal{P}_0$. Para tal, necessitamos antes de alguns lemas.

Lema C.1: Seja $(\mathcal{X}, \mathcal{A}, \lambda)$ um espaço de medida sigma finito. Vamos considerar uma ordenação parcial $\subset_\lambda$, de modo que $A \subset_\lambda B$ se $\lambda(A \cap B^c) = 0$. Nesta condições, toda coleção $\mathcal{L}$ de conjuntos de $\mathcal{A}$, fechada por uniões enumeráveis, tem um elemento máximo segundo esta ordenação. Isto é, existe $A_0 \in \mathcal{L}$ tal que $A \subset_\lambda A_0, \;\; \forall A \in \mathcal{L}$.

Prova:

Parte I - Vamos assumir que λ é uma probabilidade.

O conjunto $\{\lambda(A) : A \in \mathcal{L}\}$ é limitado e portanto, existe $\alpha = \sup_{A \in \mathcal{L}} \lambda(A)$. Além disso, $\forall n \in I\!N$, existe um $A_n \in \mathcal{L}$ tal que,

$$\lambda(A_n) > \alpha - \frac{1}{n}.$$

Seja $A_0 = \bigcup_{n=1}^{\infty} A_n$. Por hipótese, $\mathcal{L}$ é fechada por uniões enumeráveis, então $A_0 \in \mathcal{L}$ e ainda

$$\alpha - \frac{1}{n} < \lambda(A_0) \leq \alpha \quad \forall n \in I\!N.$$

Fazendo $n \to \infty$, temos que $\lambda(A_0) = \alpha$.

Considere $A \in \mathcal{A}$, é possível escrever

$$A \cup A_0 = A_0 \cup (A \cap A_0^c),$$

e então

$$\alpha \geq \lambda(A \cup A_0) = \lambda(A_0) + \lambda(A \cap A_0^c) = \alpha + \lambda(A \cap A_0^c).$$

Portanto, $\lambda(A \cap A_0^c) = 0$ o que implica que $A \subset_\lambda A_0$. Logo, A_0 é o elemento máximo de $\mathcal{L}$ segundo esta ordenação.

Parte II - Vamos considerar agora λ sendo apenas uma medida sigma finita.

Seja $(B_n)_{n\geq 1}$ uma partição de $\mathcal{X}$, tal que $\lambda(B_n) < \infty$, $\forall n \in I\!N$.

Defina,

$$\lambda_n(A) = \frac{\lambda(A \cap B_n)}{\lambda(B_n)} \quad \forall n \in I\!N, \quad \forall A \in \mathcal{A}.$$

É facil verificar que λ_n é uma medida de probabilidade. Considere agora uma outra probabilidade que não depende de n,

$$\lambda^* = \sum_{n=1}^{\infty} \frac{1}{2^n} \lambda_n.$$

Observe que $\lambda \equiv \lambda^*$. É claro que as ordenações $\subset_\lambda$ e $\subset_{\lambda^*}$ são equivalentes. Portanto, se existe um elemento máximo em $\subset_\lambda$ também existe em $\subset_{\lambda^*}$. ◇

Lema C.2: Se $\mathcal{P}$ é dominada por uma medida sigma finita λ e fechada por combinações convexas enumeráveis, então $\mathcal{P}$ é auto dominada. Isto é, existe $P_0 \in \mathcal{P}$, tal que $\mathcal{P} \ll P_0$.

Prova: Considere

$$\mathcal{L} = \{A \in \mathcal{A} : \lambda(A) > 0 \text{ e existe } \ \theta \in \Theta \ t.q. \forall x \in A, \frac{dP_\theta}{d\lambda}(x) > 0 \ \ \lambda - q.c.\}.$$

Parte I - Vamos mostrar que $\mathcal{L}$ é fechada por uniões enumeráveis.

Seja $A = \bigcup_{n=1}^{\infty} A_n$, onde $A_n \in \mathcal{L}$ e P_{θ_n} é a probabilidade associada a A_n $\forall n \in I\!N$.
Considere

$$P = \sum_{n=1}^{\infty} a_n P_{\theta_n},$$

onde $\sum_{n=1}^{\infty} a_n = 1$, de modo que P é uma probabilidade. Então,

$$\frac{dP}{d\lambda} = \sum_{n=1}^{\infty} a_n \frac{dP_{\theta_n}}{d\lambda}.$$

Se $x \in A$, então $\frac{dP}{d\lambda}(x) > 0$ $\lambda - q.c.$ Além disso, por hipótese $\mathcal{P}$ é fechada por combinações convexas, então $P \in \mathcal{P}$. Portanto, $A \in \mathcal{L}$.

Parte II - Pela parte I e o Lema C.1 concluimos que, existe um A_0 máximo em $\mathcal{L}$.
Vamos mostrar que $P_\theta(A_0^c) = 0,\ \forall \theta \in \Theta$.
Suponha que a afirmação não seja válida. Então, existe $\theta_1 \in \Theta$ tal que $P_{\theta_1}(A_0^c) > 0$.
Considere $B = \{x \in A_0^c : \frac{dP_{\theta_1}}{d\lambda}(x) > 0\}$. Então,

$$0 < P_{\theta_1}(A_0^c) = P_{\theta_1}(B) + P_{\theta_1}(A_0^c \cap B^c).$$

Além disso,

$$P_{\theta_1}(A_0^c \cap B^c) = \int_{A_0^c \cap B^c} (\frac{dP_{\theta_1}}{d\lambda}) d\lambda \leq 0.$$

Portanto, $P_{\theta_1}(B) > 0 \Rightarrow \lambda(B) > 0$. Como $B \subset A_0^c$, então $\lambda(B \cap A_0^c) > 0$. O que contradiz o fato de A_0 ser o máximo.

Parte III - Finalmente, vamos mostrar que $\mathcal{P} \ll P_{\theta_0}$, onde P_{θ_0} está associada a A_0.
Considere $N \in \mathcal{A}$ tal que $P_{\theta_0}(N) = 0$. Temos que

$$P_{\theta_0}(N) \geq P_{\theta_0}(N \cap A_0) = \int_{N \cap A_0} \frac{dP_{\theta_0}}{d\lambda} d\lambda.$$

Portanto, $\int_{N \cap A_0} \frac{dP_{\theta_0}}{d\lambda} d\lambda = 0$.

Mas, $x \in A_0 \Rightarrow \frac{dP_{\theta_0}}{d\lambda}(x) > 0 \ \lambda - q.c.$. e portanto,

$$\lambda(N \cap A_0) = 0 \Rightarrow P_\theta(N \cap A_0) = 0 \quad \forall \theta \in \Theta.$$

Utilizando a parte II temos que $P_\theta(N \cap A_0^c) = 0, \ \forall \theta \in \Theta$.

Logo,

$$P_\theta(N) = P_\theta(N \cap A_0) + P_\theta(N \cap A_0^c) = 0 \quad \forall \theta \in \Theta. \qquad \diamond$$

Teorema C.1: Se $\mathcal{P}$ é dominada, então existe uma subfamília $\mathcal{P}_0 = \{P_1, P_2, P_3, ...\}$ enumerável, tal que $\mathcal{P} \equiv \mathcal{P}_0$.

Prova: Considere uma família $\mathcal{P}^*$, de modo que se $Q \in \mathcal{P}^*$, então existem $P_{\theta_1}, P_{\theta_2}, ... \in \mathcal{P}$ e $a_1, a_2, ... \in I\!R^+$, com $\sum_{n=1}^{\infty} a_n = 1$, onde $Q = \sum_{n=1}^{\infty} a_n P_{\theta_n}$. Isto é, $\mathcal{P}^*$ é a família das combinações convexas das medidas de $\mathcal{P}$. Observe que $\mathcal{P} \subset \mathcal{P}^*$.

Como $\mathcal{P}$ é dominada, existe uma medida sigma finita λ, tal que $P_\theta \ll \lambda$, $\forall \theta \in \Theta$ e então $Q \ll \lambda$, $\forall Q \in \mathcal{P}^*$. Logo, $\mathcal{P}^* \ll \lambda$. Além disso, $\mathcal{P}^*$ é fechada por combinações convexas enumeráveis e pelo Lema C.2, $\mathcal{P}^*$ é auto dominada. Logo, existe um $Q_0 \in \mathcal{P}^*$ tal que $\mathcal{P}^* \ll Q_0$, onde $Q_0 = \sum_{n=1}^{\infty} a_n P_n$, com $\sum_{n=1}^{\infty} a_n = 1$ e $P_n \in \mathcal{P}$, $\forall n$. Portanto, $\mathcal{P} \ll Q_0$ pois $\mathcal{P} \subset \mathcal{P}^*$.

Considere $\mathcal{P}_0 = \{P_1, P_2, ...\}$, onde $P_1, P_2, ...$, são as componentes da medida Q_0. Então, $\mathcal{P} \ll \mathcal{P}_0$. Por outro lado, seja $N \in \mathcal{A}$ tal que $P_\theta(N) = 0$, $\forall \theta \in \Theta$, então $P_n(N) = 0$, $\forall P_n \in \mathcal{P}_0$. Portanto, $\mathcal{P}_0 \ll \mathcal{P}$ e então $\mathcal{P}_0 \equiv \mathcal{P}$. $\diamond$

APÊNDICE D: ANCILARIDADE E INVARIÂNCIA

Nesta seção apresentaremos um resultado de grande importância na Teoria Estatística. Através do Teorema D.1. é possível determinar se uma estatística é ancilar sem ser necessário recorrer a sua função densidade, bastando análisar seu comportamento como função das observações. Este resultado facilita muito o trabalho, pois na maioria das vezes não é facil encontrar a medida de probabilidade de determinada estatística.

Definição D.1: Considere o espaço $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$, com $\mathcal{P} = \{P_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$ onde

$$P_{\mu,\sigma}(B) = \int_B \frac{1}{\sigma^n} f\Big(\frac{x_1-\mu}{\sigma}, ..., \frac{x_n-\mu}{\sigma}\Big) d\lambda^n(x) \quad \forall B \in \mathcal{B}_n$$

e λ^n é a medida de Lebesgue do $\mathbb{R}^n$. Nestas condições, dizemos que μ é um parâmetro de locação e σ um parâmetro escala.

Definição D.2: A estatística T é invariante em relação a um grupo G se

$$T\Big(g(x)\Big) = T(x), \ \forall g \in G, \ \forall x \in \mathbb{R}^n.$$

Considerando o grupo G como

$$G = \{g_{a,b} : \mathbb{R}^n \to \mathbb{R}, t.q., g_{a,b}(x) = \Big(\frac{x_1-a}{b}, ..., \frac{x_n-a}{b}\Big), a \in \mathbb{R}, b \in \mathbb{R}^+\}.$$

O seguinte resultado é válido.

Teorema D.1: Se T é invariante sob o grupo G, então T é ancilar em relação aos parâmetros de locação e escala. Isto é, a medida induzida por T é livre de (μ, σ).

Prova: Sejam $B \in \mathcal{B}$ e Q a probabilidade induzida pela estatística T, então

$$Q(B) = P_{\mu,\sigma}\Big(T^{-1}(B)\Big) = \int_{T^{-1}(B)} \frac{1}{\sigma^n} f\Big(\frac{x_1-\mu}{\sigma}, ..., \frac{x_n-\mu}{\sigma}\Big) d\lambda(x) =$$

$$= \int_{T^{-1}(B)} \frac{1}{\sigma^n} f o g_{\mu,\sigma}(x) d\lambda(x) =$$

$$= \int_{(Tog_{\mu,\sigma})^{-1}(B)} \frac{1}{\sigma^n} f o g_{\mu,\sigma}(x) d\lambda(x) = \int_{g_{\mu,\sigma}^{-1}\left(T^{-1}(B)\right)} \frac{1}{\sigma^n} f o g_{\mu,\sigma}(x) d\lambda(x) =$$

$$= \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(x) d\lambda o g_{\mu,\sigma}^{-1}(x) = \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(x) d\lambda(\sigma x_1 + \mu, ..., \sigma x_n + \mu) =$$

$$= \int_{T^{-1}(B)} \frac{1}{\sigma^n} f(x) \sigma^n d\lambda(x) = \int_{T^{-1}(B)} f(x) d\lambda(x)$$

Portanto, $Q(B)$ não depende de (μ, σ). $\diamond$

Corolário D.1 : Se T é invariante em relação ao grupo

$$G_1 = \{g_a : \mathbb{R}^n \to \mathbb{R}, t.q., g_a(x) = (x_1 - a, ..., x_n - a),\ a \in \mathbb{R}\},$$

então T é ancilar em relação ao parâmetro locação.

Corolário D.2: Se T é invariante em relação ao grupo

$$G_2 = \{g_b : \mathbb{R}^n \to \mathbb{R}, t.q., g_b(x) = (\frac{x_1}{b}, ..., \frac{x_n}{b}),\ b \in \mathbb{R}^+\},$$

então T é ancilar em relação ao parâmetro escala.

REFERÊNCIAS BIBLIOGRÁFICAS

ASH, R.B. (1972) . *Real Analysis and Probability* . New York : Academic Press.

BAHADUR, R.R. (1955) . Statistics and subfields. *Ann.Math.Statist.*, **26**: 490-497.

BASU, D (1955). On statistics independent of a complete sufficient statistics. *Sankhyã* , **15** : 377 - 380 .

BASU, D. (1958) . On statistics independent of sufficient statistics. *Sankhyã* , **20:** 223-226 .

BASU, D. (1959) . The family of ancillary statistics . *Sankhyã* , **21** : 247-256 .

BASU, D. (1964) . Recovery of ancillary information . *Sankhyã* , **A, 26** : 3 - 16 .

BASU, D. (1975) . Statistical information and likelihood . *Sankhyã,* **A, 37** : 1 - 71 .

BASU, D. (1982) - Basu theorems . *Encyclopedia of statistical sciences,* 1, 193-196, New York : Wiley.

BASU, D. and GHOSH, J.K. (1967) . Sufficient statistics in sampling from a finite universe. *Bull.Int.Statist.Inst.* , **42** : 850 - 859 .

BASU, D. and TIWARI, R.C. (1980) . *Note on Dirichlet process.* Tallahassee, FSU, 29p. (FSU Statistics Report ,M.536) .

BASU, D. and CHENG, S.C. (1981) . A note on sufficiency in coherent models . *Int.J. Math.Math.Sci.,* **4 (3)** : 571-582.

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

DAWID, A.P. (1979a) . Conditional independence in statistical theory . *J.Royal Statist. Soc.,* **B, 41** : 1 - 31 .

DAWID, A. P. (1979b) . Some misleading arguments involving conditional independence. *J.Royal Statist.Soc.* , **B, 41** : 249 - 252 .

DURBIN, J. (1961) . Some methods of constructing exact tests . *Biometrika,* **48:** 41 - 55 .

HALMOS, P. (1970) . *Measure theory - graduate texts in mathematics* . New York: Springer.

HALMOS, P. and SAVAGE, L.J. (1949). Application of the Randon-Nikodym theorem to the theory of sufficient statistics . *Ann.Math. Statist.* , **20** : 225 - 241 .

HASEGAWA, M. and PERLMAN, M.D. (1974) . On the existence of minimal sufficient subfield . *The Ann. of Statist.*, **2 (5)** : 1049 - 1055 .

KOEHN, U. and THOMAS, D.L. (1975). On statistics independent of sufficient statistic: Basu's Lemma . *The Amer. Statist.* **29** : 40 - 42 .

LEHMANN, E. (1981) . An interpretation of completeness and Basu's theorem . *J. Amer.Statist. Ass.* , **76 (374)** : 335 - 340 .

LEHMANN, E. and SCHEFFÉ, H. (1950) . Completeness, similar regions, and unbiased estimation . *Sankhyã,* **10** : 305 - 840 .

PATHAK, P.K. (1975) . *Note on Basu's lemma* . New Mexico, UNM, 5 p. (Technical Report, 308).

PAULINO, C.D. (1988) . *Análise de dados categorizados incompletos: fundamentos, métodos e aplicações* . Tese de doutoramento , Universidade de São Paulo.

PEREIRA, C.A. (1980). *Bayesian solutions to some classical problemas of statistics* . PhD Thesis Florida State University.

PEREIRA, C.A. (1983) . Conditional independence in statistics. *Sankhyã* , **A, 45:** 324 - 337.

PICCI, G. (1974) . *Theory of sufficient statistics and structure analysis of transition probabilities with applications to identifiability* . Padova, LASEB - CNR , 50 p. (CCNRLASEB, 06/74).