UNIVERSIDADE DE ÉVORA

Mestrado em Matemática Aplicada

ANÁLISE DISCRIMINANTE, TEORIA DA DECISÃO E INFERÊNCIA BAYESIANA

Dissertação apresentada para obter o grau de Mestre em Matemática na especialidade de Estatística pela Universidade de Évora

Sandra Maria Bargão Saraiva

Évora, 2000

Dedicatórias

À D. Conceição Bargão,

ao Dário Ferreira,

à Elsa Saraiva,

ao Sr. Francisco Saraiva,

respectivamente Mãe, Namorado, Irmã e Pai

aqui citados em ordem alfabética para evitar discussões.

AGRADECIMENTOS

A realização da presente dissertação é fruto de um trabalho persistente que, todavia, está longe de me poder ser atribuído na íntegra, pois para ela contribuíram decisivamente várias personalidades com quem fui contactando e aprendendo ao longo do período da minha formação; assim, não posso deixar de expressar a minha admiração e o meu profundo reconhecimento ao Professor Doutor Tiago Mexia, Professor catedrático da Universidade Nova de Lisboa, que, para além de meu orientador científico, sempre soube transmitir à nossa relação de trabalho um ambiente de natural convivência, estando sempre disponível para transmitir os seus valiosos conhecimentos.

Ao meu namorado deixo aqui um obrigado pela atenção, motivação e conhecimentos que foram de grande valor e que muito contribuíram para a conclusão desta dissertação.

À minha colega Isabel Cristina com que há algum tempo estabeleci uma sã e duradoura relação de camaradagem queria também deixar o meu vivo agradecimento pelo bom clima de trabalho que sempre me soube proporcionar.

A todos aqueles, alguns bons amigos, que me ajudaram sobretudo a ter uma visão mais global da matemática não posso deixar de expressar também o meu agradecimento.

RESUMO

Com este trabalho procurou-se integrar a Análise Discriminante na Teoria Estatística de Decisão, interpretando esse enquadramento em termos de Inferência Bayesiana.

A nossa exposição foi feita em termos de Teoria da Medida pelo que começamos por introduzir conceitos necessários. Em seguida expusemos sucintamente os fundamentos da Inferência Bayesiana, antes de desenvolvermos a Análise Discriminante. Concluímos com uma aplicação.

A abordagem adoptada permitiu-nos flexibilizar a Análise Discriminante. Assim, através da introdução do conceito de Região de Risco à Posteriori Controlado estudamos situações em que, em vez de se atribuir um elemento a uma subpopulação, o elemento é atribuído a um conjunto de subpopulações.

ABSTRACT

We try to show that Discriminant Analysis can be considered as a branch of Statistical Decision Theory when viewed from a Bayesian approach.

First we present the necessary Measure Theory results, next we briefly outline the foundations of Bayesian Inference before developing Discriminant Analysis as an application of Bayesian Estimation.

Our approach renders Discriminant Analysis more flexible since it gives the possibility of classing an element as belonging to a group of populations. This possibility arises from the introduction of the concept of Regions of Controled Posterior Risk.

LISTA DE SÍMBOLOS E CONCEITOS

- $\mathcal{P}(\mathcal{S})$ - família de subconjuntos de $\mathcal{S}$
- $\Sigma_0(\mathcal{T})$ $[\Sigma(\mathcal{T})]$ - a álgebra [álgebra-σ] gerada por $\mathcal{T}$
- $\beta(\mathcal{S})$ - álgebra-σ dos borelianos (se $\mathcal{S}$ for espaço topológico), em particular $\beta_n = \beta(\mathrm{IR}^n)$
- $(\mathcal{S}, \Sigma)$ - espaço mensurável
- $\lambda(\mathcal{C})$ - função de conjuntos definida em Σ
- $(\mathcal{S}, \Sigma, \lambda)$ - espaço de medida
- (Ω, Σ, P^0) - espaço de probabilidade
- $I_C()$ - função indicatriz do conjunto C
- $\lambda_1 << \lambda_2$ - λ_1 é absolutamente contínua em ordem a λ_2
- $f = \frac{d\lambda_1}{d\lambda_2}$ - a função $f()$ é a derivada de RADON-NIKODIM, de λ_1 em ordem a λ_2
- $p(C_0|\vec{y}^m)$ - probabilidade condicional de $C_0 \in \Sigma$ dado $\vec{Y}^m = \vec{y}^m$
- $\mathrm{IE}\left(X^+|\vec{y}^m\right)$ - valor médio condicional de X^+ dado $\vec{Y}^m = \vec{y}^m$
- $S_1 \, \mathrm{X} \, S_2$ - produto cartesiano de S_1 por S_2
- $g \in m\Sigma$ - a função mensurável g é mensurável
- $g \in m\Sigma^+$ - função mensurável não negativa
- $g \in F\mathcal{S}^+$ - função mensurável simples não negativa
- Ω - espaço paramétrico
- $pr(A_i|B)$ - probabilidade à posteriori
- $\mathcal{D}$ - espaço das decisões possíveis
- E - espaço amostral

- $d_0 = d_0\left(\vec{\theta}^k\right)$ - decisão certa
- $W\left(d, d_0\right) \geq 0$ - Função de custo sendo d a decisão tomada e d_0 a decisão certa
- $i.i.d.$ - independentes e identicamente distribuídas
- $\vec{X}^n$ - vector aleatório cujas componentes são as nossas observações
- $\Gamma_\alpha\left(\vec{x}^n\right)$ - região de risco à posteriori controlado de nível α dada a amostra $\vec{x}^n$

INDICE

1- *INTRODUÇÃO*

1 - INTRODUÇÃO

O objectivo desta dissertação é enquadrar a Análise Discriminante dentro da Teoria de Decisão Estatística interpretando esse enquadramento em termos de Inferência Bayesiana.

Após apresentarmos alguns resultados preliminares, consideramos a Inferência Bayesiana introduzindo então o conceito de Região de Risco à Posteriori Controlado. Este conceito irá desempenhar um papel importante no nosso estudo. Com efeito, por vezes na Análise Discriminante em vez de atribuír um elemento a uma população poderá ser preferível atribuí-lo simplesmente a um conjunto de populações, as quais, como veremos, constituirão uma Região de Risco à Posteriori Controlado.

Seguir-se-á a exposição da Análise Discriminante relacionando-a com a Inferência Bayesiana. Mostraremos que uma interpretação Bayesiana da Análise Discriminante leva à minimização do custo médio da decisão, sendo pois enquadrável na Teoria de Decisão Estatística .

Existem excelentes exposições da Inferência Bayesiana, ver por exemplo, O'HAGAN (1994), e da Análise Discriminante, ver por exemplo ANDERSON (1984) que já relaciona a Análise Discriminante com a Inferência Bayesiana. No entanto o nosso objectivo levou-nos a uma apresentação distinta da apresentada nesses textos, fortemente baseada na teoria da medida. Nesta guiá-mo-nos fundamentalmente pela excelente exposição que vem em WILLIAMS (1997). Procurámos ainda pôr em evidência as possibilidades dadas pelas estatísticas suficientes quando existem.

2 - ***RESULTADOS PRELIMINARES***

2 - RESULTADOS PRELIMINARES

2.1. - MEDIDAS E INTEGRAIS

Dado um conjunto $\mathcal{S}, \mathcal{P}(\mathcal{S})$ será a família dos sub-conjuntos de $\mathcal{S}$. A família não vazia $\Sigma_0[\Sigma]$ de sub-conjuntos de $\mathcal{S}$ será uma álgebra [álgebra-σ] se for fechada para a passagem ao complementar e a união finita (numerável). Intersectando-se álgebras [álgebras-σ] obtêm-se álgebras [álgebras-σ]. Logo, para toda a família $\mathcal{T}$ de sub-conjuntos de $\mathcal{S}$, a intersecção de todas as álgebras [álgebras-σ] que a contêm será uma álgebra [álgebra-σ], a álgebra [álgebra-σ] gerada por $\mathcal{T}$, representamo-la por $\Sigma_0(\mathcal{T})$ $[\Sigma(\mathcal{T})]$. Em particular se $\mathcal{S}$ for espaço topológico, a álgebra-σ gerada pela família dos abertos será a álgebra-σ dos borelianos, representá-la-emos por $\beta(\mathcal{S})$. No caso de IR^n admitiremos que se tem a topologia baseada na métrica euclideana e poremos $\beta_n = \beta(\mathrm{IR}^n)$.

Observe-se que $P(\mathcal{S})$ é álgebra-σ e que toda a álgebra-$[\sigma]$ contém $\varnothing$ e $\mathcal{S}$. O par $(\mathcal{S},\Sigma)$ será um espaço mensurável. Sendo $\lambda(C)$ uma função de conjuntos definida em Σ, a mesma será uma medida se for não negativa, se se anular para $\varnothing$ e for σ-aditiva, isto é, dados os conjuntos $C_n \in \Sigma$, $n = 1,\ldots$, disjuntos dois a dois,

(1) $$\lambda\left(\bigcup_{n=1}^{\infty} C_n\right) = \sum_{n=1}^{\infty} \lambda(C_n);$$

o trio $(\mathcal{S},\Sigma,\lambda)$ será um espaço de medida. Se $\lambda(\mathcal{S})<+\infty \quad [=1]$, λ será uma medida finita (de probabilidade e $(\mathcal{S},\Sigma,\lambda)$ será espaço de probabilidade). Além disso, se $\mathcal{S}=\bigcup_{n=1}^{\infty} C_n$, com $C_n \in \Sigma$ e $\lambda(C_n)<+\infty$, λ será σ-finita.

O teorema de CARATHÉODORY, ver WILLIAMS (1997,pg. 20), diz-nos que duas medidas σ-finitas que coincidem numa álgebra Σ_0, coincidem em $\Sigma=\Sigma\,(\Sigma_0)$, isto é, na álgebra-$\sigma$ gerada por Σ_0. Utilizando-se este teorema pode mostrar-se que existe uma única medida, dita medida de LEBESGUE, definida em (IR^n,β_n) e tal que

(2) $$\mathcal{L}\left(\mathop{X}_{j=1}^{n}\left]a_j,b_j\right]\right)=\prod_{j=1}^{n}\left(b_j-a_j\right)\leq+\infty$$

Esta medida é σ-finita mas não é finita já que $\mathcal{L}(\mathrm{IR}^n)=+\infty$. Ao trabalhar com medidas convenciona-se que $(+\infty)+(+\infty)=+\infty$ e que $(\pm\infty)\text{ x }(+\infty)=\pm\infty$.

Além da medida de LEBESGUE interessar-nos-á considerar as chamadas medidas discretas. Seja $\zeta=\{z_j\}$ um conjunto finito ou numerável de pontos de $\mathcal{S}$ aos quais se atribuem pesos q_j, $j=1,\ldots$, não negativos. Pondo-se

(3) $$\varphi(C)=\sum_{z_j\in C} q_j$$

define-se uma medida em $\mathcal{P}(\mathcal{S})$. Se ζ for finito, φ será finita; no entanto, se ζ for numerável, φ em geral será σ-finita sendo finita apenas quando $\sum_{j=1}^{\infty} q_j<+\infty$.

Quando a soma dos pesos é 1, φ será medida de probabilidade.

Um caso particular de medida discreta é a medida $N(.)$ definida em $\mathcal{P}(\mathrm{IR})$. Neste caso, ζ é constituída pelos pontos com abcissa inteira não negativa e todos os pesos serão iguais a 1.

A função $\vec{f}^n(.)$ definida num espaço de medida $(\mathcal{S},\Sigma,\lambda)$ e assumindo valores em IR^m é mensurável, pondo-se $f \in m\Sigma$, se qualquer que seja $C \in \beta_m$, a respectiva imagem inversa por $\vec{f}^n(.)$ pertence a Σ; teremos então

(4) $$\vec{f}^n(C)^{-1} = \left\{z : \vec{f}^n(z) \in C\right\} \in \Sigma,$$

observe-se que, nesta expressão, z representa um ponto genérico de $\mathcal{S}$. Por exemplo, $\mathcal{S}$ pode ser constituído pelas matrizes simétricas $l \times l$ e, com $n = 1$, a função ser o traço das mesmas.

Se $n = 1$, poremos simplesmente f e se $f \in m\Sigma$ for não negativa, poremos $f \in m\Sigma^+$. Sendo $I_C(.)$ a função indicatriz do conjunto C, ter-se-á $\sum_{i=1}^{n} a_i I_{C_i}(.) \in m\Sigma$ sempre que $C_1,\ldots,C_m \in \Sigma$ e, além disso, $\sum_{i=1}^{n} a_i I_{C_i}(.) \in m\Sigma^+$ se $a_i \geq 0$ $,i=1,\ldots,n$. No segundo caso pomos $\sum_{i=1}^{n} a_i I_{C_i}(.) \in F\mathcal{S}^+$, dizendo que a função é mensurável simples não negativa. Supondo que se está a trabalhar com o espaço de medida $(\mathcal{S},\Sigma,\lambda)$, ter-se -á o integral

(5) $$\int\left(\sum_{i=1}^{n} a_i I_{C_i}(z)\right) d\lambda(z) = \sum_{i=1}^{n} a_i \lambda(C_i) \leq +\infty$$

e, se $f \in m\Sigma^+$, define-se

(6) $$\int f(z)\, d\lambda(z) = Sup\left\{\int g(z)\, d\lambda(z) \middle| \left(g(z) \in F\mathcal{S}^+\right) \wedge \left(g \leq f\right)\right\} \leq +\infty$$

isto é, $\forall z \in \mathcal{S}$, $g(z) \leq f(z)$. Dada $f \in m\Sigma$, $-f \in m\Sigma$ e

$$(7) \qquad \begin{cases} f^{+} = \mathrm{Max}\{f;0\} \in m\Sigma^{+} \\ f^{-} = \mathrm{Max}\{-f;0\} \in m\Sigma^{+} \end{cases}$$

tendo-se $|f| = f^{+} + f^{-} \in m\Sigma^{+}$ e $f = f^{+} - f^{-}$. Caso $\int |f(z)|\, d(z) < +\infty$ define-se

$$(8) \qquad \int f(z)\, d\lambda(z) = \int f^{+}(z)\, d\lambda(z) - \int f^{-}(z)\, d\lambda(z)$$

A razão de ser da condição $\int |f(z)|\, d(z) < +\infty$ está em evitar indeterminações da forma $(+\infty) - (+\infty)$.

É ainda fácil de se ver que se se tiver o espaço de medida $(\mathcal{S}, \mathcal{P}(\mathcal{S}), \varphi)$, toda a função real é $m\, \mathcal{P}(\mathcal{S})$ tendo-se

$$(9) \qquad \int f(z)\, d\varphi(z) = \sum_{j=1} q_j f(z_j)$$

Pode mostrar-se, ver WILLIAMS (1997,pg. 54), que se os integrais das funções $f_j()$, $j=1,\ldots,N$ estão definidos, se tem

$$(10) \qquad \int \left(\sum_{j=1}^{N} b_j f_j(z) \right) d\lambda(z) = \sum_{j=1}^{N} b_j \int f_j(z)\, d\lambda(z)$$

Por outro lado, dada $f \in m\Sigma^{+}$ e $C \in \Sigma$

$$(11) \qquad \mu(C) = \int f(z)\, I_C(z)\, d\lambda(z)$$

define, ver WILLIAMS (1997, pg. 56 e 57), uma medida em $(\mathcal{S},\Sigma)$. No que segue poremos

(12) $$\mu(C)=\int_C f(z)\,d\lambda(z)$$

Dadas duas medidas λ_1 e λ_2 definidas no espaço mensurável $(\mathcal{S},\Sigma)$, diremos que λ_1 é absolutamente contínua em ordem a λ_2 se $\lambda_2(C)=0$ implicar $\lambda_1(C)=0$, escreve-se $\lambda_1 \ll \lambda_2$. Se λ_1 e λ_2 forem σ-finitas o teorema de RADON-NIKODIM, ver WILLIAMS (1997, pg. 57) diz-nos então que existe $f \in m\Sigma^+$ tal que

(13) $$\lambda_1(C)=\int_C f(z)\,d\lambda_2(z)$$

A função $f()$ será a derivada de RADON-NIKODIM de λ_1 em ordem a λ_2, pondo-se $f=\frac{d\lambda_1}{d\lambda_2}$.

As derivadas de RADON-NIKODIM de medidas de probabilidade em ordem a $\mathcal{L}[N]$ serão as densidades (funções de probabilidade). Por exemplo, para as distribuições binomiais com parâmetros n e p, ter-se-á

(14) $$p(m|n,p)=\binom{n}{m}p^m(1-p)^{n-m} \quad ; m=0,1,\ldots,n$$

e $p(x|n,p)=0$ para todos os outros pontos em $\mathbb{R}$.

Quando todas as medidas de uma família são absolutamente contínuas em ordem à medida λ diremos que a família está dominada por λ.

Se

(15) $$\lambda_1(C) = \int_C f_2(z)\, d\lambda_2(z)$$

e

(16) $$f_1(z) = \frac{d\lambda_0}{d\lambda_1}$$

pode mostrar-se, ver WILLIAMS (1997, pg. 56 e 57), que

(17) $$\lambda_0(C) = \int_C f_1(z)\, d\lambda_1(z) = \int_C f_1(z) f_2(z)\, d\lambda_2(z)$$

isto é, tem-se $\frac{d\lambda_0}{d\lambda_1} = \frac{d\lambda_0}{d\lambda_1} \frac{d\lambda_1}{d\lambda_2}$. Esta expressão é formalmente semelhante à da derivada de funções compostas.

Suponhamos que se tem

(18) $$P(C) = \sum_{j=1}^{N} a_j P_j(C)$$

com $a_j \geq 0 \quad ,j = 1,...,N$ e $\sum_{j=1}^{N} a_j = 1$. Então se as $P_j \quad ,j = 1,\dots,N$, forem medidas de probabilidade definidas em $(\mathcal{S},\Sigma)$, P será medida de probabilidade definida no mesmo espaço mensurável. Além disso, se as $P_j()$ tiverem derivada de RADON-NIKODIM $p_j(x) \quad ,j = 1,\dots,N$, relativamente a uma medida λ definida em $(\mathcal{S},\Sigma)$, ter-se-á

(19) $$p(x) = \sum_{j=1}^{N} a_j p_j(x) = \frac{dP}{d\lambda} \; .$$

2.2. - PROBABILIDADES E VALORES MÉDIOS CONDICIONAIS

Dada a função mensurável $\vec{y}^m = \vec{t}^m(z)$ definida no espaço de medida $(\mathcal{S},\Sigma,\lambda)$ e tomando valores em IR^m, pondo-se, com $C \in \beta_m$

$$(1) \qquad \lambda_{\vec{t}^m}(C) = \lambda\left(\vec{t}^m(C)^{-1}\right)$$

onde $\vec{t}^m(C)^{-1} = \left\{z : \vec{t}^m(z) \in C\right\}$, define-se uma medida em $\left(\mathrm{IR}^m,\beta_m\right)$. Diremos que esta medida é induzida por $\vec{t}^m$.

Sendo η outra medida definida em $(\mathcal{S},\Sigma)$ a partir da mesma obtém-se a medida induzida $\eta_{\vec{t}^m}()$ também definida em $\left(\mathrm{IR}^m,\beta_m\right)$. Suponhamos que $\lambda << \eta$ e que $\eta_{\vec{t}^m}(C) = 0$, tem-se então $\eta\left(\vec{t}^m(C)^{-1}\right) = 0$, bem como

$$(2) \qquad \lambda_{\vec{t}^m}(C) = \lambda\left(\vec{t}^m(C)^{-1}\right) = 0$$

e, consequentemente, $\lambda_{\vec{t}^m} << \eta_{\vec{t}^m}$. Assim a indução de medidas respeita a relação de continuidade absoluta. Por outro lado, dado o espaço de probabilidade $(\mathcal{S},\Sigma,P)$ e $C_0 \in \Sigma$, tem-se a medida

$$(3) \qquad \gamma(C) = \int_C I_{C_0}(z)\,dP(z) = \int I_C(z)\, I_{C_0}(z)\, dP(z) = \int I_{C \cap C_0}(z)\, dP(z) = P\left(C_0 \cap C\right)$$

absolutamente contínua em ordem a P. Assim, se $\vec{t}^m$ estiver definida em $(\mathcal{S},\Sigma,P)$, ter-se-á $\gamma_{\vec{t}^m} << P_{\vec{t}^m}$ e, como $P_{\vec{t}^m}\left(\mathrm{IR}^n\right) = P(\mathcal{S}) = 1$, $P_{\vec{t}^m}$ será uma medida de probabilidade induzida. Pondo-se

(4) $$p\left(C_0 \middle| \vec{y}^m\right) = \frac{d\gamma_{\vec{t}^m}}{dP_{\vec{t}^m}}$$

ter-se-á com $C \in \beta_m$, atendendo ao teorema de RADON-NIKODIN

(5) $$\gamma_{\vec{t}^m}(C) = \int_C p\left(C_0 \middle| \vec{y}^m\right) dP_{\vec{t}^m}\left(\vec{y}^m\right)$$

Ora

(6) $$\gamma_{\vec{t}^m}(C) = \gamma\left(\vec{t}^{\,m}(C)^{-1}\right) = P\left(\vec{t}^{\,m}(C)^{-1} \cap C_0\right),$$

vindo

(7) $$P\left(\vec{t}^{\,m}(C)^{-1} \cap C_0\right) = \int_C p\left(C_0 \middle| \vec{y}^m\right) dP_{\vec{t}^m}\left(\vec{y}^m\right)$$

e, em particular

(8) $$P(C_0) = P(S \cap C_0) = P\left(\vec{t}^{\,m}\left(\mathrm{IR}^m\right)^{-1} \cap C_0\right) = \int_{\mathrm{IR}^m} p\left(C_0 \middle| \vec{y}^m\right) dP_{\vec{t}^m}\left(\vec{y}^m\right)$$

A função $p\left(C_0 \middle| \vec{y}^m\right)$ será a probabilidade condicional de $C_0 \in \Sigma$ dado $\vec{y}^m$.

Uma variável aleatória X será uma função real mensurável $f(z)$ definida num espaço de probabilidade (S, Σ, P).

Caso $\int \left|f(z)\right| \, dP(z) < +\infty$, X tem o valor médio

(9) $$\mathbb{E}(X) = \int f(z)\ dP(z)$$

Ponhamos $X^{+} = \text{Max}\ \{X;0\}$ e $X^{-} = \text{Max}\ \{-X;0\}$. Raciocinando-se como atrás para a probabilidade condicional, mostra-se que quando $\mathbb{E}(X)$ está definido, existem valores médios condicionais $\mathbb{E}\left(X^{+}\middle|\vec{y}^{m}\right)$ e $\mathbb{E}\left(X^{-}\middle|\vec{y}^{m}\right)$ e que, com

(10) $$\mathbb{E}\left(X\middle|\vec{y}^{m}\right) = \mathbb{E}\left(X^{+}\middle|\vec{y}^{m}\right) - \mathbb{E}\left(X^{-}\middle|\vec{y}^{m}\right)$$

se tem

(11) $$\mathbb{E}(X) = \int\limits_{\mathbb{R}^{m}} \mathbb{E}\left(X\middle|\vec{y}^{m}\right) dP_{\vec{t}^{m}}\left(\vec{y}^{m}\right)$$

sendo $\mathbb{E}\left(X\middle|\vec{y}^{m}\right)$ o valor esperado condicional.

Suponhamos que $\vec{t}^{m}(Z)$ está definida em (S,Σ,φ) com φ medida discreta associada ao conjunto de pontos $\zeta = \{z_j\}$ finito ou numerável. O conjunto $\{\vec{t}^{m}(z); z \in \zeta\}$ será finito ou numerável. Ponhamos $\{\vec{t}_i^{m}; i \in \zeta'\} = \{\vec{t}^{m}(z); z \in \zeta\}$. Tem-se agora uma medida induzida discreta. Os pesos q_i' do $\vec{t}_i^{m}, i \in \zeta'$, serão dados por

(12) $$q_i' = \sum_{\left(\vec{t}^{m}(z_j) = \vec{t}_i^{m}\right)} q_j, \quad i \in \zeta$$

Se φ for medida de probabilidade, $\sum\limits_{j=1} q_j = 1$ pelo que $\sum\limits_{i=1} q_i' = \sum\limits_{j=1} q_j = 1$ e $\varphi_{\vec{t}^{m}}$ também será medida de probabilidade tendo-se então

$$(13) \qquad \begin{cases} P(C_0) = \sum_{i=1} q_i^{'}\; p\left(C_0 \middle| \vec{t}_i^{\,m}\right) \\ \\ \mathrm{I\!E}(X) = \sum_{i=1} q_i^{'}\; \mathrm{I\!E}\left(X \middle| \vec{t}_i^{\,m}\right) \end{cases}$$

Seja A_i o acontecimento que se verifica quando $\vec{t}^{\,m}(Z) = \vec{t}_i^{\,m}$ $, i = 1, \ldots$. As expressões (13) podem ser rescritas com

$$(14) \qquad \begin{cases} P(C_0) = \sum_{i=1} pr(A_i)\; p\left(C_0 \middle| A_i\right) \\ \\ \mathrm{I\!E}(X) = \sum_{i=1} pr(A_i)\; \mathrm{I\!E}\left(X \middle| A_i\right) \end{cases}$$

reobtendo-se assim resultados bem conhecidos.

2.3. - MEDIDAS PRODUTO

Sejam (S_j, Σ_j), $j = 1,2$ espaços mensuráveis em que estão definidas medidas σ-finitas μ_j, $j = 1,2$. Pode mostrar-se que a família Σ_0 dos conjuntos $\bigcup_{j=1}^{N}(C_{1,j} \mathrm{X} C_{2,j})$, com $C_{l,1} \ldots, C_{l,N} \in \Sigma_l, l = 1,2$, é uma álgebra e utilizar o teorema de CARATHEODORY para mostrar que existe uma única medida $\mu_1 \mathrm{X} \mu_2$ definida em $\Sigma_1 \mathrm{X} \Sigma_2 = \sigma(\Sigma_0)$, tal que

(1) $$(\mu_1 \mathrm{X} \mu_2)\left(\bigcup_{j=1}^{N}(C_{1,j} \mathrm{X} C_{2,j})\right) = \sum_{j=1}^{N}\left(\mu_1(C_{1,j}) \times \mu_2(C_{2,j})\right)$$

quando os produtos cartesianos $C_{1,j} \mathrm{X} C_{2,j}$, $j = 1, \ldots, N$, são disjuntos dois a dois. Observe-se que $\mu_1 \mathrm{X} \mu_2$ estará definida no espaço mensurável $(S_1 \mathrm{X} S_2, \Sigma_1 \mathrm{X} \Sigma_2)$, onde $S_1 \mathrm{X} S_2$ representa o produto cartesiano de S_1 por S_2. Teremos assim os espaços mensuráveis $(S_1 \mathrm{X} S_2, \Sigma_1 \mathrm{X} \Sigma_2)$ e de medida $(S_1 \mathrm{X} S_2, \Sigma_1 \mathrm{X} \Sigma_2, \mu_1 \mathrm{X} \mu_2)$ produto, enquanto $\mu_1 \mathrm{X} \mu_2$ será a medida produto.

Seja $g(s_1, s_2) \in m(\Sigma_1 \mathrm{X} \Sigma_2)$ uma função mensurável $\Sigma_1 \mathrm{X} \Sigma_2$, escrevendo-se $g(s_1, s_2) \in l(\Sigma_1 \mathrm{X} \Sigma_2)$ se a mesma for limitada. Se, fixando-se s_1 $[s_2]$ se obtiver sempre uma função $m\Sigma_2$ $[m\Sigma_1]$ e se $\int g(s_1, s_2)\, d(\mu_1 \mathrm{X} \mu_2)$ estiver definido, o teorema de FUBINI, ver WILLIAMS (1997, pg. 77) diz-nos que

(2) $$\int g(s_1, s_2)\, d(\mu_1 \mathrm{X} \mu_2) = \int g_1(s_1) d\mu_1 = \int g_2(s_2) d\mu_2$$

com

(3) $$g_1(s_1) = \int g(s_1, s_2)\, d\mu_2 \quad \text{e} \quad g_2(s_2) = \int g(s_1, s_2)\, d\mu_1 .$$

Em particular, se μ_2 for medida discreta, sendo $s_{2,j}$, $j = 1, \ldots, N$, os pontos do respectivo conjunto associado ter-se-á

(4) $$\int g(s_1, s_2)\, d(\mu_1 \text{ X } \mu_2) = \sum_{j=1}^{N} q_j g_2(s_{2,j}) = \sum_{j=1}^{N} q_j \int g(s_1, s_{2,j})\, d\mu_1 .$$

Vamos agora obter alguns resultados que nos serão úteis.

Admitamos que (S_2, Σ_2, P^0)é espaço de probabilidade e que $P^0 << \mu_2$ com derivada de RADON NIKODIN $\frac{dP^0}{d\mu_2} = p^0(s_2)$ e estabeleçamos a

Proposição 1

Tem-se $\frac{d(\mu_1 \text{ x } P^0)}{d(\mu_1 \text{ x } \mu_2)} = I_{S_1}(s_1) p^0(s_2)$.

Demonstração:

Com $C_j \in \Sigma_j$, $j = 1,2$, vê-se que

$$(\mu_1 \text{ X } P^0)\,(C_1 \text{ X } C_2) = \mu_1(C_1)\, P^0(C_2) = \mu(C_1) \int_{C_2} p^0(s_2) d\mu_2 = \int_{C_1 X C_2} I_{S_1}(s_1)\, p^0(s_2)\, d(\mu_1 \text{ X } \mu_2)$$

logo, para reuniões disjuntas de produtos cartesianos ter-se-á

$$\left(\mu_1 \, X \, P^0\right)\left(\bigcup_{j=1}^{N}\left(C_{1,j} \, X \, C_{2,j}\right)\right) = \int_{\bigcup_{j=1}^{N}\left(C_{1,j} \, X \, C_{2,j}\right)} I_{S_1}(s_1)\, p^0(s_2)\, d\left(\mu_1 \, X \, \mu_2\right)$$

Estas reuniões constituem uma álgebra Σ_0 tal que $\left(\Sigma_1 X \Sigma_2\right) = \sigma\left(\Sigma_0\right)$. Assim, atendendo ao teorema de CARATHEODORY, existe uma única medida definida em $\Sigma_1 X \Sigma_2$, cujo valor para $C_1 \times C_2$ é $\int_{C_1 \times C_2} I_{S_1}(s_1)\, p^0(s_2)\, d\left(\mu_1 \, X \, \mu_2\right)$.

Esta medida é, para $\overline{C} \in \Sigma_1 X \Sigma_2$,

$\int_{\overline{C}} I_{S_1}(s_1)\, p^0(s_2)\, d\left(\mu_1 \, X \, \mu_2\right)$ o que estabelece a tese. $\square$

Consideremos agora uma medida de probabilidade P definida em $\Sigma_1 X \Sigma_2$ e estabeleçamos a

Proposição 2

Tendo-se $\dfrac{dP}{d\left(\mu_1 \, X \, P^0\right)} = p(s_1,s_2)$ tem-se $\dfrac{dP}{d\left(\mu_1 \, X \, \mu_2\right)} = p(s_1,s_2)\, p^0(s_2)$.

Demonstração :

Comecemos por observar que $p(s_1,s_2) \in m\left(\Sigma_1 X \Sigma_2\right)^+$ e, dado $g(s_1,s_2) \in FS\left(\Sigma_1 X \Sigma_2\right)^+$, se tem $g(s_1,s_2) = \sum_{j=1}^{N} a_j I_{\overline{C}_j}(s_1,s_2)$, vindo, qualquer que seja $\overline{C} \in \Sigma_1 X \Sigma_2$

$$\int_{\bar{C}} g(s_1,s_2)\, d(\mu_1 \mathrm{X}\, P^0) = \sum_{j=1}^{N} a_j \int_{\bar{C}_j \cap \bar{C}} I_{S_1}(s_1)\, I_{S_2}(s_2)\, d(\mu_1 \mathrm{X}\, P^0) =$$

$$= \sum_{j=1}^{N} a_j \int_{\bar{C}_j \cap \bar{C}} I_{S_1}(s_1)\, p^0(s_2)\, d(\mu_1 \mathrm{X}\, \mu_2) = \int_{\bar{C}} \left(\sum_{j=1}^{N} a_j\, I_{\bar{C}_j}(s_1,s_2) \right) p^0(s_2)\, d(\mu_1 \mathrm{X}\, \mu_2) =$$

$$= \int_{\bar{C}} g(s_1,s_2)\, p^0(s_2)\, d(\mu_1 \mathrm{X}\, \mu_2)$$

logo, tomando o supremo, para as funções de $FS(\Sigma_1 \mathrm{x} \Sigma_2)^+$ majoradas por $p(s_1,s_2)$ obtém-se

$$\int_{\bar{C}} p(s_1,s_2)\, d(\mu_1 \mathrm{x}\, P^0) = \int_{\bar{C}} p(s_1,s_2)\, p^0(s_2)\, d(\mu_1 \mathrm{x}\, \mu_2)$$

o que estabelece a tese . $\square$

Atendendo ao teorema de FUBINI, com

$$(5) \qquad \begin{cases} l_1(s_1) = \int p(s_1,s_2)\, dP^0 = \int p(s_1,s_2)\, p^0(s_2)\, d\mu \\ l_2(s_2) = \int p(s_1,s_2)\, d\mu_1 \\ h(s_2) = \int p(s_1,s_2)\, p^0(s_2)\, d\mu_1 \end{cases}$$

tem-se

(5) $$\int l_1(s_1)\, d\mu_1 = \int l_2(s_2)\, dP^0 = \int h(s_2)\, d\mu_2 = 1$$

Quando P^0 é discreta, vem

(6) $$l_1(s_1) = \sum_{j=1}^{N} q_j\, p(s_1, s_{2,j}).$$

2.4. - ESPAÇOS PARAMÉTRICOS E ESTATÍSTICAS SUFICIENTES

Suponhamos que para todo o $\vec{\theta}^{k} \in \Omega$ se tem uma medida de probabilidade definida em $(\mathrm{IR}^{n}, \beta_{n})$ absolutamente contínua em ordem à medida λ também definida em $(\mathrm{IR}^{n}, \beta_{n})$. Existem então as derivadas de RADON NIKODIN $p(\vec{x}^{n} | \vec{\theta}^{k})$. Diremos que Ω é um espaço paramétrico. A inferência paramétrica consiste em, conhecidas as derivadas de RADON NIKODIN, utilizar a informação contida em $\vec{x}^{n}$ para estudar questões relativas a $\vec{\theta}^{k}$. Assim, enquanto $\vec{x}^{n}$ é observado, $\vec{\theta}^{k}$ é "desconhecido". Quando

(1) $$\forall \vec{\theta}^{k} : p(\vec{x}^{n} | \vec{\theta}^{k}) = h(\vec{x}^{n})\, g(\vec{t}^{m}(\vec{x}^{n}) | \vec{\theta}^{k}) \quad , \quad \vec{t}^{m} = \vec{t}^{m}(\vec{x}^{n})$$

será estatística suficiente contendo, ver FRASER (1957, pg. 16 a 20), toda a informação dada por $\vec{x}^{n}$ relativa a $\vec{\theta}^{k}$.

Na Teoria de Decisão Estatística considera-se um espaço $\mathcal{D}$ de decisões possíveis. Uma regra de decisão fixa faz corresponder a cada $\vec{x}^{n}$ uma decisão d. Caso exista estatística suficiente deve-se utilizar, ver FRASER (1957, pg. 45) regras de decisão da forma

(2) $$d = d(\vec{t}^{m})$$

isto é, a decisão deve ser tomada em função do valor da estatística suficiente. Intuitivamente compreende-se a razão de ser deste resultado se pensarmos que a estatística suficiente contem toda a informação relativa a $\vec{\theta}^{k}$, dada pelas nossas observações.

3 - ***INFERÊNCIA BAYESIANA***

3 - INFERÊNCIA BAYESIANA

3.1. - O TEOREMA DE BAYES

Antes de formularmos a nossa discussão da Inferência Bayesiana duma forma abstracta convém recordarmos o bem conhecido Teorema de BAYES.

Sendo $A_1,\ldots,A_N$ acontecimentos incompatíveis dois a dois e tais que um deles se verifica sempre, suponhamos que se conhecem as probabilidades à priori $p(A_i), i=1,\ldots,N$ e as $pr(B|A_i)$, $i=1,\ldots,N$. Suponhamos que B se verificou e que a partir desse facto e das probabilidades à priori se pretende saber quais as probabilidades à posteriori $pr(A_i|B)$, $i=1,\ldots,N$ de cada um dos acontecimentos se ter verificado, dado B ter ocorrido. A resposta a esta questão é dada pelo teorema de BAYES. Como, devido ao teorema de probabilidade total se tem

(1) $$pr(B)=\sum_{i=1}^{N} pr(A_i)pr(B|A_i)$$

tem-se

(2) $$pr(A_i|B)=\frac{pr(A_i\cap B)}{pr(B)}=\frac{pr(A_i)\ pr(B|A_i)}{\sum_{l=1}^{N} pr(A_l)\ pr(B|A_l)},i=1,\ldots,N$$

É esta fórmula que traduz o teorema de BAYES. Observe-se aqui que ao conhecimento traduzido pelas probabilidades à priori se juntou o saber-se que B se tinha verificado para calcular as probabilidades à posteriori.

Se A_i e B fossem independentes ter-se-ia $pr(A_i \cap B) = pr(A_i)pr(B)$ bem como $pr(A_i|B) = pr(A_i)$ e o saber-se que B se verificou nada nos diz quanto à probabilidade de A_i se verificar. Assim o esquema que está na base do teorema de BAYES apenas "funciona" quando A_i e B não são independentes.

Sendo B um conjunto de resultados possíveis duma experiência, E o conjunto de todos os resultados e $B^C = \mathrm{E} - B$ o complementar de B, tem-se

(3) $$\Sigma(\{B\}) = \{\varnothing, \mathrm{E}, B, B^C\}$$

que será fechada para a passagem ao complementar e para a reunião numerável sendo portanto uma <u>álgebra σ</u> de sub-conjuntos de E que contêm B. Diremos que $\Sigma(\{B\})$ é a <u>álgebra σ</u> <u>gerada por B</u>. Pondo

(4) $$\begin{cases} P_i(\varnothing) = 0 & ; \ i = 1,\dots,N \\ P_i(E) = 1 & ; \ i = 1,\dots,N \\ P_i(B) = pr(B|A_i) & ; \ i = 1,\dots,N \\ P_i(B^C) = 1 - pr(B|A_i) & ; \ i = 1,\dots,N \end{cases}$$

fazemos corresponder aos A_i $,i = 1,\dots,N$ outras tantas medidas de probabilidade definidas em $(\mathrm{E}, \Sigma(\{B\}))$.

Se aos A_i fizermos corresponder parâmetros $\vec{\theta}_i^k$ $,i = 1,\dots,N$, podemos por, com $C \in \Sigma(\{B\})$

(5) $$P_i(C) = P(C|\vec{\theta}_i^k) \quad ,i = 1,\dots,N$$

3.2.-MEDIDAS À PRIORI E À POSTERIORI

O teorema de BAYES mostra como combinar informações à priori com informação recolhida para obter informação à posteriori. Vamos agora formalizar o estudo dessa combinação de informações através da noção de risco à posteriori. Nos modelos bayesianos Ω serve de suporte a um espaço de probabilidade (Ω, Σ_2, P^0) e a um espaço de medida $(\Omega, \Sigma_2, \mu_2)$ estando definida $p^0(\vec{\theta}^k) = \frac{dP^0}{d\mu_2}$.

Tem-se ainda :

a) um espaço de medida $(\mathrm{IR}^n, \beta_n, \mu_1)$;

b) uma medida de probabilidade P definida em $\beta_n \times \Sigma_2$ com derivada de RADON-NIKODIM, $p(\vec{x}^n | \vec{\theta}^k) = \frac{dP}{d(\mu_1 \times P^0)}$ e, consequentemente, tendo-se

$$(1) \qquad \frac{dP}{d(\mu_1 \times \mu_2)} = p(\vec{x}^n, \vec{\theta}^k) p^0(\vec{\theta}^k).$$

A nossa informação à priori está contida no produto $p(\vec{x}^n, \vec{\theta}^k) p^0(\vec{\theta}^k)$ que designaremos por derivada de RADON-NIKODIM conjunta. Com base nesta informação e na contida no vector $\vec{x}^n$ das observações, procuramos fundamentar decisões relativas a $\vec{\theta}^k$.

Seja $\mathcal{D}$ o espaço das decisões possíveis. No caso do teste de hipóteses tinha-se $\mathcal{D} = \{d_1, d_2\}$, com d_1 aceitar e d_2 rejeitar H_0.

Admitamos que a $\vec{\theta}^k \in \Omega$ podemos fazer corresponder $d_0 = d_0(\vec{\theta}^k)$ que representa a decisão certa. No exemplo que temos estado a considerar ter-se-ia $d_0 = d_1$ $[d_2]$caso $\vec{\theta}^k \in \omega$ $[\vec{\theta}^k \notin \omega]$. Por outro lado, uma regra de decisão fixa [aleatória] faz corresponder a $\vec{x}^n$ uma decisão $\breve{d} = \tilde{d}(\vec{x}^n)$ [uma medida de probabilidade $\tilde{P}_{\vec{x}^n}$ definida em $(\mathcal{D},\Sigma)$]. Quando se utiliza uma regra de decisão fixa a partir de $\vec{x}^n$ obtém-se directamente a decisão, ao passo que, se a regra de decisão for aleatória, a partir de $\vec{x}^n$ apenas se obtém uma distribuição de probabilidades. Por exemplo, se $\mathcal{D} = \{d_1, d_2\}$ e se $P_{\vec{x}^n}$ der a probabilidade $\frac{1}{2}$ a ambas as decisões possíveis pode-se atirar uma moeda ao ar para optar.

Ao decidir-se encorre-se em custos sempre que não se toma a decisão correcta d_0. Para quantificar esses custos utilizam-se funções de custo : $W(d,d_0) \geq 0$, onde d é a decisão tomada. Obviamente ter-se-á $W(d_0,d_0) = 0$. Ponhamos, para uma regra de decisão fixa

$$(2) \qquad \overline{W}(\vec{x}^n,\vec{\theta}^k) = W\left(\tilde{d}(\vec{x}^n), d_0(\vec{\theta}^k)\right)$$

e para uma regra de decisão aleatória

$$(3) \qquad \overline{W}(\vec{x}^n,\vec{\theta}^k) = \int_{\mathcal{D}} W\left(d, d_0(\vec{\theta}^k)\right) d\tilde{P}_{\vec{x}^n}(d)$$

Assim, no primeiro caso, $\overline{W}(\vec{x}^n,\vec{\theta}^k)$ dá-nos o valor da perda quando $\vec{x}^n$ é o vector das observações e, no segundo caso, o valor médio dessa perda.

Recordemos agora que $(\vec{x}^n,\vec{\theta}^k)$é o ponto genérico do suporte IR^n X Ω dum espaço de probabilidade. Assim, o valor médio da perda será, atendendo ao teorema de FUBINI

$$(4) \qquad \mathbb{E}(\overline{W}) = \int_{\mathbb{R}^n \times \Omega} \overline{W}(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, p^0(\vec{\theta}^{\,k})\, d(\mu_1 \times \mu_2) =$$

$$= \int_{\mathbb{R}^n} \mathbb{E}(\overline{W} | \vec{x}^{\,n})\, l_1(\vec{x}^{\,n})\, d\mu_1(\vec{x}^{\,n})$$

com

$$(5) \qquad l_1(\vec{x}^{\,n}) = \int p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, dP^0(\vec{\theta}^{\,k}) = \int p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, p^0(\vec{\theta}^{\,k})\, d\mu_2(\vec{\theta}^{\,k})$$

e

$$(6) \qquad \mathbb{E}(\overline{W} | \vec{x}^{\,n}) = \int \overline{W}(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, \frac{p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, p^0(\vec{\theta}^{\,k})}{l_1(\vec{x}^{\,n})}\, d\mu_2(\vec{\theta}^{\,k})$$

Esta última função é o risco à posteriori (dado $\vec{x}^{\,n}$) da regra de decisão $\tilde{d}$ medindo a perda média a que a mesma conduz uma vez obtido $\vec{x}^{\,n}$.

Se conseguirmos para todo o $\vec{x}^{\,n}$ minimizar $\mathbb{E}(\overline{W} | \vec{x}^{\,n})$, minimizamos $\mathbb{E}(\overline{W})$ tendo-se pois uma regra de decisão óptima. Suponhamos que existe $\check{d} = \check{d}(\vec{x}^{\,n})$, tal que, $\forall\ \vec{x}^{\,n}$

$$(7) \qquad \int W\left(\check{d}, d_0(\vec{\theta}^{\,k})\right) p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, dP^0 =$$

$$= \mathrm{Min}\left\{ \int W\left(d, d_0(\vec{\theta}^{\,k})\right) p(\vec{x}^{\,n}, \vec{\theta}^{\,k})\, dP^0 ; d \in \mathcal{D} \right\}$$

tomando-se $\tilde{d}(\vec{x}^{\,n}) = \check{d}(\vec{x}^{\,n})$, ter-se-ia a regra óptima atrás referida. Assim a solução óptima para problemas de decisão bayesiana é obtida minimizando-se o risco à posteriori. Quando o problema é de estimação tem-se $d_0\left(\vec{\theta}^k\right) = \vec{\theta}^k$ e caso exista $\check{d}\left(\vec{x}^n\right)$, $\tilde{\vec{\theta}}\left(\vec{x}^n\right) = \check{d}\left(\vec{x}^n\right)$ será um estimador Bayesiano óptimo.

Neste caso $\check{d} = \check{\vec{\theta}}^k$ e $d_0\left(\vec{\theta}^k\right) = \vec{\theta}^k$, pelo que o estimador é obtido minimizando

$$(8) \qquad R_p\left(\vec{\theta}'^k, \vec{x}^n\right) = \int W\left(\vec{\theta}'^k, \vec{\theta}^k\right) \frac{p\left(\vec{x}^n, \vec{\theta}^k\right) p^0\left(\vec{\theta}^k\right)}{l_1\left(\vec{x}^n\right)} d\mu_2\left(\vec{\theta}^k\right) =$$

$$= \int W\left(\vec{\theta}'^k, \vec{\theta}^k\right) \frac{p\left(\vec{x}^n, \vec{\theta}^k\right)}{l_1\left(\vec{x}^n\right)} dP^0\left(\vec{\theta}^k\right)$$

e esta função dá-nos o risco à posteriori de se “optar” por $\vec{\theta}'^k$ tendo-se “colhido” $\vec{x}^n$.

Claramente relacionada com a estimação Bayesiana está a noção de Região de Risco à Posteriori Controlado de nível α, formada pelos $\vec{\theta}'^k$ tais que

$$(9) \qquad \int W\left(\vec{\theta}'^k, \vec{\theta}^k\right) \frac{p\left(\vec{x}^n, \vec{\theta}^k\right) p^0\left(\vec{\theta}^k\right)}{l_1\left(\vec{x}^n\right)} d\mu_2\left(\vec{\theta}^k\right) \leq \alpha$$

Esta região será representada por $\Gamma_\alpha\left(\vec{x}^n\right)$.

4 - *ANÁLISE DISCRIMINANTE*

4-ANÁLISE DISCRIMINANTE

4.1. - DUAS POPULAÇÕES : CASO GERAL

Consideremos uma amostra de duas populações G_1 e G_2 sendo A_1 e A_2 os acontecimentos que se verificam quando um elemento, tirado ao acaso da mistura, pertence a G_1 ou a G_2. Ponhamos ainda $q_j = pr(A_j)$, $j = 1{,}2$. O problema que se tem, usualmente é: dado um elemento pertencente à mistura atribuí-lo a G_1 ou G_2 com base em "*n*" medidas realizadas nesse elemento. Uma alternativa é admitir que se "colheu" uma amostra de dimensão n e em cada um dos elementos se realizou uma medida. Se admitirmos que todos os elementos da amostra têm de pertencer à mesma população, poder-se-á admitir que as observações são *i.i.d.* o que não se poderá fazer na formulação usual do problema. Para já poremos de parte a hipótese das observações serem *i.i.d.* .

Seja então $\vec{X}^n$ o vector aleatório cujas componentes são as nossas observações. Admitamos que, quando A_j se verifica e o elemento pertence a G_j, $\vec{X}^n$ está associado a uma medida de probabilidade $P_j()$, isto é, com $C \in \beta_n$,

(1) $$pr\left(\vec{X}^n \in C\right) = P_j(C) \quad ,j = 1{,}2$$

e que estas medidas de probabilidade têm derivada de RADON-NIKODIM

(2) $$p_j\left(\vec{x}^n\right) = \frac{dP_j}{d\lambda} \quad ,j = 1{,}2 \ ,$$

com λ uma medida definida em $\left(\mathrm{IR}^n, \beta_n\right)$. Podemos fazer corresponder estas medidas e as respectivas derivadas de RADON-NIKODIM a parâmetros

$\vec{\theta}_j^k \in \Omega \quad, j = 1{,}2$. Em $(\Omega, P(\Omega))$ está definida a medida de probabilidade discreta dada por

$$(3) \qquad \varphi\left(\left\{\vec{\theta}_j^k\right\}\right) = q_j \quad, j = 1{,}2$$

vindo

$$(4) \qquad l_1\left(\vec{x}^n\right) = \sum_{j=1}^{2} q_j p\left(\vec{x}^n; \vec{\theta}_j^k\right)$$

com $p\left(\vec{x}^n; \vec{\theta}_j^k\right) = p_j\left(\vec{x}^n\right)$, e tendo-se as

$$(5) \qquad l\left(\vec{\theta}_j^k \middle| \vec{x}^n\right) = \frac{q_j p\left(\vec{x}^n; \vec{\theta}_j^k\right)}{\sum_{l=1}^{2} q_l p\left(\vec{x}^n; \vec{\theta}_l^k\right)}$$

Por outro lado, neste caso há apenas duas decisões possíveis:

d_1 - o elemento é atribuído a G_1;

d_2 - o elemento é atribuído a G_2;

o que simplifica a "estrutura dos custos". Assim, pode-se atribuír custo nulo a uma atribuição correcta e considerar os custos :

$C_{1.2}$ - quando um elemento de G_1 é atribuído a G_2;

$C_{2.1}$ - quando um elemento de G_2 é atribuído a G_1;

Para aligeirar a escrita poremos $C_1 = C_{1.2}$ e $C_2 = C_{2.1}$.

Observe-se que a correspondência atrás efectuada entre populações e parâmetros se pode realizar sempre e reduz o problema da atribuição a um problema de estimação. Assim, ao atribuirmos o elemento estamos a "escolher" o parâmetro correspondente. Teremos então

$$(6)\qquad \begin{cases} W\left(\vec{\theta}_1^{\,k}, \vec{\theta}_1^{\,k}\right) = 0 \\[2ex] W\left(\vec{\theta}_1^{\,k}, \vec{\theta}_2^{\,k}\right) = C_{1,2} = C_1 \\[2ex] W\left(\vec{\theta}_2^{\,k}, \vec{\theta}_1^{\,k}\right) = C_{2,1} = C_2 \\[2ex] W\left(\vec{\theta}_2^{\,k}, \vec{\theta}_2^{\,k}\right) = 0 \end{cases}$$

vindo

$$\begin{cases} R_p\left(\vec{\theta}_1^{\,k} \middle| \vec{x}^{\,n}\right) = C_2 \; l\left(\vec{\theta}_1^{\,k} \middle| \vec{x}^{\,n}\right) = \dfrac{q_2 \; C_2 \; p\left(\vec{x}^{\,n} \middle| \vec{\theta}_1^{\,k}\right)}{l_1\left(\vec{x}^{\,n}\right)} \\[3ex] R_p\left(\vec{\theta}_2^{\,k} \middle| \vec{x}^{\,n}\right) = C_1 \; l\left(\vec{\theta}_2^{\,k} \middle| \vec{x}^{\,n}\right) = \dfrac{q_1 \; C_1 \; p\left(\vec{x}^{\,n} \middle| \vec{\theta}_2^{\,k}\right)}{l_2\left(\vec{x}^{\,n}\right)} \end{cases}$$

Assim, se pusermos

$$(8)\qquad \begin{cases} E_1 = \left\{\vec{x}^n; R_p\left(\vec{\theta}_1^k \middle| \vec{x}^n\right) < R_p\left(\vec{\theta}_2^k \middle| \vec{x}^n\right)\right\} = \left\{\vec{x}^n; \dfrac{p\left(\vec{x}^n \middle| \vec{\theta}_2^k\right)}{p\left(\vec{x}^n \middle| \vec{\theta}_1^k\right)} < \dfrac{q_2\ C_2}{q_1\ C_1}\right\} \\ \\ E_2 = \left\{\vec{x}^n; R_p\left(\vec{\theta}_1^k \middle| \vec{x}^n\right) > R_p\left(\vec{\theta}_1^k \middle| \vec{x}^n\right)\right\} = \left\{\vec{x}^n; \dfrac{p\left(\vec{x}^n \middle| \vec{\theta}_2^k\right)}{p\left(\vec{x}^n \middle| \vec{\theta}_1^k\right)} > \dfrac{q_2\ C_2}{q_1\ C_1}\right\} \\ \\ \pi = \left\{\vec{x}^n; R_p\left(\vec{\theta}_1^k \middle| \vec{x}^n\right) = R_p\left(\vec{\theta}_2^k \middle| \vec{x}^n\right)\right\} = \left\{\vec{x}^n; \dfrac{p\left(\vec{x}^n \middle| \vec{\theta}_2^k\right)}{p\left(\vec{x}^n \middle| \vec{\theta}_1^k\right)} = \dfrac{q_2\ C_2}{q_1\ C_1}\right\} \end{cases}$$

somos levados a atribuír o elemento a G_j quando $\vec{x}^n \in E_j$, $j = 1{,}2$. Quando

$$(9)\qquad \frac{p\left(\vec{x}^n \middle| \vec{\theta}_2^k\right)}{p\left(\vec{x}^n \middle| \vec{\theta}_1^k\right)} = \frac{q_2\ C_2}{q_1\ C_1}$$

os riscos à posteriori serão iguais podendo a atribuição ser satisfeita por um processo aleatório. Ao definirmos E_1 e E_2 estamos a definir uma regra de decisão, a qual será aleatória se, quando (9) se verificar, a atribuição for feita por um processo aleatório. Então para cada $\vec{x}^n$ teremos uma medida de probabilidade $P^0\left(\mathrm{C} \mid \vec{x}^n\right)$ definida em $\{\mathcal{D}, \mathcal{P}(\mathcal{D})\}$. Observe-se que

$$(10)\qquad \begin{cases} \vec{x}^n \in E_1;\ P^0\left(\{d_1\} \middle| \vec{x}^n\right) = 1 \\ \\ \vec{x}^n \in E_2;\ P^0\left(\{d_1\} \middle| \vec{x}^n\right) = 0 \\ \\ \vec{x}^n \in \pi;\ P^0\left(\{d_1\} \middle| \vec{x}^n\right) = p' \in [0,1] \end{cases}$$

Consideremos agora um exemplo de aplicação em que as $p\left(\vec{x}^n \middle| \vec{\theta}_j^k\right)$, $j = 1{,}2$ são densidades normais

(11) $$n\left(\vec{x}^n \middle| \vec{\mu}_j^n, V_j\right) = \frac{e^{-\frac{1}{2}\left(\vec{x}^n - \vec{\mu}_j^n\right)^T V_j^{-1}\left(\vec{x}^n - \vec{\mu}_j^n\right)}}{(2\pi)^{\frac{n}{2}}\sqrt{\det\left(V_j\right)}}; \quad j = 1{,}2$$

com vectores médios $\vec{\mu}_j^n$ e matrizes de covariância V_j, $j = 1{,}2$. Representando por $x_1, \ldots, x_n$ $\left[\mu_{1,j}, \ldots, \mu_{n,j}\right]$ as componentes de $\vec{x}^n$ $\left[\vec{\mu}_j^n\right]$ e pondo V_j^{-1} $\left[\sigma_j^{i,l}\right]$, $j = 1{,}2$ temos, atendendo à simetria das matrizes V_j^{-1}, $j = 1{,}2$,

(12) $$Q\left(\vec{x}^n\right) = \left(\vec{x}^n - \vec{\mu}_1^n\right)^T V_1^{-1}\left(\vec{x}^n - \vec{\mu}_1^n\right) - \left(\vec{x}^n - \vec{\mu}_2^n\right)^T V_2^{-1}\left(\vec{x}^n - \vec{\mu}_2^n\right) =$$

$$= \sum_{i=1}^{n}\sum_{l=1}^{n}\sigma_1^{i,l}\left(x_i - \mu_{i,1}\right)\left(x_l - \mu_{l,1}\right) - \sum_{i=1}^{n}\sum_{l=1}^{n}\sigma_2^{i,l}\left(x_i - \mu_{i,2}\right)\left(x_l - \mu_{l,2}\right) =$$

$$= \sum_{i=1}^{n}\sum_{l=1}^{n}\left[\ \sigma_1^{i,l}\left(x_i x_l - x_i\mu_{l,1} - x_l\mu_{i,1} + \mu_{i,1}\mu_{l,1}\right) - \sigma_2^{i,l}\left(x_i x_l - x_i\mu_{l,2} - x_l\mu_{i,2} + \mu_{i,2}\mu_{l,2}\right)\ \right]$$

$$= \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_1^{i,l} - \sigma_2^{i,l}\right)x_i x_l + \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_2^{i,l}\mu_{l,2} - \sigma_1^{i,l}\mu_{l,1}\right)x_i + \sum_{l=1}^{n}\sum_{i=1}^{n}\left(\sigma_2^{\mathrm{i,l}}\mu_{i,2} - \sigma_1^{i,l}\mu_{i,1}\right)x_l$$

$$+ \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_1^{i,l}\mu_{i,1}\mu_{l,1} - \sigma_2^{i,l}\mu_{i,2}\mu_{l,2}\right) =$$

$$= \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_1^{i,l} - \sigma_2^{i,l}\right)x_i x_l + \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_2^{i,l}\mu_{l,2} - \sigma_1^{i,l}\mu_{l,1}\right)x_i + \sum_{i=1}^{n}\sum_{l=1}^{n}\left(\sigma_1^{\mathrm{i,l}}\mu_{i,1}\mu_{l,1} - \sigma_2^{i,l}\mu_{i,2}\mu_{l,2}\right)$$

Ora, com $k = \frac{q_2}{q_1}\frac{C_2}{C_1}$ e $U_j = \det(V_j)$, $j = 1{,}2$, tem-se

(13) $$E_1 = \left\{ \vec{x}^n;\ \frac{n\left(\vec{x}^n \middle| \vec{\mu}_2^n, V_2\right)}{n\left(\vec{x}^n \middle| \vec{\mu}_1^n, V_1\right)} \le k \right\} = \left\{ \vec{x}^n; \log \frac{n\left(\vec{x}^n \middle| \vec{\mu}_2^n, V_2\right)}{n\left(\vec{x}^n \middle| \vec{\mu}_1^n, V_1\right)} \le \log k \right\} =$$

$$= \left\{ \vec{x}^n;\ Q\left(\vec{x}^n\right) \le 2\log k + \log \frac{V_2}{V_1} \right\}$$

A função $Q\left(\vec{x}^n\right)$ será uma função discriminante quadrática. Um caso particular interessante é aquele em que $V_1 = V_2 = V$. Então, com $V^{-1} = \left[\sigma^{i,l}\right]$, tem-se

(14) $$Q\left(\vec{x}^n\right) = 2\sum_{i=1}^{n}\sum_{l=1}^{n} \sigma^{i,l}\left(\mu_{l,2} - \mu_{l,1}\right) x_i + \sum_{i=1}^{n}\sum_{l=1}^{n} \sigma^{i,l}\left(\mu_{i,1}\mu_{l,1} - \mu_{i,2}\mu_{l,2}\right)$$

$$= \sum_{i=1}^{n} a_i x_i + b$$

, com

(15) $$\begin{cases} a_i = 2\sum_{l=1}^{n} \sigma^{i,l}\left(\mu_{l,2} - \mu_{l,1}\right) \\ \\ b = \sum_{i=1}^{n}\sum_{l=1}^{n} \sigma^{i,l}\left(\mu_{i,1}\mu_{l,1} - \mu_{i,2}\mu_{l,2}\right) \end{cases}$$

Então, dada a função discriminante linear

$$(16) \qquad L(\vec{x}^n) = \sum_{i=1}^{n} a_i x_i$$

ter-se-á

$$(17) \qquad \begin{cases} E_1 = \left\{\vec{x}^n;\ L(\vec{x}^n) < k^0\right\} \\ E_2 = \left\{\vec{x}^n; L(\vec{x}^n) > k^0\right\} \\ \pi = \left\{\vec{x}^n; L(\vec{x}^n) = k^0\right\} \end{cases}$$

com $k^0 = 2\ \log k - b$.

Observe-se que, quando as $p\left(\vec{x}^n \middle| \vec{\theta}^k\right)$ são densidades,

$$(18) \qquad p\left(\vec{x}^n \in \pi \middle| \vec{\theta}_1^k\right) = pr\left(\vec{x}^n \in \pi \middle| \vec{\theta}_2^k\right) = 0$$

pelo que não se justifica tratar à parte este caso. Pode-se incluír o conjunto dos $\vec{x}^n$ que satisfazem a igualdade em E_1 ou em E_2.

Sendo Z a variável aleatória que nos dá o custo da atribuição, tem-se

$$(19) \qquad \mathbb{E}(Z) = q_1\ \mathbb{E}\left(Z \middle| A_1\right) + q_2\ \mathbb{E}\left(Z \middle| A_2\right)$$

Admitamos que se verificava A_1, isto é, que o elemento pertence a G_1. Então o custo é C_1 quando o elemento é atribuído a G_2, o que acontece com uma probabilidade p_1^+ que satisfaz as desigualdades

(20) $$P_1(E_1) \le p_1^+ \le 1 - P_1(E_2)$$

tendo-se pois

(21) $$\mathrm{IE}(Z|A_1) = p_1^+ C_1$$

Analogamente, tem-se

(22) $$\mathrm{IE}(Z|A_2) = p_2^+ C_2$$

com

(23) $$P_2(E_2) \le p_2^+ \le 1 - P_2(E_1)$$

Quando as $p(\vec{x}^n|\vec{\theta}^k)$, $j = 1,2$ são densidades tem-se

(24) $$P_j(E_1) + P_j(E_2) = 1, \; j = 1,2$$

vindo

(25) $$p_j^+ = P_j(E_j), \; j = 1,2$$

Retomemos o último exemplo. Quando o elemento é "colhido" em G_j, $L(\vec{x}^n)$ tem, ver MEXIA (1995, pg. 38), distribuição normal $\mathcal{N}(l|\eta_j, \theta)$ com valor médio $\eta_j = \sum_{i=1}^{n} \mu_{i,j} a_i$ e variância $\theta = \sum_{i=1}^{n}\sum_{l=1}^{n} \sigma_{i,l}\, a_i\, a_l$, $j = 1,2$, pelo que

$$
(26) \qquad \begin{cases} p_1^+ = \mathcal{N}\left(k^0 \middle| \eta_1, \theta\right) = \mathcal{N}\left(\dfrac{k^0 - \eta_1}{\sqrt{\theta}} \middle| \, 0{,}1\right) \\ \\ p_2^+ = 1 - \mathcal{N}\left(k^0 \middle| \eta_2, \theta\right) = 1 - \mathcal{N}\left(\dfrac{k^0 - \eta_2}{\sqrt{\theta}} \middle| \, 0{,}1\right) \end{cases}
$$

Recorde-se que a distribuição normal reduzida $\mathcal{N}(x | \, 0{,}1)$ se encontra tabelada.

4.2.-DUAS POPULAÇÕES E ESTATÍSTICA SUFICIENTE

Teremos agora

(1) $$l\left(\vec{\theta}_j^k \middle| \vec{x}^n\right) = \frac{q_j g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_j^k\right)}{\overline{g}\left(\vec{t}^{\,m}\left(\vec{x}^n\right)\right)};\ j = 1,2$$

com

(2) $$\overline{g}\left(\vec{t}^{\,m}\left(\vec{x}^n\right)\right) = \sum_{i=1}^{2} q_i\, g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_i^k\right)$$

, vindo

(3) $$\begin{cases} E_1 = \left\{ \vec{x}^n ; \dfrac{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_2^k\right)}{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_1^k\right)} < \dfrac{q_2\, C_2}{q_1\, C_1} \right\} \\[2em] E_2 = \left\{ \vec{x}^n ; \dfrac{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_2^k\right)}{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_1^k\right)} > \dfrac{q_2\, C_2}{q_1\, C_1} \right\} \\[2em] \pi = \left\{ \vec{x}^n ; \dfrac{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_2^k\right)}{g\left(\vec{t}^{\,m}\left(\vec{x}^n\right), \vec{\theta}_1^k\right)} = \dfrac{q_2 C_2}{q_1 C_1} \right\} \end{cases}$$

isto é, $\vec{x}^n$ intervem apenas através de $\vec{t}^{\,m} = \vec{t}^{\,m}\left(\vec{x}^n\right)$ na definição das regiões E_1 e E_2. Assim a distribuição é feita em função do valor da estatística suficiente o que nos permite rescrever as regiões E_1 e E_2 como

$$(4)\qquad \begin{cases} E_1^0 = \left\{ \vec{t}^{\,m}; \dfrac{g(\vec{t}^{\,m},\vec{\theta}_2^{\,k})}{g(\vec{t}^{\,m},\vec{\theta}_1^{\,k})} < \dfrac{q_2\, C_2}{q_1\, C_1} \right\} \\ \\ E_2^0 = \left\{ \vec{t}^{\,m}; \dfrac{g(\vec{t}^{\,m},\vec{\theta}_2^{\,k})}{g(\vec{t}^{\,m},\vec{\theta}_1^{\,k})} > \dfrac{q_2\, C_2}{q_1\, C_1} \right\} \\ \\ \pi^0 = \left\{ \vec{t}^{\,m}; \dfrac{g(\vec{t}^{\,m}\vec{\theta}_2^{\,k})}{g(\vec{t}^{\,m}\vec{\theta}_1^{\,k})} = \dfrac{q_2\, C_2}{q_1\, C_1} \right\} \end{cases}$$

Tal como na alínea anterior, quando as $p(\vec{x}^{\,n}|\vec{\theta}_j^{\,k})$, $j = 1{,}2$, são densidades não é necessário dar um tratamento especial ao caso em que

$$(5)\qquad \frac{g(\vec{t}^{\,m},\vec{\theta}_2^{\,k})}{g(\vec{t}^{\,m},\vec{\theta}_1^{\,k})} = \frac{q_2\, C_2}{q_1\, C_1}$$

podendo incluí-lo em E_1^0 ou E_2^0.

O tratamento simplifica-se particularmente quando m=1 e

$$(6)\qquad m(\vec{t}^{\,m}) = \frac{g(\vec{t}^{\,m},\vec{\theta}_2^{\,k})}{g(\vec{t}^{\,m},\vec{\theta}_1^{\,k})}$$

é monótona. Por exemplo, se m for crescente e $t \geq 0$, pode tomar-se

$$(7) \qquad \begin{cases} E_1 = \left[0 \,;\, m^{-1}\left(\dfrac{q_2\, C_2}{q_1\, C_1}\right)\right] \\ \\ E_2 = \left]m^{-1}\left(\dfrac{q_2\, C_2}{q_1\, C_1}\right) ;\ +\infty\right[\end{cases}$$

Vejamos um exemplo. Admitamos que, com V conhecida

$$(8) \qquad \begin{cases} p\left(\vec{x}^{\,n}, \theta_1\right) = n\left(\vec{x}^{\,n} \middle| \vec{0}^{\,n}, \theta_1\, V\right) = \dfrac{e^{-\frac{1}{2\theta_1}t}}{(2\pi)^{\frac{n}{2}}\ \theta_1^n \sqrt{\det V}} \\ \\ p\left(\vec{x}^{\,n}, \theta_2\right) = n\left(\vec{x}^{\,n} \middle| \vec{0}^{\,n}, \theta_2\, V\right) = \dfrac{e^{-\frac{1}{2\theta_2}t}}{(2\pi)^{\frac{n}{2}}\ \theta_2^n \sqrt{\det V}} \end{cases}$$

onde

$$(9) \qquad t = \vec{x}^{\,n^T} V^{-1} \vec{x}^{\,n}$$

Observe-se que, dado V e V^{-1} serem conhecidas, t pode ser calculada sendo portanto uma estatística. Teremos agora

$$(10) \qquad m(t) = \left(\frac{\theta_1}{\theta_2}\right)^n e^{\frac{\theta_2 - \theta_1}{2\theta_1\theta_2}t}$$

Se $\theta_1 < \theta_2$, esta função será crescente tendo-se, com $k = \frac{q_2\, C_2}{q_1\, C_1}$

$$(11) \qquad \begin{cases} E_1 = \left[0\ ;\ \frac{2\theta_1\,\theta_2}{\theta_2\theta_1}\left(\log k - n\ \log\frac{\theta_2}{\theta_1}\right)\right] \\ \\ E_2 = \left] \frac{2\theta_1\,\theta_2}{\theta_2 - \theta_1}\left(\log k - n\ \log\frac{\theta_1}{\theta_2}\right)\ ;\ +\infty \right[\end{cases}$$

Ponhamos para simplificar a escrita

$$(12) \qquad k^0 = \frac{2\theta_1\theta_2}{\theta_2 - \theta_1}\left(\log k - n\ \log\frac{\theta_1}{\theta_2}\right)$$

Ora, quando $\vec{X}^n$ tem a densidade $n\left(\vec{x}^n \middle| \vec{0}^n, \theta_j\, V\right)$, t distribui-se ver MEXIA (1995, pg. 52) como o produto de θ_j por um chi-quadrado central com n graus de liberdade, tendo pois ver MEXIA (1995, pg. 46) a densidade

$$(13) \qquad f_j(t) = \begin{cases} 0 & ; t < 0 \\ \frac{1}{2\Gamma\left(\frac{n}{2}\right)\theta_j}\left(\frac{t}{\theta_j}\right)^{\frac{n}{2}-1} e^{-\frac{t}{\theta_j}} & ; t > 0 \end{cases} \ ; \quad j = 1,2$$

a que corresponde a distribuição $F_j(t)$, $j = 1,2$. Neste caso, é fácil de calcular o custo médio $\mathbb{E}(Z)$. Assim tem-se

$$(14) \qquad \mathbb{E}(Z) = q_1\ \mathbb{E}\left(Z \middle| A_1\right) + q_2\ \mathbb{E}\left(Z \middle| A_2\right)$$

com

$$(15) \qquad \begin{cases} \mathbb{E}(Z|A_1) = C_1\left(1 - F_1(k^0)\right) \\ \\ \mathbb{E}(Z|A_2) = C_2\, F_2(k^0) \end{cases}$$

4.3.-MAIS DE DUAS POPULAÇÕES :CASO GERAL

Ao considerarmos mais de duas populações passa a fazer sentido utilizar Regiões de Risco à Posteriori Controlado. Assim, a par da atribuição a uma população consideremos o problema de atribuír ao conjunto das populações a que correspondem os riscos à posteriori inferiores ao nível α, fixado em função da natureza do problema em estudo.

Por outro lado, temos de considerar os custos $C_{i,j} = W(\vec{\theta}_i^k, \vec{\theta}_j^k)$, $i = 1,\dots,J,\ j = 1,\dots J$, com $C_{i,j} > 0,\ i \neq j$. Observe-se que, em geral, a matriz $C = \left[C_{i,j}\right]$ não é simétrica. Temos agora definida em Ω uma medida de probabilidade discreta que aos parâmetros $\vec{\theta}_j^k$, $j = 1,\dots,J$, associa os "pesos" q_j, $j = 1,\dots,J$, isto é, $P^0\left(\left\{\vec{\theta}_j^k\right\}\right)$, $j = 1,\dots,J$, pelo que se terá

$$(1) \qquad R_p\left(\vec{\theta}_i^k \middle| \vec{x}^n\right) = \sum_{j=1}^{J} C_{i,j} \frac{p\left(\vec{x}^n, \vec{\theta}_j^k\right)}{l_1\left(\vec{x}^n\right)} q_j, \quad i = 1,\dots,J$$

visto ter-se

$$(2) \qquad \int W\left(\vec{\theta}_i^k, \vec{\theta}^k\right) p\left(\vec{x}^n, \vec{\theta}^k\right) dP^0 = \sum_{j=1}^{J} C_{i,j}\, p\left(\vec{x}^n, \vec{\theta}_j^k\right) q_j, \quad i = 1,\dots,J$$

Para comparar as $R_p\left(\vec{\theta}_i^k \middle| \vec{x}^n\right)$, $i = 1,\dots,J$, basta comparar os

$$(3) \qquad U\left(\vec{\theta}_i^k, \vec{x}^n\right) = \sum_{j=1}^{J} C_{i,j}\, p\left(\vec{x}^n, \vec{\theta}_j^k\right) q_j, \quad i = 1,\dots,J$$

Suponhamos agora que $C_{i,j} = \left(1-\delta_{i,j}\right)C$; diremos então que estamos no caso de custos idênticos, tendo-se

(4) $$U\left(\vec{\theta}_i^k, \vec{x}^n\right) = C\left[\sum_{j=1}^{J} q_j\, p\left(\vec{x}^n, \vec{\theta}_j^k\right) - q_i\, p\left(\vec{x}^n, \vec{\theta}_i^k\right)\right],\ i = 1,\ldots,J$$

bastando, neste caso, comparar os $q_i\, p\left(\vec{x}^n, \vec{\theta}_i^k\right)$, $i = 1,\ldots,J$. Agora atribui-se o elemento à população a que corresponde o maior valor $q_i\, p\left(\vec{x}^n, \vec{\theta}_i^k\right)$.

Quando existem empates, por exemplo, tendo-se

(5) $$U\left(\vec{\theta}_i^k, \vec{x}^n\right) = U\left(\vec{\theta}_l^k, \vec{x}^n\right) = \text{Min}\ U\left(\vec{\theta}_j^k, \vec{x}^n\right),\ j = 1,\ldots,J$$

os riscos à posteriori correspondentes à atribuição a G_i ou a G_l são idênticos, pelo que as várias possibilidades são equivalentes.

Por outro lado, as Regiões de Risco à Posteriori Controlado serão dadas por

(6) $$\Gamma_\alpha\left(\vec{x}^n\right) = \left\{\vec{\theta}_i^k ; R_p\left(\vec{\theta}_i^k \middle| \vec{x}^n\right) \le \alpha\right\}$$

Observe-se que se tem de ter

(7) $$\text{Min}\left\{R_p\left(\vec{\theta}_i^k, \vec{x}^n\right),\ i = 1,\ldots,J\right\} \le \alpha$$

para que $\Gamma_\alpha\left(\vec{x}^n\right)$ não seja vazia.

4.4. - MAIS DE DUAS POPULAÇÕES:ESTATÍSTICAS SUFICIENTES

Teremos agora

(1) $$p\left(\vec{x}^{n}\middle|\vec{\theta}_{j}^{k}\right)=h\left(\vec{x}^{n}\right)g\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\middle|\vec{\theta}_{j}^{k}\right),\ j=1,\ldots,J$$

bem como

(2) $$l_{1}\left(\vec{x}^{n}\right)=\sum_{j=1}^{J}q_{j}\,h\left(\vec{x}^{n}\right)g\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\middle|\vec{\theta}_{j}^{k}\right)=h\left(\vec{x}^{n}\right)\bar{g}\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\right)$$

com

(3) $$\bar{g}\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\right)=\sum_{j=1}^{J}q_{j}\,h\left(\vec{x}^{n}\right)g\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\middle|\vec{\theta}_{j}^{k}\right)$$

Tem-se ainda

(4) $$U\left(\vec{\theta}_{i}^{k},\vec{x}^{n}\right)=\sum_{j=1}^{J}C_{i,j}\,h\left(\vec{x}^{n}\right)g\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\middle|\vec{\theta}_{j}^{k}\right)q_{j},\ i=1,\ldots,J$$

bastando comparar os

(5) $$V\left(\vec{\theta}_{i}^{k},\vec{x}^{n}\right)=\sum_{j=1}^{J}C_{i,j}\,g\left(\vec{t}^{m}\left(\vec{x}^{n}\right)\middle|\vec{\theta}_{j}^{k}\right)q_{j},\ i=1,\ldots,J$$

o que, com $\vec{t}^{m}=\vec{t}^{m}\left(\vec{x}^{n}\right)$, nos permite escrever

(6) $$V\left(\vec{\theta}_i^k,\vec{t}^m\right)=\sum_{j=1}^{J}C_{i,j}\ g\left(\vec{t}^m\middle|\vec{\theta}_j^k\right)q_j,\quad i=1,\ldots,J$$

pelo que a atribuição é feita em função da estatística suficiente.

Se estivermos no caso de custos iguais teremos

(7) $$V\left(\vec{\theta}_i^k,\vec{t}^m\right)=C\left(\sum_{j=1}^{J}q_j\ g\left(\vec{t}^m\middle|\vec{\theta}_j^k\right)-q_i\ g\left(\vec{t}^m\middle|\vec{\theta}_i^k\right)\right)=$$

$$=C\left(\bar{g}\left(\vec{t}^m\right)-q_i\ g\left(\vec{t}^m\middle|\vec{\theta}_i^k\right)\right),\quad i=1,\ldots,J$$

bastando comparar os $q_i\ g\left(\vec{t}^m\middle|\vec{\theta}_i\right)$ para realizar a atribuição.

Para obter as Regiões de Risco à Posteriori Controlado observe-se que

(8) $$R_p\left(\vec{\theta}_i^k,\vec{x}^n\right)=\frac{U\left(\vec{\theta}_i^k\middle|\vec{x}^n\right)}{l_i\left(\vec{x}^n\right)}=\frac{\sum_{j=1}^{J}C_{i,j}\ h\left(\vec{x}^n\right)g\left(\vec{t}^m\left(\vec{x}^n\right)\middle|\vec{\theta}_j^k\right)q_j}{h\left(\vec{x}^n\right)\bar{g}\left(\vec{t}^m\left(\vec{x}^n\right)\right)}=$$

$$=\frac{1}{\bar{g}\left(\vec{t}^m\left(\vec{x}^n\right)\right)}\sum_{j=1}^{J}C_{i,j}\ g\left(\vec{t}^m\left(\vec{x}^n\right)\middle|\vec{\theta}_j^k\right)q_j,\quad i=1,\ldots,J$$

vendo-se que $R_p\left(\vec{\theta}_i^k,\vec{x}^n\right)$ depende de $\vec{x}^n$ através da estatística suficiente $\vec{t}^m=\vec{t}^m\left(\vec{x}^n\right)$. Podemos pois escrever

(9) $$R_p\left(\vec{\theta}_i^k,\vec{t}^m\right)=\frac{1}{\bar{g}\left(\vec{t}^m\right)}\sum_{j=1}^{J}C_{i,j}\ g\left(\vec{t}^m\middle|\vec{\theta}_j^k\right)q_j,\quad i=1,\ldots,J$$

bem como

(10) $$\Gamma_\alpha\left(\vec{t}^{\,m}\right)=\left\{\vec{\theta}_i^{\,k};R_p\left(\vec{\theta}_i^{\,k}\middle|\vec{t}^{\,m}\right)\leq\alpha\right\}$$

voltando a ter de ter-se

(11) $$\text{Min}\left\{R_p\left(\vec{\theta}_i^{\,k},\vec{t}^{\,m}\right), i=1,\ldots,J\right\}\leq\alpha$$

para que $\Gamma_\alpha\left(\vec{t}^{\,m}\right)$ não seja vazia.

No caso dos custos iguais tem-se

(12) $$R_p\left(\vec{\theta}_i^{\,k},\vec{t}^{\,m}\right)=\frac{1}{\bar{g}\left(\vec{t}^{\,m}\right)}\left(C\,\bar{g}\left(\vec{t}^{\,m}\right)-C\,q_i\,g\left(\vec{t}^{\,m}\middle|\vec{\theta}_i^{\,k}\right)\right)=$$

$$=C\left(1-\frac{q_i\,g\left(\vec{t}^{\,m}\middle|\vec{\theta}_i^{\,k}\right)}{\bar{g}\left(\vec{t}^{\,m}\right)}\right),\quad i=1,\ldots,J$$

logo, resolvendo a desigualdade $R_p\left(\vec{\theta}_i^{\,k},\vec{t}^{\,m}\right)<\alpha$ vem

(13) $$\Gamma_\alpha\left(\vec{t}^{\,m}\right)=\left\{\vec{\theta}_i^{\,k}\middle|q_i\,g\left(\vec{t}^{\,m}\middle|\vec{\theta}_i^{\,k}\right)>\left(1-\frac{\alpha}{C}\right)\bar{g}\left(\vec{t}^{\,m}\right)\right\}.$$

4.5. - UMA APLICAÇÃO

Vamos agora considerar uma aplicação conveniente para pôr em evidência o essencial das novas possibilidades abertas pela interpretação bayesiana da Análise Discriminante. Para aligeirar os cálculos, admitiremos que as componentes de $\vec{X}^n$ são *i.i.d.* . Como vimos esta situação verifica-se quando se "colhe" uma amostra de dimensão n de elementos que se sabe provirem todos da mesma população fazendo-se uma única medida em cada elemento da amostra.

Ainda para simplificar os cálculos, admitiremos que existem 3 populações, pelo que $\left[C_{i,j}\right]$ será 3 x 3 , e que as observações têm densidades

(1) $$l(x|0,j)=\begin{cases}0 & ;x<0\\ \frac{1}{j}e^{-\frac{x}{j}} & ;x\geq 0\end{cases}\quad ;\ j=1,2,3$$

pelo que as densidades conjuntas das amostras serão

(2) $$f_j\left(\vec{x}^n\right)=\prod_{i=1}^{n}\frac{e^{-\frac{x_i}{j}}}{j}=\frac{e^{-\frac{1}{j}t}}{j^n}$$

com estatística suficiente

(3) $$t=\sum_{i=1}^{n}x_i$$

e $h\left(\vec{x}^n\right)\equiv 1$.

Neste caso, $m=1$ e

(4) $$V(i \mid t) = \sum_{j=1}^{3} C_{i,j}\, q_j \, \frac{e^{-\frac{1}{j}t}}{j^n}, \quad i = 1,2,3$$

vindo, no caso de custos iguais,

(5) $$V(i \mid t) = C \left(\bar{g}(t) - q_i \, \frac{e^{-\frac{1}{i}t}}{i^n} \right)$$

, com

(6) $$\bar{g}(t) = \sum_{j=1}^{3} q_j \, \frac{e^{-\frac{1}{j}t}}{j^n}$$

A atribuição é feita comparando as funções

(7) $$h_i(t) = a_i \, e^{-\frac{1}{i}t}, \quad i = 1,2,3$$

com

(8) $$a_i = \frac{q_i}{i^n}, \quad i = 1,2,3$$

tendo-se, com $j < l$,

(9) $$\frac{h_j(t)}{h_l(t)} \gtrless 1$$

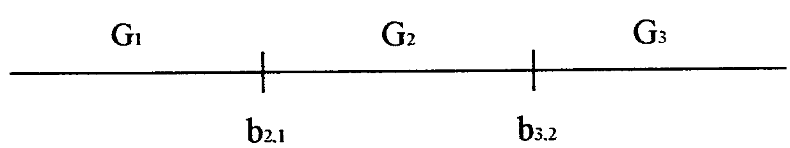
se e só se

(10) $$t \gtrless b_{j,l} = \frac{lj}{j-l} \log \frac{a_l}{a_j}$$

Observe-se que, com $j < l$, $\frac{h_j(t)}{h_l(t)}$ é crescente em t ; assim, se

(11) $$b_{2,1} < b_{3,1} < b_{3,2} \ ,$$

quando $t < b_{2,1}$ tem-se $h_3(t) < h_2(t) < h_1(t)$ devendo-se atribuír os elementos a G_1 e, quando $t > b_{3,2}$ tem-se $h_1(t) < h_2(t) < h_3(t)$, sendo G_1 a população a que os elementos devem ser atribuídos. Finalmente, quando $b_{2,1} < t < b_{3,2}$ tem-se $h_1(t) < h_2(t)$ e $h_3(t) < h_2(t)$ sendo G_2 a população a preferir. Assim, supondo-se verificada a condição (11) ter-se-á o esquema de decisão



sendo indiferente atribuir-se a G_2 ou a G_3 $\left[G_1 \text{ ou } G_2\right]$ quando $t = b_{3,2}$ $\left[t = b_{2,1}\right]$.

Ora

$$
(12)\begin{cases}
b_{2,1} = 2\ \log \dfrac{a_1}{a_2} = 2\ \log \dfrac{q_1\, 2^n}{q_2} = 2\ \log \dfrac{q_1}{q_2} + 2n\ \log 2 = 2\ \log \dfrac{q_1}{q_2} + n\ \log 4 \\[2ex]
b_{3,1} = \dfrac{3}{2}\ \log \dfrac{a_1}{a_3} = \dfrac{3}{2}\ \log \dfrac{q_1\, 3^n}{q_3} = \dfrac{3}{2}\ \log \dfrac{q_1}{q_3} + \dfrac{3}{2}\ n\ \log 3 = \dfrac{3}{2}\ \log \dfrac{q_1}{q_3} + n\ \log \sqrt{27} \\[2ex]
b_{3,2} = 6\ \log \dfrac{a_2}{a_3} = 6\ \log \dfrac{q_2\, 3^n}{q_3\, 2^n} = 6\ \log \dfrac{q_2}{q_3} + 6n\ \log \dfrac{3}{2} = 6\ \log \dfrac{q_2}{q_3} + n\ \log \dfrac{729}{64}
\end{cases}
$$

e

$$\log 4 < \log \sqrt{27} < \log \frac{729}{64},$$

logo, se $n > \dfrac{2\ \log \dfrac{q_1}{q_2} - \dfrac{3}{2}\ \log \dfrac{q_1}{q_3}}{\log \sqrt{27} - \log 4}$ $\left[n > \dfrac{\dfrac{3}{2}\ \log \dfrac{q_1}{q_3} - 6 \log \dfrac{q_2}{q_3}}{\log \dfrac{729}{64} - \log \sqrt{27}} \right]$,

tem-se $b_{2,1} < b_{3,1}$ $\left[b_{3,1} < b_{3,2}\right]$; assim (11) verifica-se desde que

$$
(13) \qquad n > \text{Max} \left\{ \frac{2\ \log \dfrac{q_1}{q_2} - \dfrac{3}{2}\ \log \dfrac{q_1}{q_3}}{\log \sqrt{27} - \log 4} \ ; \ \frac{\dfrac{3}{2}\ \log \dfrac{q_1}{q_2} - 6\ \log \dfrac{q_2}{q_3}}{\log \dfrac{729}{64} - \log \sqrt{27}} \right\}
$$

em particular, se $q_1 = q_2 = q_3$, bastará que $n \geq 1$ para que a condição se verifique.

Admitamos pois que (11) se verifica. Seja $A_{i,j}$ o acontecimento que se verifica quando um elemento de G_i é atribuído a G_j, $i = 1,2,3$, $j = 1,2,3$. Para estudarmos o comportamento da nossa regra de decisão convém-nos calcular as probabilidades desses acontecimentos para o que vamos obter as distribuições da estatística suficiente correspondentes às três populações.

Estabeleçamos a

Proposição 3

As distribuições da estatística suficiente para as três populações são

$$\mathcal{G}_j(z \mid n) = \begin{cases} 0 & ; z \leq 0 \\ 1 - e^{-\frac{z}{j}} \displaystyle\sum_{i=0}^{n-1} \frac{\left(\frac{z}{j}\right)^i}{i!} & ; z \geq 0 \end{cases} \qquad ; j=1,2,3$$

Demonstração :

Dadas as funções geradoras de momentos identificarem as distribuições, basta mostrar que as funções geradoras de momentos $\psi_j(x)$, $j = 1,2,3$ da estatística suficiente coincidem com as das distribuições indicadas. As densidades correspondentes às distribuições indicadas anulam-se para $t < 0$, tendo-se, para

$t > 0$,

$$g_j(z \mid n) = \frac{1}{(n-1)!\,j}\left(\frac{z}{j}\right)^{n-1} e^{-\frac{z}{j}},\ j = 1,2,3\,,$$

logo $\psi_j(u \mid n) = \dfrac{1}{(n-1)!\,j^n}\displaystyle\int_0^{+\infty} e^{zu}\, z^{n-1}\, e^{-\frac{z}{j}} dz\ ;\ u < \frac{1}{j};\ \ j = 1,2,3\,;$

fazendo-se a transformação $v = \dfrac{1-ju}{j}\, z,\ j = 1,2,3$, vem

$$\psi_j(u \mid n) = \frac{1}{(n-1)!\,j^n}\int_0^{+\infty}\left(\frac{j}{1-ju}\,v\right)^{n-1} e^{-v}\,\frac{j}{1-ju}\,dv =$$

$$= \frac{1}{(n-1)!}\,\frac{1}{(1-ju)^n}\int_0^{+\infty} v^{n-1}\, e^{-v}\, dv = \frac{1}{(1-ju)^n}\ ;\ u < \frac{1}{j}\ ;\ j = 1,2,3$$

visto $\displaystyle\int_0^{+\infty} v^{n-1}\, e^{-v}\, dv = \Gamma(u) = (n\text{-}1)!\,.$

Em particular as observações isoladas "colhidas" nas três populações terão funções geradoras de momentos

$$\psi_j(u \mid 1) = \frac{1}{1-ju}\ ;\ u < \frac{1}{j}\ ;\ j = 1,2,3\,,$$

visto terem as densidades $g_j(z \mid 1),\ j = 1,2,3$. Como as observações são *i.i.d*, as funções geradoras de momentos da estatística suficiente serão $\psi_j(u \mid 1)^n = \psi_j(u \mid n)\,;\ \ j = 1,2,3$, o que completa a demonstração. □

Teremos então

$$(14)\quad \begin{cases} pr\left(A_{1,j}\right) = pr\left(t \le b_{2,1}\right) = \mathcal{G}_j\left(b_{2,1}|n\right);\ \ j = 1,2,3 \\ \\ pr\left(A_{2,j}\right) = pr\left(b_{2,1} < t \le b_{3,2}\right) = \mathcal{G}_j\left(b_{3,2}|n\right) - \mathcal{G}_j\left(b_{2,1}|n\right);\ \ j = 1,2,3 \\ \\ pr\left(A_{3,j}\right) = pr\left(b_{3,2} \le t\right) = 1 - \mathcal{G}_j\left(b_{3,2}|n\right);\ \ j = 1,2,3 \end{cases}$$

O custo médio da decisão para as três populações será

$$(15)\quad \begin{cases} \mathbb{E}(W \mid 1) = C\, pr\left(A_{1,2}\right) + C\, pr\left(A_{1,3}\right) = C\left(1 - pr\left(A_{1,1}\right)\right) \\ \\ \mathbb{E}(W \mid 2) = C\, pr\left(A_{2,1}\right) + C\, pr\left(A_{2,3}\right) = C\left(1 - pr\left(A_{2,2}\right)\right) \\ \\ \mathbb{E}(W \mid 3) = C\, pr\left(A_{3,1}\right) + C\, pr\left(A_{3,2}\right) = C\left(1 - pr\left(A_{3,3}\right)\right) \end{cases}$$

visto, quando os $A_{i,i}$ se verificam o custo ser nulo, $i = 1,2,3$ e igual a C nos restantes casos.

Finalmente o custo médio será

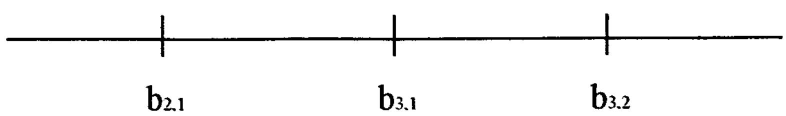
$$(16)\qquad \mathbb{E}(W) = \sum_{j=1}^{3} q_j\ \mathbb{E}(W|j)$$

visto q_j ser a probabilidade da amostra provir da j-ésima população.

Resta-nos considerar a construção das Regiões de Risco à Posteriori Controlado , para o que utilizaremos a expressão (13) da alínea precedente. Neste caso tem-se

$$(17) \qquad q_i\, g(t \mid i) = q_i\, \frac{e^{-\frac{t}{i}}}{i^n}; \quad i = 1,2,3$$

e como existem apenas três populações as Regiões de Risco à Posteriori Controlado não vazias "conterão" uma ou duas das mesmas. Observe-se agora que, quando se verifica a condição (11), se tem com $V_i = V(i \mid t)$; $i = 1,2,3$



Neste caso $h(\vec{x}^n) \equiv 1$ vindo $U(i \mid t) = V(i \mid t)$; $i = 1,2,3$ e

$$(18) \qquad \begin{cases} R_p(i \mid t) = \dfrac{V(i \mid t)}{l_1(t)} \\ \\ l_1(t) = \displaystyle\sum_{j=1}^{N} q_j \frac{e}{j^n} \end{cases}$$

Ora é natural ter-se o mesmo valor de α para o nível das Regiões de Risco à Posteriori Controlado independentemente do valor da estatística suficiente. Se não se quiser que haja Regiões de Risco à Posteriori Controlado vazias terá de se ter

$$(19)\quad \begin{cases} R_p(1 \mid t) \le \alpha; \ t \le b_{2,1} \\ R_p(2 \mid t) \le \alpha; \ b_{2,1} \le t \le b_{3,2} \\ R_p(3 \mid t) \le \alpha; \ b_{3,2} \le t \end{cases}$$

já que, atendendo a (18), sabemos que G_1 $[G_2, G_3]$ é a população a que corresponde um risco à posteriori mínimo quando $t \le b_{2,1}$ $[b_{2,1} \le t \le b_{3,2} \ ; \ b_{3,2} \le t]$. Por outro lado, se as desigualdades anteriores forem estritas ter-se-à, numa vizinhança de $b_{2,1}$

$$(21)\qquad \mathrm{Max}\left\{R_p(1,t) \ ; \ R_p(2,t)\right\} < \alpha$$

e, numa vizinhança de $b_{3,2}$,

$$(22)\qquad \mathrm{Max}\left\{R_p(2,t) \ ; \ R_p(3,t)\right\} < \alpha$$

surgindo, nessas vizinhanças $[b'_{2,1} \ ; \ b''_{2,1}]$ e $[b'_{3,2} \ ; \ b''_{3,2}]$, Regiões de Risco à Posteriori Controlado constituída por duas populações. A determinação dos limites destas vizinhanças é fácil já que bastará resolver as equações $R_p(2 \mid t) = \alpha$ para determinar $b'_{2,1}$ e $b'_{3,2}$, $R_p(1 \mid t) = \alpha$ para determinar $b''_{2,1}$ e $R_p(3 \mid t) = \alpha$ para determinar $b'_{3,2}$.

CONCLUSÕES

Mostrou-se como a Análise Discriminante se integra harmoniosamente na Estimação Bayesiana. Através desta integração podemos considerar a possibilidade de em certos casos não se atribuír o elemento a uma única população. A aplicação que apresentámos julgamos que ilustra bem estes pontos; assim, é na vizinhança de " valores de separação " para a estatística suficiente que surgem as probabilidades de atribuições múltiplas.

Procurámos também mostrar como a utilização de estatísticas suficientes simplifica grandemente os cálculos.

BIBLIOGRAFIA

ANDERSON, T. W. - An Introduction to Multivariate Statistical Analysis - Second Edition - New York - 1984

BERGER, James O. - Statistical Decision Theory and Bayesian Analysis - Second Edition - Springer Series in Statistics - Springer Verlag - New York - 1980

BOX & TIAO - Bayesian Inference in Statistical Analysis - Wiley Classics Library - New York - 1992

FRASER, D. A. S. - Nonparametrics Methods in Statistics - John Wiley & Sons, Inc. - New York - 1957

MEXIA, J. T. - Introdução à Inferência Estatística Linear - Centro de Estudos de Matemáticas Aplicadas - Edições Universitárias Lusófonas - Lisboa - 1995

O'HAGAN, Anthony - Kendall's Advanced Theory of Statistics, Volume 2 - B - Bayesian Inference - Wiley & Sons, Inc. - New York - 1994

WILLIAMS, David - Probability with Martingales - Cambridge Mathematical Textbooks - Cambridge University Press - Cambridge - 1991