

Utilidade para testes de significância

Nathália Demétrio Vasconcelos Moura

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

Área de Concentração: Estatística
Orientador: Prof. Dr. Sergio Wechsler

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, fevereiro de 2014

Agradecimentos

Poucas vezes na vida temos a oportunidade de registrar a gratidão que sentimos por todos aqueles que participaram de alguma forma dos momentos importantes da nossa vida. Aproveitando desta chance tão especial, lá vai:

À minha mãe que transformou todas as dificuldades em incentivos, mostrando sempre o quanto tudo é possível. Muito obrigada por me ensinar a encarar a vida de maneira otimista, digna e independente.

À toda a minha família, que me ajudou das mais diferentes formas, ao meu irmão Paulinho com quem construo uma história desde sempre, aqueles que completam o meu entedimento de família: Lucinha, Tina, Gui, Raffa, aos que me permitiram conhecer as formas mais puras de amor: meu avô e Muff, e Gustavo, vulgo Guga, que além de todas as anteriores possui a minha risada preferida no mundo inteiro. Em especial Paulo Moura, pessoa em que as definições de amizade, companheirismos, carinho e amor sempre se confundem de maneira matematicamente perfeita (ou seria computacionalmente?).

À todas as pessoas maravilhosas que tornaram a experiência do mestrado uma etapa nostálgica: colegas do curso de verão, disciplinas, churrascos e cafés. Gostaria de agradecer à Roberto Mangui pelo apoio incondicional, a Andrés Arias e Felipe Behring por conversas, brincadeiras e madrugadas mas principalmente pela amizade, que me sinto extremamente feliz de saber que é recíproca, e as “meninas do CRUSP” que transformaram alguns metros quadrados em um amado e barulhento lar.

Aos professores, exemplos e amigos: Alexandre Patriota e Victor Fossaluzza por todo o imenso aprendizado que permitiram por meio de palestras, debates, ou mesmo conversas de corredor.

E acima de qualquer pessoa, entidade ou agremiação alvi-negra, à Sergio Wechsler, sem dúvida devo a ele tudo de melhor que alcancei no âmbito acadêmico, me mostrou a importância de questionar, me apoiou de forma incondicional durante todas as curvas não previstas da minha vida, que não foram poucas, um orientador no sentido mais bonito que se possa atribuir à palavra. Se nosso objetivo em vida for deixar uma marca no mundo, ele sem dúvida o fez, pelo meio que mais importa: pessoas! E por isso serei eternamente grata.

“E a gente agradece, com gentileza e compaixão por todos os caminhantes, porque somente quem caminha sabe o valor, o tamanho, a conquista, de que é feita a história de cada único passo.”

Ana Jácomo

Resumo

Moura, N. D. V. **Utilidade para testes de significância**. 2014. 34 f. Dissertação (Mestrado)- Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

Este trabalho discute testes de significância sob a ótica da teoria da decisão. Diferentemente dos testes de Neyman-Pearson, testes de significância usualmente estão desprovidos de funções de perda. Funções de perda para testes de significância de Fisher, Jeffreys e FBST são analisadas.

Palavras-chave: inferência bayesiana, teoria da decisão, teste de hipóteses precisas, função de perda, testes de significância, FBST.

Abstract

Moura, N. D. V. **Utility for significance tests**. 2014. 34 f. Dissertação (Mestrado)- Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

This work approaches significances tests from a decision-theoretic viewpoint. Contraty to Neyman-Pearson tests, significance tests usually lack loss functions. Loss function for Fisher, Jeffreys, and FBST are analyzed.

Keywords: Bayesian inference, decision theory, precise null hypothesis, loss function, significance tests, FBST.

Sumário

1	Introdução	1
2	Teoria da Decisão	3
3	Testes de Hipóteses	7
3.1	Fisher x p -valor	7
3.2	Neyman-Pearson x Razão de Verossimilhanças	9
3.3	Bayes x Medidas Condicionais	10
4	Hipóteses Precisas	11
4.1	Testes de Significância	11
4.2	Razão de Verossimilhanças	11
4.3	Teste de Jeffreys	12
4.4	FBST	12
5	Funções de Perda	15
5.1	Testes de Significância	16
5.2	Razão de Verossimilhanças	16
5.3	Teste de Jeffreys	16
5.4	FBST	17
6	Interpretações	19
6.1	Testes de Significância	19
6.2	FBST	19
7	Conclusão	21
	Referências Bibliográficas	23

Capítulo 1

Introdução

Extraír informação de um conjunto de observações é o objetivo básico da inferência, existindo diferentes abordagens para tal: verossimilancista, estrutural, *fuzzy*, entre outras. A mais difundida diz respeito a compreensão frequentista, em que avaliamos o espaço definido por todas as possíveis amostras, procurando por meio do limite das frequências relativas, em uma série indefinida de replicações do evento estimar parâmetros, denominação dada a entidades desconhecidas de interesse. Porém, o fato do parâmetro ser definido como um valor fixo acarreta uma série de restrições às análises, resumidamente,

- Definição circular de probabilidade, uma vez que utiliza o conceito de independência e este é definido a partir da própria probabilidade;
- Tratar a probabilidade como um valor constante, apesar de desconhecido, situação que, salvo em alguns contextos “controlados” como simulações ou jogos de azar, é pouco razoável.
- Utilização da Lei dos Grandes Números como corroboração teórica para a definição de probabilidade, quando a aplicabilidade do teorema depende de uma sequência infinita do experimento, sob as mesmas condições, circunstância pouco plausível em termos práticos;
- Impossibilidade de modelar um experimento que foi interrompido por razões indefinidas antes do previsto;
- Violação do Princípio da Verossimilhança, apesar do Princípio da Condicionalidade e da Suficiência serem básicos¹;
- Uso indiscriminado da aleatorização, tornando a “moeda” responsável por decisões que caberiam ao analista;
- Aceitação da existência e busca de objetividade, mesmo que esta não possa ser definida.
- Análise da amostra, do conhecimento do especialista e das implicações práticas feitas separadamente, inviabilizando a consideração todo.

Como opção podemos ampliar a análise para o contexto bayesiano, em que ao encarar o parâmetro como uma entidade aleatória latente, ou seja, não observável, temos uma abordagem mais harmoniosa com o modo com que a incerteza é intuída no dia a dia. Mesmo considerando contradições probabilísticas possivelmente declaradas em situações informais, o fato é que os axiomas de coerência (*Dutch Book Argument*), pressupostos para tal abordagem (Kadane, 2011), mesmo que definidos em um contexto linear, implicando necessariamente em aditividade, são: simples, interpretáveis e intuitivos (Berry, 1997).

“Inevitably, students (and everyone else except for statisticians) give an inverse or Bayesian twist to frequentist measures such as confidence intervals and P values.” pág. 242.

¹{Princípio da Verossimilhança} \Leftrightarrow {Princípio da Condicionalidade} \wedge {Princípio da Suficiência}

De modo que, acessando a informação disponível na amostra por meio da verossimilhança, teremos o teorema de Bayes especificando o mecanismo segundo o qual os dados irão se conectar ao conhecimento a priori,

$$\textit{probabilidade a posteriori} \propto \textit{probabilidade a priori} \times \textit{verossimilhança},$$

e considerando a ótica do movimento subjetivo bayesiano², associado a DeFinetti, Savage, Lindley e DeGroot, dispomos não só da incorporação do conhecimento subjetivo prévio à análise, mas também a possibilidade de modelar qualquer entidade desconhecida do ponto de vista probabilístico. Existe também a opção “objetiva da inferência bayesiana” [sic], que apesar de tratar do mesmo processo teórico, não associa subjetividade à priori, deixando de lado o conhecimento prévio como fonte de informação e, por tal razão, não é considerada neste trabalho.

No contexto em que as evidências amostrais darão suporte à decisões, podemos realizar uma análise ainda mais abrangente, incorporando também a expectativa do pesquisador em relação aos resultados, por meio da Teoria da Decisão Estatística. Tal metodologia visa um resultado prescritivo quanto à ação a ser tomada, utilizando para tal não só as informações objetivas e subjetivas da incerteza associada, mas também a utilidade do analista quanto às possíveis consequências da ação.

Considerando ainda a incerteza que envolve a modelagem de eventos, é natural analisar conjecturas visando a determinação de padrões, no contexto estatístico realizamos tal estudo por meio dos testes de hipóteses. Devido a ampla aplicabilidade em termos práticos, os testes de hipóteses configuram papel decisivo em muitas áreas do conhecimento, tornando as discussões filosóficas associadas aos seus fundamentos ainda mais relevantes, particularmente no que diz respeito aos testes com hipóteses precisas, cenário igualmente debatido. Neste contexto, revisaremos a literatura dos testes de hipóteses, interpretando as principais leituras sob a ótica da teoria da decisão por meio das funções de perda associadas a hipóteses precisas, tendo sempre como base o entendimento bayesiano subjetivista.

²As leis de probabilidade são a estrutura segundo a qual um indivíduo coerente deve expressar sua incerteza, tendo sua necessidade teórica garantida pelo Teorema Fundamental de B. DeFinetti (Loschi e Wechsler, 2002).

Capítulo 2

Teoria da Decisão

A todo o momento somos expostos a decisões, desde questões simples e individuais, como sair de casa levando, ou não, um guarda-chuva, até ações significativamente complexas, com o poder de influenciar toda uma sociedade, como uma medida política por exemplo. De modo que se torna necessário dispormos de ferramentas que nos auxiliem a optar pela melhor ação de modo racional, levando em consideração nossos objetivos, conhecimento e preferências. Neste contexto temos disponível a **teoria da decisão**, que considera que a ação de interesse assuma valores no espaço de ações \mathcal{D} , e dependa do resultado de uma entidade que envolva incerteza, definida no conjunto Ω , denominado “estados da natureza”, de modo que, por meio das preferências do agente decisor, obtemos a escolha ótima segundo um cenário especificado. Formalizando tais preferências a partir da **função de perda**, que trata a penalidade que o agente incorre ao optar por d quando o estado da natureza obtido foi θ , teremos,

$$\begin{aligned} L : \mathcal{D} \times \Omega &\rightarrow \mathbb{R} \\ (d, \theta) &\mapsto L(d, \theta). \end{aligned}$$

No caso de uma abordagem mais otimista, considerando ganhos por exemplo, podemos trabalhar com a função de utilidade $U(d, \Theta)$, definida como o inverso aditivo da função de perda, ou seja, $U(d, \Theta) + L(d, \Theta) = 0$. Vale ressaltar que a definição acima pode ser redefinida para os reais positivos, simplificando cálculos, bastando para tal trabalharmos com a função de perda regret, que considera o infimo da função de perda para cada possível valor de Θ , sem qualquer alteração em relação as decisões ótimas (DeGroot, 1970),

$$L^*(d, \Theta) = L(d, \Theta) - \inf_{d \in \mathcal{D}} L(d, \Theta),$$

tendo como vantagem a garantia de que a penalidade será nula para algum valor dos estados da natureza. Note que a intensão da função de perda é contemplar as preferências do agente decisor em relação as possíveis consequências, assim como a distribuição a priori se refere ao grau de incerteza do indivíduo quanto ao comportamento de uma variável aleatória ¹.

Para determinar a ação a ser tomada, procuramos pela decisão tal que a penalidade associada seja mínima, porém uma vez que a escolha precisa ser feita antes de conhecermos o estado da natureza que será assumido, atribuímos ao conjunto Ω o conceito de parâmetro e, uma vez visto como variável aleatória, estimamos o seu comportamento por meio da compreensão que o agente decisor possui sobre o espaço paramétrico, ou seja o seu conhecimento a priori, que sendo representado segundo uma distribuição de probabilidade $\pi(\theta)$ permite que avaliemos a perda média para cada decisão d . Supondo que tanto a probabilidade a priori quanto a penalidade satisfaçam as suposições de coerência, definimos a **perda esperada a priori** $r_\pi(d)$, ou função de risco bayesiano a priori,

¹No contexto objetivo da abordagem bayesiana a priori é especificada segundo outros princípios: amostras passadas, máxima entropia, invariância, etc.

de modo que,

$$r_\pi(d) = \mathbb{E}_\pi[L(d, \Theta)] = \int_\Omega L(d, \theta)\pi(\theta)d\theta. \quad (2.1)$$

Note que no caso dos estados da natureza serem discretos basta substituir a integral pelo respectivo somatório. Por fim, admitindo que a integral apresentada em (2.1) é finita para cada $d \in \mathcal{D}$, teremos que a melhor decisão será aquela cuja perda a priori é minimizada para todos os elementos do espaço de decisões, ou seja,

Definição 1 *Uma decisão d^* será ótima, em relação a distribuição π de Θ , se,*

$$d_\pi^* = \arg \min_{d \in \mathcal{D}} r_\pi(d).$$

Adicionalmente teremos que qualquer ação obtida segundo a definição 1 será denominada de **decisão de Bayes contra a distribuição $\pi(\theta)$** , e o risco relacionado é qualificado como **risco de Bayes a priori**, sendo possível que nenhuma decisão do espaço \mathcal{D} seja uma decisão de Bayes.

Consideraremos agora que o agente decisor tenha acesso a amostra de uma variável aleatória X definida segundo um modelo estatístico paramétrico, ou seja, um quinteto $(\chi, \mathcal{A}, \mathcal{F}, \Omega, \pi)$, em que χ é o espaço amostral, \mathcal{A} é a sigma-álgebra de subconjuntos de χ , e \mathcal{F} especifica uma classe de distribuições sobre \mathcal{A} , sendo estas indexadas segundo o espaço paramétrico Ω , e π a densidade especificada a priori em relação a Ω , seguindo colocações anteriores. Tendo que o acesso a tal amostra no contexto da teoria da decisão receberá a denominação de **teoria da decisão estatística**.

Assumindo que para $\Theta \in \Omega$ exista a distribuição condicional de X dado $\Theta = \theta$, $f(x|\theta)$, e que esta possa ser especificada para cada possível valor do espaço paramétrico, a decisão será dada segundo uma **regra de decisão** δ , assim para cada x existirá uma estratégia associada,

$$\begin{aligned} \delta : \chi &\rightarrow \mathcal{D} \\ x &\mapsto \delta(x). \end{aligned}$$

Assim a função de perda será re-escrita em termos da regra de decisão e, conseqüentemente, a perda esperada a posteriori para δ irá possuir dois elementos aleatórios,

$$R_\pi(\delta(X)) = \mathbb{E}_\pi[L(\delta(X), \Theta)] = \int_\Omega \int_\chi L(\delta(x), \theta) f(x|\theta) \pi(\theta) dx d\theta. \quad (2.2)$$

Logo, similarmente ao caso inicial, procuramos a estratégia que minimize o risco a posteriori, ou seja,

Definição 2 *Uma regra de decisão δ^* será ótima, com respeito a $\pi(\theta)$ se,*

$$\delta_\pi^* = \arg \min_{\delta \in \mathcal{D}} R_\pi(\delta).$$

E, assumindo que podemos inverter a ordem de integração em (2.2), o que seguramente poderá ser feito se a função de perda for não-negativa e limitada, teremos que obter δ_π^* segundo a definição acima, será equivalente a encontrar a regra de decisão que, para cada valor de X , minimiza a penalidade média a posteriori, ou seja minimiza a integral,

$$\int_\chi L(\delta(x), \theta) f(\theta|x) \pi(\theta) d\theta,$$

resultando que o procedimento ótimo será $\delta_\pi^* = \mathbb{E}_\Theta[L(\delta(X), \Theta)|X = x]$. Tal simplificação é particularmente útil considerando a complexidade de calcular a integral em χ da equação (2.2), para todas as possíveis regras de decisão associadas ao problema. Note que temos o mesmo raciocínio do caso sem amostra, porém com a decisão ótima obtida em função da posteriori $\pi(\theta|x)$. Em termos gerais a função de risco indica qual penalidade deverá ser adotada, enquanto que a função de perda especifica a regra de decisão. Adicionalmente, no caso do procedimento ser obtido segundo a definição 2, tal ação será denominada **regra de decisão de Bayes contra a distribuição $\pi(\theta|x)$** e o risco relacionado, que pode não ser único, **risco de bayes a posteriori**.

Podemos também abordar o problema segundo a perspectiva da inferência frequentista, em que a amostra X é o único componente aleatório. O fato de não encararmos o parâmetro Θ como variável aleatória torna tal abordagem mais restrita, uma vez que para encontrar a decisão ótima é necessário limitar o espaço de decisões entre as não-viesadas, ou seja consideramos apenas o conjunto $\Delta' = \{\delta(X) \in \Delta : \mathbb{E}[\delta(X)|\theta] = \theta, \forall \theta \in \Omega\}$, uma vez que a esperança obtida em (2.2) não é avaliada no espaço paramétrico. Nesta leitura o estimador mais usual procura a decisão tal que a perda máxima seja minimizada, sendo por tal motivo conhecido como decisão minimax (podemos trabalhar também com a maximização do ganho mínimo, neste caso maximin).

Capítulo 3

Testes de Hipóteses

Tendo por objetivo indicar o cenário mais plausível dentre uma coleção de conjecturas que envolvam incerteza, e considerando para tal as informações disponíveis, os testes de hipóteses são procedimentos usualmente aplicados para duas premissas, de modo que configurem uma partição do espaço paramétrico (ou de um subespaço de interesse), apesar de podermos generalizar o mesmo raciocínio para cenários mais abrangentes. Nomeadas como hipótese nula e alternativa, teremos,

$$\begin{aligned}H_0 &: \Theta \in \Omega_0 \\H_1 &: \Theta \in \Omega_1,\end{aligned}$$

em que $\Omega = \Omega_0 \cup \Omega_1$, $\Omega_0 \cap \Omega_1 = \emptyset$, $\Omega_0 \neq \emptyset$ e $\Omega_1 \neq \emptyset$.

Em termos teóricos os procedimentos irão ser especificados por uma função φ , definida em $\psi = \{\varphi : \chi \rightarrow \{0, 1\}\}$, de modo que decidimos por H_0 se $\varphi = 0$, e H_1 caso contrário, sendo o valor de φ determinado por meio da região crítica (*RC*), também denominada região de rejeição, que formalmente será definida como $\varphi^{-1}(\{1\}) = \{x \in \chi : \varphi(x) = 1\}$ e, similarmente, $\varphi^{-1}(\{0\})$ será nomeada região de aceitação, sendo usual ter a hipótese nula como referência para as decisões. No contexto frequentista teremos a região de rejeição definida a partir de uma estatística, designada estatística de teste, enquanto que na leitura bayesiana dependerá também do conhecimento a priori do especialista. Em termos gerais tanto a especificação da região de rejeição quanto o modo de interpretar os resultados são suficientes para distinguir as abordagens existentes na literatura.

Adicionalmente ao especificar a hipótese, existem dois tipos de erros que podemos cometer, o erro tipo I, caso em que incorretamente indicamos a hipótese alternativa como verdadeira e, similarmente, o erro do tipo II, referente a hipótese nula, de modo que o decréscimo de um erro implica no aumento do outro. Formalmente teremos,

$$\begin{aligned}\mathbb{P}[\text{Erro do tipo I}] &= \alpha(\varphi) = \mathbb{P}[\varphi(X) = 1 | \Omega_0] \\ \mathbb{P}[\text{Erro do tipo II}] &= \beta(\varphi) = \mathbb{P}[\varphi(X) = 0 | \Omega_1].\end{aligned}$$

Em relação às características probabilísticas dos testes trabalhamos com a função poder, ou simplesmente poder do teste, que associa a probabilidade de rejeitar a hipótese nula a cada valor de Θ , ou seja, $\gamma_\varphi(\theta) = \mathbb{P}[\varphi = 1 | \theta]$. Por meio desta função, definimos o tamanho do teste como o limite superior da função poder, avaliada apenas para os valores de $\Theta \in \Omega_0$, formalmente, $\alpha = \sup_{\theta \in \Theta_0} \gamma_\varphi(\theta)$. Por fim, chamaremos o valor α_0 de nível de significância da família ψ^* se este for o limitante superior para os tamanhos de teste do referido conjunto, ou seja, $\psi^* = \{\varphi \in \psi : \gamma_\varphi(\theta_0) \leq \alpha_0\}$, para $0 < \alpha_0 < 1$. Segue a descrição das principais leituras dos testes de hipóteses.

3.1 Fisher x *p*-valor

A leitura mais disseminada no contexto da inferência clássica diz respeito a abordagem de Sir Karl Popper, difundido no meio estatístico por Ronald Fisher, segundo a qual focamos em apenas uma das hipóteses e como conclusões possíveis dispomos da rejeição da hipótese especificada ou a falta de evidências para tal, ou seja, trabalhamos com um raciocínio indutivo em que independente

da quantidade de evidências a favor da hipótese testada, esta nunca deverá ser aceita. Tal perspectiva pode ser justificada pela leitura da amostra como uma ferramenta limitada quando comparada com a possível abrangência do fenômeno estudado, de modo que adotamos uma postura cautelosa considerando apenas a eliminação de premissas, ao invés da verificação destas, segundo o próprio Fisher (1935),

“Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis”. pág. 18.

Em termos aplicados o nível de significância observado, introduzido por Karl Pearson, também denominado p -valor, se apresenta como a ferramenta ideal para tal leitura, uma vez que retrata a probabilidade de obtermos amostras menos prováveis do que a observada sob H_0 . Para aplicar esta abordagem fixamos o experimento que fornecerá os dados, e examinamos as amostras tão extremas quanto a que foi observada, considerando para tal a ordenação do espaço amostral por meio de uma distribuição que supomos conhecida e é indexada de acordo com a hipótese testada. Formalmente,

Definição 3 *Seja $T(X)$ uma estatística, tal que seus valores extremos indiquem evidência contra H_0 , se observarmos $X = x$ com o correspondente $t = T(x)$ o p -valor será dado por,*

$$\alpha^* = \mathbb{P}_{\Omega_0}(T(X) \geq t),$$

Apesar de não se fazer necessária a indicação de qual hipótese estamos tratando, já que nesta leitura trabalhamos apenas com uma, é usual utilizar H_0 como referência e, por tal razão, procuramos especificar a hipótese nula de modo que represente o pior caso possível. Em termos aplicados, temos que valores “pequenos” do p -valor indicam a discordância entre o modelo especificado e os dados, formalmente, rejeitamos a hipótese se $\varphi^{-1}(\{1\}) = \{x \in \chi : \alpha^* < \alpha^+\}$, em que α^+ é a referência de quão extrema a amostra pode ser, tal valor deve ser discriminado pelo analista, conforme argumento do próprio Fisher (1956): “In choosing the grounds upon which a general hypothesis should be rejected, personal judgement may and should properly be exercised”, pág. 50.

Apesar do p -valor, ou *tail area*, ser um instrumental coêrente com a proposta do Fisher, além de amplamente utilizado, esta medida possui várias características indesejáveis, dentre as mais citadas temos:

- Sua grandeza depende do tamanho da amostra, assim como sua interpretação depende da verossimilhança assumida e da regra de parada utilizada (Wagenmakers, 2007), coexistindo diferentes relações na literatura (Pereira e Wechsler, 1993);
- O significado da rejeição da hipótese indica apenas que há algo de errado com o modelo utilizado porém sem especificar qual é o problema, como exemplo o caso da rejeição em um teste pode ser referente à distribuição assumida, independência, homocedasticidade, etc (Berger e Delampady, 1987);
- A falta de coerência entre a rejeição de uma hipótese e seus subconjuntos, uma vez que não se trata de uma medida de avaliação do espaço paramétrico, mas apenas do subespaço definido pela hipótese testada (Patriota, 2013);
- No contexto de duas hipóteses definidas segundo uma partição de Θ , podemos conjecturar exemplos em que a falta de exigência quanto a incorporação da hipótese alternativa na análise resulta em conclusões contraditórias (Pereira e Wechsler, 1993);
- Dificuldade de interpretação, uma vez sua definição condicional, $\mathbb{P}(x|H_0)$, é sumariamente intuita como a probabilidade condicional $\mathbb{P}(H_0|x)$ (van der Pas, 2010).

Devido a diferença de abordagem desta classe de testes (procurar evidências contra H_0) quando comparada aos testes de hipóteses usuais (aceitar ou rejeitar as hipóteses), esta leitura é usualmente denominada **testes de significância**, sendo particularmente interessante no contexto em que a contradição da hipótese testada não é absoluta.

3.2 Neyman-Pearson x Razão de Verossimilhanças

Temos ainda a perspectiva defendida por Jerzy Neyman e Egon Pearson (N-P) como uma das principais abordagens da inferência frequentista, porém de modo parcial, uma vez que o raciocínio dos autores é desenvolvido considerando a aceitação ou rejeição de todas as hipóteses envolvidas, a partir de um nível de significância pré-especificado, diferentemente das aplicações usuais, em que prevalece a conclusão de apenas rejeitar a hipótese, seguindo as recomendações dos testes de significância.

Para introduzir tal teoria iremos inicialmente considerar que o teste seja constituído por hipóteses simples, ou seja, $H_0 : \Theta = \theta_0$ e $H_1 : \Theta = \theta_1$, de modo que a especificação da região crítica será dada pela razão das funções de verossimilhança avaliadas nos respectivos subespaços de Ω , $\lambda(x) = f(x|\theta_0)/f(x|\theta_1)$. Posteriormente, dado a impossibilidade de controlarmos simultaneamente os dois tipos de erros envolvidos, iremos restringir a análise à família ψ^* , composta pelos testes de nível de significância α_0 , $0 < \alpha_0 < 1$, de modo que,

$$\varphi^*(x) = \begin{cases} 1 & \text{se } \lambda(x) \leq k \\ 0 & \text{se } \lambda(x) \geq k, \end{cases} \quad (3.1)$$

com $k \geq 0$, será tal que $\gamma_{\varphi^*}(\theta_1) \geq \gamma_{\varphi}(\theta_1)$, $\forall \varphi \in \psi^*$ e, pelo Lema de Neyman Pearson, este será o teste mais poderoso de nível α_0 .

Para o caso de uma das hipóteses ser composta, digamos H_1 , restringimos o domínio para os testes uniformemente mais poderosos (*UMP*), em que é necessário que o teste seja o mais poderoso para cada possível valor da hipótese composta, ou seja, a região crítica não deverá depender de nenhum particular valor de $\Theta \in \Omega_1$. O mesmo raciocínio é requerido para tratarmos de hipóteses compostas unilaterais, entretanto para casos mais gerais como hipóteses bilaterais, ou um vetor de parâmetros desconhecidos, trabalhamos com os testes da razão de verossimilhanças generalizada (*TRVG*), definida a partir dos supremos das verossimilhanças, tendo conclusões limitadas a uma família mais restrita, em que apenas os procedimentos tais que o erro tipo I é igual a α_0 , ou seja, $\psi^{**} = \{\varphi \in \psi : \gamma_{\varphi^{**}}(\theta_1) = \gamma_{\varphi}(\theta_1)\}$. Logo o teste $\varphi^{**}(x) \in \psi^{**}$, com a região crítica,

$$RC = \left\{ x : \frac{\sup_{\Theta \in \Omega_0} L(\theta|x)}{\sup_{\Theta \in \Omega_1} L(\theta|x)} \leq k, \right.$$

$k \geq 0$, será o teste uniformemente mais poderoso (*UMP*), ou seja, possuirá a função poder uniformemente superior aos demais procedimentos da família ψ^{**} . Em termos gerais para ampliarmos as análises sempre será necessário restringir o domínio dos testes, para só então termos determinadas propriedades garantidas, como exemplo teremos os testes não viesados uniformemente mais poderosos, testes invariantes e uniformemente mais poderosos, família exponencial, etc.

O fato de especificarmos o teste com ênfase no valor do erro tipo I, está relacionado com o conceito de repetibilidade da amostra, sugerindo que cometeremos a inapropriada rejeição de H_0 em $\alpha\%$ das decisões, e por tal razão especificamos valores pequenos para o α . Adicionalmente, o resultado do procedimento não será uma inferência numérica, mas sim um comportamento, no caso o de aceitar ou rejeitar a hipótese de referência. Assim como na leitura anterior, existem pontos controversos para a análise de N-P,

- Após a obtenção dos resultados do experimento a relevância do α se torna questionável, já que reflete amostras que não foram e não virão a ser observadas, além da própria conclusão que diz respeito a uma sequência de experimentos, quando o real interesse reside em tomar decisões pontuais;
- Adotar valores pré-especificados para o nível de significância, tipicamente $\alpha = 0,05$ e $\beta < 0,20$, sem qualquer crítica quanto a natureza do evento ou o modelo assumido, tendo como única justificativa a incidência literária;
- Em termos aplicados é possível conjecturar exemplos em que, simplesmente por aumentar o

tamanho da amostra, há um incremento na relação entre os tipos de erros, sendo possível obter até a inversão da relação especificada inicialmente para α e β ;

- Para variáveis discretas existe ainda a eminência de trabalhar com testes aleatorizados, em que definimos aleatoriamente sobre os pontos referentes a vizinhança da região de rejeição fazerem parte deste, ou não. Caso em que atribuímos para a aleatoriedade o cargo de agente decisor, em nome apenas de uma precisão matemática;
- Dificuldade em especificar a distribuição amostral, sendo conduta usual recorrer a distribuições assintóticas.

É interessante notar que tal teoria não se trata de uma concorrente aos testes de significância, uma vez que aborda o problema segundo uma perspectiva diferente. Apesar disso, as análises frequentistas usualmente realizam os cálculos e adotam as propriedades da metodologia de Neyman-Pearson, e a conclusão é dada segundo a perspectiva de prova por contradição do Fisher, não aceitando a hipótese nula e tomando tal decisão de acordo com a comparação do nível de significância pré-especificado e o p -valor obtido, o que se mostra confuso já que são abordagens desenvolvidas segundo princípios diferentes e visando respostas distintas, e de fato podem resultar em conclusões antagônicas (Christensen, 2005), (Berger, 2003).

Por fim temos que na abordagem frequentista, o fato de tratarmos apenas os dados como relevantes, pode ser considerado uma limitação, conforme argumentado em Jeffreys (1961) “An hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred.” pág. 385.

3.3 Bayes x Medidas Condicionais

Em contraste com as teorias frequentistas usuais, que pedem a definição de procedimentos *ad hoc*, na inferência bayesiana tratamos simplesmente de avaliar a probabilidade da hipótese, de modo direto, de modo que no contexto em que existe acesso ao resultado de um experimento, teremos $\mathbb{P}(H_0|x) = \int_{\Omega_0} \pi(\theta|x) d\theta$. Porém, visando englobar a situação como um todo, podemos trabalhar com a razão conhecida como *fator de Bayes* (Jeffreys, 1961),

$$B_{H_0} = \frac{\mathbb{P}(H_0|x)}{\mathbb{P}(H_1|x)} = \frac{\mathbb{P}(x|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(x|H_1) \cdot \mathbb{P}(H_1)},$$

e, portanto, $B_{H_0} > 1$ significa que, no contexto da amostra observada, H_0 é relativamente mais plausível que H_1 . Neste caso, para a rejeição da hipótese nula, não será suficiente que esta seja pouco provável, temos que obter uma hipótese alternativa com probabilidade ainda menor, trazendo certa afinidade com a estatística de N-P, porém de modo mais eficaz, uma vez que aqui tratamos das probabilidades diretamente. Para esta leitura existem várias caracterizações que permitem “regras de bolso” que, desprovidas de crítica, podem originar críticas similares as pré-especificação, usuais na compreensão frequentista. De modo geral, teremos como conclusão aceitar ou rejeitar a hipótese de referência de modo subjetivo.

Note que nesta leitura não damos prioridade para a hipótese nula, apenas avaliamos as relações existentes entre as hipóteses. Adicionalmente, vale ressaltar que testes bayesiano em geral podem retornar resultados incoerentes se trabalharmos com distribuições inapropriadas, um caso relevante de ser citado é o famoso paradoxo de Lindley (1957), segundo o qual os resultados encontrados por meio do p -valor frequentista e a probabilidade a posteriori da hipótese nula apontam para ações opostas, entretanto, apesar de tal caso ser referido como um paradoxo, a diferença entre os resultados deve ser vista como consequência dos fundamentos diferentes e não como um paradoxo em si.

Capítulo 4

Hipóteses Precisas

No caso de um teste em que o espaço Ω_0 possua uma dimensão inferior ao de Ω_1 , teremos $H_0 : \Theta = \theta_0$ versus $H_1 : \Theta \neq \theta_0$ se o espaço paramétrico for definido na reta real. Sendo tal abordagem relevante para responder questões mais pontuais, como o descarte de covariáveis de um modelo por exemplo, cenário usualmente recorrente no contexto experimental, entretanto, a aceitação deste subcaso não é absoluta, já que alguns autores consideram as hipóteses precisas como utópicas em termos práticos, o argumento pode ser resumido por meio do seguinte parágrafo de Tukey (1991),

“All we know about the world teaches us that the effects of A and B are always different - in some decimal place - for any A and B. Thus asking ‘Are the effects different?’ is foolish.” pág. 100,

à medida que, segundo esta visão, a questão deveria ser “em qual direção o efeito de A difere do efeito de B” tendo como opções: positivo, negativo ou incerto. Apesar desta argumentação ser válida, entendemos a especificação das hipóteses como um problema de decisão, em que nossa penalidade em considerar uma hipótese precisa, no lugar de uma diferença de “tamanho ϵ ”, por exemplo, é pequena e, portanto, pode ser tratada como uma ação cabível.

4.1 Testes de Significância

Para os procedimentos definidos segundo a perspectiva do Fisher a única mudança diz respeito a avaliação do modelo segundo uma menor subespaço, o que dá particular significado para a filosofia de Popper, já a questão da restritividade que a hipótese possui em relação ao espaço como um todo é evidenciada. Um exemplo recorrente na literatura diz respeito à hipótese “todos os cisnes são brancos”, em que Popper argumenta que independentemente da quantidade de cisnes brancos observados, não podemos corroborar tal afirmação, enquanto que com apenas um cisne preto a conjectura poderia ser classificada como falsa.

4.2 Razão de Verossimilhanças

Em termos matemáticos temos que, conforme apresentado no capítulo 3.2, para o caso de uma hipótese precisa versus composta, trabalhamos com os elementos definidos no espaço dos testes uniformemente mais poderosos. Quanto a compreensão desta abordagem temos, em oposição a interpretação dada por Fisher, a célebre frase “it is not only not right, it is not even wrong!”, popularmente atribuída ao físico Wolfgang Pauli, salientando um ponto chave no contexto dos testes de hipótese, a falta de uma leitura universal, dado a variedade de situações possíveis.

4.3 Teste de Jeffreys

No contexto de hipóteses precisas, temos que a probabilidade a posteriori do subconjunto Ω_0 é nula, de modo que, salvo os casos em que a priori é discreta ou degenerada no ponto θ_0 , não podemos trabalhar com as abordagens bayesianas usuais, apresentadas no capítulo 3.3. Devido a tal dificuldade, Jeffreys (1961) reescreveu a distribuição a priori como uma mistura de uma massa pontual atribuída a $\theta = \theta_0$ e a, já especificada, densidade $\pi(\theta)$ para os demais valores do espaço paramétrico, por meio de uma função convexa, formalmente teremos a priori dada por,

$$\pi^*(\theta) = \begin{cases} \zeta & \text{se } \Theta = \theta_0 \\ (1 - \zeta)\pi(\theta) & \text{se } \Theta \neq \theta_0, \end{cases}$$

em que $0 < \zeta < 1$. Logo, a probabilidade a posteriori será,

$$\pi^*(\theta|x) = \begin{cases} \zeta f(x|\theta_0)/f(x) & \text{se } \Theta = \theta_0 \\ (1 - \zeta) \frac{f(x|\theta)\pi(\theta)}{f(x)} & \text{se } \Theta \neq \theta_0. \end{cases}$$

De modo que a rejeição ou aceitação das hipóteses, a partir deste ponto, pode ser feita seguindo os raciocínios usuais para avaliação das hipóteses a posteriori. Apesar do teste de Jeffreys ser difundido a ponto de, em alguns livros, ser nomeado como “teste de Bayes para hipóteses precisas”, o fato é que ao atribuir uma massa de probabilidade positiva ζ para a hipótese nula estamos diminuindo a chance de ocorrência dos demais pontos do espaço paramétrico na proporção de $(1 - \zeta)$, o que não seria um problema se esta fosse de fato a opinião do analista quanto ao comportamento do evento, o que usualmente não é o caso. Como resultado, se torna necessária uma maior quantidade de evidências contra a hipótese nula para possibilitar a rejeição. Adicionalmente, a natureza contínua do parâmetro é completamente distorcida, uma vez que a densidade sob um espaço paramétrico contínuo é um sistema de ordenação, em que atribuímos probabilidade nula para cada ponto.

4.4 FBST

Considerando a falta de uma solução bayesiana que, no contexto das hipóteses precisas, preservasse a natureza contínua do parâmetro Pereira e Stern (1999) introduziram o FBST (Full Bayesian Significance Test), uma medida de evidência a favor da hipótese nula, desenvolvida com base no princípio da menor surpresa¹, que apesar de permitir a aplicação em espaços gerais, tem como principal propósito os testes de significância para hipóteses precisas. Considerando como ordenação a probabilidade a posteriori dos pontos do espaço paramétrico, com densidade inferior a probabilidade máxima obtida em H_0 , região denominada por conjunto tangente à hipótese nula ou ainda conjunto de credibilidade, a probabilidade pontual nula é respeitada. Formalmente,

Definição 4 Considerando o interesse em testar o subconjunto Ω_0 , e tendo que foi obtida o conjunto tangente $T(x) = \{\theta : \pi(\theta|x) > \sup_{\Theta_0} \pi(\theta|x)\}$. A medida de evidência desenvolvida por Pereira e Stern é definida como

$$EV(\Theta_0, x) = 1 - \mathbb{P}(\Theta \in T(x)|x) = \int_{T(x)} \pi(\theta|x)d\theta.$$

De modo que para valores altos de $EV(\Omega_0, x)$, teremos θ_0 entre os pontos mais prováveis a posteriori e, portanto, favorecendo a hipótese nula. Também tratado como *e-value*, tal procedimento apresenta como principais vantagens,

- Lógica intuitiva com interpretação geométrica, além de se tratar de uma probabilidade no espaço paramétrico de interesse;

¹Construção de um subconjunto do espaço paramétrico, cujos pontos são mais inesperados do que o que foi observado.

- Desconstrução do paradoxo de Lindley;
- Aplicabilidade mesmo sem uma distribuição a priori disponível, bastando para tal substituir a posteriori pela verossimilhança normalizada (Pereira *et al.*, 2008);
- Conformidade com o princípio da verossimilhança, além de ser consistente e invariante sob transformações um-a-um do parâmetro (Madrugá *et al.*, 2003);
- O fato de ser um procedimento exato, sem a necessidade de aproximações assintóticas.

Adicionalmente, tal medida, também conhecida como e -valor se trata de uma decisão de Bayes, ou seja, é o resultado da minimização de uma função de perda, conforme será discutido no próximo capítulo.

Capítulo 5

Funções de Perda

A variedade de abordagens existentes para os testes de hipóteses justifica a ampla literatura voltada para a comparação entre as diferentes filosofias encontradas (Berger e Sellke, 1987), (Diniz *et al.*, 2012). Porém, como cada abordagem é construída segundo estruturas e objetivos diferentes, variando inclusive os espaços em que são avaliados (espaço amostral ou paramétrico), sempre haverá muitos pontos discutíveis, limitando as conclusões aos cenários especificados.

Uma alternativa interessante é trabalhar com as diferentes propostas segundo uma única linguagem, no caso a teoria da decisão, permitindo não só uma comparação mais realista dos desempenhos mas principalmente a identificação de quais raciocínios são condizentes com cada ferramenta. Adicionalmente, temos que minimizar a penalidade esperada é a única maneira de especificarmos uma decisão compatível com os axiomas de coerência.

Para permitir tal relação iremos considerar o espaço de decisão \mathcal{D} dado por $\{d_0, d_1\}$, em que d_i denota ação de aceitar a hipótese $H_i : \Theta \in \Omega_i$, com $i \in \{0, 1\}$ e, conforme apresentado no capítulo 3, devemos especificar uma penalidade para cada par $\mathcal{D} \times \Omega$. Considerando inicialmente uma função de perda em que apenas os erros são relevantes, e seguindo a notação utilizada em Parmigiani e Inoue (2009), teremos,

	Ω_0	Ω_1
d_0	0	L_0
d_1	L_1	0

Tabela 5.1: Função de perda genérica para testes de hipóteses.

De modo que para uma amostra x teremos a regra de decisão $\delta(x)$ definida no contexto binário. Portanto o risco será dado por,

$$\begin{aligned} R_\pi(\delta(X)) &= \int_\Omega \int_{\mathcal{X}} L(\delta(x), \theta) \pi(x|\theta) dx d\theta \\ &= \begin{cases} L_1 \mathbb{P}(\delta(X) = d_1 | \theta) & \text{se } \Theta \in \Omega_0 \\ L_0 \mathbb{P}(\delta(X) = d_0 | \theta) & \text{se } \Theta \in \Omega_1, \end{cases} \end{aligned}$$

e, conseqüentemente, a perda esperada a posteriori,

$$\mathbb{E}_\Theta[L(\delta(X), \Theta) | x] = \int_\Omega L(\delta(x), \theta) \pi(\theta | x) d\theta = \begin{cases} L_0(1 - \pi(\Theta \in \Omega_0 | x)) & \text{se } d = d_0 \\ L_1 \pi(\Theta \in \Omega_0 | x) & \text{se } d = d_1. \end{cases}$$

Logo a regra de Bayes será tal que decidimos por d_0 para o caso de,

$$\frac{L_0}{L_1} < \frac{\pi(\Theta \in \Omega_0 | x)}{1 - \pi(\Theta \in \Omega_0 | x)}.$$

Resultando no já conhecido Fator de Bayes, entretanto, no caso de caracterizarmos a penali-

dade segundo um diferente critério, possivelmente obteríamos outra decisão ótima, coerente com as novas peculiaridades da função de perda. E, com base neste fato, iremos fazer o caminho reverso, procurando identificar as penalidades que atribuem às abordagens apresentadas a denominação de decisão de Bayes, discutindo suas particularidades em termos aplicados. Tal destaque se deve principalmente à abrangência que a combinação da inferência bayesiana com a teoria da decisão possibilitam, tanto do ponto de vista matemático quanto subjetivo, além da minimização da penalidade esperada ser a única maneira de garantir que a decisão será compatível com os axiomas de coerência.

5.1 Testes de Significância

Visando especificar uma penalidade que refletisse os testes de significância sob a perspectiva de Fisher, Rice (2010) decompôs a decisão em duas etapas de modo que, inicialmente, escolhemos entre reportar ou não os resultados obtidos, representando tais ações por $h = 0$ e $h = 1$, respectivamente, e no caso de decidirmos pela divulgação das conclusões, o valor d será apresentado, sendo este definido no espaço paramétrico. Deste modo teremos a função de perda dada segundo uma combinação de penalidades.

Quando optamos por $h = 1$, o que implica em fornecer resultados por meio do estimador d , avaliaremos o custo segundo a perda quadrática usual, e no caso de decidir por $h = 0$, a penalidade será dada em termos das conclusões que deixaram de ser conhecidas e de quão informativas estas seriam, neste caso, representamos tal prejuízo proporcionalmente à distância entre θ e θ_0 . Observando que ambas penalidades possuem a mesma escala de medida, podemos considerar a combinação destas segundo uma combinação linear, adicionalmente o agente decisor irá especificar a sua opinião quanto a relação entre as consequências por meio de um fator γ . Assim a função de perda será,

$$L_\gamma = \gamma^{1/2} (1 - h) (\theta_0 - \theta)^2 + \gamma^{-1/2} h (\hat{\theta} - \theta)^2. \quad (5.1)$$

compatível com a compreensão dos testes de significância de Fisher. A regra de Bayes obtida para (5.1) será tal que se $h = 1$ então teremos a decisão $\mathbb{E}[\Theta|X]$, e para $h = 0$ qualquer decisão d obterá a mesma utilidade. Logo, reportaremos resultados apenas se,

$$\frac{\mathbb{E}[\Theta - \theta_0 | x]^2}{\text{Var}[\Theta | x]} \geq \frac{1 - \gamma}{\gamma}.$$

Como nota, é interessante observar que modelar a filosofia “fisheriana” segundo a perspectiva da teoria da decisão pode ser vista uma ironia teórica, considerando o pouco crédito que Ronald Fisher atribuía à tal ciência, argumentando ser uma técnica inútil para cientistas e aproveitável apenas para comerciantes, conjectura pouco próxima a realidade.

5.2 Razão de Verossimilhanças

Originalmente N-P especificaram o procedimento de modo tal que consideramos os erros I e II separadamente, entretanto, podemos construir um teste de modo que para uma constante positiva k a combinação linear dos erros seja minimizada, mantendo as mesmas propriedades apresentadas. Logo para o teste definido em (3.1), teremos que,

$$\gamma_\varphi^*(\theta_0) + k(1 - \gamma_\varphi^*(\theta_1)) \leq \gamma_\varphi(\theta_0) + k(1 - \gamma_\varphi(\theta_1)), \quad \forall \varphi \in \psi,$$

ou seja, $(\alpha(\varphi^*) + k\beta(\varphi^*)) \leq \alpha(\varphi) + k\beta(\varphi)$.

5.3 Teste de Jeffreys

No que tange a inferência bayesiana temos uma relação quase direta, principalmente por ambas ciências incorporarem meios não experimentais na teoria para resolução dos problemas. Entretanto

a solução apresentada por Jeffreys não apresenta qualquer relação com a teoria utilizada para as demais abordagens conforme discutido

5.4 FBST

Apesar do FBST fazer uso apenas da informação contida na densidade a posteriori, e por tal razão ser classificada como *full bayesian*, foi apenas com o trabalho de [Madruga et al. \(2001\)](#) que esta medida foi analisada do ponto de vista da teoria da decisão, sendo obtida através da minimização da função de perda,

Definição 5 A função de perda L , definida sob o espaço $\mathcal{D} \times \Omega \times \chi$, é dada por $L(\text{Rejeitar } H_0, \Theta) = a[1 - \mathbb{I}\{\Theta \in T(x)\}]$ e $L(\text{Aceitar } H_0, \Theta) = b + c[\mathbb{I}\{\Theta \in T(x)\}]$, sendo $a, b, c > 0$.

Note que, diferentemente da teoria clássica em que a penalidade é definida em termos do espaço de decisões e do espaço paramétrico, aqui consideramos uma classe mais abrangente em que a amostra observada também é incorporada à penalidade, uma vez que considera o espaço tangente, que é definido a partir da densidade a posteriori, ou seja, a partir da amostra, de modo que teremos $L(d, \theta, x)$.

Verificando que o FBST é resultado da minimização da função de perda 5, o risco a posteriori tendo como decisão aceitar a hipótese é dado por,

$$\begin{aligned} \mathbb{E}_\pi[L(d_0, \Theta)|x] &= \mathbb{E}_\pi[L(\text{Aceitar } H_0, \Theta)|x] \\ &= \int_\Omega [b + c\mathbb{I}\{\theta \in T(x)\}]\pi(\theta|x)d\theta \\ &= b + c(1 - EV(\Omega_0, x)). \end{aligned}$$

Enquanto que no caso de rejeitar H_0 , o risco será,

$$\begin{aligned} \mathbb{E}_\pi[L(d_1, \theta)|x] &= \mathbb{E}_\pi[L(\text{Rejeitar } H_0, \Theta)|x] \\ &= \int_\Omega a[1 - \mathbb{I}\{\theta \in T(x)\}]\pi(\theta|x)d\theta \\ &= a EV(\Omega_0, x). \end{aligned}$$

Portanto a aceitação da hipótese H_0 se dará se, e somente se,

$$EV(\Omega_0, x) > \frac{b+c}{a+c}.$$

Uma questão interessante apresentada pelos autores da referida função de perda é que a inserção de uma massa de probabilidade positiva na hipótese nula, nos moldes do teste de Jeffreys, faz com que não obtenhamos o FBST por meio da minimização da função de perda apresentada em (5). Como característica perdemos a separabilidade.

Capítulo 6

Interpretações

A teoria da decisão nos permite quantificar questões subjetivas, tão importantes quanto a amostra em si, através da função de perda. Sendo possível alinhar a penalidade utilizada não só com as expectativas, mas com a postura desejada como um todo, e por tal razão é necessário conhecer as características e raciocínios condizentes com a função de perda utilizada. Neste contexto iremos discutir as peculiaridades das penalidades apresentadas.

6.1 Testes de Significância

No contexto inferencial, em que o objetivo é aprender sobre o parâmetro, não reportar resultados soa pouco apropriado, entretanto, para determinados cenários tal postura pode ser interessante, ou simplesmente uma filosofia adotada pelo analista, principalmente pelo fato das técnicas estatísticas serem aplicadas nas mais diferentes áreas do conhecimento, de modo que se torna de suma importância oferecer técnicas que abranjam diferentes objetivos e circunstâncias.

E visando cobrir esta possibilidade temos na estrutura apresentada no capítulo 5.1 tanto o ceticismo em relação a possíveis conclusões sobre o parâmetro quanto a opinião do agente decisor em relação as consequências envolvidas, incorporados na função de perda por meio do coeficiente γ . Para tal, supomos uma correspondência entre as unidades de imprecisão da estimativa d e a perda inerente a falta de resultados, considerando a distância de θ_0 , sendo a ideia básica relacionar a inacurácia da estimativa reportada com o efeito que a falta de conclusões pode resultar, procurando o ponto de equilíbrio entre esses dois tipos de perdas.

Tomando o p -valor como a medida de referência para os testes de significância, podemos ainda agregar algumas das suas características, por exemplo, o fato da evidência contra a hipótese nula crescer com o tamanho da amostra pode ser englobado na distribuição priori, considerando para tal uma densidade em que θ e θ_0 estejam próximos, além da relação com o tamanho da amostra, que aqui também será refletida na função de perda, já que para uma amostra grande estaremos pouco inclinados a concluir nada (e bastante para isso especificar um valor alto para γ).

É interessante notar que ao atribuir à H_0 o pior cenário, desejando rejeitar esta, estamos indiretamente tratando do conceito de utilidade.

6.2 FBST

Apesar das penalidades usualmente contemplarem a relação entre as decisões e os estados da natureza, existem outros fatores que podem influenciar as nossas preferências, como por exemplo a amostra, elemento capaz de acarretar reflexos de caráter psicológico, financeiro e até mesmo social. De modo que contemplar tal entidade na função de perda permite uma compreensão plena da situação analisada. A penalidade associada ao FBST é um exemplo desta classe de funções, bem como [Bernardo e Smith \(1994\)](#) que formalizaram uma penalidade para a região de credibilidade com maior densidade a posteriori.

Ainda que soe inusitado especificar a função de perda com base na densidade a posteriori, talvez por culturalmente “utilizar a amostra mais de uma vez” ser uma prática não recomendada, apesar de utilizarmos, por exemplo, a função de perda total ¹, ou o próprio cálculo da função de risco em que trabalhamos com a já especificada posteriori $\pi(\theta|x)$. Em relação a este tópico, a questão correta não é por que usar, mas sim o porquê não usar, afinal não há restrições teóricas relacionadas a tal conduta.

De modo geral teremos que quanto maior a quantidade de pontos no conjunto tangente, menos provável será a região a qual H_0 pertence e, conseqüentemente, maior será a penalidade incorrida caso a decisão seja por H_0 . Em contraponto, para o caso de rejeitarmos a hipótese nula, a penalidade será proporcional a quantidade de pontos pertencentes ao complementar de $T(x)$. Em relação aos riscos, temos que as situações limites são intuitivas, de modo que, independente das constantes da função de perda, teremos que d_0 terá a penalidade mínima se $EV(\Omega_0, x) = 1$, ou seja, quando não houverem pontos no conjunto tangente, enquanto que para $EV(\Omega_0, x) = 0$ a função de perda relacionada a d_1 será sempre menor igual a de d_0 . Para os casos não extremos, é necessário discutir as constantes da função de perda, visando um raciocínio mais didático, iremos considerar inicialmente uma versão mais simples, em que $b = 0$ implicando a aceitação da hipótese nula se $EV(\Omega_0, x) > c/(a + c)$, de modo que a constante a irá refletir a preferência do analista quanto a aceitação de H_0 e, similarmente, a constante c irá contemplar a rejeição de H_0 .

Considerando um valor positivo para a constante b , teremos uma maior penalização para o erro tipo II, de modo que,

- Se $a < b$ o agente decisor irá sempre rejeitar a hipótese nula; (d_1 é preferível e $T(x)$ é embaraçador)
- Se $a \gg b$ e c for “pequeno” então teremos a aceitação de H_0 com maior facilidade. (d_0 é preferível e $T(x)$ não é embaraçador)

Para refletir cortes de 0,90 e 0,95 seria necessário que as constantes respeitassem as relações $9a - 10b = c$ e $19a - 20b = c$, respectivamente.

¹Função de perda que tem o tamanho da amostra incorporado a penalidade original.

Capítulo 7

Conclusão

Neste trabalho discutimos as principais leituras dos testes de hipóteses da inferência estatística, podendo observar que os testes frequentistas não devem ser vistos como opções para um mesmo problema, tampouco comparáveis, uma vez que abordam circunstâncias diferentes, sendo a abordagem bayesiana usual a técnica mais interessante. Porém, ao revisar tais metodologias no contexto da teoria da decisão, por meio das respectivas penalidades associadas, pudemos contornar as principais críticas quanto a compreensão frequentista, além de acoplar aspectos subjetivos à análise e, principalmente, unificar as leituras segundo uma única abordagem, evitando todas as incoerências práticas atualmente encontradas no meio acadêmico.

Em relação a perspectiva de Fisher, a filosofia por trás da falta de conclusões passaram a ser suportadas pelo fator de Bayes, útil para determinados cenários, o que é particularmente interessante, já que nesta leitura temos o alinhamento entre as expectativas e as conclusões. Enquanto que a medida de significância, *e-value*, pode passar a ser aplicada no contexto da teoria da decisão, incorporando todas as nuances subjetivas envolvidas.

Para trabalhos futuros, entender melhor a proposta de outras escolas, como exemplo da lógica *fuzzy* em que os axiomas de coerência não são uma premissa, e estudar a relação entre a especificação das penalidades a priori e a posteriori.

Referências Bibliográficas

- Berger (2003)** James O. Berger. Could Fisher, Jeffreys and Neyman have agreed on testing? *Statistical Science*, 18(1):1–12. Citado na pág. [10](#)
- Berger e Delampady (1987)** James O. Berger e Mohan Delampady. Testing precise hypotheses. *Statistical Science*, 2(2):317–352. Citado na pág. [8](#)
- Berger e Sellke (1987)** James O. Berger e Thomas Sellke. Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association*, 82(397):112–122. Citado na pág. [15](#)
- Bernardo e Smith (1994)** José M. Bernardo e Adrian F. M. Smith. *Bayesian Theory*. John Wiley & Sons, New York. ISBN 0-471-92416-4. Citado na pág. [19](#)
- Berry (1997)** Donald A. Berry. Teaching elementary bayesian statistics with real applications in science. *The American Statistician*, 51(3):241–246. Citado na pág. [1](#)
- Christensen (2005)** Ronald Christensen. Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59:121–126. URL <http://EconPapers.repec.org/RePEc:bes:amstat:v:59:y:2005:m:may:p:121-126>. Citado na pág. [10](#)
- DeGroot (1970)** Morris Herman DeGroot. *Optimal Statistical Decisions*. McGraw-Hill. ISBN 0-471-68029-X. Citado na pág. [3](#)
- Diniz et al. (2012)** Márcio Diniz, Carlos Alberto de Bragança Pereira, Adriano Polpo, Julio Michael Stern e Sergio Wechsler. Relationship between bayesian and frequentist significance indices. *International Journal for Uncertainty Quantification*, 2(2):161–172. Citado na pág. [15](#)
- Fisher (1935)** Ronald Aylmer Fisher. *The design of experiments*. Edinburgh: Oliver and Boyd. Citado na pág. [8](#)
- Fisher (1956)** Ronald Aylmer Fisher. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd. Citado na pág. [8](#)
- Jeffreys (1961)** Harold Jeffreys. *The theory of probability*. Oxford University Press., 3 ed. Citado na pág. [10](#), [12](#)
- Kadane (2011)** Joseph B. Kadane. *Principles of Uncertainty*. Chapman & Hall. Citado na pág. [1](#)
- Lindley (1957)** Dennis V. Lindley. A statistical paradox. *Biometrika*, 44(1-2):187–192. Citado na pág. [10](#)
- Loschi e Wechsler (2002)** Rosangela H. Loschi e Sergio Wechsler. Coherence, bayes's theorem and posterior distributions. *Brazilian Journal of Probability and Statistics*, 16:169–185. Citado na pág. [2](#)
- Madruza et al. (2001)** Maria Regina Madruza, Luis Gustavo Esteves e Sergio Wechsler. On the bayesianity of pereira-stern tests. *Sociedad de Estadística e Investigación Operativa*, 10:291–299. Citado na pág. [17](#)

- Madruga et al. (2003)** Maria Regina Madruga, Carlos Alberto de Bragança Pereira e Julio Michael Stern. Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference*, 117:185–198. Citado na pág. 13
- Parmigiani e Inoue (2009)** Giovanni Parmigiani e Lurdes Y. T. Inoue. *Decision Theory: Principles and Approaches*. John Wiley & Sons. Citado na pág. 15
- Patriota (2013)** A.G. Patriota. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*. URL <http://dx.doi.org/10.1016/j.fss.2013.03.007>. Citado na pág. 8
- Pereira e Stern (1999)** Carlos Alberto de Bragança Pereira e Julio Michael Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110. ISSN 1099-4300. doi: 10.3390/e1040099. URL <http://www.mdpi.com/1099-4300/1/4/99>. Citado na pág. 12
- Pereira e Wechsler (1993)** Carlos Alberto de Bragança Pereira e Sergio Wechsler. On the concept of p-value. *Revista Brasileira de Probabilidade e Estatística*, 7:159–177. Citado na pág. 8
- Pereira et al. (2008)** Carlos Alberto de Bragança Pereira, Julio Michael Stern e Sergio Wechsler. Can a significance test be genuinely bayesian. *Bayesian Analysis*, 3(3):79–100. Citado na pág. 13
- Rice (2010)** Kenneth Rice. A decision-theoretic formulation of fisher’s approach to testing. *The American Statistician*, 64(4):345–349. Citado na pág. 16
- Tukey (1991)** John W. Tukey. The philosophy of mutiple comparisons. *Statistical Science*, 6(1): 100–116. Citado na pág. 11
- van der Pas (2010)** Stéphanie L. van der Pas. Much ado about the p-value. Bachelor thesis, *Mathematisch Instituut, Universiteit Leiden*. Citado na pág. 8
- Wagenmakers (2007)** Eric-Jan Wagenmakers. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5):779–804. Citado na pág. 8