**Propriedades lógicas de
classes de testes de hipóteses**

Gustavo Miranda da Silva

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Sergio Wechsler

Coorientador: Prof. Dr. Luís Gustavo Esteves

São Paulo, setembro de 2014

Propriedades lógicas de classes de testes de hipóteses

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 03/11/2014. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Luís Gustavo Esteves (coorientador) - IME-USP
- Prof. Dr. Julio Michael Stern - IME-USP
- Prof. Dr. Marcelo Esteban Coniglio - UNICAMP
- Prof. Dr. José Galvão Leite - UFSCar
- Prof. Dr. Marcio Alves Diniz - UFSCar

Agradecimentos

Agradeço a Deus; a meus familiares, noiva e amigos por incentivarem a realização deste trabalho. Ao Professor Carlinhos por seus conselhos sempre precisos, aos colegas Rafael Izbicki e Thiago Feitosa pela discussão de alguns pontos do trabalho e ao Professor Alexandre Patriota pelas diversas sugestões dadas na qualificação e aos Professores Julio Stern, Marcelo Coniglio, José Galvão e Marcio Diniz pelas sugestões durante a defesa.

Também e principalmente, ao Professor Sérgio Wechsler e ao Professor Luís Gustavo Esteves pela dedicação e paciência em todos estes anos.

Resumo

SILVA, G. M. **Propriedades Lógicas de Classes de Testes de Hipóteses**. 2014. 68 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

Ao realizar testes de hipóteses simultâneos espera-se que as decisões obtidas neles sejam logicamente consistentes entre si. Neste trabalho, verifica-se sob quais condições testes de Bayes simultâneos atendem às condições lógicas isoladamente ou em conjunto. Demonstra-se que as restrições para que os testes simultâneos atendam essas condições isoladamente são bastante intuitivas. No entanto, ao tentar obedecer as condições conjuntamente, perde-se otimalidade. Além disso, avalia-se a relação entre esses testes de Bayes simultâneos e os testes gerados por estimadores, isto é, mostra-se que, sob algumas condições, tomar uma decisão baseado em um estimador de Bayes é equivalente a tomar uma decisão baseada em um teste de Bayes. Por fim, mostra-se que, se tomamos uma decisão baseada em Estimadores de Máxima Verossimilhança, então essa decisão deve ser igual à tomada por um teste de Bayes e concluímos que essas decisões são admissíveis e obedecem ao Princípio da Verossimilhança.

Palavras-chave: Teoria da Decisão, Testes de hipóteses, Testes simultâneos, Propriedades lógicas.

Abstract

SILVA, G. M. **Logical Properties of Classes of Hypotheses Tests**. 2014. 68 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

When performing simultaneous hypotheses testing is expected that the decisions obtained therein are logically consistent with each other. In this work, we find restrictions under which simultaneous Bayes tests meet logical conditions separately or jointly. It is shown that the conditions for the simultaneous tests meet these conditions alone are quite intuitive. However, when trying to obey the conditions jointly, we lose optimality. Furthermore, we evaluate the relationship between these tests and simultaneous Bayes tests generated by estimators, ie, we show that, under some conditions, to choose an estimator based on Bayes decision is equivalent to choosing a decision based on a Bayes test. Finally, we show that if we take a decision based on Maximum Likelihood Estimators, then that decision should be equal to taking a Bayes test and concluded that these decisions are admissible and obey the Likelihood Principle.

Keywords: Decision theory, hypotheses testing, simultaneous tests, logical properties.

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Considerações preliminares	1
1.2 Definições básicas	2
1.3 Classes de testes de hipóteses	6
1.4 Propriedades Lógicas	8
1.4.1 Monotonicidade	8
1.4.2 Invertibilidade	9
1.4.3 Consonância da união	12
1.5 Desideratas	14
1.6 Objetivos	16
2 Classes de testes de Bayes e suas propriedades lógicas	17
2.1 Classes de testes monótonas	20
2.2 Classes de testes invertíveis	23
2.3 Classes de testes consoantes com a união	27
2.4 Classes de Testes que (não) atendem às Desideratas	31
3 Classes de testes geradas por estimadores	41
3.1 Estimadores de Bayes e EMVs	42
3.2 Princípio da Verossimilhança	48
3.3 Desiderata D2 <i>versus</i> Otimalidade	49

4 Conclusões	51
4.1 Considerações Finais	51
4.2 Sugestões para Pesquisas Futuras	52
Referências Bibliográficas	53

Lista de Figuras

2.1	Distribuições a priori para as quais $\varphi_A(0) = 0$	19
2.2	Distribuições a priori para as quais $\varphi_A(1) = 0$	19
2.3	Conjunto $\mathcal{B}$	34
2.4	Projeção de $\mathcal{B}$ em $u \times v$	35

Lista de Tabelas

1.1	Função de perda 0-1	2
1.2	Função de perda do exemplo deGroot	6
1.3	Exemplo de função de perda para teste da Normal	8
1.4	Exemplo de função de perda para teste da Normal	9
1.5	Exemplo de função de perda para teste da Normal	10
1.6	Exemplo de função de perda para teste da Normal	10
1.7	Exemplo de função de perda para teste da Normal	11
1.8	Exemplo de função de perda para teste da Normal	11
2.1	Distribuições a priori para as quais $\varphi_A(0) = 0$	18
2.2	Distribuições a priori para as quais $\varphi_A(1) = 0$	18
2.3	Função de perda 0-2	20
2.4	Exemplo de famílias de funções de perda atendendo (2.1)	23
2.5	Exemplo de famílias de funções de perda atendendo (2.1)	23
2.6	Exemplo de classe de funções de perda que não atende 2.1	23
2.7	Exemplo de classe de funções de perda que não atende 2.1	24
2.8	Exemplo de famílias de funções de perda atendendo (2.1)	26
2.9	Exemplo de famílias de funções de perda atendendo (2.1)	26
2.10	Exemplo de classe de funções de perda que não atende (2.5)	27
2.11	Exemplo de classe de funções de perda que não atende (2.5)	27
2.12	Exemplo de função de perda utilizando uma medida finita.	30
2.13	Exemplo de função de perda de classe que não atendem (2.9)	30
2.14	Exemplo de função de perda de classe que não atendem (2.9)	30
2.15	Exemplo de função de perda de classe que não atendem (2.9)	31
2.16	Exemplo de função de perda utilizando uma medida finita.	32

3.1	Calculo dos riscos esperados para A	46
-----	---	----

Capítulo 1

Introdução

1.1 Considerações preliminares

Desde que [Pearson \(1900\)](#) sistematizou o uso de testes de hipóteses, inúmeras discussões a respeito deste tópico ocuparam papel importante no desenvolvimento de várias teorias de Inferência Estatística ([Bera e Premaratne \(2001\)](#)). Mais adiante, os testes também protagonizaram papel importante na discussão e comparação entre as escolas de inferência.

A maior parte dos textos a respeito de testes de hipóteses trata de métodos que, de acordo com determinado critério de otimalidade, tragam a melhor solução para o problema de decisão. Em menor escala, no entanto, existem textos que tratam de testes simultâneos e da concordância lógica de suas soluções ótimas.

Deste segundo grupo, podemos destacar trabalhos como [Gabriel \(1969\)](#), por exemplo, que afirma que os testes simultâneos deveriam obedecer a seguinte propriedade: quando uma hipótese testada não é rejeitada a partir de uma observação (amostral), todas as hipóteses implicadas por ela também não deveriam ser rejeitadas. A essa propriedade ele dá o nome de *coerência*. Ainda neste mesmo trabalho, ele também propõe alternativas aos testes de hipóteses vigentes em Análise de Variância (ANOVA) que respeitem esta propriedade.

Outros trabalhos, como [Betz e Levin \(1982\)](#) e [Rom e Holland \(1995\)](#), falam sobre problemas de concordância lógica ao realizar testes para modelos ANOVA fatoriais hierárquicas. [Lehmann \(1957a\)](#) mostra casos em que não se consegue chegar a uma conclusão conjunta quando se pretende responder a uma série de questionamentos em sequência.

Em [Schervish \(1996\)](#) e [Lavine e Schervish \(1999\)](#), os autores mostram que o p-valor e o fator de Bayes não são coerentes (no sentido proposto por [Gabriel \(1969\)](#)). Além disso, [Lavine e Schervish \(1999\)](#) propõem restrições para as funções de perda de testes simultâneos de modo que eles obedeçam a esta propriedade. [Hommel e Bretz \(2008\)](#) mostram que o teste de Bonferroni-Holm leva a decisões com inconsistências lógicas para modelos de regressão linear.

[Fossaluza \(2008\)](#) discute a importância da propriedade introduzida por [Gabriel \(1969\)](#) através de alguns exemplos. Em [Fossaluza \(2008\)](#), no entanto, tal propriedade, denominada *monotonicidade*, é estendida para uma σ -álgebra de subconjuntos do espaço paramétrico ([Gabriel \(1969\)](#) havia definido para filtros de hipóteses) a fim de contemplar testes bayesianos baseados em probabilidades a posteriori. Tal extensão é formalizada em [Silva \(2010\)](#) através da introdução do conceito de *classes de testes* que, de certa maneira, é uma generalização dos testes de hipóteses. [Silva \(2010\)](#) também analisa, a partir dessa definição, quais dos testes mais usuais da literatura estatística apresentam monotonicidade e algumas condições (suficientes) que devem ser impostas às funções de perda para que uma classe de testes de Bayes satisfaça essa propriedade.

Finalmente, em [Izbicki \(2010\)](#), são propostos alguns axiomas que procedimentos de teste simultâneos poderiam obedecer e algumas condições para que estes procedimentos possam obedecer a estes axiomas. Uma revisão da literatura de testes de hipóteses simultâneos pode ser encontrada em [Izbicki e Esteves \(2014\)](#).

Neste trabalho, serão abordadas, sob o ponto de vista da *Teoria da Decisão Bayesiana*, condições necessárias e suficientes para que testes de Bayes simultâneos tenha monotonicidade e satisfaçam outras propriedades lógicas introduzidas em [Izbicki \(2010\)](#).

1.2 Definições básicas

Os primeiros conceitos de teoria estatística da decisão - uma tentativa formal de dar uma interpretação matemática da maneira como aprendemos a partir dos dados - se devem a Abraham Wald ([Wald \(1945\)](#), [Wald \(1949\)](#)). No seu artigo de 1949 sobre *funções estatísticas de decisão*, Wald propôs uma base unificadora para grande parte da teoria estatística, tratando a inferência estatística como um caso especial da teoria dos jogos. ([Neumann e Morgenstern \(1944\)](#))

Um dos primeiros livros sobre Teoria da Decisão Estatística é [Raiffa e Schlaifer \(1961\)](#), que contribuiu enormemente para as definições de estatística Bayesiana e a tomada de decisões nos primeiros anos de desenvolvimento da área. Além dele, [Ferguson \(1967\)](#), [Berger \(1980\)](#), [DeGroot \(1970\)](#) entre outros, também contribuíram para a discussão extensiva da Teoria da Decisão Bayesiana no século passado.

De acordo com [Parmigiani e Inoue \(2009\)](#), o problema da tomada de decisões sob incertezas trata de “como fazer escolhas cujas consequências não são completamente previsíveis, porque eventos que acontecerão no futuro afetarão as consequências dessas decisões tomadas neste momento”.

Para lidar com estes problemas de tomada de decisão sob incertezas, onde se deseja descobrir o valor do parâmetro θ , os elementos usualmente considerados são: o espaço paramétrico Θ (que contém os possíveis valores do parâmetro), o espaço de decisões D , uma função a valores reais não negativos L , que é definida no espaço produto $D \times \Theta$ e uma medida de probabilidade π definida no espaço $(\Theta, \sigma(\Theta))$, onde $\sigma(\Theta)$ é uma σ -álgebra de subconjuntos de Θ . Para qualquer ponto $(d, \theta) \in D \times \Theta$, o número $L(d, \theta)$ representa a penalidade (ou perda) quando o decisor escolhe a decisão d e o valor do parâmetro é θ . É assumido que, para cada $d \in D$, a perda $L(d, \cdot)$ é uma função $\sigma(\Theta)$ -mensurável no espaço Θ . Um exemplo de função de perda é apresentado a seguir.

Exemplo 1.1. Seja $\{\Theta_0, \Theta_1\}$ uma partição do espaço paramétrico Θ , isto é, $\Theta = \Theta_0 \cup \Theta_1$, com $\Theta_0 \cap \Theta_1 = \emptyset$ e seja $D = \{0 \text{ (decisão por } \theta \in \Theta_0), 1 \text{ (decisão por } \theta \in \Theta_1)\}$ um espaço de decisões. A função de perda $L : D \times \Theta \rightarrow \mathbb{R}^+$ dada pela Tabela 1.1 é conhecida por função de perda 0-1.

L_A	$\theta \in \Theta_0$	$\theta \in \Theta_1$
0	0	1
1	1	0

Tabela 1.1: Função de perda 0-1

A função de perda 0-1 apresentada no exemplo acima considera que a perda ao decidir por $\theta \in \Theta_0$ quando na verdade $\theta \in \Theta_1$ (chamado erro do tipo II) é igual a perda ao decidir por $\theta \in \Theta_1$ quando na verdade $\theta \in \Theta_0$ (chamado erro do tipo I), isto é, a perda de cometer o erro do tipo II é igual à perda de cometer o erro do tipo I. Nesse caso, ao tomar uma decisão correta, a perda é sempre igual a zero.

Se π representa a distribuição de probabilidade para o parâmetro θ , para qualquer decisão $d \in D$, a perda esperada ou risco dessa decisão, $\rho(d, \pi)$, é dado por

$$\rho(d, \pi) = \int_{\Theta} L(d, \theta) d\pi(\theta). \quad (1.1)$$

Assume-se que a integral na equação acima é finita para todo $d \in D$. Qualquer decisão d que tenha risco infinito pode ser “eliminada” do conjunto D . A função do decisor é escolher a decisão d que minimize o risco $\rho(d, \pi)$ (DeGroot (1970)).

Em muitos casos é possível observar um vetor aleatório X para auxiliar a tomada de decisão. Nestes casos, além dos elementos definidos anteriormente, teríamos ainda um espaço amostral χ de possíveis realizações do experimento e uma família de distribuições de probabilidade $\mathbb{P}$, descrita a seguir.

Seja $\mathbb{P} = \{P_{\theta} : \theta \in \Theta\}$ uma família de distribuições de probabilidade de um vetor aleatório X , cujo valor estará disponível para a tomada de decisão, indexada pelo parâmetro θ . Nesse caso, o agente decisor deve escolher uma regra de decisão que minimize o risco baseado não só na função de perda, mas também na distribuição do parâmetro θ atualizada de acordo com os valores de $x \in \chi$ observados. A função $P_{\theta}(\cdot)$ quando vista como função de θ para um $x \in \chi$ fixado é chamada de função de verossimilhança e pode ser denotada por $V_x(\cdot)$. Podemos agora definir a escolha da melhor decisão com base em uma função da variável observável X .

Definição 1.1. Uma regra de decisão é uma função $\phi : \chi \rightarrow D$, que para cada possível $x \in \chi$, especifica a decisão $\phi(x) \in D$.

Quando $D = \{0, 1\}$, uma regra de decisão é chamada também de *função de teste*. A função de risco de uma regra de decisão ϕ associa a cada valor θ de Θ , o risco frequentista¹ dado por

$$R(\theta, \phi) = \int_{\chi} L(\phi(x), \theta) dP_{\theta}(x).$$

Como os valores de X e de θ são desconhecidos antes da realização do experimento, a perda associada à decisão ϕ também é desconhecida. Seja Φ a classe de todas as possíveis regras de decisão. Para $\phi \in \Phi$, definimos o risco a posteriori de ϕ por

$$\rho(\phi, \pi_x) = E[L(\phi, \theta)] = \int_{\Theta} \int_{\chi} L(\phi(x), \theta) dP_{\theta}(x) d\pi(\theta) = \int_{\Theta} R(\theta, \phi) d\pi(\theta).$$

Uma regra de decisão ótima é a regra $\phi^* \in \Phi$ que minimiza o risco definido acima.

Definição 1.2. Se uma regra de decisão ϕ^* tem risco $\rho(\phi^*, \pi_x)$ finito para todo $x \in \chi$ e $\rho(\phi^*, \pi_x) \leq \rho(\phi, \pi_x)$ para todas regras de decisão ϕ , então ϕ^* é uma regra de decisão de Bayes contra π .

Em geral, não é fácil escolher uma regra de decisão em Φ que minimize o risco. Porém, alternativamente, pode-se aguardar a observação de $x \in \chi$ e considerar somente o problema de escolher a decisão $d^* \in D$ apropriada neste caso. Encontrar a regra de decisão que minimiza esse risco a posteriori (definido abaixo) é mais simples e conduz a resultados idênticos (DeGroot (1970)).

Após a observação de $X = x$, o decisor atualiza seu conhecimento sobre o parâmetro θ através da distribuição a posteriori $\pi(\theta|x) \propto P_{\theta}(x)\pi(\theta)$. Deste modo, observado $x \in \chi$, o risco a posteriori da decisão $d \in D$ é dado por

$$\rho(d, \pi_x) = \int_{\Theta} L(d, \theta) d\pi(\theta|x).$$

¹ Sob a perspectiva frequentista, deve-se encontrar a função de teste que minimiza este risco.

O agente decisor deve então escolher a regra ϕ^* que, para cada $x \in \mathcal{X}$, associa $\phi^*(x) = d^*$ que minimiza o risco posterior dado $X = x$. A regra escolhida de forma a minimizar, para cada $x \in \mathcal{X}$, o risco posterior da regra de decisão de Bayes e a justificativa de sua otimalidade é baseada em axiomas de utilidade (DeGroot (1970)). Ver Neumann e Morgenstern (1944).

Como foi dito, em geral, não é necessário escolher a função de decisão para todo $x \in \mathcal{X}$. Desta forma, se $D = \{d_0, d_1\}$, por exemplo, após a observação de $x \in \mathcal{X}$, basta escolher a decisão apropriada, isto é: se $\rho(d_0, \pi_x) < \rho(d_1, \pi_x)$, então $\phi(x) = d_0$ é a decisão de Bayes e, se $\rho(d_1, \pi_x) < \rho(d_0, \pi_x)$, então $\phi(x) = d_1$ é a decisão de Bayes. No caso de empate, pode-se escolher qualquer umas das decisões (DeGroot (1970)). Neste trabalho, especificamente, focaremos em um caso particular dos problemas de decisão, o de teste de hipóteses. Neste caso, em geral, o espaço de decisões D tem apenas dois pontos que significam aceitar (não rejeitar) e rejeitar a hipótese nula especificada. A exceção ocorre quando são considerados os testes aleatorizados.

Vale ressaltar que quando uma quantidade observável X está disponível para tomada de decisão, as penalidades (perdas) associadas ao problema de decisão podem depender não só da ação d , e do estado da natureza θ , mas também dos dados x . Tais funções de perda dependentes da amostra, embora não tão comuns na literatura, são discutidas em Berger (1980), dentre outros.

Definição 1.3. *Suponha que $\{\Theta_0, \Theta_1\}$ é uma partição do espaço paramétrico Θ . A afirmação que $\theta \in \Theta_0$ é chamada hipótese estatística nula e é rotulada H_0 . A correspondente afirmação que $\theta \in \Theta_1$ é chamada hipótese alternativa e é rotulada H_1 . Uma regra de decisão é chamada de teste de hipóteses se $D = \{0, 1\}$ e a função de perda $L : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ satisfaz*

$$L(1, \theta) \geq L(0, \theta), \text{ para } \theta \in \Theta_0 \text{ e } L(1, \theta) \leq L(0, \theta), \text{ para } \theta \in \Theta_1. \quad (1.2)$$

A decisão $d = 0$ é chamada de aceitar (não rejeitar) a hipótese (nula) e a decisão $d = 1$ é chamada de rejeitar a hipótese (nula).

Em alguns textos, o teste de hipóteses é definido sem a condição (1.2). Schervish (1995) impõe essa condição que, apesar de bastante intuitiva, pode ser questionada, pois em tese um teste de hipóteses pode ter uma função de perda que penalize mais a decisão correta do que a decisão errada, embora funções de perda deste tipo possam induzir a testes bastante estranhos, como o do exemplo a seguir.

Exemplo 1.2. *Seja $X|\theta \sim \text{Uniforme}(0, \theta)$, onde $\Theta = \{1, 2\}$ e a distribuição a priori de θ é $\pi(1) = \pi(2) = 1/2$. Suponha que se deseja testar $H_0 : \theta = 1$ contra $H_1 : \theta = 2$ e a função de perda é dada por $L(0, \theta) = I_{\{1\}}(\theta)$, onde $I_A(x)$ é a função indicadora do elemento x no conjunto A , isto é, $I_A(x) = 1$, se $x \in A$ e $I_A(x) = 0$, se $x \notin A$, e $L(1, \theta) = I_{\{2\}}(\theta)$. Portanto, essa função de perda não satisfaz a condição (1.2).*

O teste de Bayes ϕ , gerado por essa função de perda, é dado por

$$\phi(x) = 0 \text{ se, e somente se, } 1 < x \leq 2.$$

Ou seja, mesmo observando $x = 2$, aceitamos a hipótese nula $H_0 : \theta = 1$, isto é, aceitamos que $X|\theta \sim \text{Uniforme}(0, 1)$.

Alguns autores, como Popper (1959) e Mayo e Spanos (2006) questionam se um experimento científico poderia “aceitar a hipótese nula” ou se a denominação correta seria “não rejeitar a hipótese nula”. Particularmente, sob o ponto de vista da Teoria da Decisão, o problema de testes de hipóteses pode ser visto como um

problema em que duas ações podem ser tomadas, decidir-se por H_0 (chamada indistintamente de “aceitar H_0 ”) e decidir-se por H_1 (chamado indistintamente de “rejeitar H_0 ”). Na sequência do texto seguiremos essa notação. [Johnson \(2013\)](#), [Jeffreys \(1939\)](#) e [Howson e Urbach \(2005\)](#) discutem mais profundamente sobre o significado de uma decisão num problema de testes de hipóteses.

Os testes aleatorizados podem ser vistos como uma generalização do problema de testes hipóteses. Pode-se pensar em um teste aleatorizado ϕ da seguinte forma: primeiro, observe $X = x$ e, então, jogue uma moeda com probabilidade de cara igual a $\phi(x)$. Se der cara, rejeita-se a hipótese H_0 . Neste trabalho, quando nos referirmos a teste de hipóteses, estaremos falando de testes de hipóteses não-aleatorizados.

Outras definições importantes acerca de teste de hipóteses são dadas a seguir. A *função poder* β_ϕ de um teste ϕ , $\beta_\phi : \Theta \rightarrow [0, 1]$, é dada por $\beta_\phi(\theta) = E_\theta[P(\phi(X) = 1 | \theta)]$. O *tamanho* de ϕ é $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$. Um teste é chamado de *nível* α , para $0 \leq \alpha \leq 1$, se seu tamanho é no máximo α . A hipótese nula (alternativa) é *simples* se Θ_0 (Θ_1) é um conjunto unitário. A hipótese nula (alternativa) é *composta* se não é simples.

Uma propriedade importante para avaliar a qualidade de uma regra de decisão é a *admissibilidade*. A admissibilidade é uma propriedade aceita tanto pela Escola Frequentista como pela Bayesiana e é definida a seguir:

Definição 1.4. *Uma regra de decisão $\phi : \mathcal{X} \rightarrow \{0, 1\}$ é inadmissível se existe alguma regra ϕ_1 tal que $R(\theta, \phi_1) \leq R(\theta, \phi)$ para todo θ , com a desigualdade estrita para algum $\theta_0 \in \Theta$. Se existe uma tal regra como ϕ_1 , dizemos que ϕ_1 domina ϕ . Se não existe ϕ_1 como acima, então dizemos que ϕ é admissível.*

Para Θ finito, uma regra de decisão admissível é regra de decisão de Bayes contra alguma priori para Θ . Uma demonstração rigorosa deste fato pode ser encontrada em [Ferguson \(1967\)](#) e [French e Insua \(2000\)](#).

Além disso, sob certas condições, uma regra de decisão de Bayes é admissível.

Teorema 1.1. *Seja Θ finito e π uma medida de probabilidade que tenha medida positiva para todos elementos de Θ . Se $\phi : \mathcal{X} \rightarrow \{0, 1\}$ é um teste de Bayes contra π , então ϕ é admissível.*

Demonstração. Vamos mostrar que um teste de Bayes nestas condições não pode ser inadmissível. Seja ϕ um teste de Bayes contra π e suponha que ele seja inadmissível, isto é, existe um teste ϕ^* tal que $R(\theta, \phi^*) \leq R(\theta, \phi)$ para todo θ , com a desigualdade estrita para algum $\theta_0 \in \Theta$. Desta forma,

$$\rho(\phi^*, \pi_x) = \sum_{\theta} R(\theta, \phi^*) \pi(\theta) < \sum_{\theta} R(\theta, \phi) \pi(\theta) = \rho(\phi, \pi_x),$$

a inequação sendo estrita pois, para algum $\theta_0 \in \Theta$, $R(\theta_0, \phi^*) < R(\theta_0, \phi)$ e $\pi(\theta_0)$ é positivo. Então, ϕ não poderia ser teste de Bayes. Portanto, ϕ é admissível. $\square$

O exemplo a seguir mostra que de fato uma regra de decisão admissível não é necessariamente regra de decisão de Bayes, quando Θ é infinito (enumerável).

Exemplo 1.3. ([DeGroot \(1970\)](#)) *Considere um problema de decisão onde Θ e D tem infinitos elementos. Sejam $\Theta = \{\theta_1, \theta_2, \dots\}$, $D = \{d^*, d_1, d_2, \dots\}$ e a função de perda $L : D \times \Theta \rightarrow \mathbb{R}^+$ dada pela Tabela 1.2.*

Podemos notar que, para todo $n \in \mathbb{N}$, $L(d^, \theta_{n+1}) = 1/2 < 1 = L(d_n, \theta_{n+1})$. Logo d^* é admissível. Além disso, para todo $n \in \mathbb{N}$, d_{n+1} domina d_n . Deste modo, d^* é a única decisão admissível do conjunto D .*

Contudo, seja $\pi(\theta_i)$, $i = 1, 2, \dots$ uma medida qualquer sobre Θ . Então, $\rho(d^, \pi) = 1/2$ e $\rho(d_n, \pi) = 1 - \sum_{i=1}^n \pi(\theta_i) = \sum_{i=n+1}^{\infty} \pi(\theta_i)$. Com $\sum_{i=1}^{\infty} \pi(\theta_i) = 1$, existe $n_0 \in \mathbb{N}$ tal que $\rho(d_{n_0}, \pi) < 1/2$ e portanto d^* não é decisão de Bayes contra nenhuma π .*

L	θ_1	θ_2	θ_3	θ_4	$\dots$
d^*	1/2	1/2	1/2	1/2	$\dots$
d_1	0	1	1	1	$\dots$
d_2	0	0	1	1	$\dots$
d_3	0	0	0	1	$\dots$
$\dots$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\dots$

Tabela 1.2: Função de perda do exemplo de Groot

Os resultados acima apresentam, de certa forma, uma justificativa da importância de uma regra de decisão de Bayes para um decisor que aceite que regras de decisão admissíveis são razoáveis.

No Capítulo 3, a admissibilidade será relacionada com as propriedades lógicas propostas na sequência deste capítulo. Para o estudo de tais propriedades em testes simultâneos, é necessário generalizar a definição de teste de hipóteses para que seja possível trabalhar com dois ou mais testes simultaneamente.

1.3 Classes de testes de hipóteses

A seguir, é definida classe de testes de hipóteses, que será utilizada em todos os demais resultados deste trabalho.

Definição 1.5. *Seja $\psi = \{\phi : \mathcal{X} \rightarrow \{0, 1\} \text{ tal que } \phi \text{ é } \sigma(\mathcal{X})\text{-mensurável}\}$, o conjunto de todas as funções de teste. Uma classe de testes de hipóteses é uma função $\varphi : \sigma(\Theta) \rightarrow \psi$ que, para cada hipótese $A \in \sigma(\Theta)$ associa o teste $\varphi_A \in \psi$ para testar $H_0 : \theta \in A$ contra $H_1 : \theta \notin A$. Para $A \in \sigma(\Theta)$, $\varphi_A(x) = 0$ representa decidir pela hipótese $\theta \in A$ a partir da observação $x \in \mathcal{X}$ e $\varphi_A(x) = 1$ representa decidir pela hipótese $\theta \notin A$.*

Definido desta forma, um teste φ_A poderia ser qualquer teste em ψ para testar $H_0 : \theta \in A$. A definição 1.5 engloba classes especiais como uma classe de testes de Razão de Verossimilhança Generalizada (RVG) (Silva (2010)) ou uma classe de testes de Bayes, fazendo com que para cada $A \in \sigma(\Theta)$, φ_A seja um teste de Bayes para $H_0 : \theta \in A$ contra $H_1 : \theta \notin A$, fixada uma perda $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$.

Outros exemplos de classe de testes são dados a seguir.

Exemplo 1.4. *(Classe de testes baseados em probabilidades a posteriori) Suponha que $\Theta = \mathbb{R}^n$ e que $\sigma(\Theta) = \mathcal{B}(\Theta)$, o conjunto dos Borelianos de Θ . Suponha ainda que a distribuição a priori para θ é dada por π . Para cada $A \in \mathcal{B}(\Theta)$, seja $\varphi_A : \mathcal{X} \rightarrow \{0, 1\}$ definida por*

$$\varphi_A(x) = \begin{cases} 1 & \text{se } x \text{ é tal que } \pi(A|x) < \frac{1}{2}, \\ 0 & \text{c.c.} \end{cases}$$

onde $\pi(\cdot|x)$ é a distribuição a posteriori de Θ , dado $X = x$. Ou seja, para cada hipótese $A \in \mathcal{B}(\Theta)$, esta é a classe que associa o teste que decide por $\theta \in A$ quando a probabilidade a posteriori de A é maior ou igual a $1/2$.

Uma abordagem alternativa para o problema de testes de hipóteses foi proposta em Pereira e Stern (1999), onde é definida uma medida de evidência (e-valor) para testes de significância (testes de hipóteses nulas precisas), ou seja, com probabilidade a posteriori (priori) nula (Madruga et al. (2001)). Mais informações sobre este teste podem ser encontradas em Pereira e Stern (1999), Pereira et al. (2008), Madruga et al. (2001) e Stern (2008).

Exemplo 1.5. (Classe de testes FBST) Suponha que $\Theta = \mathbb{R}^n$ e que $\sigma(\Theta) = \mathcal{B}(\Theta)$. Para cada $x \in \mathcal{X}$, seja $\pi(\cdot|x)$ a densidade a posteriori de θ , dado $X = x$. Para cada hipótese $\theta \in A$, seja $T_A(x) = \{\theta : \pi(\theta|x) > \sup_{A^c}(\pi(\theta|x))\}$, o conjunto tangente à hipótese nula $\theta \in A$ e seja

$$EV_A(x) = 1 - \pi(\theta \in T_A(x)|x),$$

a medida de evidência de Pereira-Stern ([Madruga et al. \(2001\)](#)) para a hipótese $\theta \in A$. Para $A = \emptyset$, convencionou-se que $\sup_A(\pi(\theta|x)) = 0$.

Para cada $A \in \mathcal{B}(\Theta)$, pode-se definir a classe de testes como:

$$\varphi_A(x) = \begin{cases} 1 & \text{se } x \text{ é tal que } EV_A(x) \leq c, \\ 0 & \text{c.c.} \end{cases}$$

onde $c \in [0, 1]$ é fixado. Ou seja, decide-se pela hipótese $\theta \in A$, ao observar $x \in \mathcal{X}$ sempre que $EV_A(x)$ for maior que c .

Em geral, para encontrarmos um teste de Bayes, consideramos uma função de perda $L : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ e minimizamos, para cada $x \in \mathcal{X}$, a perda esperada a posteriori baseada nesta função. Ao generalizarmos o conceito de teste de Bayes para uma classe de testes de Bayes, também deveríamos considerar uma família de funções de perda e não apenas uma única função. Deste modo, para cada hipótese nula $A \in \sigma(\Theta)$, associamos uma função de perda $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ a fim de derivarmos o testes de Bayes φ_A . Isto é, para cada possível hipótese nula que se deseja testar, podemos ter uma função de perda diferente.

Definição 1.6. (Classe de testes de Bayes) Seja $\{L_A\}_{A \in \sigma(\Theta)}$ uma família de funções de perda onde, para cada $A \in \sigma(\Theta)$, $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$. Seja π medida de probabilidade sobre Θ . Uma classe de testes de Bayes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π é qualquer classe de testes de hipóteses definida sobre os elementos de $\sigma(\Theta)$ tal que cada elemento φ_A é um teste de Bayes para a hipótese nula $\theta \in A$ contra π considerando a penalidade L_A , $A \in \sigma(\Theta)$.

Exemplo 1.6. Seja $X|\theta \sim U(0, \theta)$, onde $\Theta = \{1, 2, 3, 4\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$ o conjunto das partes de Θ . Além disso, seja $\pi(\theta) = 1/4$ para todo $\theta \in \Theta$. Para cada $A \in \sigma(\Theta)$, define-se $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ como a função de perda 0-1, para cada $A \in \sigma$. Desta forma, a classe de testes definida por

$$\varphi_A(x) = \begin{cases} 1 & \text{se } x \text{ é tal que } \pi(A|x) < 1/2, \\ 0 & \text{c.c.} \end{cases}$$

é uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π . Pois, os riscos esperados da hipótese nula e alternativa são, $\rho(0, \pi_x) = \pi(\theta \notin A|x) = 1 - \pi(A|x)$ e $\rho(1, \pi_x) = \pi(A|x)$, respectivamente. Desta forma, a decisão $d=1$, será a decisão de Bayes se $\rho(1, \pi_x) < \rho(0, \pi_x)$, isto é, se $\pi(A|x) < 1/2$.

No exemplo acima, para cada $A \in \sigma(\Theta)$, temos a correspondente função de perda 0-1. Outros exemplos de classes de testes podem ser encontrados em [Izbicki \(2010\)](#) e [Izbicki e Esteves \(2014\)](#).

A seguir, relembremos o conceito de estimador, bastante comum na literatura estatística.

Definição 1.7. Um estimador para $g(\theta)$ é uma função $W : \mathcal{X} \rightarrow g(\Theta)$, onde $g(\cdot)$ é uma função definida em Θ .

É comum adotar uma definição menos restritiva do que a definição 1.7, onde W é uma função em $\mathbb{R}$ definida sobre o espaço amostral. No entanto, em alguns casos é desejável restringir a função W a possíveis valores de $g(\theta)$.

A definição a seguir será importante no restante do trabalho, principalmente no Capítulo 3.

Definição 1.8. (Classe de testes gerada por um estimador) *Seja $W : \mathcal{X} \rightarrow \Theta$ um estimador para θ . A classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada pelo estimador W é a classe que associa, para cada $A \in \sigma(\Theta)$, o teste $\varphi_A(x) = 1 - I_A(W(x))$.*

Assim, um teste da classe de testes gerada por um estimador W é o teste que aceita a hipótese nula a partir de $x \in \mathcal{X}$, se a estimativa $W(x)$ pertencer à hipótese nula, como podemos observar no exemplo abaixo.

Exemplo 1.7. *Seja $X|\theta$ uma variável aleatória com distribuição P_θ . Seja $W : \mathcal{X} \rightarrow \Theta$ o estimador de máxima verossimilhança (EMV) de θ , isto é, o estimador que associa, para cada $x \in \mathcal{X}$, o valor $W(x)$ tal que $V_x(W(x)) = \sup_{\theta \in \Theta} V_x(\theta)$. A classe de testes definida por $\varphi_A(x) = 1 - I_A(W(x))$, para cada $A \in \sigma(\Theta)$ e cada $x \in \mathcal{X}$, é a classe de testes gerada pelo EMV W .*

A seguir, são definidas algumas propriedades lógicas que podem ser desejáveis em testes simultâneos.

1.4 Propriedades Lógicas

Nesta seção, falaremos sobre três propriedades lógicas que um decisor poderia desejar que testes simultâneos atendessem. Essas propriedades exigem que os testes simultâneos tenham consistência entre suas decisões quando analisados conjuntamente.

1.4.1 Monotonicidade

A primeira destas propriedades é a monotonicidade. A propriedade de monotonicidade aparece em Gabriel (1969) com o nome de coerência (mas não para uma σ -álgebra de Θ de hipóteses de interesse) e, a partir de então, foi explorada em diversos textos, como Rom e Holland (1995), Betz e Levin (1982), Schervish (1996), Lavine e Schervish (1999) e mais recentemente em Hommel e Bretz (2008) e Raviv (2013). Em Fossaluza (2008), Silva (2010), Izbicki (2010) e Izbicki e Esteves (2014), ela aparece com o nome de monotonicidade e é definida para uma σ -álgebra do espaço paramétrico. A sua importância pode ser notada no exemplo a seguir:

Exemplo 1.8. *Seja $X|\theta \sim N(\theta, 2)$ e suponha que se deseja testar $H_0 : \theta \leq 0$ contra $H_1 : \theta > 0$. Suponha que $\theta \sim N(0, 2)$ e que a função de perda é dada pela tabela abaixo,*

L	$\theta \leq 0$	$\theta > 0$
0	0	1
1	2	0

Tabela 1.3: Exemplo de função de perda para teste da Normal

Note que $\theta|x \sim N(x/2, 1)$. Da tabela acima, aceita-se H_0 , se e somente se, $\pi(\theta \leq 0|x) > 1/3$.

Suponha agora que deseja-se testar se $H'_0 : \theta \leq 1$ contra $H'_1 : \theta > 1$ com função de perda dada pela tabela abaixo

L'	$\theta \leq 1$	$\theta > 1$
0	0	6
1	1	0

Tabela 1.4: Exemplo de função de perda para teste da Normal

Novamente, a posteriori $\theta|x \sim N(x/2, 1)$ e, aceita-se H'_0 , se e somente se, $\pi(\theta \leq 1|x) > 6/7$.

Suponha que foi observado $x = 0$, então $\pi(\theta \leq 0|0) = 0,50$ e, conseqüentemente, aceita-se $H_0 : \theta \leq 0$. Mas, rejeita-se que $H'_0 : \theta \leq 1$, pois $\pi(\theta \leq 1|0) = 0,84 < 6/7$.

O exemplo mostra que ao realizar dois testes de hipóteses simultâneos, decidimos por hipóteses contraditórias, o que não é desejável: $x = 0$ traz evidências para rejeição de que $\theta \leq 1$, mas não traz evidência estatística de que $\theta \leq 0$. Podemos encontrar diversos exemplos de testes clássicos definidos na literatura estatística que não atendem a essa propriedade como, por exemplo, testes baseados em p -valores, testes RVG, entre outros.

Definição 1.9. (Monotonicidade) Uma classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona, se para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, $\varphi_A(x) \geq \varphi_B(x)$, para todo $x \in \chi$.

Isto é, caso tenha-se aceitado uma hipótese, deveria-se aceitar toda hipótese implicada por ela. A seguir, dois exemplos de classes de testes monótonas.

Exemplo 1.9. Suponha que $\Theta = \mathbb{R}^n$ e que $\sigma(\Theta) = \mathcal{B}(\Theta)$. Para cada $x \in \chi$, seja $\pi(\cdot|x)$ a densidade a posteriori de θ , dado $X = x$. Para cada hipótese $\theta \in A$, sejam $T_A(\cdot)$ e $EV_A(\cdot)$ definidos como no Exemplo 1.5. Se $A \subseteq B$, então $T_B(x) = \{\theta : \pi(\theta|x) > \sup_B \pi(\theta|x)\} \subseteq \{\theta : \pi(\theta|x) > \sup_A \pi(\theta|x)\} = T_A(x)$, para todo $x \in \chi$, de forma que, $EV_A(x) \leq EV_B(x)$. Portanto, fixado $c \in [0, 1]$, $\varphi_A(x) \geq \varphi_B(x)$ para todo $x \in \chi$, isto é, a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona.

Exemplo 1.10. Suponha que $\Theta = \mathbb{R}^n$ e $\sigma(\Theta) = \mathcal{B}(\Theta)$. Seja $R : \chi \rightarrow \mathcal{B}(\Theta)$ uma estimativa por região para θ . O teste

$$\varphi_A(x) = \begin{cases} 1 & \text{se } A \cap R(x) = \emptyset, \\ 0 & \text{se } A \cap R(x) \neq \emptyset \end{cases}$$

conhecido como Teste de Aitchison ou teste de Lindley (Zellner (1971)), é o teste que rejeita uma hipóteses nula se, e somente se, uma específica estimativa por região não intersecciona a hipótese nula A .

Para todo $x \in \chi$, tal que $A \cap R(x) \neq \emptyset$, teremos que $B \cap R(x) \neq \emptyset$, se $A \subseteq B$. Ou seja, para todo $x \in \chi$, com $A \subseteq B$, se $\varphi_A(x) = 0$, então $\varphi_B(x) = 0$. Portanto, a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona.

Mais exemplos de classes monótonas e não monótonas, inclusive para testes clássicos, podem ser encontrados em Silva (2010), Izbicki (2010), Schervish (1996), Fossaluza (2008) e Izbicki e Esteves (2014).

Dentre as várias propriedades lógicas desejáveis em testes simultâneos, a monotonicidade (coerência) é indubitavelmente a mais discutida. Vimos acima que algumas classes de testes bayesianos podem não apresentar essa propriedade, o que nos leva a questionar quais famílias de funções de perda confeririam monotonicidade aos testes Bayesianos. A resposta a essa pergunta será dada no Capítulo 2.

1.4.2 Invertibilidade

Sob a perspectiva da Teoria da Decisão Bayesiana, é de se esperar que os rótulos “nula” e “alternativa” dados às hipóteses $\theta \in A$ e $\theta \in A^c$ não influenciem a decisão por uma dessas hipóteses. Em outras palavras,

se especificarmos $H_0 : \theta \in A$ e $H_1 : \theta \notin A$ e decidirmos por $H_0(H_1)$ a partir da observação do ponto amostral x , talvez seja natural imaginar que ao testarmos $H'_0 : \theta \in A^c$ e $H'_1 : \theta \notin A^c$, a decisão baseado no mesmo x deva ser decidir por $H'_1(H'_0)$.

Schervish (1995) comenta que “existe uma dualidade entre hipóteses nulas e alternativas que não são respeitadas na maioria da literatura de testes de hipóteses clássicos”. Segundo ele, um teste pode satisfazer um critério de otimalidade clássico para um par de hipóteses nula e alternativa específico, mas quando se inverte as hipóteses nula e alternativa, o teste resultante não mais satisfaz este critério. Um exemplo desse caso, pode ser encontrado em Izbicki (2010). Além disso, o autor define um teste Uniformemente Mais Cauteloso (UMC), que a grosso modo seria um teste Uniformemente Mais Poderoso (UMP) indiferente a essa troca da hipótese nula com a alternativa. Schervish (1995) avalia ainda condições sobre funções de risco frequentistas de modo que a dualidade entre hipóteses nulas e alternativas seja respeitada.

Robert (2007) diz que “os testes uniformemente mais poderosos produzem situações de assimetria entre a hipótese nula e alternativa, o que induz os procedimentos a comportamentos não naturais”. Ou seja, ele critica a maneira como estes testes clássicos tratam de forma diferente as hipóteses nula e alternativa, pois o natural em um procedimento de testes de hipóteses seria que a decisão indicada pelo teste não fosse vulnerável à especificação de H_0 e H_1 .

Novamente é possível encontrar exemplos de testes clássicos, como testes baseados em p -valores e testes RVG, além do UMP comentado acima, que não atendem tal dualidade, daqui em diante denominada *invertibilidade*. Silva (2010), Izbicki (2010) e Izbicki e Esteves (2014) também discutem, através de exemplos, este problema em outros testes de hipóteses clássicos comumente usados na literatura. A seguir um exemplo de testes bayesianos em que a dualidade descrita por Schervish não é respeitada.

Exemplo 1.11. Suponhamos que $X|\theta \sim N(\theta, 2)$ e que deseja-se testar as seguintes hipóteses (nulas) $H_0 : \theta \leq 0$ e $H'_0 : \theta > 0$. Considere a priori que $\theta \sim N(0, 2)$. Suponha ainda que as funções de perda para os dois testes são dadas a seguir

L	$\theta \leq 0$	$\theta > 0$
0	0	5
1	95	0

Tabela 1.5: Exemplo de função de perda para teste da Normal

L'	$\theta > 0$	$\theta \leq 0$
0	0	5
1	95	0

Tabela 1.6: Exemplo de função de perda para teste da Normal

Note que a posteriori $\theta|x \sim N(x/2, 1)$. Das tabelas acima, a decisão de Bayes é aceitar H_0 , se e somente se, $\pi(\theta \leq 0|x) > 0,05$ e aceitar H'_0 , se e somente se, $\pi(\theta > 0|x) > 0,05$.

Considerando que foi observado $x = 0$, aceita-se H_0 e aceita-se H'_0 , pois

$$\pi(\theta \leq 0|0) = \pi(\theta > 0|0) = 0,5 > 0,05.$$

Isto é, dependendo de qual a hipótese definida como hipótese nula, essa não será rejeitada.

O exemplo acima mostra que ao realizar dois testes, trocando a hipótese nula com a alternativa, decidimos em ambos os casos pela hipótese nula. A seguir, definiremos formalmente a propriedade de invertibilidade.

Definição 1.10. (*Invertibilidade*) Uma classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é invertível, se para todo $A \in \sigma(\Theta)$, $\varphi_A(x) = 1 - \varphi_{A^c}(x)$, para todo $x \in \mathcal{X}$.

Isto é, não se deveria esperar que um procedimento de testes indique decisões diferentes em função da hipótese de interesse ser considerada hipótese nula ou hipótese alternativa.

A seguir um exemplo de classe de testes de Bayes invertíveis.

Exemplo 1.12. Novamente, suponha que $X|\theta \sim N(\theta, 2)$ e que deseja-se testar as seguintes hipóteses (nulas) $H_0 : \theta \leq 0$ e $H'_0 : \theta > 0$. Considere a priori que $\theta \sim N(0, 2)$. Suponha agora que as funções de perda para os dois testes são dadas a seguir

L	$\theta \leq 0$	$\theta > 0$
0	0	1
1	1	0

Tabela 1.7: Exemplo de função de perda para teste da Normal

L'	$\theta > 0$	$\theta \leq 0$
0	0	1
1	1	0

Tabela 1.8: Exemplo de função de perda para teste da Normal

Note que a posteriori $\theta|x \sim N(x/2, 1)$. Das tabelas 1.7 e 1.8, a decisão de Bayes é aceitar H_0 , se e somente se, $\pi(\theta \leq 0|x) > 0,5$ e aceitar H'_0 , se e somente se, $\pi(\theta > 0|x) > 0,5$.

Ou seja, para qualquer $x \in \mathcal{X}$, não aceitamos simultaneamente H_0 e H'_0 . É possível provar que o mesmo vale para qualquer H_0 e H'_0 , portanto a classe de testes é invertível.

A partir dos exemplos 1.11 e 1.12, verificamos que há testes simultâneos que obedecem e não obedecem à invertibilidade, o que também nos leva a questionar quais famílias de funções de perda confeririam invertibilidade aos testes Bayesianos.

A construção de testes clássicos simultâneos, em geral, não levam em conta a propriedade de invertibilidade, pois sob o enfoque frequentista, a decisão $\varphi_A(x) = 0$ é de “não rejeitar” H_0 ao invés de “aceitar H_0 ”, como falamos anteriormente. Por isso, poderia-se afirmar que esta propriedade não seria importante no contexto frequentista. No entanto, [Izbicki e Esteves \(2014\)](#) argumentam que, no contexto de classes de testes de hipóteses, a ausência da invertibilidade em testes simultâneos pode conduzir à violação do *Princípio da Compatibilidade do tipo I* ([Lehmann \(1957b\)](#)).

Este princípio estabelece que, para todo x , a intersecção dos complementos das regiões rejeitadas não deve ser vazio, isto é,

$$\bigcap_{B \in \sigma(\Theta): \varphi_B(x)=1} B^c \neq \emptyset,$$

como no exemplo abaixo.

Exemplo 1.13. Sejam $\Theta = \{-3, 3\}$ e $X|\theta \sim N(\theta, 1)$ e suponha que desejamos testar as seguintes hipóteses nulas $H_0 : \theta \in A = \{-3\}$ e $H'_0 : \theta \in A^c = \{3\}$. Pelo Lema de Neyman-Pearson (Robert (2007)), os testes mais poderosos (MP) de nível $\alpha = 0,05$ devem ser

$$\phi(x) = \begin{cases} 1 & \text{se } x > -1,65 \\ 0 & \text{caso contrário.} \end{cases}$$

$$\phi'(x) = \begin{cases} 1 & \text{se } x < 1,65 \\ 0 & \text{caso contrário.} \end{cases}$$

Observado $x = 0$, rejeitamos H_0 e H'_0 . Assim,

$$\bigcap_{B \in \sigma(\Theta) : \phi_B(x)=1} B^c \subseteq A^c \cap A = \emptyset,$$

o que viola o princípio da compatibilidade do tipo I.

Mais detalhes sobre o Princípio da Compatibilidade do tipo I podem ser encontrados em Izbicki e Esteves (2014).

1.4.3 Consonância da união

A próxima propriedade, chamada consonância da união, afirma que se aceitarmos uma hipótese formada pela união de dois conjuntos A_1 e A_2 , $A_1 \cup A_2$, então ao menos um dos A_i s não deve ser rejeitado. O exemplo abaixo ilustra uma situação onde isso não ocorre.

Exemplo 1.14. Seja $X|\theta \sim \text{Multinomial}(410, \theta)$, onde $\theta = (\theta_1, \theta_2, \theta_3)$ e, a priori, $\theta \sim \text{Dirichlet}(1, 1, 1)$. Suponha que deseja-se testar as seguintes hipóteses nulas $H_0^U : \cup_{i=1}^3 \theta_i > 1/2$, $H_0^1 : \theta_1 > 1/2$, $H_0^2 : \theta_2 > 1/2$ e $H_0^3 : \theta_3 > 1/2$. Se foi observado $x = (200, 200, 10)$, a distribuição a posteriori é dada por $\theta|x \sim \text{Dirichlet}(201, 201, 11)$, de modo que

$$\pi(\cup_{i=1}^3 \{\theta_i > 1/2\} | x) = 0,588;$$

$$\pi(\{\theta_1 > 1/2\} | x) = 0,294;$$

$$\pi(\{\theta_2 > 1/2\} | x) = 0,294 \text{ e}$$

$$\pi(\{\theta_3 > 1/2\} | x) = 0,000.$$

Considerando que, para todas as hipóteses nulas testadas, são adotadas função de perda 0-1, aceita-se H_0^U e rejeita-se H_0^i , $i = 1, 2, 3$. Isto é, aceita-se que θ está no conjunto $\cup_{i=1}^3 \{\theta_i > 1/2\}$ e, ao mesmo tempo, rejeita-se que ele está em alguma parte deste conjunto, em algum dos $\{\theta_i > 1/2\}$.

A seguir, definiremos a propriedade de consonância com a união. Vamos considerar dois tipos de consonância da união, a finita e a enumerável. Em Izbicki e Esteves (2014) são definidas outros tipos de consonância que não serão utilizadas nestes trabalho.

Definição 1.11. (Consonância da união finita) Uma classe de testes $\{\phi_A\}_{A \in \sigma(\Theta)}$ é consonante com a união finita, se para todo $A, B \in \sigma(\Theta)$, $\phi_{A \cup B}(x) \geq \phi_A(x) \phi_B(x)$, para todo $x \in \mathcal{X}$.

Isto é, caso tenha-se aceitado a união de dois subconjuntos A e B , então deveria-se aceitar pelo menos um dos dois subconjuntos. Observação: É possível verificar, por indução, que uma classe de testes é consonante com a união se, e somente se, para todo $A_1, \dots, A_n$, com $n \geq 1$, $\varphi_{\cup_{i=1}^n A_i} \geq \prod_{i=1}^n \varphi_{A_i}$.

Definição 1.12. (*Consonância da união enumerável*) Uma classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é consonante com a união enumerável, se para todo $A_1, A_2, \dots \in \sigma(\Theta)$, $\varphi_{\cup_{i=1}^\infty A_i} \geq \prod_{i=1}^\infty \varphi_{A_i}$, para todo $x \in \mathcal{X}$.

A seguir, um exemplo de classe de testes consonante com a união finita e enumerável.

Exemplo 1.15. Suponha que $\Theta = \mathbb{R}$ e $\sigma(\Theta)$ é o conjunto dos Borelianos (Lima (1977)) de $\mathbb{R}$, $\mathbb{B}(\mathbb{R})$. Além disso, seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes tal que

$$\varphi_A(x) = 1 - I_A(W(x)),$$

onde $W(\cdot)$ é o EMV para θ . Sejam $A, B \in \sigma(\Theta)$, para todo $x \in \mathcal{X}$, tal que $\varphi_A(x) = 1$ e $\varphi_B(x) = 1$, teremos $W(x) \notin A$ e $W(x) \notin B$ e portanto, $W(x) \notin A \cup B$, ou seja, $\varphi_{A \cup B}(x) = 1$.

Analogamente, sejam $A_1, A_2, \dots \in \sigma(\Theta)$, para todo $x \in \mathcal{X}$, tal que $\varphi_{A_1}(x) = 1$, $\varphi_{A_2}(x) = 1$, $\dots$, teremos $W(x) \notin A_1$, $W(x) \notin A_2$, $\dots$ e portanto, $W(x) \notin \cup_{i=1}^\infty A_i$, ou seja, $\varphi_{\cup_{i=1}^\infty A_i}(x) = 1$.

Exemplo 1.16. Novamente suponha que $\Theta = \mathbb{R}$, $\sigma(\Theta) = \mathbb{B}(\mathbb{R})$ e $W(\cdot)$ é o EMV para θ , além disso, $X|\theta \sim N(\theta, 1)$. Seja a distância de um ponto y a um conjunto C definida como $\text{dist}(y, C) = \inf_{z \in C} d(y, z)$, onde $d(y, z)$ é a distância euclidiana entre y e z (Deza e Deza (2012)). Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes tal que

$$\varphi_A(x) = \begin{cases} 1 & \text{se } \text{dist}(W(x), A) > 0, \\ 0 & \text{caso contrário.} \end{cases}$$

Seja $A_1, \dots, A_n \in \sigma(\Theta)$, para todo $x \in \mathcal{X}$, tal que $\varphi_{A_i}(x) = 1$, $i = 1, \dots, n$, teremos que

$$\text{dist}(W(x), A_i) > 0, i = 1, \dots, n$$

e, portanto, $\text{dist}(W(x), \cup_{i=1}^n A_i) > 0$, ou seja, $\varphi_{\cup_{i=1}^n A_i}(x) = 1$. De forma que, a classe de testes é consonante com a união finita.

Por outro lado, seja

$$A_n = \left(-\infty, -\frac{1}{n} \right].$$

Observado $x = 0$, o EMV é $W(0) = 0$, logo, $\text{dist}(0, (-\infty, 1/n]) = 1/n > 0$, para todo $n \geq 1$. Mas,

$$\text{dist}(0, \cup_{n=1}^\infty (-\infty, -1/n]) = \text{dist}(0, \cup_{n=1}^\infty (-\infty, 0]) = 0$$

e, portanto, $\varphi_{\cup_{n=1}^\infty A_n}(x) = 0$, isto é a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ não é consonante com a união enumerável.

A propriedade de consonância da união aparece também em Gabriel (1969), onde é definida informalmente da seguinte forma: “alguns procedimentos de teste simultâneos podem rejeitar uma hipótese sem rejeitar todos os componentes contidos nessa hipótese, esses procedimentos são chamados de não-consonantes”. Além disso, Izbicki e Esteves (2014) afirmam que “a consonância com a união não foi formalmente definida na literatura, embora tenha sido sugerida *en passant*. Por exemplo, a interpretação dada por Finner e Strassburger (2002) às decisões finais obtidas em um procedimento de testes simultâneos, sugere implicitamente que a consonância com a união é razoável.”

Em [Izbicki \(2010\)](#), essa propriedade é explorada mais profundamente e são mostrados outros casos de consonância como, por exemplo, a consonância da intersecção e as relações entre essas propriedades.

1.5 Desideratas

Os exemplos mostrados nas subseções 1.4.1, 1.4.2 e 1.4.3 indicam que somente as restrições (1.2) sobre penalidades utilizadas na definição de testes de hipóteses não bastam para produzir testes simultâneos atendendo algumas propriedades lógicas, de modo que é preciso encontrar algumas outras condições para que isso aconteça.

Em [Izbicki \(2010\)](#), são definidas desideratas (de propriedades lógicas) que uma classe de testes de hipóteses deveria atender. Essas desideratas são conjuntos de postulados lógicos que se espera que testes simultâneos devam obedecer. Neste trabalho serão consideradas duas desideratas, D1 e D2 descritas a seguir:

D1 A classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ deve satisfazer

1. Invertibilidade: Para todo $A \in \sigma(\Theta)$, $\varphi_A = 1 - \varphi_{A^c}$.
2. Monotonicidade: Para todo $A, B \in \sigma(\Theta)$, com $A \subseteq B \Rightarrow \varphi_A \geq \varphi_B$.

Existem classes de testes que atendem a desiderata D1, por exemplo, classes de testes de Bayes geradas por famílias de funções de perda 0-1. Segue abaixo um resultado importante acerca de classes monótonas e invertíveis.

Teorema 1.2. ([Izbicki \(2010\)](#)) Se a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ atende a desiderata D1, então para toda partição $\{A_1, \dots, A_n\}$ de Θ ,

$$\sum_{i=1}^n [1 - \varphi_{A_i}(x)] \leq 1, \text{ para todo } x \in \chi. \quad (1.3)$$

Isto é, aceita-se no máximo um elemento da partição.

Demonstração. Seja $\{A_1, \dots, A_n\}$ uma partição finita mensurável de Θ e seja $x \in \chi$. Suponha que exista $i \in \{1, \dots, n\}$ tal que $\varphi_{A_i}(x) = 0$. Pela invertibilidade $\varphi_{A_i^c}(x) = 1$ e como $A_i \cap A_j = \emptyset$ para todo $j \neq i$, $A_j \subseteq A_i^c$ e, pela monotonicidade, $\varphi_{A_j}(x) = 1$, para todo $j \neq i$, de modo que $\sum_{i=1}^n (1 - \varphi_{A_i}(x)) = 1$.

Além disso, se não existe $i \in \{1, \dots, n\}$ tal que $\varphi_{A_i}(x) = 0$, então $\sum_{i=1}^n (1 - \varphi_{A_i}(x)) = 0$. □

É possível mostrar que o Teorema 1.2 também é válido para qualquer partição enumerável de Θ .

Exemplo 1.17. Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes tal que, para cada $A \in \sigma(\Theta)$, φ_A é teste de Bayes gerado pela seguinte função de perda 0-1. Podemos escrever

$$\varphi_A(x) = \begin{cases} 1 & \text{se } x \text{ é tal que } \rho_A(0, \pi_x) - \rho_A(1, \pi_x) > 0, \\ 0 & \text{caso contrário,} \end{cases}$$

onde $\rho_A(d, \pi_x) = E[L_A(d, \theta) | X = x]$, $d = 0, 1$. Se $\pi(\cdot | x)$ denota a posteriori de θ dado $X = x$, para cada $A \in \sigma(\Theta)$, $\rho_A(0, \pi_x) - \rho_A(1, \pi_x) = \pi(A^c | x) - \pi(A | x)$. Segue que, para todo $x \in \chi$ e todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, $\rho_A(0, \pi_x) - \rho_A(1, \pi_x) = \pi(A^c | x) - \pi(A | x) \geq \pi(B^c | x) - \pi(B | x) = \rho_B(0, \pi_x) - \rho_B(1, \pi_x)$ e, portanto, $\varphi_A(x) \geq \varphi_B(x)$, pois, se $\rho_A(0, \pi_x) - \rho_A(1, \pi_x) < 0$, então $\rho_B(0, \pi_x) - \rho_B(1, \pi_x) < 0$.

Para cada $A \in \sigma(\Theta)$ e cada $x \in \chi$, $\rho_A(0, \pi_x) - \rho_A(1, \pi_x) = -\rho_{A^c}(0, \pi_x) + \rho_{A^c}(1, \pi_x)$. Segue que, $\varphi_A(x) = 1 - \varphi_{A^c}(x)$. Devemos ressaltar que caso de empate, isto é, se $\rho_A(0, \pi_x) = \rho_A(1, \pi_x)$ pode-se tomar d como decisão para A e $1 - d$ como decisão para A^c , $d \in \{0, 1\}$.

Portanto, $\{\varphi_A\}_{A \in \sigma(\Theta)}$ atende D1.

D2 A classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ deve satisfazer

1. Invertibilidade: Para todo $A \in \sigma(\Theta)$, $\varphi_A = 1 - \varphi_{A^c}$.
2. Monotonicidade e consonância da união finita: Para todo $A, B \in \sigma(\Theta)$, $\varphi_{A \cup B} = \varphi_A \varphi_B$.

Observação: Note que a monotonicidade também pode ser caracterizada por $\varphi_{A \cup B} \leq \varphi_A \varphi_B$, pois, por definição, $\varphi_{A \cup B} \leq \varphi_A$ e $\varphi_{A \cup B} \leq \varphi_B$. Dessa caracterização, junto com a definição de consonância com a união finita, segue a condição 2 da desiderata D2. A seguir, uma outra caracterização dessa desiderata.

Teorema 1.3. (*Izbicki (2010)*) Uma classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ atende a desiderata D2 se, e somente se, para toda partição finita $\{A_1, \dots, A_n\}$ de Θ ,

$$\sum_{i=1}^n [1 - \varphi_{A_i}(x)] = 1, \text{ para todo } x \in \chi. \quad (1.4)$$

Isto é, aceita-se exatamente um elemento da partição $\{A_1, \dots, A_n\}$, para cada $x \in \chi$.

Demonstração. Primeiro provaremos que uma classe de testes que atende D2, aceita somente um elemento de uma partição finita qualquer.

Seja $\{A_1, \dots, A_n\}$ uma partição finita mensurável de Θ e seja $x \in \chi$. Pela consonância da união,

$$\prod_{i=1}^n \varphi_{A_i}(x) \leq \varphi_{\cup A_i}(x) = \varphi_{\Theta}(x) = 0,$$

onde a última igualdade vale pela monotonicidade ($\varphi_{\Theta}(x) \leq \varphi_{\emptyset}(x)$) e pela invertibilidade ($\varphi_{\Theta}(x) = 1 - \varphi_{\emptyset}(x)$). Desta forma, existe $i_0 \in \{1, \dots, n\}$ tal que $\varphi_{A_{i_0}}(x) = 0$. Mas, para todo $j \neq i_0$, pelo mesmo argumento usado no Teorema 1.2 vale que $\varphi_{A_j}(x) = 1$ para todo $j \neq i_0$, pois $A_j \cap A_{i_0} = \emptyset$ para todo $j \neq i_0$, então $\sum_{i=1}^n (1 - \varphi_{A_i}) = 1$.

Provaremos agora a recíproca. Seja $A \in \sigma(\Theta)$ e $x \in \chi$. Considere, para $n = 2$, a partição $A_1 = A$ e $A_2 = A^c$. Vale que $1 - \varphi_A(x) + 1 - \varphi_{A^c}(x) = 1$, então $\varphi_A(x) = 1 - \varphi_{A^c}(x)$, isto é, vale a invertibilidade.

Sejam $A, B \in \sigma(\Theta)$ com $A \subseteq B$ e $x \in \chi$. Considere a partição $A_1 = A$, $A_2 = B - A = B \cap A^c$ e $A_3 = (A \cup B)^c$. Vale que $1 - \varphi_A(x) + 1 - \varphi_{B-A}(x) + 1 - \varphi_{(A \cup B)^c}(x) = 1$. Supondo que $\varphi_A(x) = 0$, então $\varphi_{(A \cup B)^c}(x) = 1$ e, pela invertibilidade, $\varphi_{A \cup B}(x) = 0$. Mas $\varphi_{A \cup B}(x) = \varphi_B(x) = 0$ e, portanto, vale a monotonicidade.

Por fim, sejam $A, B \in \sigma(\Theta)$. Considerando a mesma partição $A_1 = A$, $A_2 = B - A = B \cap A^c$ e $A_3 = (A \cup B)^c$, mas agora supondo que $\varphi_{A \cup B}(x) = 0$ temos, pela invertibilidade $\varphi_{(A \cup B)^c}(x) = 1$. Assim, ou $\varphi_A(x) = 0$ ou $\varphi_{B-A}(x) = 0$. Mas, pela monotonicidade, $\varphi_{B-A}(x) = 0$ implica que $\varphi_B(x) = 0$. Deste modo, se $\varphi_{A \cup B}(x) = 0$, então $\varphi_A(x) \varphi_B(x) = 0$, isto é, vale a consonância da união finita. $\square$

Do mesmo modo, prova-se resultado similar para a consonância da união enumerável. A seguir, um exemplo de classe de testes que atende D2.

Exemplo 1.18. Considere o modelo estatístico com família de distribuição $\mathbb{P} = \{P_{\theta} : \theta \in \Theta\}$. Para cada $x \in \chi$, seja $V_x(\cdot)$ a função de verossimilhança para θ gerada por x . Seja W um EMV para θ . Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes tal que, para todo $A \in \sigma(\Theta)$,

$$\varphi_A(x) = 1 - I_A(W(x)),$$

para todo $x \in \mathcal{X}$. Isto é, decide-se pela hipótese (nula) $\theta \in A$ se o EMV de θ pertence a A . É possível provar que a classe $\{\varphi_A\}_{A \in \sigma(\Theta)}$ atende à desiderata D2. E, por consequência, também atende à desiderata D1.

Com efeito, será mostrado no exemplo a seguir que classes de testes geradas por estimadores, como a do exemplo anterior, sempre atendem D2.

Exemplo 1.19. *Seja $W : \mathcal{X} \rightarrow \Theta$ um estimador para θ e $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada pelo estimador W . Para cada $x \in \mathcal{X}$ e qualquer partição $\{A_1, \dots, A_k\}$ de Θ , $W(x)$ estará em um, e somente um, elemento dessa partição.*

Portanto, pelo Teorema 1.3, uma classe de testes gerada por um estimador atende à desiderata D2.

Na verdade, provaremos no Capítulo 3, que se Θ for finito, então a classe de testes atender D2 implica que ela é gerada por algum estimador W .

Nos próximos capítulos serão derivados alguns resultados com o objetivo de caracterizar, sob o ponto de vista da Teoria da Decisão Bayesiana, classes de testes Bayesianos que são monótonos, invertíveis e consoantes com a união.

1.6 Objetivos

Resumidamente, os objetivos deste trabalho são:

Sob a perspectiva da Teoria da Decisão Bayesiana, caracterizar as propriedades de monotonicidade, invertibilidade e consonância da união, isto é, especificar condições necessárias e suficientes, sobre uma família de funções de perda, para que a correspondente *classe de testes de Bayes* seja, respectivamente, monótona, invertível e consonante com a união.

Além disso, também sob a perspectiva da Teoria da Decisão Bayesiana, iremos caracterizar, se existir, os testes bayesianos que atendem a desiderata D1 e desiderata D2. Também iremos avaliar a relação entre testes monótonos, invertíveis e consonantes com a união com testes gerados por estimadores. Em especial, associar classes de testes de Bayes atendendo D2 e classes de testes gerados por estimadores de Bayes.

Capítulo 2

Classes de testes de Bayes e suas propriedades lógicas

Neste capítulo, iremos caracterizar as propriedades lógicas definidas no capítulo anterior. Isto é, mostraremos como devem ser as famílias de funções de perda para que as classes de testes de Bayes geradas por elas atendam monotonicidade, invertibilidade e consonância da união. Além disso, iremos mostrar, caso existam, como devem se portar essas perdas para que os correspondentes testes atendam essas propriedades conjuntamente.

Por fim, mostraremos que, sob certas condições, para uma classe de testes qualquer $\{\varphi_A\}_{A \in \sigma(\Theta)}$ atendendo D2, existem uma distribuição de probabilidade μ em $\Theta \times \mathcal{X}$ e uma família de funções de perda tais que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes gerada por essa família de funções de perda contra a distribuição a priori (marginal de μ) para θ . Isso sugere que uma classe atendendo D2 é a classe de testes de algum decisor bayesiano.

Começaremos nossa análise mostrando que a família de funções de perda 0-1 gera classes de testes de Bayes com algumas propriedades lógicas.

Exemplo 2.1. *Seja $X|\theta \sim \text{Bernoulli}(\theta)$, onde $\Theta = \{1/4, 1/3, 1/2\}$, $\sigma(\Theta) = \mathcal{P}(\Theta)$, e a distribuição a priori para θ dada por $\pi(1/4) = \alpha$, $\pi(1/3) = \beta$ e $\pi(1/2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ tais que $\alpha + \beta \leq 1$. Suponha que $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ é uma família de funções de perda tal que, para cada $A \in \sigma(\Theta)$, L_A é a função de perda 0-1. Uma classe de testes de hipóteses $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π é tal que, para todo $A \in \sigma(\Theta)$, $\varphi_A(x) = 0$ se, e somente se, $\pi(\theta \in A|x) > 1/2$.*

As tabelas 2.1 e 2.2 mostram os valores de (α, β) para os quais o teste φ_A aceita a hipótese nula para $x = 0$ e para $x = 1$, respectivamente, $A \in \mathcal{P}(\Theta) \setminus \{\emptyset, \Theta\}$.

Além disso, $\varphi_\emptyset(x) = 1$ e $\varphi_\Theta(x) = 0$, para todo $x \in \mathcal{X}$ e todo $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$. Os valores de (α, β) são ilustrados nas Figura 2.1 e Figura 2.2.

É fácil ver que para todo $A, B \in \mathcal{P}(\Theta)$ com $A \subseteq B$, $\varphi_A(x) \geq \varphi_B(x)$ e $\varphi_A(x) = 1 - \varphi_{A^c}(x)$ para todo $x \in \mathcal{X}$, isto é, a classe de testes gerada pela família de funções de perda definida aqui é monótona e invertível para toda distribuição a priori para θ (veja Exemplo 1.17).

No entanto, se $x = 0$, $\alpha = 0, 2$ e $\beta = 0, 2$, por exemplo, então $\varphi_{\{1/4, 1/3\}}(0) = 0 < 1 = \varphi_{\{1/4\}}(0)\varphi_{\{1/3\}}(0)$, ou seja, a classe de testes não é consonante com a união.

Analogamente, se $x = 1$, $\alpha = 0, 2$ e $\beta = 0, 2$, então $\varphi_{\{1/4, 1/3\}}(1) = 0 < 1 = \varphi_{\{1/4\}}(1)\varphi_{\{1/3\}}(1)$, ou seja, a classe de testes não é consonante com a união.

De fato, para qualquer $(\alpha, \beta) \in C_3$ (Figura 2.1), quando $x = 0$, ou qualquer $(\alpha, \beta) \in C'_3$ (Figura 2.2),

A	$(\alpha, \beta) \in$
$\{1/4\}$	$C_1 = \{(\alpha, \beta) : \alpha > (2\beta + 6)/15\}$
$\{1/3\}$	$C_2 = \{(\alpha, \beta) : \alpha < (14\beta - 6)/3\}$
$\{1/2\}$	$C_3 = \{(\alpha, \beta) : \alpha < (6 - 14\beta)/15\}$
$\{1/4, 1/3\}$	C_3
$\{1/4, 1/2\}$	C_2
$\{1/3, 1/2\}$	C_1

Tabela 2.1: Distribuições a priori para as quais $\varphi_A(0) = 0$

A	$(\alpha, \beta) \in$
$\{1/4\}$	$C'_1 = \{(\alpha, \beta) : \alpha > (6 - 2\beta)/9\}$
$\{1/3\}$	$C'_2 = \{(\alpha, \beta) : \alpha > (6 - 10\beta)/3\}$
$\{1/2\}$	$C'_3 = \{(\alpha, \beta) : \alpha < (6 - 10\beta)/9\}$
$\{1/4, 1/3\}$	C'_3
$\{1/4, 1/2\}$	C'_2
$\{1/3, 1/2\}$	C'_1

Tabela 2.2: Distribuições a priori para as quais $\varphi_A(1) = 0$

quando $x = 1$, a classe de testes gerada pelas funções de perda 0-1 não é consonante com a união.

A classe de testes do exemplo acima é invertível e monótona, porém não é consonante com a união finita, isto é, atende à desiderata D1, mas não atende à desiderata D2. Na sequência deste capítulo, iremos investigar quais classes de testes de Bayes atendem D1, quais atendem D2 e as condições para que isso aconteça sob o enfoque da Teoria da Decisão Bayesiana.

O teorema abaixo mostra que sempre que a família de funções de perda tiver restrições como as que ocorrem no Exemplo 2.1, a classe de testes de Bayes gerada por ela atenderá à desiderata D1.

Teorema 2.1. *Seja $\{L_A\}_{A \in \sigma(\Theta)}$ uma família de funções de perda tal que*

- (i) *para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, $L_A(0, \theta) - L_A(1, \theta) \geq L_B(0, \theta) - L_B(1, \theta)$, para todo $\theta \in \Theta$.*
- (ii) *para todo $A \in \sigma(\Theta)$, $L_A(0, \theta) = L_{A^c}(1, \theta)$, para todo $\theta \in \Theta$.*

Então, uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ atende à desiderata D1, para qualquer distribuição a priori para θ .

Demonstração. Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ atendendo (i) e (ii). Primeiro, mostraremos que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona e, em seguida, mostraremos que é invertível.

Sejam $A, B \in \sigma(\Theta)$ tais que $A \subseteq B$ e suponha que, para todo $x \in \mathcal{X}$ e todo $d \in D$, $\rho_A(d, \pi_x) < \infty$. Pela condição (i) da hipótese do teorema, como $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes, então $\rho_A(0, \pi_x) - \rho_A(1, \pi_x) \geq \rho_B(0, \pi_x) - \rho_B(1, \pi_x)$, para todo $x \in \mathcal{X}$ e contra toda priori π para θ .

Desta forma, seja $x \in \mathcal{X}$ tal que $\varphi_A(x) = 0$, isto é, $0 \geq \rho_A(0, \pi_x) - \rho_A(1, \pi_x) \geq \rho_B(0, \pi_x) - \rho_B(1, \pi_x)$, isto é, $\varphi_B(x) = 0$. Logo, $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona.

Agora, mostraremos que $\varphi_A(x) = 1 - \varphi_{A^c}(x)$, para todo $A \in \sigma(\Theta)$ e todo $x \in \mathcal{X}$. Seja $x \in \mathcal{X}$ e suponhamos que $\varphi_A(x) = 0$. Se $\rho_A(0, \pi_x) < \rho_A(1, \pi_x)$, então, pela condição (ii), $\rho_{A^c}(1, \pi_x) > \rho_{A^c}(0, \pi_x)$ e, portanto, $\varphi_{A^c}(x) = 1$. Do mesmo modo, prova-se que se $\varphi_A(x) = 1$ (com $\rho_A(0, \pi_x) > \rho_A(1, \pi_x)$), então $\varphi_{A^c}(x) = 0$. Como foi comentado anteriormente, no caso de igualdade, isto é, $\rho_A(0, \pi_x) = \rho_A(1, \pi_x)$, toma-se d como decisão para A e $1 - d$ como decisão para A^c , de modo que $\varphi_A(x) = 1 - \varphi_{A^c}(x)$. $\square$

Figura 2.1: Distribuições a priori para as quais $\varphi_A(0) = 0$

Figura 2.2: Distribuições a priori para as quais $\varphi_A(1) = 0$

A condição (i) do Teorema 2.1 diz que a diferença entre as perdas de aceitação e rejeição ao testar uma hipótese A deve ser maior que a correspondente diferença ao testar uma hipótese B implicada por A .

De fato, sempre que $A \subseteq B$, parece razoável, para $\theta \in B^c$, que a perda de aceitar uma hipótese A seja maior que a perda de aceitar uma hipótese B , pois, aparentemente, é um erro de maior gravidade. Analogamente, para $\theta \in A$, a perda ao rejeitar a hipótese A deve ser menor que a perda de rejeitar uma hipótese B . A condição (i) somente exige que a diferença entre as perdas de aceitar e ao rejeitar A seja maior ou igual à diferença entre as perdas de aceitar e rejeitar B , o que parece ser uma restrição bastante intuitiva.

A condição (ii) estabelece que a perda de aceitar uma hipótese A deve ser igual à perda de rejeitar uma hipótese A^c . A princípio, não parece haver motivos para que essas perdas sejam diferentes, a não ser que o decisor avalie que rejeitar $\theta \in A$ não seja em algum sentido, equivalente a aceitar $\theta \in A^c$.

A seguir mostraremos através de um exemplo, que a recíproca do Teorema 2.1 não é válida, isto é, se a classe de testes atende D1, é possível que uma família de funções de perda não atenda às condições (i) e

(ii) do Teorema 2.1.

Exemplo 2.2. Novamente, seja $X|\theta \sim \text{Bernoulli}(\theta)$, onde $\Theta = \{1/4, 1/3, 1/2\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$ com distribuição a priori para θ dada por $\pi(1/4) = \alpha$, $\pi(1/3) = \beta$ e $\pi(1/2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ tais que $\alpha + \beta \leq 1$. Suponha que $\{L_A\}_{A \in \sigma(\Theta)}$ é uma família de funções de perda, onde L_A , para todo $A \in \sigma(\Theta) \setminus \{1/4, 1/3\}$, é dada pela Tabela 2.3 e $L_{\{1/4, 1/3\}}$ é a função de perda 0-1.

L_A	$\theta \in A$	$\theta \notin A$
0	0	2
1	2	0

Tabela 2.3: Função de perda 0-2

A classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ é idêntica à classe do Exemplo 2.1, isto é, a classe de testes atende à desiderata D1 para todo $x \in \mathcal{X}$, porém

$$L_{\{\frac{1}{3}\}}(0, 1/3) - L_{\{\frac{1}{3}\}}(1, 1/3) = -2 < -1 = L_{\{\frac{1}{4}, \frac{1}{3}\}}(0, 1/3) - L_{\{\frac{1}{4}, \frac{1}{3}\}}(1, 1/3),$$

ou seja, a família de funções de perda não atende a condição (i) do Teorema 2.1.

Portanto, as condições (i) e (ii) do Teorema 2.1 não são necessárias para conferir monotonicidade e invertibilidade aos testes. Este fato nos leva a questionar quais são as condições necessárias sobre as funções de perda para que uma classe de testes de Bayes gerada por elas atenda a cada uma das propriedades lógicas apresentadas no Capítulo 1. Tais caracterizações são enunciadas nos resultados a seguir.

2.1 Classes de testes monótonas

Primeiramente, caracterizaremos a propriedade de monotonicidade, isto é encontraremos as condições sobre as famílias de funções de perda de forma a gerar classes de testes de Bayes monótonas.

Daqui em diante, para não sobrecarregar a notação, denotaremos, para cada $\theta \in \Theta$, a diferença entre as perdas ao aceitar e rejeitar a hipótese nula $H_0 : \theta \in A$, $L_A(0, \theta) - L_A(1, \theta)$, por $\Delta_A(\theta)$.

Teorema 2.2. Suponha que, para todo $\theta_1, \theta_2 \in \Theta$, existe $x \in \mathcal{X}$ tal que $V_x(\theta_1)$ e $V_x(\theta_2)$ são positivos.

Uma classe de testes de Bayes gerada por uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ é monótona para toda priori sobre Θ se, e somente se, para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, $\{L_A\}_{A \in \sigma(\Theta)}$ satisfaz,

$$\Delta_A(\theta_1)\Delta_B(\theta_2) \geq \Delta_B(\theta_1)\Delta_A(\theta_2), \quad (2.1)$$

para todo $\theta_1 \in A$ e todo $\theta_2 \in B^c$.

Demonstração. Primeiro, vamos mostrar que uma classe de testes de Bayes monótona é gerada por uma família de funções de perda com a restrição acima. Para isso, faremos a prova pela contra-positiva. Suponha que existem $A, B \in \sigma(\Theta)$ com $A \subseteq B$ e existem $\theta_1 \in A$ e $\theta_2 \in B^c$ tais que

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} > \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}. \quad (2.2)$$

Pelas condições em (1.2), isso é equivalente a

$$\Delta_A(\theta_1)\Delta_B(\theta_2) < \Delta_A(\theta_2)\Delta_B(\theta_1).$$

Se $\Delta_B(\theta)$ é zero, para qualquer $\theta \in \Theta$, podemos começar direto da segunda inequação. Multiplicando por -1 e somando $\Delta_A(\theta_2)\Delta_B(\theta_2)$ a ambos os lados da inequação acima temos,

$$\Delta_A(\theta_2)\Delta_B(\theta_2) - \Delta_A(\theta_1)\Delta_B(\theta_2) > \Delta_A(\theta_2)\Delta_B(\theta_2) - \Delta_A(\theta_2)\Delta_B(\theta_1).$$

Logo,

$$\Delta_B(\theta_2)[\Delta_A(\theta_2) - \Delta_A(\theta_1)] > \Delta_A(\theta_2)[\Delta_B(\theta_2) - \Delta_B(\theta_1)].$$

Desta forma,

$$1 \geq \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) - \Delta_B(\theta_1)} > \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) - \Delta_A(\theta_1)} \geq 0 \text{ e existe } \alpha_0 \in (0, 1) \text{ tal que}$$

$$\frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) - \Delta_B(\theta_1)} > \alpha_0 > \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) - \Delta_A(\theta_1)}. \quad (2.3)$$

Considere que $x \in \mathcal{X}$, tal que $V_x(\theta_1) > 0$ e $V_x(\theta_2) > 0$, é observado e que, a priori de θ é dada por

$$\pi(\theta_1) = \frac{\alpha_0 V_x(\theta_2)}{\alpha_0 V_x(\theta_2) + (1 - \alpha_0) V_x(\theta_1)} \text{ e } \pi(\theta_2) = 1 - \pi(\theta_1).$$

Deste modo, a posteriori de θ , dado x , é dada por $\pi(\theta_1|x) = \alpha_0^1$ e $\pi(\theta_2|x) = 1 - \alpha_0$. Segue que $\varphi_A(x) = 0$, se e somente se,

$$\rho_A(0, \pi_x) - \rho_A(1, \pi_x) < 0 \Leftrightarrow \alpha_0 > \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) - \Delta_A(\theta_1)}$$

e $\varphi_B(x) = 0$, se e somente se,

$$\rho_B(0, \pi_x) - \rho_B(1, \pi_x) < 0 \Leftrightarrow \alpha_0 > \frac{\Delta_B(\theta_2)}{\Delta_B(\theta_2) - \Delta_B(\theta_1)}.$$

Pela inequação (2.3), temos que $\varphi_A(x) = 0$, mas $\varphi_B(x) = 1$ e, portanto, existe uma distribuição a priori para a qual a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ não é monótona.

Provaremos agora que uma classe de testes gerada pela função de perda com a restrição do enunciado é monótona. Suponha que, para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, a família $\{L_A\}_{A \in \sigma(\Theta)}$ satisfaz

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} \leq \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)},$$

ou $\Delta_A(\theta_1)\Delta_B(\theta_2) \geq \Delta_A(\theta_2)\Delta_B(\theta_1)$, para todo $\theta_1 \in A$ e todo $\theta_2 \in B^c$.

Integrando em θ_1 no subconjunto A contra qualquer medida de probabilidade π , temos

$$\int_A \Delta_A(\theta_1)\Delta_B(\theta_2)d\pi(\theta_1) \geq \int_A \Delta_A(\theta_2)\Delta_B(\theta_1)d\pi(\theta_1), \forall \theta_2 \in B^c.$$

Assim,

$$\Delta_B(\theta_2) \int_A \Delta_A(\theta_1)d\pi(\theta_1) \geq \Delta_A(\theta_2) \int_A \Delta_B(\theta_1)d\pi(\theta_1), \forall \theta_2 \in B^c.$$

¹Aqui, tomamos a priori de sorte que, observado $X = x$, a posteriori de θ_1 é exatamente igual a α_0 . Fazemos isso para que a distribuição a posteriori coincida com os valores α e $1 - \alpha$.

Do mesmo modo, integrando em θ_2 no subconjunto B^c , contra a mesma π , temos

$$\int_{B^c} \Delta_B(\theta_2) d\pi(\theta_2) \int_A \Delta_A(\theta_1) d\pi(\theta_1) \geq \int_{B^c} \Delta_A(\theta_2) d\pi(\theta_2) \int_A \Delta_B(\theta_1) d\pi(\theta_1), \quad (2.4)$$

Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$, com $\{L_A\}_{A \in \sigma(\Theta)}$ satisfazendo (2.1). Assim, para todo $x \in \mathcal{X}$, $\varphi_A(x) = 0$ se, e somente se,

$$\int_{\Theta} \Delta_A(\theta) d\pi_x(\theta) = \int_A \Delta_A(\theta) d\pi_x(\theta) + \int_{A^c \cap B} \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) < 0,$$

onde $\pi_x(\theta)$ denota a distribuição a posteriori para θ dado x . Multiplicando a desigualdade acima pela integral $\int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \geq 0$, temos que $\varphi_A(x) = 0$ implica

$$\int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_A \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) < 0.$$

Usando a desigualdade (2.4) na primeira parcela acima, segue que

$$\int_A \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) < 0.$$

Trocando $\int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{A^c \cap B} \Delta_A(\theta) d\pi_x(\theta) \geq 0$ por um produto negativo, temos

$$\int_A \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) + \int_{A^c \cap B} \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \int_{B^c} \Delta_A(\theta) d\pi_x(\theta) < 0.$$

Desta forma,

$$\int_{B^c} \Delta_A(\theta) d\pi_x(\theta) \left\{ \int_A \Delta_B(\theta) d\pi_x(\theta) + \int_{A^c \cap B} \Delta_B(\theta) d\pi_x(\theta) + \int_{B^c} \Delta_B(\theta) d\pi_x(\theta) \right\} < 0,$$

de onde

$$\int_{\Theta} \Delta_B(\theta) d\pi_x(\theta) < 0$$

e, portanto, $\varphi_B(x) = 0$, isto é, a classe de testes é monótona. $\square$

Vale ressaltar que se a família de funções de perda do enunciado do teorema acima depende também da variável observável X , isto é, se para cada $A \in \sigma(\Theta)$, $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$, mas ainda assim satisfaz a inequação (2.1) para todo $x \in \mathcal{X}$, então uma classe de testes de Bayes gerada por esta (nova) família continuará sendo monótona para qualquer distribuição a priori para θ . No entanto, essa dependência de X faz com que a recíproca desse teorema não valha, em geral, para todo $x \in \mathcal{X}$.

A desigualdade (2.1) do Teorema 2.2 corresponde a uma generalização da condição (i) do Teorema 2.1: com efeito, se $\theta \in A$, segue, de (i), que $\Delta_A(\theta_1) \geq \Delta_B(\theta_1)$, isto é, $\Delta_A(\theta_1)/\Delta_B(\theta_1) \leq 1$ e, do mesmo modo, $\Delta_A(\theta_2) \geq \Delta_B(\theta_2)$, isto é, $\Delta_A(\theta_2)/\Delta_B(\theta_2) \geq 1$ para $\theta_2 \in B^c$. Assim, é razoável que valha a desigualdade (2.1). A condição do Teorema 2.2 é menos restritiva que a condição (i) do Teorema 2.1 e portanto existem muitas famílias de funções de perda que atendem a essa propriedade. Considere o exemplo desse fato a seguir.

Exemplo 2.3. Sejam $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Considere a família de funções de perda $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ dada pelas tabelas abaixo.

Além disso, considere $L_{\Theta}(0, \theta) = L_{\emptyset}(1, \theta) = 0$ e $L_{\Theta}(1, \theta) = L_{\emptyset}(0, \theta) = 1$ para todo $\theta \in \Theta$. Seja $A = \{\theta_1\}$ e $B = \{\theta_1, \theta_2\}$, então

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-2}{-1} \leq \frac{3}{1} = \frac{\Delta_A(\theta_3)}{\Delta_B(\theta_3)}.$$

Observando $x \in \mathcal{X}$, $\varphi_A(x) = 0$ se, e somente se, $\pi(\theta_1|x) \geq 3/5$ e $\varphi_B(x) = 0$ se, e somente se, $\pi(\theta_3|x) \leq$

$L_{\{\theta_1\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_2\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_3\}}$	θ_1	θ_2	θ_3
0	0	3	3	0	1	0	1	0	1	1	0
1	2	0	0	1	0	1	0	1	0	0	1

Tabela 2.4: Exemplo de famílias de funções de perda atendendo (2.1)

$L_{\{\theta_1, \theta_2\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_2, \theta_3\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_1, \theta_3\}}$	θ_1	θ_2	θ_3
0	0	0	1	0	1	0	0	0	0	1	0
1	1	1	0	1	0	1	1	1	1	0	1

Tabela 2.5: Exemplo de famílias de funções de perda atendendo (2.1)

1/2. Assim, se aceitamos que $\theta \in A$, então aceitamos que $\theta \in B$. Na verdade, é possível mostrar que para quaisquer $A, B \in \mathcal{P}(\Theta)$ com $A \subseteq B$, a família de funções de perda satisfaz a condição (2.1) e, portanto, pelo Teorema 2.2, a classe de testes é monótona. No entanto, essa família de funções de perda não atende a condição (i) do Teorema 2.1, pois $\Delta_{\{\theta_1\}}(\theta_1) = -2 < -1 = \Delta_{\{\theta_1, \theta_2\}}(\theta_1)$.

Exemplo 2.4. Seja $X|\theta \sim \text{Bernoulli}(\theta)$, $\Theta = \{1/5, 2/5, 3/5, 4/5\}$, $\sigma(\Theta) = \mathcal{P}(\Theta)$ e seja π , a priori de θ , a distribuição uniforme em Θ . Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$. Suponhamos que $A = \{1/5\}$, $B = \{1/5, 2/5\}$ e que L_A e L_B são dados por

L_A	1/5	2/5	3/5	4/5
0	0	1	1	1
1	2	0	0	0

Tabela 2.6: Exemplo de classe de funções de perda que não atende 2.1

Como

$$\frac{\Delta_A(1/5)}{\Delta_B(1/5)} = \frac{-2}{-1} > \frac{1}{3} = \frac{\Delta_A(4/5)}{\Delta_B(4/5)},$$

a família $\{L_A\}_{A \in \sigma(\Theta)}$ não atende a condição (2.1).

Observando $x \in \mathcal{X}$, $\varphi_A(x) = 0$ se, e somente se, $\pi(\theta \in A|x) > 1/3$ e $\varphi_B(x) = 0$ se, e somente se, $\pi(\theta \in B|x) > 3/4$. Observando $x = 0$, $\pi(\theta \in A|0) = 4/10$ e $\pi(\theta \in B|0) = 7/10$, resultando que $\varphi_A(0) = 0$ e $\varphi_B(0) = 1$, isto é, a classe de testes não é monótona.

2.2 Classes de testes invertíveis

A seguir, caracterizaremos a propriedade de invertibilidade, isto é encontraremos as condições sobre as famílias de funções de perda de forma a gerar classes de testes de Bayes invertíveis.

Teorema 2.3. Suponha que para todo $\theta_1, \theta_2 \in \Theta$, existe $x \in \mathcal{X}$ tal que $V_x(\theta_1)$ e $V_x(\theta_2)$ são positivos.

Uma classe de testes de Bayes gerada por uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ é invertível para toda priori sobre Θ se, e somente se, para todo $A \in \sigma(\Theta)$, $\{L_A\}_{A \in \sigma(\Theta)}$ satisfaz,

$$\Delta_A(\theta_1)\Delta_{A^c}(\theta_0) = \Delta_A(\theta_0)\Delta_{A^c}(\theta_1), \quad (2.5)$$

para todo $\theta_0 \in A$ e todo $\theta_1 \in A^c$.

L_B	1/5	2/5	3/5	4/5
0	0	0	3	3
1	1	1	0	0

Tabela 2.7: Exemplo de classe de funções de perda que não atende 2.1

Demonstração. Primeiro vamos provar que se a classe é invertível contra toda priori, então a função de perda que gera essa classe atende às condições do enunciado. Faremos a prova pela contra-positiva. Suponha que existe $A \in \sigma(\Theta)$ e existem $\theta_0 \in A$ e $\theta_1 \in A^c$ tais que

$$\frac{\Delta_A(\theta_1)}{-\Delta_{A^c}(\theta_1)} \neq \frac{-\Delta_A(\theta_0)}{\Delta_{A^c}(\theta_0)}.$$

Segue que

$$\frac{\Delta_A(\theta_1)}{\Delta_A(\theta_0)} \neq \frac{\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0)}.$$

Invertendo os dois lados, temos

$$\frac{\Delta_A(\theta_0)}{\Delta_A(\theta_1)} \neq \frac{\Delta_{A^c}(\theta_0)}{\Delta_{A^c}(\theta_1)}.$$

Multiplicando novamente por -1 e somando 1 dos dois lados, segue que

$$1 - \frac{\Delta_A(\theta_0)}{\Delta_A(\theta_1)} \neq 1 - \frac{\Delta_{A^c}(\theta_0)}{\Delta_{A^c}(\theta_1)}.$$

Logo,

$$\frac{\Delta_A(\theta_1) - \Delta_A(\theta_0)}{\Delta_A(\theta_1)} \neq \frac{\Delta_{A^c}(\theta_1) - \Delta_{A^c}(\theta_0)}{\Delta_{A^c}(\theta_1)}.$$

Novamente invertendo ambos os lados, segue que

$$\frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)} \neq \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)}.$$

Portanto, ou

$$0 \leq \frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)} < \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)} \leq 1$$

ou

$$1 \geq \frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)} > \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)} \geq 0,$$

de forma que existe $\alpha_0 \in (0, 1)$ tal que

$$\frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)} < \alpha_0 < \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)} \quad (2.6)$$

ou

$$\frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)} > \alpha_0 > \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)}. \quad (2.7)$$

Suponha que (2.6) é válido (o argumento abaixo é similar quando (2.7) é válido) e considere que $x \in \mathcal{X}$ tal que $V_x(\theta_1) > 0$ e $V_x(\theta_2) > 0$ é observado e que a priori de θ é dada por

$$\pi(\theta_0) = \frac{\alpha_0 V_x(\theta_1)}{\alpha_0 V_x(\theta_1) + (1 - \alpha_0) V_x(\theta_0)} \text{ e } \pi(\theta_1) = 1 - \pi(\theta_0).$$

Deste modo, a posteriori de θ , dado x , é dada por $\pi(\theta_0|x) = \alpha_0$ e $\pi(\theta_1|x) = 1 - \alpha_0$. Então, $\varphi_A(x) = 0$ se, e somente se,

$$\alpha_0 > \frac{\Delta_A(\theta_1)}{\Delta_A(\theta_1) - \Delta_A(\theta_0)}$$

($\varphi_A(x) = 1$ se, e somente se, $\alpha_0 < \Delta_A(\theta_1)/[\Delta_A(\theta_1) - \Delta_A(\theta_0)]$).

Analogamente, $\varphi_{A^c}(x) = 0$, se e somente se,

$$\alpha_0 < \frac{-\Delta_{A^c}(\theta_1)}{\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)}$$

($\varphi_{A^c}(x) = 1$ se, e somente se, $\alpha_0 > -\Delta_{A^c}(\theta_1)/[\Delta_{A^c}(\theta_0) - \Delta_{A^c}(\theta_1)]$). Mas, pela equação (2.6), existe α_0 tal que $\varphi_A(x) = 0$, mas $\varphi_{A^c}(x) = 0$ ($\varphi_A(x) = 1$ e $\varphi_{A^c}(x) = 1$ se (2.7) vale). De todo modo, $\varphi_A(x) \neq 1 - \varphi_{A^c}(x)$ e, portanto, a classe gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ não é invertível.

Provaremos agora que uma classe de testes gerada por uma família de funções de perda satisfazendo (2.5) é invertível. Suponha que para todo $A \in \sigma(\Theta)$,

$$\frac{\Delta_A(\theta_1)}{-\Delta_{A^c}(\theta_1)} = \frac{-\Delta_A(\theta_0)}{\Delta_{A^c}(\theta_0)},$$

para todo $\theta_0 \in A$ e todo $\theta_1 \in A^c$. Então $\Delta_A(\theta_1)\Delta_{A^c}(\theta_0) = \Delta_A(\theta_0)\Delta_{A^c}(\theta_1)$, para todo $\theta_0 \in A$ e todo $\theta_1 \in A^c$.

Integrando em θ_1 no subconjunto A^c contra qualquer π em $\sigma(\Theta)$, temos

$$\int_{A^c} \Delta_A(\theta_1)\Delta_{A^c}(\theta_0)d\pi(\theta_1) = \int_{A^c} \Delta_A(\theta_0)\Delta_{A^c}(\theta_1)d\pi(\theta_1), \text{ para todo } \theta_0 \in A.$$

Do mesmo modo, integrando em θ_0 no subconjunto A ,

$$\int_A \Delta_{A^c}(\theta_0)d\pi(\theta_0) \int_{A^c} \Delta_A(\theta_1)d\pi(\theta_1) = \int_A \Delta_A(\theta_0)d\pi(\theta_0) \int_{A^c} \Delta_{A^c}(\theta_1)d\pi(\theta_1). \quad (2.8)$$

Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ que satisfaz a condição acima. Então, $\varphi_A(x) = 0$ se, e somente se,

$$\int_{\Theta} \Delta_A(\theta)d\pi_x(\theta) = \int_A \Delta_A(\theta)d\pi_x(\theta) + \int_{A^c} \Delta_A(\theta)d\pi_x(\theta) < 0$$

Multiplicando ambos os lados por $\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \geq 0$, $\varphi_A(x) = 0$ é equivalente a

$$\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \int_A \Delta_A(\theta)d\pi_x(\theta) + \int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \int_{A^c} \Delta_A(\theta)d\pi_x(\theta) < 0.$$

Da igualdade em (2.8), segue que

$$\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) \int_A \Delta_A(\theta)d\pi_x(\theta) + \int_A \Delta_A(\theta)d\pi_x(\theta) \int_{A^c} \Delta_{A^c}(\theta)d\pi_x(\theta) < 0.$$

Desta forma, $\varphi_A(x) = 0$ se, e somente se,

$$\int_A \Delta_A(\theta)d\pi_x(\theta) \left\{ \int_A \Delta_{A^c}(\theta)d\pi_x(\theta) + \int_{A^c} \Delta_{A^c}(\theta)d\pi_x(\theta) \right\} < 0.$$

Então $\varphi_A(x) = 0$ se, e somente se,

$$\int_A \Delta_{A^c}(\theta)d\pi_x(\theta) + \int_{A^c} \Delta_{A^c}(\theta)d\pi_x(\theta) > 0,$$

pois $\int_A \Delta_A(\theta) d\pi_x(\theta) \leq 0$. Assim,

$$\varphi_A(x) = 0 \Leftrightarrow \int_{\Theta} \Delta_{A^c}(\theta) d\pi_x(\theta) > 0 \Leftrightarrow \varphi_{A^c}(x) = 1$$

e, portanto, a classe de testes é invertível. $\square$

Vale ressaltar que se a família de funções de perda do enunciado do teorema acima depende também da variável observável X , isto é, se para cada $A \in \sigma(\Theta)$, $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$, mas ainda assim satisfaz a condição (2.5) para todo $x \in \mathcal{X}$, então a classe de testes de Bayes gerada por esta (nova) família continuará sendo invertível para qualquer distribuição a priori para θ . No entanto, essa dependência de X faz com que a recíproca desse Teorema não valha em geral, para todo $x \in \mathcal{X}$.

A condição (2.5) estabelece que a razão entre a diferença entre penalidades de erro e acerto quando testamos $H_0 : \theta \in A$ e a diferença entre perdas de erro e acerto quando testamos $H_0' : \theta \in A^c$ deve ser constante (para todo $\theta \in \Theta$) para conferir invertibilidade aos testes. Várias famílias de funções de perda atendem a essa condição como o exemplo a seguir.

Exemplo 2.5. Sejam $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Considere a família de funções de perda $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ dada pelas tabelas abaixo.

$L_{\{\theta_1\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_2\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_3\}}$	θ_1	θ_2	θ_3
0	0	1	1	0	4	0	6	0	1	1	0
1	1	0	0	1	1	3	0	1	0	0	2

Tabela 2.8: Exemplo de famílias de funções de perda atendendo (2.1)

$L_{\{\theta_1, \theta_2\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_2, \theta_3\}}$	θ_1	θ_2	θ_3	$L_{\{\theta_1, \theta_3\}}$	θ_1	θ_2	θ_3
0	0	0	2	0	1	0	0	0	0	1	0
1	1	1	0	1	0	1	1	1	1	0	2

Tabela 2.9: Exemplo de famílias de funções de perda atendendo (2.1)

Além disso, considere $L_{\emptyset}(0, \theta) = L_{\emptyset}(1, \theta) = 0$ e $L_{\Theta}(1, \theta) = L_{\Theta}(0, \theta) = 1$ para todo $\theta \in \Theta$. Seja $A = \{\theta_1\}$, então

$$\frac{\Delta_A(\theta_1)}{-\Delta_{A^c}(\theta_1)} = \frac{-1}{-1} = \frac{-1}{-1} = \frac{-\Delta_A(\theta_3)}{\Delta_{A^c}(\theta_3)}.$$

Observando $x \in \mathcal{X}$, $\varphi_A(x) = 0$ se, e somente se, $\pi(\theta_1|x) \geq 1/2$ e $\varphi_{A^c}(x) = 0$ se, e somente se, $\pi(\theta_1|x) \leq 1/2$. Assim, se aceitamos que $\theta \in A$, então rejeitamos que $\theta \in A^c$. Na verdade, é possível mostrar que para qualquer $A \in \mathcal{P}(\Theta)$, a família de funções de perda satisfaz a condição (2.5) e, portanto, pelo Teorema 2.3, a classe de testes é invertível. No entanto, essa família de funções de perda não atende a condição (ii) do Teorema 2.1, pois $L_{\{\theta_2\}}(0, \theta_1) = 4 \neq 1 = L_{\{\theta_1, \theta_3\}}(1, \theta_1)$.

Exemplo 2.6. Seja $X|\theta \sim \text{Bernoulli}(\theta)$, $\Theta = \{1/5, 2/5, 3/5, 4/5\}$, $\sigma(\Theta) = \mathcal{P}(\Theta)$ e seja π , a priori de θ , a distribuição uniforme em Θ . Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$. Suponhamos que $A = \{1/5\}$ e que L_A e L_{A^c} são dados pelas Tabelas 2.10 e 2.11.

Como

$$\frac{\Delta_A(3/5)}{-\Delta_{A^c}(3/5)} = \frac{1}{3} \neq \frac{2}{1} = \frac{-\Delta_A(1/5)}{\Delta_{A^c}(1/5)},$$

L_A	1/5	2/5	3/5	4/5
0	0	1	1	1
1	2	0	0	0

Tabela 2.10: Exemplo de classe de funções de perda que não atende (2.5)

L_{A^c}	1/5	2/5	3/5	4/5
0	1	0	0	0
1	0	3	3	3

Tabela 2.11: Exemplo de classe de funções de perda que não atende (2.5)

a família $\{L_A\}_{A \in \sigma(\Theta)}$ não atende a condição (2.5).

Observando $x \in \chi$, $\varphi_A(x) = 0$ se, e somente se, $\pi(\theta \in A|x) > 1/3$ e $\varphi_{A^c}(x) = 0$ se, e somente se, $\pi(\theta \in A^c|x) < 3/4$. Observando $x = 0$, $\pi(A|0) = 4/10$ e $\pi(A^c|0) = 6/10$, resultando que $\varphi_A(0) = 0$ e $\varphi_{A^c}(0) = 0$, isto é, a classe de testes não é invertível.

2.3 Classes de testes consoantes com a união

Por fim, caracterizaremos a propriedade de consonância com a união, isto é encontraremos as condições sobre as famílias de funções de perda de forma a gerar classes de testes de Bayes consoantes com a união.

Teorema 2.4. Suponha que, para todo $\theta_1, \theta_2 \in \Theta$ existe $x \in \chi$ tal que $V_x(\theta_1)$ e $V_x(\theta_2)$ são positivos.

(a) Se uma classe de testes de Bayes gerada por uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ é consoante com a união finita para toda priori sobre Θ , então para todo $A, B \in \sigma(\Theta)$ disjuntos, $\{L_A\}_{A \in \sigma(\Theta)}$ satisfaz,

$$0 \geq \Delta_A(\theta_2)\Delta_{A \cup B}(\theta_1) \geq \Delta_{A \cup B}(\theta_2)\Delta_A(\theta_1) \text{ ou } 0 \geq \Delta_B(\theta_2)\Delta_{A \cup B}(\theta_1) \geq \Delta_B(\theta_1)\Delta_{A \cup B}(\theta_2) \quad (2.9)$$

para todo $\theta_1 \in A \cup B$ e todo $\theta_2 \notin A \cup B$.

(b) Se para todo $A, B \in \sigma(\Theta)$ disjuntos, uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ satisfaz,

$$[\Delta_A(\theta_1) + \Delta_B(\theta_1)]\Delta_{A \cup B}(\theta_2) \leq [\Delta_A(\theta_2) + \Delta_B(\theta_2)]\Delta_{A \cup B}(\theta_1) \quad (2.10)$$

para todo $\theta_1 \in A \cup B$ e todo $\theta_2 \notin A \cup B$, então uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ é consoante com a união finita para toda priori sobre Θ ,

Demonstração. Primeiro vamos mostrar a parte (a), isto é, que uma classe de testes consonante é gerada por uma família de funções de perda com a restrição (2.9). Para isso faremos a prova pela contra-positiva. Suponha que existem $A, B \in \sigma(\Theta)$ disjuntos e existem $\theta_1 \in A \cup B$ e $\theta_2 \in (A \cup B)^c$ tais que

$$\frac{\Delta_A(\theta_1)}{\Delta_{A \cup B}(\theta_1)} < \frac{\Delta_A(\theta_2)}{\Delta_{A \cup B}(\theta_2)} \text{ ou } \frac{\Delta_B(\theta_1)}{\Delta_{A \cup B}(\theta_1)} < \frac{\Delta_B(\theta_2)}{\Delta_{A \cup B}(\theta_2)}, \quad (2.11)$$

Pelas condições (1.2), a primeira condição de (2.11) é equivalente a

$$\frac{\Delta_A(\theta_1)}{\Delta_A(\theta_2)} > \frac{\Delta_{A \cup B}(\theta_1)}{\Delta_{A \cup B}(\theta_2)}$$

Desta forma, após algumas manipulações,

$$1 \geq \frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) - \Delta_A(\theta_1)} > \frac{\Delta_{A \cup B}(\theta_2)}{\Delta_{A \cup B}(\theta_2) - \Delta_{A \cup B}(\theta_1)} \geq 0 \text{ e existe } \alpha_0 \in (0, 1) \text{ tal que}$$

$$\frac{\Delta_A(\theta_2)}{\Delta_A(\theta_2) - \Delta_A(\theta_1)} > \alpha_0 > \frac{\Delta_{A \cup B}(\theta_2)}{\Delta_{A \cup B}(\theta_2) - \Delta_{A \cup B}(\theta_1)}. \quad (2.12)$$

Considere $x \in \mathcal{X}$, tal que $V_x(\theta_1) > 0$ e $V_x(\theta_2) > 0$, é observado e que, a priori de θ é dada por

$$\pi(\theta_1) = \frac{\alpha_0 V_x(\theta_2)}{\alpha_0 V_x(\theta_2) + (1 - \alpha_0) V_x(\theta_1)} \text{ e } \pi(\theta_2) = 1 - \pi(\theta_1).$$

Deste modo, a posteriori de θ , dado x , é dada por $\pi(\theta_1|x) = \alpha_0$ e $\pi(\theta_2|x) = 1 - \alpha_0$. Segue que $\varphi_A(x) = 1$, se e somente se,

$$\rho_A(0, \pi_x) - \rho_A(1, \pi_x) > 0 \Leftrightarrow \alpha_0 < \Delta_A(\theta_2)/[\Delta_A(\theta_2) - \Delta_A(\theta_1)].$$

Pela inequação (2.12), temos que $\varphi_A(x) = 1$, mas $\varphi_{A \cup B}(x) = 0$. Analogamente, partindo da segunda inequação de (2.11) teremos que $\varphi_B(x) = 1$, mas $\varphi_{A \cup B}(x) = 0$. Ou seja, se existem $\theta_1 \in A \cup B$ e $\theta_2 \in (A \cup B)^c$ tais que as inequações da condição (2.11) valem ao mesmo tempo, então existe uma distribuição a priori para a qual a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ não é consonante com a união.

Provaremos agora a parte (b), isto é, que uma classe de testes gerada pela função de perda com a restrição (2.10) do enunciado é consonante.

Suponha que, para todo $A, B \in \sigma(\Theta)$ disjuntos, $\{L_A\}_{A \in \sigma(\Theta)}$ é uma família de funções de perda que satisfaz (2.10) ou seja, $-\Delta_{A \cup B}(\theta_2)[\Delta_A(\theta_1) + \Delta_B(\theta_1)] \geq -\Delta_{A \cup B}(\theta_1)[\Delta_A(\theta_2) + \Delta_B(\theta_2)]$, para todo $\theta_1 \in A \cup B$ e todo $\theta_2 \notin A \cup B$.

Integrando em θ_1 no subconjunto $A \cup B$ contra qualquer medida de probabilidade π , temos

$$\int_{A \cup B} -\Delta_{A \cup B}(\theta_2)[\Delta_A(\theta_1) + \Delta_B(\theta_1)] d\pi(\theta_1) \geq \int_{A \cup B} -\Delta_{A \cup B}(\theta_1)[\Delta_A(\theta_2) + \Delta_B(\theta_2)] d\pi(\theta_1),$$

para todo $\theta_2 \notin A \cup B$. Assim,

$$-\Delta_{A \cup B}(\theta_2) \int_{A \cup B} [\Delta_A(\theta_1) + \Delta_B(\theta_1)] d\pi(\theta_1) \geq -[\Delta_A(\theta_2) + \Delta_B(\theta_2)] \int_{A \cup B} \Delta_{A \cup B}(\theta_1) d\pi(\theta_1).$$

Do mesmo modo, integrando em θ_2 no subconjunto $(A \cup B)^c$, contra a mesma π , temos

$$\int_{(A \cup B)^c} -\Delta_{A \cup B}(\theta_2) d\pi(\theta_2) \int_{A \cup B} \Delta_A(\theta_1) + \Delta_B(\theta_1) d\pi(\theta_1) \geq \int_{(A \cup B)^c} -[\Delta_A(\theta_2) + \Delta_B(\theta_2)] d\pi(\theta_2) \int_{A \cup B} \Delta_{A \cup B}(\theta_1) d\pi(\theta_1)$$

se, e somente se,

$$\int_{(A \cup B)^c} \Delta_{A \cup B}(\theta_2) d\pi(\theta_2) \int_{A \cup B} \Delta_A(\theta_1) + \Delta_B(\theta_1) d\pi(\theta_1) \leq \int_{(A \cup B)^c} [\Delta_A(\theta_2) + \Delta_B(\theta_2)] d\pi(\theta_2) \int_{A \cup B} \Delta_{A \cup B}(\theta_1) d\pi(\theta_1). \quad (2.13)$$

Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$, com $\{L_A\}_{A \in \sigma(\Theta)}$ satisfazendo (2.10). Assim, para todo $x \in \mathcal{X}$, $\varphi_A(x) = 1$ e $\varphi_B(x) = 1$ se, e somente se,

$$\int_{\Theta} \Delta_A(\theta) d\pi_x(\theta) > 0 \text{ e } \int_{\Theta} \Delta_B(\theta) d\pi_x(\theta) > 0.$$

onde $\pi_x(\cdot)$ denota a distribuição a posteriori para θ , dado x . Logo,

$$\int_{\Theta} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) > 0 \Rightarrow \int_{A \cup B} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) + \int_{(A \cup B)^c} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) > 0$$

Multiplicando o resultado acima por $\int_{(A \cup B)^c} \Delta_{A \cup B}(\theta) d\pi_x(\theta) \geq 0$, temos

$$\int_{(A \cup B)^c} \Delta_{A \cup B}(\theta) d\pi_x(\theta) \int_{A \cup B} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) + \int_{(A \cup B)^c} \Delta_{A \cup B}(\theta) d\pi_x(\theta) \int_{(A \cup B)^c} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) > 0$$

Usando a desigualdade (2.13) na primeira parcela acima, segue que

$$\int_{(A \cup B)^c} \Delta_A(\theta_2) + \Delta_B(\theta_2) d\pi(\theta_2) \int_{A \cup B} \Delta_{A \cup B}(\theta_1) d\pi(\theta_1) + \int_{(A \cup B)^c} \Delta_{A \cup B}(\theta) d\pi_x(\theta) \int_{(A \cup B)^c} \Delta_A(\theta) + \Delta_B(\theta) d\pi_x(\theta) > 0$$

Desta forma,

$$\int_{(A \cup B)^c} \Delta_A(\theta_2) + \Delta_B(\theta_2) d\pi(\theta_2) \left\{ \int_{A \cup B} \Delta_{A \cup B}(\theta_1) d\pi(\theta_1) + \int_{(A \cup B)^c} \Delta_{A \cup B}(\theta) d\pi_x(\theta) \right\} > 0$$

de onde

$$\int_{\Theta} \Delta_{A \cup B}(\theta) d\pi_x(\theta) > 0$$

e, portanto, $\varphi_{A \cup B}(x) = 1$, isto é, a classe de testes é consoante. □

Vale ressaltar que se a família de funções de perda do enunciado da parte (b) do teorema acima depende também da variável observável X , isto é, se para cada $A \in \sigma(\Theta)$, $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$, mas ainda assim satisfaz a condição (2.10) para todo $x \in \mathcal{X}$, então a classe de testes de Bayes gerada por esta (nova) família continuará sendo consoante com a união finita para qualquer distribuição a priori para θ .

A desigualdade (2.9) do Teorema 2.4 (a) corresponde a uma generalização de uma “negação” da condição (i) do Teorema 2.1, para A, B disjuntos: com efeito, se $\theta_1 \in A$, segue, de uma negação de (i), que $\Delta_A(\theta_1) \leq \Delta_{A \cup B}(\theta_1)$, isto é, $\Delta_A(\theta_1)/\Delta_{A \cup B}(\theta_1) \geq 1$ e, do mesmo modo, $\Delta_A(\theta_2) \leq \Delta_{A \cup B}(\theta_2)$, isto é, $\Delta_A(\theta_2)/\Delta_{A \cup B}(\theta_2) \leq 1$ para $\theta_2 \in A \cup B$. Analogamente quando $\theta_1 \in B$. Assim, se um decisor tem uma função de perda como essa negação de (i), é razoável que valha a desigualdade (2.9).

Além disso, a desigualdade (2.10) do Teorema 2.4 (b) corresponde a uma generalização da condição $\Delta_{A \cup B}(\theta) \geq \Delta_A(\theta) + \Delta_B(\theta)$ (condição que estabelece que a diferença entre as perdas de aceitação e rejeição ao testar uma hipótese $\theta \in A \cup B$ deve ser maior ou igual à soma desta diferença para as partes dessa hipótese): se $\theta_1 \in A \cup B$, segue que $\Delta_{A \cup B}(\theta_1) \geq \Delta_A(\theta_1) + \Delta_B(\theta_1)$, isto é, $[\Delta_A(\theta_1) + \Delta_B(\theta_1)]/\Delta_{A \cup B}(\theta_1) \geq 1$ e, do mesmo modo, $\Delta_{A \cup B}(\theta_2) \geq \Delta_A(\theta_2) + \Delta_B(\theta_2)$, isto é, $[\Delta_A(\theta_2) + \Delta_B(\theta_2)]/\Delta_{A \cup B}(\theta_2) \leq 1$, para $\theta_2 \in A \cup B$.

As condições (2.9) e (2.10), por se tratarem, em certo sentido, de negações das condições de monotonicidade, não são intuitivas do ponto de vista lógico. No entanto, existem muitas famílias de funções de perda que atendem à essa propriedade. Considere o exemplo desse fato a seguir.

Exemplo 2.7. *Seja $X|\theta \sim \text{Bernoulli}(\theta)$ com $\Theta = \{1/4, 1/3, 1/2\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Seja $\pi(1/4) = \alpha$, $\pi(1/3) = \beta$ e $\pi(1/2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ tais que $\alpha + \beta \leq 1$, a distribuição a priori para θ . Seja $\lambda(\cdot)$ uma medida finita qualquer em $\mathcal{P}(\Theta)$ com $\lambda(\theta)$ positivo para todo $\theta \in \Theta$. Considere que para $A \in \sigma(\Theta)$ a função de perda L_A é dada pela tabela 2.16.*

Agora, suponha que $\lambda(A)$ é o número de elementos do conjunto mensurável A . Por exemplo, $\lambda(\Theta) = 3$, $\lambda(\emptyset) = 0$ e se $A = \{1/2\}$, então $\lambda(A) = 1$. Note que, neste caso, $\lambda(\cdot)$ não é uma medida de probabilidade.

L_A	$\theta \in A$	$\theta \notin A$
0	0	$\lambda(A)$
1	$\lambda(A^c)$	0

Tabela 2.12: Exemplo de função de perda utilizando uma medida finita.

Para todo $A, B \in \sigma(\Theta)$ disjuntos, temos

$$\frac{\Delta_A(\theta_1) + \Delta_B(\theta_1)}{\Delta_{A \cup B}(\theta_1)} = \frac{-\lambda(A^c) + \lambda(B)}{-\lambda(A^c \cap B^c)} = \frac{-2 + 1}{-1} = 1 = \frac{1 + 1}{2} = \frac{\lambda(A) + \lambda(B)}{\lambda(A \cup B)} = \frac{\Delta_A(\theta_2) + \Delta_B(\theta_2)}{\Delta_{A \cup B}(\theta_2)},$$

para todo $\theta_1 \in A$ e todo $\theta_2 \notin A \cup B$. Analogamente, vale a mesma igualdade, se $\theta_1 \in B$. Portanto, pelo Teorema 2.4 (b), a classe de testes de Bayes gerada por essas funções de perda é consonante com a união para qualquer $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$.

De fato, famílias de funções de perda do tipo da tabela 2.12, conferem consonância contra qualquer distribuição a priori para θ , para qualquer θ (não apenas finito como no Exemplo 2.7). Para isso, basta escrever a condição (2.10) na forma produto como aparece na demonstração do Teorema 2.4 (b) e não como quociente (isso para evitar divisões por zero).

Exemplo 2.8. Seja $X|\theta \sim N(\theta, 1)$ onde $\Theta = \{-1, 0, 1\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Suponha ainda que a distribuição a priori para θ é dada por $\pi(-1) = \alpha$, $\pi(0) = \beta$ e $\pi(1) = 1 - \alpha - \beta$. Desta forma, a distribuição a posteriori de θ , dado x é dada por $\pi(-1|x) \propto \alpha f(x|-1)$, $\pi(0|x) \propto \beta f(x|0)$ e $\pi(1|x) \propto (1 - \alpha - \beta)f(x|1)$, onde $f(x|\theta) = (\sqrt{2\pi})^{-1} e^{-(x-\theta)^2/2}$ é a função densidade de probabilidade da distribuição Normal($\theta, 1$).

Sejam $A = \{-1\}$ e $B = \{0\}$, a família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ é dada pelas tabelas abaixo.

$L_{\{-1\}}$	$\theta = -1$	$\theta = 0$	$\theta = 1$
0	0	2	4
1	1	0	0

Tabela 2.13: Exemplo de função de perda de classe que não atendem (2.9)

$L_{\{0\}}$	$\theta = -1$	$\theta = 0$	$\theta = 1$
0	2	0	4
1	0	1	0

Tabela 2.14: Exemplo de função de perda de classe que não atendem (2.9)

Para $\theta_1 = -1$ e $\theta_2 = 1$, vale que

$$\frac{\Delta_A(\theta_2)}{\Delta_{A \cup B}(\theta_2)} = \frac{4}{4} > \frac{-1}{-2} = \frac{\Delta_A(\theta_1)}{\Delta_{A \cup B}(\theta_1)} \text{ e } \frac{\Delta_B(\theta_2)}{\Delta_{A \cup B}(\theta_2)} = \frac{4}{4} > \frac{2}{-2} = \frac{\Delta_B(\theta_1)}{\Delta_{A \cup B}(\theta_1)}$$

isto é, não atende a condição do Teorema 2.4.

Observado $x \in \mathcal{X}$, $\phi_{\{-1\}}(x) = 0$ se, e somente se,

$$\alpha > \frac{4f(x|1) + \beta[2f(x|0) - 4f(x|1)]}{f(x|-1) + 4f(x|1)},$$

$L_{\{-1,0\}}$	$\theta = -1$	$\theta = 0$	$\theta = 1$
0	0	0	4
1	2	2	0

Tabela 2.15: Exemplo de função de perda de classe que não atendem (2.9)

$\varphi_B(x) = 0$ se, e somente se,

$$\alpha < \frac{\beta[f(x|0) - 4f(x|1)] - 4f(x|1)}{2f(x|-1) - 4f(x|1)}$$

e $\varphi_{A \cup B}(x) = 0$ se, e somente se,

$$\alpha > \frac{4f(x|1) - \beta[4f(x|1) - 2f(x|0)]}{2f(x|-1) + 4f(x|1)}.$$

Suponha que $\alpha = 0,36$, $\beta = 0,60$ e $x = -1$, então $\varphi_A(-1) = 1$, $\varphi_B(-1) = 1$, mas $\varphi_{A \cup B}(-1) = 0$. Portanto, uma classe de testes de Bayes gerada por essa $\{L_A\}_{A \in \sigma(\Theta)}$ não atende a consonância com a união.

2.4 Classes de Testes que (não) atendem às Desideratas

Os Teoremas 2.2, 2.3 e 2.4 mostram condições necessárias e suficientes que uma família de funções de perda deve atender de modo que uma classe de testes de Bayes gerada por esta família seja monótona, invertível e consonante com a união, respectivamente, para toda distribuição a priori do parâmetro. Consequentemente, se existirem famílias de funções de perda que atendam as condições (2.1) e (2.5) ao mesmo tempo, tais perdas gerarão classes de testes de Bayes que atenderão D1 para toda priori e, se existirem funções de perda que atendam simultaneamente as condições (2.1), (2.5) e (2.10), gerarão classes de testes de Bayes que atenderão D2 para toda priori.

A seguir, veremos que é possível que uma família de funções de perda atenda mais de uma dessas condições ao mesmo tempo.

Por exemplo, se, para cada $A \in \sigma(\Theta)$, a função de perda é dada por $L_A(0, \theta) = I(\theta \in A^c)$ e $L_A(1, \theta) = I(\theta \in A)$, isto é, se para cada hipótese nula $\theta \in A$, L_A é a função de perda 0-1, então a classe de testes de Bayes gerada por esta família de funções é monótona e invertível para toda distribuição a priori para θ , ou seja, atende a desiderata D1.

Na sequência, daremos exemplos de classes de testes de Bayes geradas por famílias de funções de perda que atendem as condições (2.1) e (2.5). Essas famílias de funções de perda são bastante intuitivas. No entanto, apesar de gerarem testes com as propriedades de monotonicidade e invertibilidade, as classes de testes resultantes não atendem à propriedade de consonância da união, isto é, atendem a desiderata D1, mas não atendem D2.

Exemplo 2.9. Seja $X|\theta \sim \text{Bernoulli}(\theta)$ com $\Theta = \{1/4, 1/3, 1/2\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Seja $\pi(1/4) = \alpha$, $\pi(1/3) = \beta$ e $\pi(1/2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ tal que $\alpha + \beta \leq 1$, a distribuição a priori para θ . Seja $\lambda(\cdot)$ uma medida finita qualquer em $\mathcal{P}(\Theta)$ com $\lambda(\theta)$ positivo para todo $\theta \in \Theta$. Considere que para $A \in \sigma(\Theta)$ a função de perda L_A é dada pela tabela 2.16.

L_A	$\theta \in A$	$\theta \notin A$
0	0	$\lambda(A^c)$
1	$\lambda(A)$	0

Tabela 2.16: Exemplo de função de perda utilizando uma medida finita.

Para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, temos

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-\lambda(A)}{-\lambda(B)} \leq 1 \leq \frac{\lambda(A^c)}{\lambda(B^c)} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)},$$

para todo $\theta_1 \in A$ e todo $\theta_2 \in B^c$. Portanto, pelo Teorema 2.2, a classe de testes de Bayes gerada por essas funções de perda é monótona para qualquer $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$. Na verdade, isso vale para qualquer Θ , $\sigma(\Theta)$ e $\lambda : \sigma(\Theta) \rightarrow \mathbb{R}^+$ finita com $\lambda(\theta) > 0$.

Além disso, para todo $A \in \sigma(\Theta)$, temos

$$\frac{\Delta_A(\theta_0)}{-\Delta_{A^c}(\theta_0)} = \frac{\lambda(A^c)}{\lambda(A)} = 1 = \frac{\lambda(A)}{\lambda(A^c)} = \frac{-\Delta_A(\theta_1)}{\Delta_{A^c}(\theta_1)},$$

para todo $\theta_1 \in A$ e todo $\theta_0 \in A^c$. Portanto, pelo Teorema 2.3, a classe de testes de Bayes gerada pela família $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ é invertível para qualquer $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$.

Suponha novamente que $\lambda(A)$ é o número de elementos do conjunto mensurável A . Deste modo, observado $x = 0$, $\varphi_{\{1/4\}}(0) = 0$ se, e somente se, $\alpha > (12 + 4\beta)/21$, $\varphi_{\{1/3\}}(0) = 0$ se, e somente se, $\alpha < (10\beta - 6)/3$ e $\varphi_{\{1/2\}}(0) = 0$ se, e somente se, $\alpha < (3 - 11\beta)/12$.

Se $\alpha = 0,4$ e $\beta = 0,4$ por exemplo, teremos $\varphi_{\{1/4\}}(0) = 1$, $\varphi_{\{1/3\}}(0) = 1$ e $\varphi_{\{1/2\}}(0) = 1$, e pelo Teorema 1.3 a classe de testes não respeita D2.

Em resumo, a classe de testes é monótona e invertível, mas não atende D2, ou seja, ela não é consonante com a união.

Exemplo 2.10. Ainda considerando $X|\theta \sim \text{Bernoulli}(\theta)$ com $\Theta = \{1/4, 1/3, 1/2\}$ e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Seja $\pi(1/4) = \alpha$, $\pi(1/3) = \beta$ e $\pi(1/2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ tal que $\alpha + \beta \leq 1$, a distribuição a priori para θ . Para cada $A \in \mathcal{P}(\Theta)$, considere a função de perda L_A definida da seguinte forma:

$$L_A(0, \theta) = \text{dist}(\theta, A) \text{ e } L_A(1, \theta) = \text{dist}(\theta, A^c), \theta \in \Theta,$$

onde $\text{dist}(y, C)$ é a distância euclidiana de um ponto y a um conjunto C , definida como no Exemplo 2.10.

Para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, temos

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-\text{dist}(\theta_1, A^c)}{-\text{dist}(\theta_1, B^c)} \leq 1 \leq \frac{\text{dist}(\theta_2, A)}{\text{dist}(\theta_2, B)} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)},$$

para todo $\theta_1 \in A$ e todo $\theta_2 \in B^c$. Portanto, pelo Teorema 2.2, a classe de testes de Bayes gerada por essa família de funções de perda é monótona para qualquer $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$. Na verdade, é fácil ver que isso sempre vale quando (Θ, d) é espaço métrico (Lima (1977)).

Além disso, para todo $A \in \sigma(\Theta)$, temos

$$\frac{\Delta_A(\theta_0)}{-\Delta_{A^c}(\theta_0)} = \frac{\text{dist}(\theta_0, A)}{\text{dist}(\theta_0, A^c)} = 1 = \frac{\text{dist}(\theta_1, A^c)}{\text{dist}(\theta_1, A)} = \frac{-\Delta_A(\theta_1)}{\Delta_{A^c}(\theta_1)},$$

para todo $\theta_1 \in A$ e todo $\theta_0 \in A^c$. Portanto, pelo Teorema 2.3, a classe de testes de Bayes gerada por $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ é invertível para qualquer $\alpha, \beta \in [0, 1]$ com $\alpha + \beta \leq 1$.

Deste modo, observado $x = 0$, $\varphi_{\{1/4\}}(0) = 0$ se, e somente se, $\alpha > (18 - 10\beta)/27$, $\varphi_{\{1/3\}}(0) = 0$ se, e somente se, $\alpha > (12 - 20\beta)/3$ e $\varphi_{\{1/2\}}(0) = 0$ se, e somente se, $\alpha > (12 - 28\beta)/39$.

Para $\alpha = 0, 2$ e $\beta = 0, 4$, por exemplo, teremos $\varphi_{\{1/4\}}(0) = 1$, $\varphi_{\{1/3\}}(0) = 1$ e $\varphi_{\{1/2\}}(0) = 1$ e, pelo Teorema 1.3, a classe de testes não atende D2.

Em resumo, a classe de testes é monótona e invertível, mas não atende D2, ou seja, ela não é consonante com a união.

Os Exemplos 2.9 e 2.10 mostram que mesmo para famílias de funções de perda bastante intuitivas, a classe de testes gerada não atende simultaneamente as três propriedades.

Como foi mostrado, existem famílias de funções de perda tais que a classe de testes gerada por elas atende D1 para toda distribuição a priori para θ . No entanto, como veremos a seguir, o mesmo não ocorre para a desiderata D2.

Teorema 2.5. *Se Θ e $\sigma(\Theta)$ são tais que $|\Theta| \geq 3$ e existe uma partição mensurável de tamanho três $\{A_1, A_2, A_3\}$ deste espaço. Suponha ainda que $\{\theta\} \in \sigma(\Theta)$ para todo $\theta \in \Theta$ e que a condição (1.2) sobre as perdas que definem um teste de hipóteses é satisfeita estritamente. Então, não existe uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ tal que, a classe de testes de Bayes gerada por essa família atenda à desiderata D2 para toda distribuição a priori para θ .*

Demonstração. Para provar este Teorema utilizaremos um argumento geométrico.

Sejam $\theta_1, \theta_2, \theta_3 \in \Theta$ e $\{A_1, A_2, A_3\}$ uma partição não vazia de Θ , de modo que $\theta_i \in A_i$, $i = 1, 2, 3$.

Note inicialmente que para todo $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A} = \{(a, b, c) \in \mathbb{R}^+ : a + b + c = 1\}$, existe uma distribuição a priori π para θ tal que $\pi(\theta_i|x) = \alpha_i$, $i = 1, 2, 3$. Com efeito, seja π dada por

$$\pi(\theta_1) = \frac{\alpha_1 \alpha_2 V_x(\theta_2) V_x(\theta_3) - \alpha_1 \alpha_2 V_x(\theta_3) V_x(\theta_2) + \alpha_1 V_x(\theta_3)}{V_x(\theta_1) V_x(\theta_2) - \alpha_1 V_x(\theta_1) V_x(\theta_2) + \alpha_1 V_x(\theta_2) V_x(\theta_3) - \alpha_2 V_x(\theta_1) V_x(\theta_2)},$$

$$\pi(\theta_2) = \frac{\pi(\theta_1)[\alpha_2 V_x(\theta_1) - \alpha_2 V_x(\theta_3)] + \alpha_2 V_x(\theta_3)}{(1 - \alpha_2) V_x(\theta_2) + \alpha_2 V_x(\theta_3)} \text{ e } \pi(\theta_3) = 1 - \pi(\theta_1) + \pi(\theta_2).$$

Deste modo, a posteriori de θ , dado x , é dada por $\pi(\theta_1|x) = \alpha_1$, $\pi(\theta_2|x) = \alpha_2$ e $\pi(\theta_3|x) = \alpha_3 = 1 - \alpha_1 - \alpha_2$.

Seja $\{L_A\}_{A \in \sigma(\Theta)}$ uma família qualquer de funções de perda nas condições do enunciado. Para todo $(\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A}$, a diferença entre os riscos de aceitar e rejeitar $H_0^{(i)} : \theta \in A_i$ é dada por

$$\rho_{A_i}(0, \pi_x) - \rho_{A_i}(1, \pi_x) = c_{i1} \alpha_1 + c_{i2} \alpha_2 + c_{i3} \alpha_3, \text{ onde } c_{ij} = L_{A_i}(0, \theta_j) - L_{A_i}(1, \theta_j) \text{ com } c_{ii} < 0 \text{ e } c_{ij} > 0, i \neq j.$$

Para avaliar os três testes simultâneos entre todas as possíveis prioris, consideremos a transformação $T : \mathcal{A} \rightarrow \mathbb{R}^3$ dada por

$$(\alpha_1, \alpha_2, \alpha_3) \mapsto T(\alpha_1, \alpha_2, \alpha_3) = (\rho_{A_1}(0, \pi_x) - \rho_{A_1}(1, \pi_x), \rho_{A_2}(0, \pi_x) - \rho_{A_2}(1, \pi_x), \rho_{A_3}(0, \pi_x) - \rho_{A_3}(1, \pi_x)),$$

a transformação que associa a cada priori $(\alpha_1, \alpha_2, \alpha_3)$ as diferenças de riscos entre aceitar e rejeitar $H^{(i)} : \theta \in A_i$, $i = 1, 2, 3$.

Seja $\mathcal{B} = T(\mathcal{A}) = \{T(\alpha_1, \alpha_2, \alpha_3) : (\alpha_1, \alpha_2, \alpha_3) \in \mathcal{A}\}$. Primeiramente, vamos provar que esse conjunto $\mathcal{B}$ é convexo.

Sejam $q_1 = (q_{11}, q_{12}, q_{13}) \in \mathcal{B}$ e $q_2 = (q_{21}, q_{22}, q_{23}) \in \mathcal{B}$. Então, existem $\pi'_x = (\alpha'_1, \alpha'_2, \alpha'_3)$ e $\pi''_x = (\alpha''_1, \alpha''_2, \alpha''_3)$ tais que $q_1 = T(\pi'_x)$ e $q_2 = T(\pi''_x)$, isto é,

$$q_1 = \left(\sum_{j=1}^3 c_{1j} \alpha'_j, \sum_{j=1}^3 c_{2j} \alpha'_j, \sum_{j=1}^3 c_{3j} \alpha'_j \right) \text{ e } q_2 = \left(\sum_{j=1}^3 c_{1j} \alpha''_j, \sum_{j=1}^3 c_{2j} \alpha''_j, \sum_{j=1}^3 c_{3j} \alpha''_j \right).$$

Então, para $\gamma \in [0, 1]$,

$$\begin{aligned} \gamma q_1 + (1 - \gamma) q_2 &= \left(\gamma \sum_{j=1}^3 c_{1j} \alpha'_j + (1 - \gamma) \sum_{j=1}^3 c_{1j} \alpha''_j, \gamma \sum_{j=1}^3 c_{2j} \alpha'_j + (1 - \gamma) \sum_{j=1}^3 c_{2j} \alpha''_j, \gamma \sum_{j=1}^3 c_{3j} \alpha'_j + (1 - \gamma) \sum_{j=1}^3 c_{3j} \alpha''_j \right) \\ &= \left(\sum_{j=1}^3 c_{1j} [\gamma \alpha'_j + (1 - \gamma) \alpha''_j], \sum_{j=1}^3 c_{2j} [\gamma \alpha'_j + (1 - \gamma) \alpha''_j], \sum_{j=1}^3 c_{3j} [\gamma \alpha'_j + (1 - \gamma) \alpha''_j] \right) \\ &= T(\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3), \end{aligned}$$

onde $\bar{\alpha}_j = \gamma \alpha'_j + (1 - \gamma) \alpha''_j$, $j = 1, 2, 3$. Isto é, existe $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3) \in \mathcal{A}$ tal que $\gamma q_1 + (1 - \gamma) q_2 = T(\bar{\alpha})$, então $\gamma q_1 + (1 - \gamma) q_2 \in T(\mathcal{A}) = \mathcal{B}$. Logo, $\mathcal{B}$ é convexo.

Provaremos agora o resultado propriamente dito. Pelo Teorema 1.3, uma classe de testes atende D2 contra a priori π se, e somente se, aceitarmos exatamente um elemento da partição, para cada $x \in \mathcal{X}$. Isto significa que, para toda distribuição a priori, exatamente uma das três coordenadas de $T(\alpha_1, \alpha_2, \alpha_3)$ pode ser negativa e as outras duas devem ser positivas.

Tomando prioris degeneradas, temos $T(1, 0, 0) = (c_{11}, c_{21}, c_{31})$, $T(0, 1, 0) = (c_{12}, c_{22}, c_{32})$ e $T(0, 0, 1) = (c_{13}, c_{23}, c_{33})$. Sabendo que o conjunto $\mathcal{B}$ é convexo, o esboçamos na Figura 2.3.

Figura 2.3: Conjunto $\mathcal{B}$

Da Figura 2.3, para que a classe de testes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ (determinada pela matrix $C = (c_{ij})_{3 \times 3}$) atenda à desiderata D2 contra toda distribuição a priori π , todos os pontos do “triângulo” $\mathcal{B}$ formado na figura deverão estar nos octantes onde temos somente uma coordenada negativa e as outras duas positivas (2° , 4° e 5° octantes). Isto é, para toda matriz C , com $c_{ii} < 0$ e $c_{ij} > 0$, $i \neq j$, se existir algum ponto do “triângulo” fora desses três octantes, então existe uma distribuição a priori para θ tal que a classe de testes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra essa priori não atende D2.

O problema agora se resume em provar que, qualquer triângulo como o da Figura 2.3 intersecciona um dos outros cinco octantes (1° , 3° , 6° , 7° , 8°).

Notemos, inicialmente, que esses pontos formam de fato um triângulo, isto é, não estão alinhados, pois, caso estivessem, um deles seria combinação linear dos outros dois, por exemplo, existiria $\gamma \in [0, 1]$ tal que $(c_{11}, c_{12}, c_{13}) = \gamma(c_{21}, c_{22}, c_{23}) + (1 - \gamma)(c_{31}, c_{32}, c_{33})$. No entanto, como $c_{11} < 0$ e $c_{21}, c_{31} > 0$ não existe $\gamma \in [0, 1]$ tal que $c_{11} = \gamma c_{21} + (1 - \gamma)c_{31}$. O mesmo argumento vale para as outras duas possibilidades.

Agora mostraremos que, obrigatoriamente, algum ponto desse triângulo passa por um dos cinco octantes. Para auxiliar a visualização do argumento, nas figuras 2.4 mostramos possíveis projeções do conjunto $\mathcal{B}$.

Figura 2.4: *Projeção de $\mathcal{B}$ em $u \times v$*

No que se refere a reta entre os pontos (c_{11}, c_{12}, c_{13}) e (c_{21}, c_{22}, c_{23}) e o octante de localização dos três pontos, o conjunto $\mathcal{B}$ é de um dos tipos da Figura 2.4, onde são feitas possíveis projeções do conjunto

$\mathcal{B}$ sobre (α_1, α_2) . Se forem como na primeira figura (caso em que $c_{12}/(c_{12} - c_{11}) < -c_{22}/(c_{21} - c_{22})$) e considerando o segmento que une os pontos (c_{11}, c_{12}, c_{13}) e (c_{21}, c_{22}, c_{23}) contido em $\mathcal{B}$, existe $\gamma \in [0, 1]$ tal que $\gamma c_{1j} + (1 - \gamma)c_{2j} < 0$, $j = 1, 2$ e, portanto, existe uma priori π concentrada em $\{\theta_1, \theta_2, \theta_3\}$ contra a qual $H_0^{(1)}$ e $H_0^{(2)}$ são aceitas e, portanto, a classe gerada não atende D2.

Analogamente, se for como na terceira figura, (caso em que $c_{12}/(c_{12} - c_{11}) > -c_{22}/(c_{21} - c_{22})$) existe $\gamma \in [0, 1]$ tal que, $\gamma c_{1j} + (1 - \gamma)c_{2j} > 0$, $j = 1, 2, 3$ e, nesse caso, as três hipóteses $H_0^{(1)}$, $H_0^{(2)}$, $H_0^{(3)}$ são rejeitadas e, portanto, a classe não atende D2 para a correspondente priori.

Por fim, se forem como na segunda figura (caso em que $c_{12}/(c_{12} - c_{11}) = -c_{22}/(c_{21} - c_{22})$), o segmento que liga (c_{11}, c_{12}, c_{13}) a (c_{21}, c_{22}, c_{23}) intercepta o eixo w em $(0, 0, t^*)$. Considerando o segmento que une $(0, 0, t^*)$ a (c_{31}, c_{32}, c_{33}) , existe $\gamma \in [0, 1]$ tal que

$$\begin{cases} \gamma 0 + (1 - \gamma)c_{13} > 0 \Rightarrow \gamma < 1 \\ \gamma 0 + (1 - \gamma)c_{23} > 0 \Rightarrow \gamma < 1 \\ \gamma t^* + (1 - \gamma)c_{33} > 0 \Leftrightarrow \gamma > \frac{-c_{33}}{t^* - c_{33}}. \end{cases}$$

Nesse caso, rejeitamos as três hipóteses nulas. Portanto, sempre existe uma distribuição a priori π para θ tal que a classe de testes de Bayes gerada por $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ contra π não atende D2. $\square$

O Teorema acima mostra que não é possível encontrar restrições sobre uma família de funções de perda, de forma que a classe de testes gerada por ela atenda a desiderata D2 para toda priori sob o parâmetro.

Em outras palavras: enquanto um decisor qualquer (com qualquer priori sobre Θ) realiza, a partir de funções de perda bastante intuitivas, testes de Bayes simultâneos atendendo a desiderata D1, foi mostrado (Teorema 2.5) que o mesmo não ocorre em relação a D2.

O resultado do Teorema 2.5 se deve, possivelmente, ao fato de que a exigência simultânea de várias condições lógicas, embora intuitivas, sobre testes múltiplos seja muito restritiva. Daí, mesmo testes simultâneos que gozam de otimalidade em algum sentido (aqui otimalidade sob a perspectiva da Teoria da Decisão Bayesiana) não atendem à desiderata D2.

Ao exigirmos que os testes sejam ao mesmo tempo monótonos, invertíveis e consonantes com a união finita para toda priori em Θ limitamos a classe de testes a um ponto que inviabiliza sua otimalidade, o que pode sugerir uma fragilidade do processo de testes de hipóteses simultâneos como uma ferramenta científica lógica.

Ainda que não existam classes de funções de perda que façam com que uma classe de testes de Bayes atenda à desiderata D2 para todas as prioris sobre Θ fixada a família de distribuições para X indexadas em θ , $\mathbb{P} = \{P_\theta : \theta \in \Theta\}$, é possível mostrar que uma classe de testes que atende a desiderata D2, sob certas condições, é uma classe de testes de Bayes de algum agente decisor.

Teorema 2.6. *Sejam Θ e $\mathcal{X}$ finitos e $\sigma(\Theta) = \mathcal{P}(\Theta)$. Se uma classe de testes atende a desiderata D2, então existem uma medida de probabilidade $\mu : \sigma(\Theta \times \mathcal{X}) \rightarrow [0, 1]$ e uma família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ satisfazendo (1.2) restrita, tais que essa classe é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra a μ -marginal de θ .*

Demonstração. Sejam $\mathcal{X} = \{x_1, \dots, x_n\}$, $\Theta = \{\theta_1, \dots, \theta_k\}$ e seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes atendendo D2. Para cada $x \in \mathcal{X}$, define-se $\xi_x : \sigma(\Theta) \rightarrow [0, 1]$ por $\xi_x(A) = \pi(\theta \in A | x) \equiv 1 - \varphi_A(x)$. Primeiro, será mostrado que $\xi_x(\cdot)$ é medida de probabilidade:

1. Para todo $A \in \sigma(\Theta)$, por definição, $\xi_x(A) \in \{0, 1\}$, logo $\xi_x(A) \geq 0$.

2. Verifiquemos que $\xi_x(\Theta) = 1$. Como $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ é invertível, então para todo $A \in \sigma(\Theta)$, $\varphi_A(x) + \varphi_{A^c}(x) = 1$ e, em particular, $\varphi_\Theta(x) + \varphi_\emptyset(x) = 1$, para todo $x \in \mathcal{X}$. Pela monotonicidade, segue que $\varphi_\Theta(x) = 0$ (e $\varphi_\emptyset(x) = 1$), para todo $x \in \mathcal{X}$ e, portanto, $\xi_x(\Theta) = 1$ (e $\xi_x(\emptyset) = 0$).
3. Vamos mostrar que se $A_n \downarrow \emptyset$, então $\xi_x(A_n) \downarrow 0$. (continuidade no vazio, James (1996)). Suponha que $A_1 \supseteq A_2 \supseteq \dots$. Então, $\cap_{i=1}^n A_i = A_n$, para todo $n \geq 1$. Além disso, suponha que $\lim_{n \rightarrow \infty} A_n = \cap_{n=1}^\infty A_n = \emptyset$. Por definição, $\xi_x(A_n) = 1 - \varphi_{A_n}(x)$. Para algum $m \in \mathbb{N}$, $A_n = \emptyset$ para todo $n \geq m$, pois Θ é finito. Desta forma, $\varphi_{A_m}(x) = 1$ e, para todo $n \geq m$, temos que $\varphi_{A_n}(x) = 1$, para todo $x \in \mathcal{X}$.

Logo, $\lim_{n \rightarrow \infty} \xi_x(A_n) = \lim_{n \rightarrow \infty} 1 - \varphi_{A_n}(x) = 0$.

4. Por fim, mostraremos que $\xi_x(A \cup B) = \xi_x(A) + \xi_x(B)$, $\forall A, B \in \sigma(\Theta)$ disjuntos.

Se $A, B \in \sigma(\Theta)$ são disjuntos, então, se $\varphi_A(x) = 0$, então, $\varphi_B(x) = 1$, para todo $x \in \mathcal{X}$, pois $\varphi_{A^c}(x) = 1 - \varphi_A(x) = 1$ e, como $B \subseteq A^c$, então $\varphi_B(x) \geq \varphi_{A^c}(x) = 1$, para todo $x \in \mathcal{X}$. Analogamente, se $\varphi_B(x) = 0$, então $\varphi_A(x) = 1$. Além disso, $\varphi_{A \cup B} = \varphi_A \varphi_B$ (da monotonicidade e consonância da união) e, portanto, os possíveis valores de $\varphi_{A \cup B}$, φ_A e φ_B , com A, B disjuntos são:

- $\varphi_{A \cup B}(x) = 0$, $\varphi_A(x) = 1$ e $\varphi_B(x) = 0$.
Então $\xi_x(A \cup B) = 1 - \varphi_{A \cup B}(x) = 1$, $\xi_x(A) = 1 - \varphi_A(x) = 0$ e $\xi_x(B) = 1 - \varphi_B(x) = 1$. Logo, $\xi_x(A \cup B) = \xi_x(A) + \xi_x(B)$.
- $\varphi_{A \cup B}(x) = 0$, $\varphi_A(x) = 0$ e $\varphi_B(x) = 1$.
Análogo ao item anterior.
- $\varphi_{A \cup B}(x) = 1$, $\varphi_A(x) = 1$ e $\varphi_B(x) = 1$.
Análogo ao item anterior.

Portanto, $\forall A, B \in \sigma(\Theta)$ disjuntos $\xi_x(A \cup B) = \xi_x(A) + \xi_x(B)$. Assim, para todo $x \in \mathcal{X}$, $\xi_x : \mathcal{P}(\Theta) \rightarrow [0, 1]$ é medida de probabilidade.

Além disso, seja P uma medida de probabilidade em $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$. Definindo $\mu : \mathcal{P}(\Theta \times \mathcal{X}) \rightarrow [0, 1]$ por $\mu(B) = \sum_{(\theta, x) \in B} P(x) \xi_x(\theta)$, $B \in \mathcal{P}(\Theta \times \mathcal{X})$, é fácil ver que μ é medida de probabilidade em $\mathcal{P}(\Theta \times \mathcal{X})$ e que $\xi_x(\cdot)$ é a distribuição de θ , dado x .

Por fim, vamos provar que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes contra a μ marginal de θ .

Seja $\{L_A\}_{A \in \sigma(\Theta)}$ uma classe de perdas atendendo a condição (1.2) restrita, isto é, $L_A(0, \theta) < L_A(1, \theta)$ para $\theta \in A$ e $L_A(0, \theta) > L_A(1, \theta)$ para $\theta \notin A$, para todo $A \in \sigma(\Theta)$ e todo $\theta \in \Theta$.

Após observar $x \in \mathcal{X}$, seja $\rho_A(d, \xi_x)$ a perda esperada a posteriori ao escolher a decisão $d = 0, 1$. Então,

$$\rho_A(0, \xi_x) - \rho_A(1, \xi_x) = \sum_{\theta_j \in A} [L_A(0, \theta_j) - L_A(1, \theta_j)] \xi_x(\theta_j) + \sum_{\theta_j \in A^c} [L_A(0, \theta_j) - L_A(1, \theta_j)] \xi_x(\theta_j).$$

Se $\varphi_A(x) = 0$, então $\xi_x(A) = 1$ e, portanto, $\rho_A(0, \xi_x) - \rho_A(1, \xi_x) = \sum_{\theta_j \in A} [L_A(0, \theta_j) - L_A(1, \theta_j)] < 0$. Analogamente, se $\varphi_A(x) = 1$, então $\rho_A(0, \xi_x) > \rho_A(1, \xi_x)$. Assim, $\varphi_A(x) = 0$ se, e somente se, $\rho_A(0, \xi_x) < \rho_A(1, \xi_x)$. Portanto, $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra a μ -marginal de θ . $\square$

O Teorema 2.6 mostra que uma classe de testes que atende D2, tal que $\sigma(\Theta) = \mathcal{P}(\Theta)$, é sempre uma classe de testes de Bayes de algum decisor. Podemos imaginar um paralelo entre esse Teorema e o fato de que toda decisão admissível d^* é decisão de Bayes contra alguma distribuição sobre Θ finito. Aqui, toda classe de testes que atende D2 (ou seja, que atende propriedade lógicas razoáveis) é classe de testes de Bayes

contra alguma distribuição para θ (Θ finito). Assim, impor tais propriedades lógicas a testes simultâneos parece ser razoável no sentido de corresponder à ação de algum agente decisor bayesiano.

É interessante notar que o Teorema acima fala da existência de uma medida de probabilidade em $\sigma(\Theta \times \chi)$ e não da existência de uma priori para θ , fixada a família de distribuição $\mathbb{P}$. No casos em que $\mathbb{P}$ é fixada de antemão, nem sempre existe priori sobre Θ tal que uma dada classe de testes atendendo D2 seja uma classe de testes bayesianos.

Vale ressaltar que o resultado do Teorema 2.6 é válido para Θ enumerável se a condição de consonância da união finita é substituída pela consonância da união enumerável.

A seguir, mostraremos que, se uma classe de testes atende a desiderata D2, pode não existir uma priori tal que essa classe seja classe de testes de Bayes, uma vez fixada a família de distribuição $\mathbb{P} = \{P_\theta : \theta \in \Theta\}$.

Exemplo 2.11. *Seja $X|\theta \sim \text{Uniforme}\{1, \dots, \theta\}$, onde $\Theta = \{2, 3, 4\}$. Suponha que a distribuição a priori para θ é $\pi(4) = \alpha$, $\pi(3) = \beta$ e $\pi(2) = 1 - \alpha - \beta$, com $\alpha, \beta \in [0, 1]$ e $\alpha + \beta \leq 1$. Seja $W : \chi \rightarrow \Theta$ um estimador tal que $W(1) = 2$, $W(2) = W(3) = 3$ e $W(4) = 4$ e $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ uma classe de testes onde $\varphi_A(x) = 1 - I_A(W(x))$, para $A \in \sigma(\Theta)$. No Exemplo 1.19, foi mostrado que essa classe atende D2.*

Pela construção da classe $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$, $\varphi_{\{3\}}(2) = 0$ e $\varphi_{\{3\}}(1) = 1$. Seja $\{L_A\}_{A \in \mathcal{P}(\Theta)}$, $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ para $A \in \sigma(\Theta)$, uma família de funções de perda qualquer. Se $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ for uma classe de testes de Bayes, devemos ter:

$$\rho_{\{3\}}(0, \pi_2) < \rho_{\{3\}}(1, \pi_2) \text{ e } \rho_{\{3\}}(0, \pi_1) > \rho_{\{3\}}(1, \pi_1). \quad (2.14)$$

Mas, as distribuições a posteriori de θ , dado $X = 2$ e dado $X = 1$, π_2 e π_1 , respectivamente, coincidem para todo $\alpha, \beta \in [0, 1]$ e $\alpha + \beta \leq 1$. Assim, não existem $\{L_A\}_{A \in \mathcal{P}(\Theta)}$, $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ para $A \in \mathcal{P}(\Theta)$ e priori sobre Θ tais que as desigualdades em (2.14) sejam verificadas.

Na demonstração do Teorema 2.6, consideramos uma medida em $\sigma(\Theta \times \chi)$ tal que as distribuições condicionais de θ dado x são degeneradas. No exemplo a seguir, mostraremos que existe uma outra medida de probabilidade que pode ser considerada com distribuições condicionais tais que $\pi(\theta_i|x) > 0$, $i = 1, \dots, k$ e todo $x \in \chi$.

Exemplo 2.12. *Sejam $\Theta = \{\theta_1, \dots, \theta_k\}$ e χ finitos. Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ uma classe de testes que atende a desiderata D2. Assim, para cada $x \in \chi$, existe um único $\theta(x) \in \Theta$ tal que $\varphi_{\{\theta(x)\}}(x) = 0$, de modo que podemos escrever $\varphi_A(x) = 1 - I_A(\theta(x))$. Seja P a distribuição uniforme sobre χ . Para cada $x \in \chi$, considere a medidade de probabilidade $\xi_x : \sigma(\Theta) \rightarrow [0, 1]$ tal que*

$$\xi_x(\{\theta(x)\}) = \frac{2^{k-1}}{1+2+4+\dots+2^{k-1}} \text{ e } \xi_x(\{\theta_i(x)\}) = \frac{2^{i-1}}{1+2+4+\dots+2^{k-1}}, \quad i = 1, \dots, k-1$$

onde $(\theta_1(x), \dots, \theta_{k-1}(x))$ é uma enumeração dos elementos de $\Theta \setminus \{\theta(x)\}$.

Além disso, seja $\mathbb{P}$ uma medida de probabilidade em $(\chi, \mathcal{P}(\chi))$. Definindo $\mu : \mathcal{P}(\Theta \times \chi) \rightarrow [0, 1]$ por $\mu(B) = \sum_{(\theta, x) \in B} \mathbb{P}(x) \xi_x(\theta)$, $B \in \mathcal{P}(\Theta, \chi)$, é fácil ver que μ é medida de probabilidade em $\mathcal{P}(\Theta \times \chi)$ e que $\xi_x(\cdot)$ é a distribuição de θ , dado x .

Considerando a família de funções de perda 0-1, temos, para cada $A \in \mathcal{P}(\Theta)$, que as perdas esperadas a posteriori, dado $X = x$ são tais que $\rho_A(0, \xi_x) - \rho_A(1, \xi_x) = \pi(\theta \in A^c|x) - \pi(\theta \in A|x)$.

Se $\varphi_A(x) = 0$, então $\theta_0(x) \in A$ e, portanto $\pi(A|x) > 2^{k-1}/(1+2+\dots+2^{k-1}) > 1/2$, de modo que $\rho_A(0, \xi_x) - \rho_A(1, \xi_x) < 0$. Analogamente, se $\varphi_A(x) = 1$, segue que $\rho_A(0, \xi_x) - \rho_A(1, \xi_x) > 0$. Assim, $\varphi_A(x) = 0$ se, e somente se, $\rho_A(0, \xi_x) < \rho_A(1, \xi_x)$ e, portanto, $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ é classe de testes de Bayes gerada por $\{L_A\}_{A \in \mathcal{P}(\Theta)}$ contra a μ -marginal de θ .

Ou seja, aqui a medida de probabilidade no espaço $\Theta \times \mathcal{X}$ é diferente da utilizada na demonstração do Teorema 2.6. No próximo capítulo mostraremos como devem ser as classes de testes para que atendam D2, já que não é possível caracterizá-las de forma geral pelas classes de testes de Bayes.

Capítulo 3

Classes de testes geradas por estimadores

Classes de testes geradas por estimadores são tipos especiais de classes de testes e, num certo sentido, permitem relacionar estimadores pontuais com testes de hipóteses. Além disso, como mostramos no Exemplo 1.19, a classe gerada por um estimador sempre atende a desiderata D2, embora, muitas vezes, tais testes não atendam critérios de otimalidade usuais, como, por exemplo, minimização de perda esperada ou maximização (uniforme) da função poder na hipótese alternativa.

Um questionamento que poderia surgir dos resultados obtidos anteriormente é como deveria ser uma classe de testes para atender a desiderata D2. Izbicki (2010) provou, em situações bem gerais, que uma classe de testes é monótona, invertível e consoante com a união enumerável se, e somente se, existe um estimador $W : \mathcal{X} \rightarrow \Theta$ que gera tal classe. A seguir, reproduzimos tal resultado para os casos em que Θ é finito ou enumerável.

Teorema 3.1. *Seja Θ finito (enumerável). Uma classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é monótona, invertível e consoante com a união finita (enumerável) se, e somente se, existe $W : \mathcal{X} \rightarrow \Theta$ tal que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é uma classe de testes gerada pelo estimador W .*

Demonstração. Que uma classe de testes gerada por um estimador atende D2, está provado no Exemplo 1.19. Agora, provaremos que uma classe que atende D2 é uma classe gerada por um estimador.

Pelo Teorema 1.3, para todo $x \in \mathcal{X}$, existe e é único, $\theta_0 = \theta_0(x) \in \Theta$ finito, tal que $\varphi_{\{\theta_0\}}(x) = 0$. Assim, para $A \in \sigma(\Theta)$, se $\theta_0 \in A$, então $\{\theta_0(x)\} \subseteq A$ e, pela monotonicidade, $\varphi_A(x) = 0$. Por outro lado, se $\theta_0 \notin A$, então $\{\theta_0(x)\} \subseteq A^c$ e, pela monotonicidade, $\varphi_{A^c}(x) = 0$, de onde, pela invertibilidade, $\varphi_A(x) = 1$. Assim, para cada $A \in \sigma(\Theta)$, $\varphi_A(x) = 1$ se, e somente se, $\theta(x) \notin A$, isto é, se $1 - I_A(\theta(x)) = 1$.

Definindo $W : \mathcal{X} \rightarrow \Theta$ por $W(x) = \theta(x)$, segue que W é um estimador para θ e que para todo $A \in \sigma(\Theta)$, $\varphi_A(x) = 1 - I_A(W(x))$, isto é, $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é gerada pelo estimador W .

A prova é análoga quando Θ é enumerável e a consonância da união finita é substituída pela consonância da união enumerável. $\square$

Na sequência, mostraremos uma conexão entre testes e estimadores bayesianos, isto é, apresentaremos condições que uma família de funções de perda deve obedecer para que classes de testes geradas por estimadores bayesianos caracterizem classes de testes de Bayes. Ou seja, estabeleceremos uma relação entre estimadores pontuais de Bayes e testes de hipóteses de Bayes.

3.1 Estimadores de Bayes e EMVs

Em princípio, parece ser razoável que um agente bayesiano que realiza testes simultâneos que satisfazem a desiderata D2 o faça através de um estimador de Bayes. Do mesmo modo, se um agente constrói um estimador de Bayes ótimo e o utiliza para fazer testes simultâneos, tais testes devem ser ótimos sob a perspectiva bayesiana. Isto será estudado no próximo Teorema.

Teorema 3.2. *Sejam χ e Θ finitos e seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada por um estimador $W : \chi \rightarrow \Theta$ e que, portanto, atende D2. Suponha que $V_x(W(x))$ é positiva, para todo $x \in \chi$.*

- (a) *Se existem $\pi : \sigma(\Theta) \rightarrow [0, 1]$, medida de probabilidade a priori para θ com $\pi(W(x))$ positivo para todo $x \in \chi$, e $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$, com $L(\theta, \theta) = 0$ e $L(d, \theta)$ positivo para $d \neq \theta$, tais que W é estimador de Bayes gerado por L contra π , então existe $\{L_A\}_{A \in \sigma(\Theta)}$, com $L_A : \{0, 1\} \times \Theta \times \chi \rightarrow \mathbb{R}^+$ para cada $A \in \sigma(\Theta)$, tal que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π .*
- (b) *Se existe $\pi : \sigma(\Theta) \rightarrow [0, 1]$, medida de probabilidade a priori para θ com $\pi(W(x))$ positivo para todo $x \in \chi$, e existe $\{L_A\}_{A \in \sigma(\Theta)}$, com $L_A : \{0, 1\} \times \Theta \rightarrow \mathbb{R}^+$ satisfazendo $L_A(0, \theta) < L_A(1, \theta)$, $\theta \in A$ e $L_A(0, \theta) > L_A(1, \theta)$, $\theta \notin A$ para cada $A \in \sigma(\Theta)$ tais que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π , então existe $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$, com $L(\theta, \theta) = 0$ e $L(d, \theta)$ positivo para $d \neq \theta$, tal que W é estimador de Bayes gerado por L contra π .*

Demonstração. Primeiramente, vamos provar a parte (b).

Suponha que $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π . Como $\{\varphi_A\}_{A \in \sigma(\Theta)}$ é classe de testes gerada por um estimador, ela atende D2. Segue, pelo Teorema 1.3, que para cada $x \in \chi$ essa classe aceita somente uma hipótese simples da partição de Θ em conjuntos unitários. Assim, para cada $x \in \chi$, $\varphi_{\{W(x)\}}(x) = 0$ e $\varphi_{\{d\}}(x) = 1$, para $d \in \Theta \setminus \{W(x)\}$.

Nestas condições, sendo $\Delta_A(\theta) = L_A(0, \theta) - L_A(1, \theta)$,

$$\sum_{\theta \in \Theta} \Delta_{\{W(x)\}}(\theta) \pi(\theta|x) < 0 < \sum_{\theta \in \Theta} \Delta_{\{d\}}(\theta) \pi(\theta|x),$$

para todo $d \in \Theta \setminus \{W(x)\}$ e $\pi(\cdot|x)$ distribuição a posteriori de θ obtida a partir da priori π e da verossimilhança gerada por x , para todo $x \in \chi$.

Logo, para todo $d \in \Theta \setminus \{W(x)\}$,

$$\sum_{\theta \in \Theta} \Delta_{\{W(x)\}}(\theta) \pi(\theta|x) < \sum_{\theta \in \Theta} \Delta_{\{d\}}(\theta) \pi(\theta|x).$$

Definindo $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$ por $L(d, \theta) = \Delta_{\{d\}}(\theta) - \min_{d \in \Theta} \{\Delta_{\{d\}}(\theta)\} = \Delta_{\{d\}}(\theta) - \Delta_{\{\theta\}}(\theta)$, isto é, para cada $d \in \Theta$ e cada $\theta \in \Theta$, toma-se a diferença entre as penalidades de aceitar e rejeitar a hipótese $\{d\}$ quando o estado da natureza é θ e retira-se o mínimo para garantir que essa perda seja positiva para quaisquer d e θ , com $d \neq \theta$, e zero quando $d = \theta$. Segue que, para $x \in \chi$,

$$\sum_{\theta \in \Theta} L(W(x), \theta) \pi(\theta|x) < \sum_{\theta \in \Theta} L(d, \theta) \pi(\theta|x),$$

para todo $d \in \Theta \setminus \{W(x)\}$.

Consequentemente, $W(x)$ é a estimativa que minimiza a perda esperada contra $\pi(\cdot|x)$ e, portanto, W é um estimador de Bayes gerado por L contra π .

Agora vamos provar a parte (a). Para isso, escreveremos uma família de função de perda tal que a classe de testes de Bayes gerada por ela atenda a desiderata D1 e, em seguida, mostraremos que essa classe também atende D2 e, por fim, mostraremos que essa classe coincide com a classe de testes gerada pelo estimador W .

Definimos a seguir a família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$, $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$ é tal que, para $\theta \in A^c$ e $x \in \mathcal{X}$, $L_A(1, \theta, x) = 0$ e

$$L_A(0, \theta, x) = \min \left\{ \min \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_A(W(x)) + \max \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_{A^c}(W(x)) : d \in A \right\}, \quad (3.1)$$

e, para $\theta \in A$ e $x \in \mathcal{X}$, $L_A(0, \theta, x) = 0$ e

$$L_A(1, \theta, x) = \min \left\{ \frac{1}{C} \min \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_{A^c}(W(x)) + C \max \left\{ L(d, \theta); \frac{1}{L(d, \theta)} \right\} I_A(W(x)) : d \in A^c \right\}, \quad (3.2)$$

onde $C \geq 1$ é qualquer constante tal que

$$C > \max \left\{ \frac{E[L(W(x), \theta)|x]}{\pi(W(x)|x)} : x \in \mathcal{X} \right\}.$$

A dependência das funções de perda acima nos dados e no estimador W visa impor que a decisão de menor risco no problema de estimação, $W(x)$, seja também a única hipótese simples a não ser rejeitada através dos testes de Bayes quando $x \in \mathcal{X}$ é observado. A constante C deve ser gerada a ponto de que, para cada $x \in \mathcal{X}$, a hipótese simples $\{W(x)\}$ seja aceita com base em x .

Funções de perda dependendo da amostra, embora não tão frequentes na literatura, são discutidas, por exemplo, em [Berger \(1985\)](#) e [Madruga et al. \(2001\)](#), [Evans e Jang \(2011\)](#), dentre outros. Em particular, [Berger \(1985\)](#) discute regras de penalidade que dependem da amostra e também de um estimador por região (região de confiança) para o parâmetro de interesse. As funções de perda propostas aqui também dependem de um estimador, no caso, pontual.

Essa família de funções de perda satisfaz a condição (2.1) do Teorema 2.2, para todo $A, B \in \sigma(\Theta)$ com $A \subseteq B$, todo $\theta_1 \in A$ e todo $\theta_2 \in B^c$, para qualquer $x \in \mathcal{X}$. Mostraremos isso na sequencia, em três casos. Sejam $A, B \in \sigma(\Theta)$ com $A \subseteq B$ e sejam $\theta_1 \in A$ e $\theta_2 \in B^c$.

(1) Para $W(x) \in A$,

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{-C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

(2) Para $W(x) \in B \setminus A$ (lembrando que $C \geq 1$),

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{-C \min \left\{ \max \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \min \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

(3) Para $W(x) \in B^c$,

$$\frac{\Delta_A(\theta_1)}{\Delta_B(\theta_1)} = \frac{-\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in A^c \right\}}{-\frac{1}{C} \min \left\{ \min \left\{ L(d, \theta_1); \frac{1}{L(d, \theta_1)} \right\} : d \in B^c \right\}} \leq 1 \leq \frac{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in A \right\}}{\min \left\{ \max \left\{ L(d, \theta_2); \frac{1}{L(d, \theta_2)} \right\} : d \in B \right\}} = \frac{\Delta_A(\theta_2)}{\Delta_B(\theta_2)}.$$

Assim, uma classe de testes de Bayes gerada por essa família de funções de perda é monótona pelo Teorema 2.2, se $C \geq 1$, contra qualquer priori para θ .

Essa família de funções de perda também satisfaz a condição (2.5) do Teorema 2.3, pois, para $A \in \sigma(\Theta)$,

$$\frac{\Delta_A(\theta_1)}{-\Delta_{A^c}(\theta_1)} = \frac{\min\{\min\{L(d, \theta_1), \frac{1}{L(d, \theta_1)}\} : d \in A\}}{\frac{1}{C} \min\{\min\{L(d, \theta_1), \frac{1}{L(d, \theta_1)}\} : d \in A\}} = \frac{1}{\frac{1}{C}} = C = \frac{C \min\{\max\{L(d, \theta_0), \frac{1}{L(d, \theta_0)}\} : d \in A^c\}}{\min\{\max\{L(d, \theta_0), \frac{1}{L(d, \theta_0)}\} : d \in A^c\}} = \frac{-\Delta_A(\theta_0)}{\Delta_{A^c}(\theta_0)},$$

para todo $\theta_0 \in A$ e todo $\theta_1 \in A^c$, quando $W(x) \in A$. Analogamente, ela satisfaz (2.5) quando $W(x) \notin A$. Deste modo, uma classe de testes de Bayes gerada por essa família de funções de perda é invertível pelo Teorema 2.3.

Seja $\{\varphi'_A\}_{A \in \sigma(\Theta)}$ uma classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ definida acima. Para cada $x \in \mathcal{X}$, os riscos a posteriori, dado x , para a hipótese $\{W(x)\}$ são dados por

$$\begin{aligned} \rho_{\{W(x)\}}(0, \pi_x) &= \sum_{\theta_i \neq W(x)} L_{\{W(x)\}}(0, \theta_i, x) \pi(\theta_i | x) \\ &= \sum_{\theta_i \neq W(x)} \min \left\{ L(W(x), \theta_i); \frac{1}{L(W(x), \theta_i)} \right\} \pi(\theta_i | x), \end{aligned} \quad (3.3)$$

e

$$\begin{aligned} \rho_{\{W(x)\}}(1, \pi_x) &= L_{\{W(x)\}}(1, W(x), x) \pi(W(x) | x) \\ &= C \min \left\{ \max \left\{ L(d, W(x)); \frac{1}{L(d, W(x))} \right\} : d \neq W(x) \right\} \pi(W(x) | x), \end{aligned} \quad (3.4)$$

onde

$$C > \max \left\{ \frac{E[L(W(x), \theta) | x]}{\pi(W(x) | x)} : x \in \mathcal{X} \right\} \geq \frac{\sum_{\theta_i \neq W(x_0)} L(W(x_0), \theta_i) \pi(\theta_i | x_0)}{\pi(W(x_0) | x_0)},$$

para qualquer $x_0 \in \mathcal{X}$.

Desta forma, utilizando a desigualdade acima em (3.4), temos

$$\rho_{\{W(x)\}}(1, \pi_x) > \frac{\sum_{\theta_i \neq W(x)} L(W(x), \theta_i) \pi(\theta_i | x)}{\pi(W(x) | x)} \min \left\{ \max \left\{ L(d, W(x)); \frac{1}{L(d, W(x))} \right\} : d \neq W(x) \right\} \pi(W(x) | x),$$

ou seja,

$$\rho_{\{W(x)\}}(1, \pi_x) > \min \left\{ \max \left\{ L(d, W(x)); \frac{1}{L(d, W(x))} \right\} : d \neq W(x) \right\} \sum_{\theta_i \neq W(x)} L(W(x), \theta_i) \pi(\theta_i | x).$$

De (3.3) temos que,

$$\sum_{\theta_i \neq W(x)} L(W(x), \theta_i) \pi(\theta_i | x) \geq \sum_{\theta_i \neq W(x)} \min \left\{ L(W(x), \theta_i); \frac{1}{L(W(x), \theta_i)} \right\} \pi(\theta_i | x) = \rho_{\{W(x)\}}(0, \pi_x).$$

De modo que,

$$\rho_{\{W(x)\}}(1, \pi_x) > \min \left\{ \max \left\{ L(d, W(x)); \frac{1}{L(d, W(x))} \right\} : d \neq W(x) \right\} \rho_{\{W(x)\}}(0, \pi_x).$$

Mas, para qualquer $d \neq W(x)$, $L(d, W(x))$ é positiva. Logo, $\max \left\{ L(d, W(x)); [L(d, W(x))]^{-1} \right\} \geq 1$, para todo $d \neq W(x)$, e, por sua vez, $\min \left\{ \max \left\{ L(d, W(x)); [L(d, W(x))]^{-1} \right\} : d \neq W(x) \right\} \geq 1$.

Consequentemente, $\rho_{\{W(x)\}}(1, \pi_x) > \rho_{\{W(x)\}}(0, \pi_x)$, isto é, $\varphi'_{\{W(x)\}}(x) = 0$.

Foi mostrado anteriormente que a classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π , $\{\varphi_A\}_{A \in \sigma(\Theta)}$

atende a desiderata D1, quando $C \geq 1$, de modo que, pelo Teorema 1.2, aceita-se no máximo uma hipótese simples. Logo, para cada $x \in \mathcal{X}$, se $\varphi'_{\{W(x)\}}(x) = 0$, então $\varphi'_{\{d\}}(x) = 1$, para todo $d \in \Theta \setminus \{W(x)\}$. Assim, como, para cada $x \in \mathcal{X}$, aceitamos exatamente uma hipótese simples, segue pelo Teorema 1.3, que a classe de testes de Bayes $\{\varphi'_A\}_{A \in \sigma(\Theta)}$ atende D2. Pelo Teorema 3.1, segue que $\{\varphi'_A\}_{A \in \sigma(\Theta)}$ é gerada por um estimador, a saber, W . Portanto, para essa distribuição a priori para θ , a classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ coincide com a classe de testes gerada pelo estimador de Bayes W . $\square$

O Teorema 3.2 relaciona testes baseados em estimadores de Bayes com classes de testes de Bayes. Sob algumas condições para a família de funções de perda, é equivalente avaliar para cada $x \in \mathcal{X}$ se a estimativa de Bayes está no conjunto que caracteriza uma hipótese nula ou se o teste de Bayes aceita essa hipótese nula.

A seguir, um exemplo que ilustra o Teorema acima.

Exemplo 3.1. Seja $\Theta = \{0, 1/1000, 2/1000, \dots, 999/1000, 1\}$ e $X|\theta \sim \text{Bin}(10, \theta)$. Suponha que $W_1 : \mathcal{X} \rightarrow \Theta$ seja tal que $W_1(x) = x/10$, isto é, W_1 é o EMV para θ . Além disso, considerando a distribuição a priori para θ , π , uniforme em Θ e seja $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$ tal que $L(\theta, \theta) = 0$ e $L(d, \theta) = 1$, $d \neq \theta$, W_1 também é estimador de Bayes para θ .

Além disso, como $W_1(x)$ é moda da distribuição de θ , dado x , então $\pi(W_1(x)|x) \geq 1/1001$, para todo $x \in \mathcal{X}$ (caso contrário, a distribuição a posteriori não somaria 1) e $E[L(W_1(x), \theta)|X = x] = 1 - \pi(W_1(x)|x) \leq 1 - 1/1001 = 1000/1001$. Segue que

$$\max_{x \in \mathcal{X}} \frac{E[L(W_1(x), \theta)|x]}{\pi(W_1(x)|x)} \leq \frac{1000}{1001} 1001 = 1000.$$

Assim, pela parte (b) do Teorema 3.2, tomando $C = 1001$, a classe de testes gerada por W_1 , $\varphi_A(x) = 1 - I_A(W_1(x))$, $A \in \mathcal{P}(\Theta)$, é uma classe de testes de Bayes contra, por exemplo, $\{L_A\}_{A \in \sigma(\Theta)}$ dada por

$$L_A(0, \theta, x) = 0, \theta \in A \text{ e } L_A(1, \theta, x) = 0, \theta \in A^c,$$

$$L_A(0, \theta, x) = 1, \theta \notin A \text{ e } L_A(1, \theta, x) = \frac{1}{1001} I_{A^c}(W_1(x)) + 1001 I_A(W_1(x)), \theta \in A.$$

A distribuição a posteriori para θ é

$$\pi\left(\theta = \frac{j}{1000} | x\right) \propto \frac{\frac{1}{1001} \binom{10}{x} \left(\frac{j}{1000}\right)^x \left(\frac{1000-j}{1000}\right)^{10-x}}{\frac{1}{1001} \sum_{i=0}^{1000} \binom{10}{x} \left(\frac{i}{1000}\right)^x \left(\frac{1000-i}{1000}\right)^{10-x}},$$

e, após algumas simplificações, temos

$$\pi\left(\theta = \frac{j}{1000} | x\right) = \frac{j^x (1000-j)^{10-x}}{\sum_{i=0}^{1000} i^x (1000-i)^{10-x}}.$$

Se $A = \{1/1000, 2/1000, \dots, 330/1000\}$, por exemplo, os riscos a posteriori $\rho_A(0, \pi_x) = \sum_{\theta \notin A} \pi(\theta|x)$ e $\rho_A(1, \pi_x) = \sum_{\theta \in A} L_A(1, \theta) \pi(\theta|x)$, para cada $x \in \mathcal{X}$, são calculados na tabela abaixo.

Pela Tabela 3.1, o teste de Bayes gerado por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π é dado por

$$\varphi_A(x) = \begin{cases} 1 & \text{se } x > 4 \\ 0 & \text{c.c.} \end{cases}$$

x	$\rho_A(0, \pi_x)$	$\rho_A(1, \pi_x)$
0	0,012	989
1	0,078	923
2	0,240	761
3	0,481	520
4	0,718	0,000
5	0,882	0,000
6	0,963	0,000
7	0,992	0,000
8	0,999	0,000
9	1,000	0,000
10	1,000	0,000

Tabela 3.1: Cálculo dos riscos esperados para A

que coincide com o teste gerado pelo estimador de Bayes, W_1 , para o mesmo subconjunto A .

Deste exemplo, derivamos uma consequência do Teorema 3.2.

Corolário 3.1. Suponha Θ e $\mathcal{X}$ finitos. Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada por um estimador de máxima verossimilhança (EMV) para θ , $W : \mathcal{X} \rightarrow \Theta$, com $V_x(W(x))$ positivo para todo $x \in \mathcal{X}$. Então, existem uma família de funções de perda $\{L_A\}_{A \in \mathcal{P}(\Theta)}$, com $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$, para cada $A \in \mathcal{P}(\Theta)$ e uma distribuição a priori π , com $\pi(\theta)$ positivo, para todo $\theta \in \Theta$, tal que a classe $\{\varphi_A\}_{A \in \mathcal{P}(\Theta)}$ é classe de testes de Bayes contra π .

Demonstração. Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada pelo estimador $W : \mathcal{X} \rightarrow \Theta$, onde W é um EMV para θ , ou seja, para cada $x \in \mathcal{X}$, aceitamos a hipótese $\theta \in A$ se, e somente se, $W(x) \in A$.

Suponha que $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ e seja $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$ dada por $L(\theta, \theta) = 0$ e $L(d, \theta) = 1$, $d \neq \theta$. Desta forma, o EMV é o estimador de Bayes para a perda 0-1 contra uma priori uniforme para θ .

Do Teorema 3.2, basta tomar

$$C > \max_{x \in \mathcal{X}} \frac{E[L(d, \theta)|x]}{\pi(W(x)|x)} = \max_{x \in \mathcal{X}} \frac{1 - \pi(W(x)|x)}{\pi(W(x)|x)}$$

nas funções de perda então construídas.

Como $W(x)$ é moda da distribuição de θ , dado x , segue que $\pi(W(x)|x) \geq 1/k$, para todo $x \in \mathcal{X}$. Além disso, como $g : (0, 1] \rightarrow \mathbb{R}^+$ dada por $g(t) = (1 - t)/t$ é estritamente decrescente, segue que

$$\max_{x \in \mathcal{X}} \frac{1 - \pi(W(x)|x)}{\pi(W(x)|x)} \leq g(1/k) = \frac{1 - \frac{1}{k}}{\frac{1}{k}} = k - 1.$$

Assim, tomando qualquer $C > k - 1$, por exemplo, $C = k > 1$, temos que a classe de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$, com $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$ dada por $L_A(1, \theta, x) = 0$ e $L_A(0, \theta, x) = 1$ para $\theta \in A^c$ e $L_A(0, \theta, x) = 0$ e $L_A(1, \theta, x) = kI_A(W(x)) + (1/k)I_{A^c}(W(x))$ para $\theta \in A$, gera a classe de testes $\{\varphi_A\}_{A \in \sigma(\Theta)}$, isto é, $\{\varphi_A\}_{A \in \sigma(\Theta)}$, a classe de testes gerada pelo estimador W , é classe de testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra a distribuição uniforme em Θ . $\square$

O Exemplo 3.1 ilustra esse corolário, pois W_1 é também EMV para θ , a seguir exibiremos mais um exemplo de distribuição a priori e família de funções de perda tais que a classe de testes gerada pelo EMV coincide com a classe de testes de Bayes gerada por esta família.

Exemplo 3.2. Seja $\Theta = \{1/4, 1/3, 1/2\}$ e $X|\theta \sim \text{Bernoulli}(\theta)$. O EMV de θ é dado por $W(0) = 1/4$ e $W(1) = 1/2$ e portanto, $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada pelo estimador W é caracterizada por

$$\varphi_{\{1/4\}}(x) = \begin{cases} 1 & \text{se } x = 1 \\ 0 & \text{se } x = 0; \end{cases}$$

$$\varphi_{\{1/3\}}(x) = 1, \text{ se } x = 0, 1;$$

$$\varphi_{\{1/2\}}(x) = \begin{cases} 1 & \text{se } x = 0 \\ 0 & \text{se } x = 1; \end{cases}$$

Seja π distribuição uniforme em Θ , então a distribuição a posteriori para θ é dada por $\pi(1/4|x) \propto 3^{1-x}/12$, $\pi(1/3|x) \propto 2^{1-x}/9$ e $\pi(1/2|x) \propto 1/6$ e seja $L(\theta, \theta) = 0$ e $L(d, \theta) = 1$, $d \neq \theta$. O estimador de Bayes gerado por L contra π coincide com o EMV.

Pelo Corolário 3.1 assumindo, por exemplo, que $C = k = 3$ e tomando, para cada $x \in \mathcal{X}$ a família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ onde $L_A(1, \theta, x) = 0$ e $L_A(0, \theta, x) = 1$, para $\theta \in A^c$ e $L_A(0, \theta, x) = 0$ e $L_A(1, \theta, x) = kI_A(W(x)) + (1/k)I_{A^c}(W(x))$, para $\theta \in A$, teremos que $\{\varphi'_A\}_{A \in \sigma(\Theta)}$, uma classe testes de Bayes gerada por $\{L_A\}_{A \in \sigma(\Theta)}$ contra π . Observado $x = 0$, $\{\varphi'_A\}_{A \in \sigma(\Theta)}$ pode ser caracterizada a seguir. Para $A = \{1/4\}$,

$$\begin{aligned} \rho_{\{1/4\}}(0, \pi_x) - \rho_{\{1/4\}}(1, \pi_x) &= \sum_{\theta \in \Theta} [L_{\{1/4\}}(0, \theta, 0) - L_{\{1/4\}}(1, \theta, 0)] \pi(\theta|0) \\ &\propto -(3)3/12 + 2/9 + 1/6 = -1/4, \end{aligned}$$

ou seja, $\varphi'_{\{1/4\}}(0) = 0$. Para $A = \{1/3\}$,

$$\begin{aligned} \rho_{\{1/3\}}(0, \pi_x) - \rho_{\{1/3\}}(1, \pi_x) &= \sum_{\theta \in \Theta} [L_{\{1/3\}}(0, \theta, 0) - L_{\{1/3\}}(1, \theta, 0)] \pi(\theta|0) \\ &\propto 3/12 - (1/3)2/9 + 1/6 = 37/108, \end{aligned}$$

ou seja, $\varphi'_{\{1/3\}}(0) = 1$. Para $A = \{1/2\}$,

$$\begin{aligned} \rho_{\{1/2\}}(0, \pi_x) - \rho_{\{1/2\}}(1, \pi_x) &= \sum_{\theta \in \Theta} [L_{\{1/2\}}(0, \theta, 0) - L_{\{1/2\}}(1, \theta, 0)] \pi(\theta|0) \\ &\propto 3/12 + 2/9 - (1/3)1/6 = 1/3, \end{aligned}$$

ou seja, $\varphi'_{\{1/2\}}(0) = 1$.

Analogamente, quando observamos $x = 1$, temos $\varphi'_{\{1/4\}}(1) = 1$, $\varphi'_{\{1/3\}}(1) = 1$ e $\varphi'_{\{1/2\}}(1) = 0$. Podemos ver que a classe de testes $\{\varphi'_A\}_{A \in \sigma(\Theta)}$ gerada pela família de funções de perda $\{L_A\}_{A \in \sigma(\Theta)}$ coincide com a classe $\{\varphi_A\}_{A \in \sigma(\Theta)}$ gerada pelo EMV, para os testes de hipóteses simples, e como consequência do Corolário, essas classes são idênticas.

Um outro possível resultado que podemos obter do Teorema 3.2 é que para um espaço amostral finito e uma medida de probabilidade π tal que $\pi(\theta)$ é positivo, para todo $\theta \in \Theta$, um teste de Bayes é admissível. Portanto, nessas condições, todos os testes de classes geradas por estimadores de máxima verossimilhança são admissíveis.

Corolário 3.2. Nas condições do Teorema 3.2 (a), se π é tal que $\pi(\theta)$ é positivo, para todo $\theta \in \Theta$, então, para todo $A \in \mathcal{P}(\Theta)$, o teste φ_A é admissível considerando a perda L_A .

Demonstração. Pelo Teorema 3.2 (a), para cada $A \in \sigma(\Theta)$, existe $L_A : \{0, 1\} \times \Theta \times \mathcal{X} \rightarrow \mathbb{R}^+$ tal que φ_A é teste de Bayes contra π .

Para um teste φ_A qualquer para testar $\theta \in A$, seja $R(\theta, \varphi_A)$ o risco frequentista de φ_A quando o valor do parâmetro é θ , isto é,

$$R(\theta, \varphi_A) = \sum_{x \in \mathcal{X}} L_A(\varphi_A(x), \theta, x) P(X = x | \theta).$$

Suponhamos, por absurdo, que φ_A é inadmissível. Então, existe teste φ_0 tal que

$$R(\theta, \varphi_0) \leq R(\theta, \varphi_A) \text{ para todo } \theta \in \Theta \text{ e}$$

$$R(\theta_0, \varphi_0) < R(\theta_0, \varphi_A) \text{ para algum } \theta_0 \in \Theta.$$

Nessas condições, o risco do teste φ_0 é tal que $\rho_A(\varphi_0, \pi_x) = \sum_{\theta \in \Theta} R(\theta, \varphi_0) \pi(\theta)$ que é igual a

$$R(\theta_0, \varphi_0) \pi(\theta_0) + \sum_{\theta \in \Theta \setminus \{\theta_0\}} R(\theta, \varphi_0) \pi(\theta) < R(\theta_0, \varphi_A) \pi(\theta_0) + \sum_{\theta \in \Theta \setminus \{\theta_0\}} R(\theta, \varphi_A) \pi(\theta) = \rho_A(\varphi_A, \pi_x),$$

o risco do teste φ_A . Logo, φ_A não é teste de Bayes, pois $\rho_A(\varphi_0, \pi_x) < \rho_A(\varphi_A, \pi_x)$, o que é absurdo. Portanto, φ_A é admissível considerando a perda L_A . $\square$

Pelo resultado acima, os testes das classes estabelecidas nos exemplos 3.1 e 3.2 são todos admissíveis.

3.2 Princípio da Verossimilhança

Podemos ainda verificar se esses testes gerados por um estimador atendem ao Princípio da Verossimilhança (Berger (1985)). Utilizando a notação de Wechsler *et al.* (2008), consideramos um experimento E como sendo uma tripla $(\mathcal{X}, \Theta, \mathbb{P})$, onde $\mathcal{X}$ é o espaço amostral dos possíveis resultados, Θ espaço paramétrico e $\mathbb{P} = \{P_\theta : \theta \in \Theta\}$ família de distribuições para X indexada em Θ .

Consideramos uma distribuição a priori π para θ e $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$ uma função de perda a partir da qual é derivado o estimador de Bayes $W : \mathcal{X} \rightarrow \Theta$ para θ contra π , o qual suporemos único. Por fim, suponhamos que testes de hipóteses simultâneos são conduzidos segundo a classe de testes gerada pelo estimador W , $\{\varphi_A^{(E)}\}_{A \in \mathcal{P}(\Theta)}$

Sejam $E_1 = (\mathcal{X}, \Theta, \mathbb{P}_1)$ e $E_2 = (\mathcal{Y}, \Theta, \mathbb{P}_2)$ dois experimentos acerca de θ . Dizemos que as realizações $(E_1, \mathcal{X})$ e $(E_2, \mathcal{Y})$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ são equivalentes $((E_1, \mathcal{X}) \cong (E_2, \mathcal{Y}))$, na notação de Wechsler *et al.* (2008)) se, para todo $A \in \mathcal{P}(\Theta)$,

$$\varphi_A^{(E_1)}(x) = \varphi_A^{(E_2)}(y),$$

isto é, se os resultados de todos os testes de hipóteses coincidem ao conduzir o experimento E_1 e observar $x \in \mathcal{X}$ e ao conduzir o experimento E_2 e observar $y \in \mathcal{Y}$.

Nessas condições, temos o seguinte resultado.

Teorema 3.3. *Seja Θ finito e π a distribuição a priori para θ . Seja $\{\varphi_A\}_{A \in \sigma(\Theta)}$ a classe de testes gerada pelo estimador de Bayes $W : \mathcal{X} \rightarrow \Theta$ para θ obtido a partir de $L : \Theta \times \Theta \rightarrow \mathbb{R}^+$ contra π . Isto é, para cada $A \in \mathcal{P}(\Theta)$, $\varphi_A(x) = 1 - I_A(W(x))$, $x \in \mathcal{X}$.*

Sejam $E_1 = (\mathcal{X}_1, \Theta, \mathbb{P}_1)$ e $E_2 = (\mathcal{X}_2, \Theta, \mathbb{P}_2)$ dois experimentos com $\mathcal{X}_1$ e $\mathcal{X}_2$ enumeráveis, e $x_1 \in \mathcal{X}_1$ e $x_2 \in \mathcal{X}_2$ tais que $P(X_1 = x_1 | \theta) = h(x_1, x_2) P(X_2 = x_2 | \theta)$ para todo $\theta \in \Theta$ com $h(x_1, x_2)$ positivo. Então, $(E_1, x_1) \cong (E_2, x_2)$.

Demonstração. Sejam $x_1 \in \chi_1$ e $x_2 \in \chi_2$ tais que $P(x_1|\theta) = h(x_1, x_2)P(x_2|\theta)$ para todo $\theta \in \Theta$. Seja $W_i : \chi_i \rightarrow \Theta$ o estimador de Bayes para θ obtido no experimento E_i , $i = 1, 2$, considerando perda L contra π .

Para $x_i \in \chi_i$, $E[L(W_i(x_i), \theta)|x_i] < E[L(d, \theta)|x_i]$, para todo $d \neq W_i(x_i)$, supondo W_i único, $i = 1, 2$.

Assim,

$$\sum_{\theta \in \Theta} L(W_1(x_1), \theta) \pi(\theta|x_1) < \sum_{\theta \in \Theta} L(d, \theta) \pi(\theta|x_1), \text{ para todo } d \neq W_1(x_1),$$

então

$$\sum_{\theta \in \Theta} L(W_1(x_1), \theta) P(x_1|\theta) \pi(\theta) < \sum_{\theta \in \Theta} L(d, \theta) P(x_1|\theta) \pi(\theta), \text{ para todo } d \neq W_1(x_1),$$

que implica

$$\sum_{\theta \in \Theta} L(W_1(x_1), \theta) h(x_1, x_2) P(x_2|\theta) \pi(\theta) < \sum_{\theta \in \Theta} L(d, \theta) h(x_1, x_2) P(x_2|\theta) \pi(\theta), \text{ para todo } d \neq W_1(x_1).$$

Logo

$$\sum_{\theta \in \Theta} L(W_1(x_1), \theta) \pi(\theta|x_2) < \sum_{\theta \in \Theta} L(d, \theta) \pi(\theta|x_2), \text{ para todo } d \neq W_1(x_1). \quad (3.5)$$

Por outro lado, pela unicidade de $W_2 : \chi_2 \rightarrow \Theta$, temos

$$\sum_{\theta \in \Theta} L(W_2(x_2), \theta) \pi(\theta|x_2) < \sum_{\theta \in \Theta} L(d, \theta) \pi(\theta|x_2), \text{ para todo } d \neq W_2(x_2). \quad (3.6)$$

De (3.5) e (3.6), segue que $W_1(x_1) = W_2(x_2)$. Logo, para todo $A \in \mathcal{P}(\Theta)$, $\phi_A^{E_1}(x_1) = 1 - I_A(W_1(x_1)) = 1 - I_A(W_2(x_2)) = \phi_A^{E_2}(x_2)$ e, portanto, $(E_1, x_1) \cong (E_2, x_2)$. $\square$

Deste modo, os testes das classes estabelecidas nos exemplos 3.1 e 3.2 também respeitam o Princípio da Verossimilhança.

3.3 Desiderata D2 versus Otimalidade

Diante dos resultados dos capítulos 2 e 3, verificamos que a imposição da desiderata D2 é muito restritiva, ou seja, ao exigirmos que testes simultâneos atendam à desiderata D2, fazemos com que tais testes sejam sempre testes gerados por estimadores (Teorema 3.1), o que sugere que para atender critérios de otimalidades usuais em testes de hipóteses simultâneos, devemos utilizar menos restrições lógicas para os testes. No entanto, apesar de bastante simples, os testes gerados por estimadores são logicamente consistentes no sentido de atenderem D2 admissíveis e respeitam o Princípio da Verossimilhança, que são propriedades desejáveis para testes de hipóteses.

Capítulo 4

Conclusões

4.1 Considerações Finais

Neste trabalho, sob a perspectiva bayesiana, relembramos os conceitos de classes de testes de hipóteses que generaliza um teste de hipótese e expandimos este conceito para funções de perda, definindo famílias de funções de perda, que são também generalizações das funções de perda. Utilizando estas famílias, definimos classes de testes de Bayes geradas por estas famílias, que são de uma certa maneira um conjunto de testes de Bayes indexado pela possível hipótese de interesse.

Partindo do conceito de classes de testes, relembramos algumas propriedades lógicas e conjuntos dessas propriedades (chamados de desideratas em [Izbicki \(2010\)](#)) que desejaríamos que classes de testes atendessem.

Na sequência, derivamos teoremas com condições necessárias e suficientes que as famílias de funções de perda devem atender (Teoremas [2.2](#), [2.3](#) e [2.4](#)), para que classes de testes de Bayes tenham cada uma das propriedades lógicas isoladamente e quando possível em conjunto. Também foi mostrado que não existem famílias de funções de perda que fazem com que classes de testes de Bayes geradas por ela atendam a desiderata D2 para toda distribuição a priori (Teorema [2.5](#)). No entanto, sob algumas condições, uma classe de testes que atende a desiderata D2 é sempre uma classe de testes de Bayes de um decisor bayesiano (Teorema [2.6](#)).

No Capítulo [3](#) utilizamos o resultado obtido em [Izbicki \(2010\)](#) de que uma classe de testes que atende D2 sempre pode ser escrita como uma classe de testes gerada por um estimador (Teorema [3.1](#)), para mostrar que sob algumas condições para a família de funções de perda, é equivalente avaliar, para cada $x \in \mathcal{X}$, se a estimativa de Bayes está no conjunto da hipótese nula ou se o teste de Bayes aceita essa hipótese. Em certo sentido, mostramos sob quais condições uma tomada de decisão baseada em um estimador de Bayes é similar a tomada de decisão baseada em um teste de Bayes (Teorema [3.2](#)).

Na sequência, baseado neste último resultado, mostramos que sob algumas condições, avaliar uma estimativa de máxima verossimilhança, para cada $x \in \mathcal{X}$, implica em avaliar se a estimativa de Bayes está no conjunto da hipótese nula (Corolário [3.1](#)) e ainda mostramos que os testes dessa classe são todos admissíveis (Corolário [3.2](#)).

Por fim, mostramos que os testes de uma classe de testes gerada por um estimador atendem ao Princípio da Verossimilhança (Teorema [3.3](#)).

Observando os resultados obtidos nos Capítulos [3](#), vemos que a imposição da desiderata D2 faz com que classes de testes que atendam a essa condição sejam sempre classes de testes geradas por estimadores, o que indica que condições lógicas muito restritivas podem ter esse efeito na otimalidade dos testes. No

entanto, mostramos que esses testes tem algumas propriedades interessantes.

4.2 Sugestões para Pesquisas Futuras

- Caracterizar outras propriedades lógicas para classes de testes de hipóteses.
- Avaliar mais a fundo a relação entre a admissibilidade com as propriedades lógicas para testes simultâneos.
- Relacionar outros princípios importantes, como o Princípio da Verossimilhança, com essas propriedades lógicas.
- Generalizar os resultados obtidos para espaços paramétricos não enumeráveis.
- Estudar as propriedades lógicas para problemas de decisão com decisões sequenciais.

Referências Bibliográficas

- Bera e Premaratne(2001)** A. K. Bera e G. Premaratne. General hypothesis testing. *A companion to theoretical econometrics*. Citado na pág. 1
- Berger(1980)** J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics, 2º edição. Citado na pág. 2, 4
- Berger(1985)** J.O. Berger. In defense of the likelihood principle: Axiomatics and coherency. *Bayesian Statistics*, 2:33–66. Citado na pág. 43, 48
- Betz e Levin(1982)** M.A. Betz e J.R. Levin. Coherent analysis-of-variance hypothesis-testing strategies: A general model. *Journal of Educational Statistics*, 7:193–206. Citado na pág. 1, 8
- DeGroot(1970)** M. DeGroot. *Optimal Statistical Decisions*. Garland Science, New York, 4 edição. Citado na pág. 2, 3, 4, 5
- Deza e Deza(2012)** M. Deza e E. Deza. *Encyclopedia of Distances*, volume 2. Springer. Citado na pág. 13
- Evans e Jang(2011)** M. Evans e G.H. Jang. Inferences from prior-based loss functions, Abril 2011. URL <http://arxiv.org/abs/1104.3258>. Citado na pág. 43
- Ferguson(1967)** T. S. Ferguson. *Mathematical statistics: A decision theoretic approach*, volume 7. Academic Press New York. Citado na pág. 2, 5
- Finner e Strassburger(2002)** H. Finner e K. Strassburger. The partitioning principle: a powerful tool in multiple decision theory. *Annals of Statistics*, 30(4):1194–1213. Citado na pág. 13
- Fossaluza(2008)** V. Fossaluza. Testes de hipóteses em eleições majoritárias. Dissertação de Mestrado, IME-USP, Brasil. Citado na pág. 1, 8, 9
- French e Insua(2000)** S. French e D. R. Insua. *Statistical decision theory*. Kendall's Library of Statistics Series. Arnold. ISBN 9780340614600. Citado na pág. 5

- Gabriel(1969)** K.R. Gabriel. Simultaneous test procedures - some theory of multiple comparisons. *The Annals of Mathematical Statistics*, 40:224–250. Citado na pág. [1](#), [8](#), [13](#)
- Higham(1998)** Nicholas J. Higham. *Handbook of Writing for the Mathematical Sciences*. SIAM: Society for Industrial and Applied Mathematics, second edição. Citado na pág.
- Hommel e Bretz(2008)** G. Hommel e F. Bretz. Aesthetics and power considerations in multiple testing – a contradiction? *Biometrical Journal*, 50(5):657–666. Citado na pág. [1](#), [8](#)
- Howson e Urbach(2005)** C. Howson e P. Urbach. *Scientific Reasoning: The Bayesian Approach*. Open Court, 3ª edição. Citado na pág. [5](#)
- Izbicki(2010)** R. Izbicki. Classes de testes de hipóteses. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil. Citado na pág. [2](#), [7](#), [8](#), [9](#), [10](#), [14](#), [15](#), [41](#), [51](#)
- Izbicki e Esteves(2014)** R. Izbicki e L.G. Esteves. Logical consistency in simultaneous test procedures. Relatório técnico, IME-USP. Citado na pág. [2](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#)
- James(1996)** B. R. James. *Probabilidade: Um curso de nível intermediário*. LTC, Rio de Janeiro, 2 edição. Citado na pág. [37](#)
- Jeffreys(1939)** H. Jeffreys. *Theory of Probability*. Cambridge University Press, Cambridge. Citado na pág. [5](#)
- Johnson(2013)** V.E. Johnson. Uniformly most powerful bayesian tests. *The Annals of Statistics*, 41(4): 1716–1741. Citado na pág. [5](#)
- Lavine e Schervish(1999)** M. Lavine e M. J. Schervish. Bayes factors: What they are and what they are not. *The American Statistician*, 53(2):119–122. Citado na pág. [1](#), [8](#)
- Lehmann(1957a)** E.L. Lehmann. A theory of some multiple decision problems i. *The Annals of Mathematical Statistics*, 28(1):1–25. Citado na pág. [1](#)
- Lehmann(1957b)** E.L. Lehmann. A theory of some multiple decision problems ii. *The Annals of Mathematical Statistics*, 28(1):547–572. Citado na pág. [11](#)
- Lima(1977)** E. L. Lima. *Espaços Métricos*. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 5 edição. Citado na pág. [13](#), [32](#)
- Madrugá et al.(2001)** M.R. Madrugá, L.G. Esteves e S. Wechsler. On the bayesianity of pereira-stern tests. *Sociedad de Estadística e Investigación Operativa Test*, 10:291–299. Citado na pág. [6](#), [7](#), [43](#)

- Mayo e Spanos(2006)** D. G. Mayo e A. Spanos. Severe testing as a basic concept in a neyman–pearson philosophy of induction. *The British Journal for the Philosophy of Science*, 57(2):323–357. Citado na pág. 4
- Neumann e Morgenstern(1944)** J. Von Neumann e O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton university press. Citado na pág. 2, 4
- Parmigiani e Inoue(2009)** G. Parmigiani e L. Inoue. *Decision Theory: Principles and Approaches*. Wiley, Baltimore, first edição. Citado na pág. 2
- Pearson(1900)** K. Pearson. On the criterion that a given system of deviations form the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, Series 5*, 50:157–175. Citado na pág. 1
- Pereira e Stern(1999)** C.A.B. Pereira e J.M. Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy*, 1:104–115. Citado na pág. 6
- Pereira et al.(2008)** C.A.B. Pereira, J.M. Stern e S. Wechsler. Can a significance test be genuinely bayesian? *Bayesian Analysis*, 3:79–100. Citado na pág. 6
- Popper(1959)** K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson e Co., London, 6ª edição. Citado na pág. 4
- Raiffa e Schlaifer(1961)** H. Raiffa e R. Schlaifer. *Applied statistical decision theory*. Studies in managerial economics. Division of Research, Graduate School of Business Administration, Harvard University. ISBN 9780875840178. URL <http://books.google.com.br/books?id=wPBLAAAAMAAJ>. Citado na pág. 2
- Raviv(2013)** E. Raviv. On p-value, Fevereiro 2013. URL <http://eranraviv.com/blog/on-p-value/>. Citado na pág. 8
- Robert(2007)** C. P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, volume 2. Springer. Citado na pág. 10, 12
- Rom e Holland(1995)** D.M. Rom e B. Holland. A new closed multiple testing procedure for hierarchical families of hypotheses. *Journal of Statistical Planning and Inference*, 46:265–275. Citado na pág. 1, 8
- Schervish(1995)** M. J. Schervish. *Theory of Statistics*. Springer Series in Statistics, 3º edição. Citado na pág. 4, 10
- Schervish(1996)** M.J. Schervish. P values: What they are and what they are not. *The American Statistician*, 50:203–206. Citado na pág. 1, 8, 9

- Silva(2010)** G. M. Silva. Monotonicidade em testes de hipóteses. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil. Citado na pág. [1](#), [6](#), [8](#), [9](#), [10](#)
- Stern(2008)** J.M. Stern. *Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses*. 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, São Paulo. Citado na pág. [6](#)
- Vivaldi(2013)** F. Vivaldi. *Mathematical writing for undergraduate students*. School of Mathematical Sciences. Queen Mary, University of London. Citado na pág.
- Wald(1945)** A. Wald. Statistical decision functions which minimize the maximum risk. *The Annals of Mathematics*, 46(2):265–280. Citado na pág. [2](#)
- Wald(1949)** A. Wald. Statistical decision functions. *Annals of Mathematical Statistics*, 20(2):165–205. Citado na pág. [2](#)
- Wechsler et al.(2008)** S. Wechsler, C.A.B. Pereira e P.C. Marques F. Birnbaum’s theorem redux. *AIP Conference Proceedings*, 1073:96–100. Citado na pág. [48](#)
- Zellner(1971)** A. Zellner. *An Introduction to Bayesian inference in econometrics*. Wiley series in probability and mathematical statistics, New York. Citado na pág. [9](#)