Valores Epistêmicos Bayesianos: Foca na Surpresa, Mede Probabilidade!

Julio Michael Stern* and Carlos Alberto de Bragança Pereira

Instituto de Matemática e Estatística da Universidade de São Paulo.

Resumo O e-valor ou valor epistêmico, $\text{ev}(H)$, mede a significância estatística de H, uma hipótese sobre o parâmetro θ de um modelo Bayesiano. O e-valor é obtido através de uma transformação probabilidade-possibilidade da medida a posteriori do modelo, $p(\theta)$, e pode, por sua vez, ser utilizado para definir o FBST ou Teste Bayesiano Completo de Significância. Este artigo investiga a relação desta nova abordagem com transformações padrão de probabilidade-possibilidade. Em particular, mostramos como e por que o e-valor foca na ou conforma-se com $s(\theta) = p(\theta)/r(\theta)$, a função surpresa em relação à densidade de referência, $r(\theta)$, enquanto mantém-se consistente com a medida de probabilidade a posteriori do modelo. Ademais, investigamos objeções tradicionalmente levantadas na teoria estatística contra medidas de significância engendradas por transformações de probabilidade-possibilidade.

Keywords: Construtivismo cognitivo; Funções surpresa e verdade; Hipóteses precisas e complexas; Modelos Bayesianos; Raciocínio possibilístico e probabilístico; Testes de significância; Transformações entre cálculos credais; Valores epistêmicos.

Shackle [51] interpreted the possibility of an event as the absence of surprise felt when it occurs... The possibilistic interpretation of histograms may be carried out in several ways, at first glance, at least... It is supposed that the events [are] sufficient in order to equalize frequencies and probabilities.
Didier Dubois e Henri Prade [19, p.177].

1 Introdução

Este artigo compara o conceito de significância estatística de hipóteses precisas em três escolas de pensamento da estatística, a saber: (1) Estatística clássica ou freqüentista; (2) Estatística Bayesiana baseada na teoria da decisão; (3) Estatística Bayesiana baseada no arcabouço epistemológico do Construtivismo Cognitivo (Cog-Con). Outros tópicos correlatos são abordados em [61].

O primeiro objetivo deste artigo é clarificar alguns aspectos formais e enfatizar o papel que cabe à teoria da possibilidade e às transformações de probabilidade-possibilidade. A referência básica para teoria de possibilidade utilizada neste

* jstern@ime.usp.br

artigo é Dubois e Prade [19]. Este artigo pioneiro é direto, intuitivo e conciso, cobrindo no entanto todos os conceitos pertinentes.

O segundo objetivo deste artigo é explicar algumas objeções tradicionalmente levantadas na estatística Bayesiana baseada na teoria da decisão contra medidas de significância estatística engendradas por transformações de probabilidade-possibilidade. Especificamente, identificamos a analisamos quatro itens percebidos como grandes obstáculos, a saber: (1) Dificuldades técnicas em obter procedimentos invariantes. (2) Fidelidade a procedimentos de eliminação de parâmetros molestos (nuisance ou perturbadores). (3) Rejeição por parte da teoria da decisão de um ponto ótimo (máximo-a-posteriori ou máxima-verossimilhança restrita à hipótese) como um representante legítimo para uma hipótese composta. (4) Entendimentos tradicionais de testes de significância como cobertura (ou não) de uma hipótese pontual por um intervalo de credibilidade de tamanho determinado.

Para cumprir com os objetivos anteriormente mencionados, este artigo é organizado como se segue: A Seção 2 apresenta uma curta revisão e estabelece a linguagem concernente a modelos estatísticos. A Seção 3 revê conceitos básicos de teoria da possibilidade e, estritamente dentro deste arcabouço, define o valor epistêmico da hipótese H, $\text{ev}(H)$, que, por seu turno, é utilizado para definir o FBST - o Teste Bayesiano Completo de Significância. O FBST é uma nova teoria estatística de significância desenvolvida dentro do arcabouço epistemológico do construtivismo cognitivo. Para uma revisão geral do FBST, enfatizando suas conexões com lógica e outros aspectos relevantes ao escopo deste artigo, vide Borges e Stern [8]. A Seção 4 explica o papel desempenhado no FBST pela densidade de referência e pela função surpresa, e como elas são usadas para se alcançar a propriedade fundamental de invariância da resultante medida de significância. Em contraste, esta seção revê os argumentos apresentados por Box e Tiao [9] em prol de sacrificar a propriedade de invariancia em testes de significância. A Seção 5 apresenta os p-valores freqüentistas e discute suas características pseudo-possibilísticas. Esta seção também discute a lógica tradicional de procedimentos eliminação de parâmetros molestos. A Seção 6 apresenta os fatores de Bayes, a solução 'probabilidade para toda obra' da teoria de decisão para teste de hipótese. Esta seção também revê alguns argumentos de teoria da decisão contra o uso de regras lógicas de composição possibilística em procedimentos estatísticos. A Seção 7 discute o método de Lindley, uma abordagem para teste de hipóteses baseada na sua cobertura (ou não) por intervalos de credibilidade de um tamanho prescrito. O método de Lindley é uma abordagem de compromisso probabilistico-possibilística que pode ser vista como um precursor do FBST. Esta Seção também discute tradicionais requisitos sobre as regiões de credibilidade usadas no método de Lindley, como conectividade topológica, ou sobre a densidade de probabilidade subjacente, como unimodalidade e monotonicidade. A Seção 8 apresenta nossos comentários finais e direções para pesquisa futura.

2 Modelos Estatísticos Bayesianos e Freqüentistas

O modelo padrão da estatística Bayesiana paramétrica concerne a uma variável aleatória (vetorial), x, que tem uma *distribuição amostral* com uma forma funcional especificada, $p(x \mid \theta)$, indexada pelo parâmetro (vetorial) θ. A mesma forma funcional, considerada como função da variável livre θ com um argumento fixo x, é a *função de verosimilhança* do modelo.

Em estatística *freqüentista*, θ deve ser tomado como uma 'quantidade desconhecida mas fixa'. Assim, no arcabouço freqüentista, permite-se o uso do cálculo de probabilidade no espaço amostral, mas proíbe-se terminantemente o seu uso no espaço paramétrico, isto é, x deve ser considerado como uma variável aleatória, enquanto θ não deve ser considerado aleatório de forma alguma. Outras conseqüências da proibição freqüentista a enunciados probabilísticos no espaço paramétrico são examinadas na Seção 6.

No contexto Bayesiano, o parâmetro θ é considerado como uma variável aleatória *latente* (não observável). Desta forma, o mesmo formalismo utilizado para expressar credibilidade ou (in)certeza, a saber, probabilidade como um cálculo de crença abstrato, é utilizado em ambos os espaços, amostral e paramétrico. Conseqüentemente, a distribuição conjunta, $p(x, \theta)$, deveria sumarizar toda a informação disponível em um modelo estatístico, vide Wechsler et al. [63].

Seguido as regras do cálculo de probabilidades, a distribuição conjunta de x e θ pode ser fatorizada tanto como a função de verosimilhança do parâmetro dada a observação vezes a distribuição a *priori* de θ, ou como a distribuição a *posteriori* do parâmetro vezes a densidade marginal da observação,

$$p(x, \theta) = p(x \mid \theta) p(\theta) = p(\theta \mid x) p(x) \ .$$

A distribuição de probabilidade a *priori*, $p_0(\theta)$, representa a informação inicial disponível sobre o parâmetro. Neste contexto, a distribuição *preditiva* para a variável aleatória observada, x, é representada por uma mistura (ou superposição) de processos estocásticos, todos eles com a forma funcional da distribuição amostral, de acordo com a mistura (ou ponderação) especificada pela distribuição a priori.

$$p(x) = \int_{\Theta} p(x \mid \theta) p_0(\theta) d\theta \ .$$

Se em seguida observarmos um único evento, x, segue da fatoração da distribuição conjunta acima que a distribuição a *posteriori* para θ, representando a informação disponível sobre o parâmetro após a observação, é dada por

$$p_1(\theta) \propto p(x \mid \theta) p_0(\theta) \ .$$

O subscrito n em $p_n(\theta)$ conta o número de eventos observados. Se quisermos substituir o símbolo de 'proporcional a', $\propto$, pelo de igualdade, $=$, será necessário dividir o lado direito pela constante de normalização, $c_1 = \int_{\Theta} p(x \mid \theta) p_0(\theta) d\theta$.

Esta é a *Regra de Bayes* - o mecanismo básico de aprendizado da estatística Bayesiana, que fornece a probabilidade (inversa) do parâmetro dadas a observações. Computar constantes de normalização pode ser difícil ou enfadonho. Assim, especialmente em modelos grandes, é costume trabalhar com densidades não normalizadas ou *potenciais* tanto quanto possível, computando constantes de normalização apenas ao final se necessário. É interessante observar que a função de distribuição conjunta, tomada com uma parâmetro fixo, x, e um parâmetro livre, θ, é um potencial para a distribuição a posteriori.

O aprendizado Bayesiano é um processo recursivo, onde a distribuição a posteriori após um passo no aprendizado torna-se a distribuição a priori para o passo seguinte. Assumindo que as observação são c.i.i.d., condicionalmente (aos parâmetros) independentes e identicamente distribuídas, a distribuição a posteriori após n observações, $x^{(1)}, \ldots, x^{(n)}$, será,

$$p_n(\theta) \ \propto \ p(x^{(n)} \,|\, \theta) p_{n-1}(\theta) \ \propto \ \prod\nolimits_{i=1}^{n} p(x^{(i)} \,|\, \theta) p_0(\theta) \ .$$

Sempre que possivel, é muito conveniente utilizar uma *priori conjugada*, isto é, uma distribuição de mistura cuja forma funcional é invariante pela operação Bayesiana no modelo estatístico em tela, vide Castillo et al. [11, Sec.13.8], [12].

Os 'princípios e os fins' do processo de aprendizado Bayesiano merecem uma discussão mais profunda, isto é, deveríamos apresentar argumentos racionais para escolher a priori utilizada para dar partida no processo de aprendizado, e algums teoremas de convergência para a posteriori para um número crescente de observações. Para alcançar este propósito, temos que medir o conteúdo informacional de uma distribuição a priori. Stern [60] fornece uma curta revisão explicando como e por que o conceito de entropia é a chave que destranca estes mistérios.

2.1 Hipóteses Estatísticas

Uma hipótese estatística, H, (por vezes chamada de hipótese nula) afirma que o parâmetro θ do modelo estatístico encontra-se dentro do *conjunto da hipótese*, Θ_H. A bem da simplicidade, poderemos, de agora em diante, usar uma notação relaxada, escrevendo H no lugar de Θ_H. O conjunto da hipótese é geralmente definido por restrições de desidualdade e igualdade definidas por funções vetoriais, $g = [g_1(\theta), g_2(\theta), \ldots, g_l(\theta)]'$ e $h = [h_1(\theta), h_2(\theta), \ldots, h_k(\theta)]'$ no espaço paramétrico,

$$H : \theta \in \Theta_H \ , \quad \Theta_H = \{\theta \in \Theta \,|\, g(\theta) \leq 0, h(\theta) = 0\} \ .$$

Ao longo deste artigo assumiremos que o vetor de restrições é regular e não degenerado, isto é, que em parte alguma as restrições implicam em uma singularidade no conjunto da hipótese, $H \subseteq \Theta$, ou na variedade algébrica que o representa. Portanto, a dimensão da hipótese é especificada pelo número de restrições de igualdade. Utilizaremos a seguinte notação: $t = \dim(\Theta)$, é a dimensão do espaço paramétrico, k, a codimensão de H, conta o número de equações escalares restringindo H; $\dim(H) = h = t - k$, a dimensão de H, conta o número de graus de liberdade para uma partícula movendo-se em H. Estamos particularmente interessados em hipóteses *precisas*, i.e., aquelas nas quais $\dim(H) < \dim(\Theta)$, ou seja, $k > 0$. Uma hipótese *pontual* tem dimensão zero, isto é, a hipótese é um singleto, $H = \{\theta^0\}$. Ao longo deste artigo, assumiremos, sempre que necessário, condições apropriadas de regularidade topológica e analítica, como continuidade, diferenciabilidade, e a existência de argumentos maximais, restritos e irrestritos.

3 ev($\boldsymbol{H}$) - Uma Transformação Probabilidade-Possibilidade

O primeiro objetivo desta seção é definir $\text{ev}(H)$, o e-valor ou *valor epistêmico* de uma hipótese $H \subseteq \Theta$, em um modelo estatístico Bayesiano como definido na última seção, com densidade a posteriori $p_n(\theta)$ e densidade de referência $r(\theta)$. Para algumas aplicações interessantes ilustrando o uso de e-valores e do FBST em aplicações práticas, vide Diniz et al. [16], Lauretto et al. [32], Irony et al. [25], Johnson et al. [26], Pereira e Stern [43], Pereira et al. [44], Rifo e Torres [47] e Rodrigues [48].

A *função surpresa*, $s(\theta)$, indica a mudança na densidade a posteriori, $p_n(\theta)$, relativamente à densidade de referência, $r(\theta)$, representando uma situação inicial de informação mínima, vide seção 4. Os superscritos 'chapéu' e 'estrela' indicam argumentos maximais e valores ótimos, respectivamente irrestritos e restritos à hipótese, como segue:

$$s(\theta) = \frac{p_n(\theta)}{r(\theta)} \ , \quad \begin{array}{ll} \widehat{s} = \sup_{\theta \in \Theta} s(\theta) \ , & \widehat{\theta} = \arg\max_{\theta \in \Theta} s(\theta) \ , \\ s^* = \sup_{\theta \in H} s(\theta) \ , & \theta^* = \arg\max_{\theta \in H} s(\theta) \ . \end{array}$$

O *v-corte* (superior e fechado) de uma função $s(\theta)$, $T(v)$; seu complemento, o *conjunto de máxima função surpresa* (HSFS) acima do nível v, $\overline{T}(v)$, (também chamado v-corte inferior e aberto); e sua borda (também chamado curva de nível v), são definidos como

$$T(v) = \{\theta \in \Theta \,|\, s(\theta) \leq v\} \ , \quad M(v) = \{\theta \in \Theta \,|\, s(\theta) = v\} \ , \quad \overline{T}(v) = \Theta - T(v) \ .$$

A *função verdade* do modelo estatístico, $W(v)$, é a distribuição de probabilidade a posteriori acumulada até o nível de surpresa v. O complemento e a derivada de $W(v)$ também são definidos como segue. Se $W(v)$ é descontinua, $m(v)$ deve ser interpretada como uma função generalizada no sentido de Schwartz [50].

$$W(v) = \int_{T(v)} p_n(\theta)\, d\theta \ , \quad \overline{W}(v) = 1 - W(v) \ , \quad m(v) = \frac{d}{dv} W(v) \ .$$

Muitas propriedades importantes de $W(v)$ seguem diretamente da propriedade de *aninhamento* exibida pelos v-cortes que, por ser turno, fornecem o escopo das integrais que definem a função verdade, vide Klir e Folger [31, Ch.4], e Zadeh [67],

$$u \leq v \Rightarrow T(u) \subseteq T(v) \Rightarrow W(u) \leq W(v) \ .$$

Finalmente, o e-valor e seu complemento para uma hipótese $H \subseteq \Theta$, são definidos como segue.

$$\mathrm{ev}(H) = W(v^*) \ , \quad \overline{\mathrm{ev}}(H) = 1 - \mathrm{ev}(H) \ .$$

Assim como foi definido, o e-valor é uma função de conjunto. Todavia, a bem da simplicidade, poderemos utilizar uma notação relaxada, escrevendo $\mathrm{ev}(\{\theta^0\}) = \mathrm{ev}(\theta^0)$ como uma função pontual, no caso da hipótese pontual $H = \{\theta^0\}$.

O e-valor de uma hipótese H é baseado no caso mais favorável, $\mathrm{ev}(H) = \mathrm{ev}(\theta^*)$, uma propriedade que caracteriza $\mathrm{ev}(H)$ como um cálculo absrato de crença de tipo *possibilístico*, vide Darwiche [14], Darwiche e Ginsberg [15], e Borges e Stern [8]. Utilizando a propriedade de aninhamento dos v-cortes, é fácil estabelecer que $\mathrm{ev}(H)$ também tem as mui desejadas propriedades de *consistência* com a medida subjacente de probabilidade e *conformidade* (similaridade em forma) com a função surpresa subjacente, isto é,

$$\text{Consistência:} \quad \mathrm{ev}(H) \geq p_n(H) \ , \quad \forall \ H \subseteq \Theta \ ;$$

$$\text{Conformidade:} \quad \mathrm{ev}(\theta) \geq \mathrm{ev}(\tau) \ \Leftrightarrow \ s(\theta) \geq s(\tau) \ , \quad \forall \ \theta, \tau \in \Theta \ .$$

Uma *medida de plausibilidade*, $\mathrm{Pl}(H)$, é definida por sua *atribuição básica de probabilidade*, $m : 2^{\Theta} \mapsto [0,1]$, tal que $\int_{S \subseteq \Theta} m(S) = 1$. Os *elementos focais* de m são os sub-conjuntos do universo com atribuição básica de probabilidade não nula, $\mathcal{F} = \{E \subseteq \Theta \,|\, m(E) > 0\}$. Finalmente, a plausibilidade de $H \subseteq \Theta$, $\mathrm{Pl}(H)$, é definida como

$$\mathrm{Pl}(H) = \int_{E \in \mathcal{F} \,|\, E \cap H \neq \emptyset} m(E) \ .$$

Portanto, $\mathrm{ev}(H)$ pode ser caracterizado como uma função de plausibilidade tendo, como elementos focais, v-cortes da função surpresa, enquanto que a densidade básica de probabilidade atribuída a $T(v)$ é obtida integrando a densidade de probabilidade a posteriory sobre sua borda, $m(v) = \int_{M(v)} p_n(\theta) d\theta$.

Uma função de plausibilidade define, como sua dual, uma *função de crença* através das relações

$$\mathrm{Bel}(\overline{H}) = \int_{E \in \mathcal{F} \,|\, E \subseteq \overline{H}} m(E) \ = \ 1 - \mathrm{Pl}(H) \ .$$

3.1 Onus Probandi Epistêmico

Esta sub-seção tece comentários adicionais sobre a interpretação das funções de plausibilidade e (des)crença nos contextos jurídico e de epistemologia. Dubois e Prade [19, p.12] (adaptando a notação segundo as convenções adotadas neste artigo) fornecem as seguintes intuições sobre o significado das funções de crença e plausibilidade:

> *A massa $m(E_i)$ pode ser interpretada como a alocação global de probabilidade para todo o conjunto de eventos elementares que constituem E_i, sem*

no entanto especificar como esta massa é distribuída entre cada um dos próprios eventos elementares...
Os eventos E_i são chamados elementos focais, Shafer [52], e podem ser utilizados para modelar observações imprecisas. Nesta situação, a probabilidade de um evento A será imprecisa, isto é, estará contida no intervalo $[Bel(A), Pl(A)]$.
$Bel(A)$ é calculada considerando todos os elementos focais que tornam a ocorrência de A necessária, (i.e. que implicam A). $Pl(A)$ obtida considerando todos os elementos focais que tornam a ocorrência de A possível.

Isto é, *apenas* os elementos focais completamente contidos em $\overline{H}$ contribuem para a sua credibilidade, enquanto *todos* os elementos que interceptam H contribuem para a sua plausibidade.

No contexto legal, as observações no parágrafo precedente validam a interpretação de $\overline{\text{ev}}(H) = \text{Bel}(\overline{H})$ como a crença de que H seja uma afirmação falsa, isto é, de que alguém afirmando H seja culpado de perjúrio (mentir com implicações legais). No contexto legal, acusações válidas precisam estar conformes com dois princípios jurídicos básicos conhecidos como *onus probandi* e *in dubito pro reo*, para uma análise detalhada, vide Stern [54, 55].

O fundamento epistemológico natural da estatística freqüentista é o Falsificacionismo Popperiano, onde medidas de significância são utilizadas para falsificar uma teoria em um 'tribunal científico'. Para uma análise detalhada da metáfora do tribunal científico e seu papel na análise estatística, vide Stern [61].

3.2 A Transformação Padrão de Probabilidade-Possibilidade

Dubois e Prade [19, p.178] definem (para variáveis discretas) uma medida de possibilidade *padrão*, $\pi(H)$. Generalizando a mesma definição para variáveis contínuas, esta coincide com $\text{ev}(H)$ no caso especial de termos uma medida de referência trivial, isto é, a densidade de referência (imprópria) $r(\theta) = 1$. Neste caso, $s(\theta) = p_n(\theta)$ e os HSFSs são os mais conhecidos *conjuntos de máxima densidade de probabilidade* (HPDSs). Considerando que a transformação padrão de probabilidade-possibilidade já foi extensivamente estudada ao longo de muito tempo nas áreas de inteligência artificial e ciência da computação, seria natural considerar a sugestão intuitiva de utilizar $\pi(H)$ como uma medida de *significância estatística* para a hipótese H. Ocorre porém, que $\pi(H)$ não é apropriada para esta tarefa.

O e-valor, $\text{ev}(H)$, é mais flexível que a medida de possibilidade padrão, $\pi(H)$, pois permite o uso de uma densidade de referência não trivial, $r(\theta)$, e uma função surpresa, $s(\theta)$, que tem uma 'forma distinta' da densidade a posteriori, $p_n(\theta)$. Portanto, $\text{ev}(H)$ pode ser mantida consistente com $p_n(H)$, enquanto $\text{ev}(\theta)$ mantém-se conforme com $s(\theta)$. Melhor dito, a medida de possibilidade $\text{ev}(H)$ tem elementos focais que são definidos pelas curvas de nível da função surpresa, enquanto sua atribuição básica de probabilidade é obtida integrando a densidade de probabilidade a posteriori. A próxima seção irá examinar como e por que esta flexibilidade adicional pode gerar uma medida de possibilidade que pode ser utilizada como uma medida invariante de significância estatística.

4 Invariância e Geometria de Referência

Invariância é uma propriedade chave de medidas de significância bem definidas. Uma medida invariante é independente da parametrização (regular) sendo utilizada para descrever o modelo estatístico. Por exemplo, uma medida de significância invariante não pode depender do particular sistema de coordenadas que está sendo utilizado como sistema de referência no espaço paramétrico, ou da particular forma algébrica das equações utilizadas para descrever o conjunto da hipótese.

No FBST, o papel da densidade de referência, $r(\theta)$, é tornar $\text{ev}(H)$ explicitamente invariante sob transformações regulares do sistema de coordenadas. A escolha natural para uma densidade de referência é uma priori não-informativa, interpretada como uma representação de pouca ou mínima informação no estaço paramétrico, ou a priori limite para zero observações, ou um estado fundamental neutro para a operação Bayesiana. Prioris não informativas padrão (possivelmente impróprias) incluem a uniforme, a priori de Jeffrey, e densidades de máxima entropia, vide Stern [60] para uma discussão detalhada.

Invariância, como utilizada em estatística, é um conceito métrico. A densidade de referência pode ser interpretada como induzida pela métrica informacional no espaço paramétrico, $dl^2 = d\theta' G(\theta) d\theta$. A priori invariante de Jeffrey é dada por $p(\theta) = \sqrt{\det G(\theta)}$, vide Amari [1], Amari et al. [2] e Stern [60] para mais interpretações. Para uma demonstração formal das propriedades de invariância do e-valor, vide Borges e Stern [8, p.405-406].

O operador utilizado para expandir a medida de possibilidade pontual para uma medida de conjunto é a maximização (ou, mais tecnicamente, o supremo). O operador $\max_H$ é essencialmente independente da descrição algébrica usada para descrever o conjunto da hipótese. Portanto, $\text{ev}(H)$ é invariante por parametrizações alternativas do conjunto da hipótese.

O operador de maximização é utilizado em vários procedimentos da estatística clássica ou freqüentista. A estatística Bayesiana baseada na teoria da decisão, todavia, favorece operações de integração ou de 'tomar a média'. Estas preferências não são fortuitas, mas profundamente embasadas no fundamento epistemológico destes bem estabelecidos arcabouços, para uma extensa investigação sobre o tema, vide Stern [61].

4.1 Renunciando à Invariância?

Esta seção analisa a não invariância de HPDSs, e algumas implicações desta não-invariância em procedimentos Bayesianos que utilizam tais conjuntos, como o método de Lindley a ser discutido na seção 7. HPDSs são objetos não invariantes, como reconhecido em Box e Tiao [9, p.1469]:

> *Efeito de transformações: Seja $\varphi = \phi(\theta)$ uma transformação bijetora dos parâmetros θ em φ. Mas é obvio que, por sua própria definição, os regiões de mais alta densidade em θ não serão, em geral, regiões de mais alta densidade em φ.*

Procedimentos para eliminação de parâmetros molestos (nuisance) são uma técnica importante utilizada tanto na estatística freqüentista como na estatística Bayesiana baseada na teoria da decisão, como discutido nas seções 5, 6, e 7. Estes procedimentos são baseados em manobras de reparametrização, que são feitas sob-medida para um dado modelo e hipótese estatística. Portanto, qualquer não invariância em relação à escolha de coordenadas para o espaço paramétrico fatalmente contaminará procedimentos de eliminação de parâmetros molestos. Esta situação lastimável é amplamente reconhecida em Box e Tiao [9, p.1475-1477]:

Suponha que em geral tenhamos k parâmetros, $\theta = [\theta_1, \ldots, \theta_k]$. Devemos definir $(k-1)$ comparações não-redundantes como $(k-1)$ funções independentes, $\phi_i = f_i(\theta)$, $k = 1 \ldots (k-1)$, que são todas iguais a zero se e somente se $\theta_1 = \ldots = \theta_k$. Claramente, há uma vasta gama de escolha para funções deste tipo. Como regiões de mais alta densidade, como regiões de confiança, não são invariantes por transformações não lineares, temos que pensar com cuidado sobre como parametrizar estas comparações.
Em particular, estamos interessados em descobrir se $\varphi_0 = 0$ está ou não incluido [em uma regiões de mais alta densidade de nível $(1-\alpha)$]. O ponto $\varphi_0 = 0$ corresponde à situação onde $\theta_1 = \ldots = \theta_k$, e é de especial interesse na comparação da localização de distribuições.

Box e Tiao [9, p.1470] advogam o uso de métodos estatísticos baseados em HPDs, vide seção 7. Todavia, eles subestimam a importância da invariância, abrandando e atenuando a importância do assunto. De qualquer modo, uma propriedade que não se pode obter acaba sendo considerada como uma característica não essencial:

Parece que não podemos ter esperança de obter uma medida de significância ou credibilidade genuinamente invariante. Devemos nos lembrar que invariância por transformações e virtude não são sinônimos. Para problemas que não deveriam ser invariantes por transformação, uma busca por invariância serve apenas para garantir soluções inapropriadas.

Tanto quanto saibamos, a opção de Box e Tiao [9] de renunciar à invariância quando do uso de regiões de credibilidade, permaneceu consensual na corrente principal da estatística Bayesiana. Todavia, pedimos licença para discordar fortemente da conclusão final de Box e Tiao sobre a importância das propriedades de invariância em procedimentos estatísticos. Uma análise em profundidade deste assunto é dada em Stern [60].

5 *p*-valores Freqüentistas e sua Desconstrução

A idéia geral de um p-valor é computar a probabilidade de, ao repetir um experimento aleatório sob um dado modelo estatístico e uma dada hipótese H, obter uma observação mais extrema, isto é, menos provável, que aquela que foi efetivamente observada.

Na seção 2 afirmamos que, no arcabouço freqüentista, o parâmetro (vetorial) θ do modelo estatístico é considerado uma quantidade 'fixa mas desconhecida'. Ademais, no paradigma freqüentista, θ não deve ser considerado de forma alguma como uma variável aleatória. Assim, a linguagem de probabilidades fica terminantemente proibida em qualquer descrição ou manipulação envolvendo a incerteza existente sobre θ.

Neste contexto, é fácil argumentar que o estimador de *máxima verossimilhança* (ML), $\widehat{\theta}$, é a melhor escolha para fixar parâmetros livres. Da mesma forma, arumenta-se que o ML restrito, θ^*, é a melhor escolha para fixar os parâmetros sujeitos às restrições impostas por uma dada hipótese H.

Nestas condições, a distribuição preditiva, $p_n(x \mid \theta^*)$, pode ser utilizada para induzir uma ordem no espaço amostral, de modo a tornar a até agora vaga idéia de p-valor em um conceito bem definido, vide Pereira e Wechsler [45]. $\overline{\mathrm{pv}}(H)$, o complemento do p-valor da hipótese H, é definido como segue:

$$\overline{\mathrm{pv}}(H) = \int_{C(y_1,\ldots,y_n)} \prod_{i=1}^{n} p(x_i \mid \theta^*) \, dx^n \ , \quad \theta^* = \arg\max_{\theta \in H} \prod_{i=1}^{n} p(y_i \mid \theta) \ ,$$

$$C(y_1,\ldots,y_n) = \left\{ x_1,\ldots,x_n \mid \prod_{i=1}^{n} p(x_i \mid \theta^*) \leq \prod_{i=1}^{n} p(y_i \mid \theta^*) \right\} \ .$$

5.1 A Natureza Pseudo-possibilística dos *p*-valores

$\mathrm{pv}(H) = 1 - \overline{\mathrm{pv}}(H)$ mede a probabilidade de eventos 'extremos', extremos de acordo com a ordem estabelecida no espaço amostral, que por sua vez é engendrada pelo parâmetro ótimo (mais verossimilhante), θ^*. Portanto, $\mathrm{pv}(H)$ considera a probabilidade dos piores resultados sob os melhores parâmetros. Neste sentido, p-valores têm uma aparência possibilística. Todavia, embora a hipótese H seja enunciada no espaço paramétrico, $H \subseteq \Theta$, $\mathrm{pv}(H)$ 'empurra o problema' computando uma probabilidade no espaço amostral. Portanto, $\mathrm{pv}(H)$ tem apenas uma semelhança superficial com as transformações de probabilidade-possibilidade estudadas na Seção 3.

5.2 Construção de *p*-valores Práticos.

Assim como acima definido, $\overline{\mathrm{pv}}(H)$ pode ser muito difícil de calcular. Note-se que, na equação que define o p-valor, y_i e x_i, $i = 1 \ldots n$, representam, respectivamente, o banco de dados observado e outro possível. Cada banco de dados é codificado em forma matricial, $x_{i,j}, i = 1 \ldots n, j = 1 \ldots m$, consistindo de n observação singulares, $x_{i,\bullet}$, cada uma destas codificada como um vetor de dimensão m. Assim, a dimensão do espaço de integração, $m * n$, aumenta linearmente com o número de observações, acarretando um aumento exponencial do trabalho computacional decorrente. Ademais, a geometria da região de corte, $C(y_1,\ldots,y_n)$, é especificada por restrições não lineares que podem ser difíceis de manejar tanto analítica quanto numericamente. Portanto, é compreensível que a equação dando a definição geral de um p-valor seja raramente empregada na prática. Alternativas práticas de implementação fazem uso de várias técnicas de aproximação, manobras de reparametrização, estratégias de redução de dimensionalidade, e tabelas e algoritmos pré-compilados que muito simplificam as subseqüentes rotinas computacionais.

A seguir, poremos em perspectiva alguns destes métodos e como eles se (inter)relacionam com implementações práticas de um p-valor. Esperamos que esta análise desconstrutiva ajude a entender porque p valores são uma ferramenta tão bem sucedida e, mais adiante no artigo, também entender porque e-valores podem alcançá-los ou superá-los na prática. Uma importante fonte de inspiração para o desenvolvimento do e-valor foi um desafio feito por Oscar Kempthorne ao segundo autor, pedindo uma medida de significância Bayesiana que competisse com os p-valores freqüentistas, inclusive no caso de hipóteses precisas. A Seção 7 irá revisitar este assunto, explicando como, na opinião dos autores, o e-valor é uma resposta à altura do desafio lançado por Kempthorne.

(a) Aproximações assintóticas: Sob restrições de regularidade comunmente assumidas na prática da estatística, e para bancos de dados grandes, $n \rightarrow \infty$, é razoável aproximar o comportamento de certas variáveis aleatórias de interesse por distribuições estatísticas que são bem conhecidas e de manipulação conveniente. O teorema central do limite, afirmando a convergência de médias (vetoriais) para variáveis aleatórias Normais (multivariadas), fornece a mais conhecida destas aproximações.

(b) Estatísticas suficientes: Seria muito útil poder reduzir toda a amostra, $[x_1, \ldots, x_n]$, para uma estatística $S(x_1, \ldots, x_n)$ que fosse compacta e condensada. Tipicamente, embora a pré-imagem do mapa $S(\ )$ seja um espaço de dimensão $m * n$, crescendo linearmente com o tamanho da amostra, sua imagem é um espaço de dimensão fixa. Talvez o exemplo mais conhecido seja a redução de uma amostra de n valores escalares a sua média e variância, $S(x_1, \ldots, x_n) = [m, s^2]$, onde $m = (1/n) \sum x_i$ e $s^2 = (1/n) \sum (x_i - \mu)^2$. Em geral, muita informação relevante seria perdida em tal redução. No entanto, em condicções especialmente apropriadas, para um dado mapa, S, em um modelo estatístico específico, nenhuma informação relevante é perdida. Nestas circunstâncias, S é denominada uma *estatística suficiente*, vide Kempthorne e Folks [30]. Em aplicações práticas, estatísticas (aproximadamente) suficientes são obtidas em conjunção com aproximações assintóticas.

(c) Algoritmos numéricos para distribuições Gaussianas: Ao lidar com a Normal multivariada e outras distribuições relacionadas, como a Chi-2, F e T central e não central, etc., muitas computações úteis podem ser feitas com o auxílio de algoritmos pré-compilados, tabelas, ou até dispositivos analógicos, vide Pickett [46]. Acreditamos que estas técnicas simples mas poderosas ofereçam uma explicação, ao menos parcial, para o extraordinário sucesso da estatística freqüentista nos tempos que antecedem a disponibilidade de computadores pessoais. Na seção 7, discutimos como estas técnicas podem ser aplicadas com sucesso no contexto de estatística Bayesiana.

(d) Eliminação de parâmetros molestos: Na corrente principal da matemática estatística é usual, sempre que possível, reparametrizar um model utilizando novas coordenadas, $[\delta, \lambda]$, $\dim(\delta) = k$, $\dim(\lambda) = h$, de modo que *parâmetros de interesse,* δ, sejam completamente especificados pela hipótese, enquanto os *parâmetros molestos*, λ, fiquem totalmente livres. Tal decomposição $\Theta = \Delta \times \Lambda$ pode ser seguida de um procedimento de *eliminação de parâmetros molestos*, isto é, um mapeamento ou 'projeção' $\mathcal{D}(\Theta) = \Delta$, que reduz a hipótese composta original a uma hipótese pontual, $\mathcal{D}(H) = \{\delta^0\}$. Basu e Ghosh [5] fornecem uma lista com mais de 10 categorias de procedimentos para alcançar este objetivo, incluindo o uso dos operadores max_λ ou $\int d\lambda$, maximização ou integração, para obter uma verossimilhança projetada de 'perfil' ou uma densidade posteriori marginal, $p(\delta \,|\, x)$, vide também Pereira e Lindley [42]. Operadores de maximização e verossimilhanças de perfil são especialmente importantes no arcabouço freqüentista, enquanto operações de integração e posterioris marginais são especialmente importantes em estatística Bayesiana baseada na teoria da decisão, como discutido na próxima seção.

Podemos agora começar a contrastar as características de p-valores e e-valores, uma análise de contraste que irá, nas seções seguintes, ser estendida à estatística Bayesiana baseada na teoria de decisão. O FBST faz cálculos probabilísticos no espaço paramétrico e, assim fazendo, cai dentro do arcabouço Bayesiano. No entanto, o FBST não segue o paradigma de eliminação de variáveis molestas, trabalhando no espaço paramétrico original, em sua plena dimensão. A este respeito, o FBST rompe tanto com a tradição freqüentista quanto com a tradição Bayesiana. Ademais, o FBST é definido por um primeiro passo de otimização, uma operação possibilística sobre a função surpresa, seguido de um segundo passo de integração na medida de probabilidade a posteriori. A combinação destes dois passos é estranha a ambos os arcabouços tradicionais da teoria estatística. Finalmente, com os algoritmos numéricos de otimização e integração correntemente disponíveis, o FBST tem pouca necessidade de recorrer a procedimentos baseados em técnicas de aproximação assintótica. As conseqüências de todos estes desvios em relação aos meios e métodos desenvolvidos pelas duas escolas na corrente principal da estatística será explorada em mais detalhe nas seções seguintes.

6 Teoria da Decisão e Probabilidade para Toda Obra

Esta seção apresenta os fatores de Bayes, a solução para teste de hipótese fornecida pela escola Bayesiana baseada em teoria da decisão, onde probabilidade é a ferramenta para toda e qualquer obra. Dadas duas hipóteses (ou modelos) alternativos H_1 e H_2, e a (matriz) de dados observados, $X = [x_1, x_2, \ldots, x_n]$, o *fator de Bayes* $B_{1,2}$ transforma a *razão de chances a priori* na *razão de chances a posteriori*, isto é,

$$\frac{\Pr(H_1 \,|\, X)}{\Pr(H_2 \,|\, X)} = \frac{\Pr(X \,|\, H_1)}{\Pr(X \,|\, H_2)} \frac{\Pr(H_1)}{\Pr(H_2)} = B_{1,2} \frac{\Pr(H_1)}{\Pr(H_2)} \, .$$

Esta transformação decorre diretamente da regra de Bayes,

$$\Pr(H_k \mid X) = \frac{\Pr(X \mid H_k)\Pr(H_k)}{\Pr(X \mid H_1)\Pr(H_1) + \Pr(X \mid H_2)\Pr(H_2)} \ .$$

O cálculo de fatores de Bayes para hipóteses parametrizadas é explicado em Kass [29, p.776] como segue (onde g_k é a probabilidade a priori do parâmetro θ_k sob a hypotese H_k):

No caso mais simples, de duas hipóteses são distribuições simples sem parâmetros livres (caso hipótese simples versus hipótese simples), $B_{1,2}$ é a razão de verossimilhança. Nos outros casos, quando há parâmetros desconhecidos em algumas das ou em ambas as hipóteses, o fator de Bayes ainda é dado por [a formula acima], e, em certo sentido, continua tendo a forma de uma razão de verossimilhança. Neste caso, todavia, as densidades $p(X \mid H_k)$ são obtidas integrando *(não maximizando) sobre o espaço paramétrico,*

$$B_{1,2} \ = \ \frac{\int_{H_1} p(X \mid H_1, \theta_1) g_1(\theta_1 \mid H_1) d\theta_1}{\int_{H_2} p(X \mid H_2, \theta_2) g_2(\theta_2 \mid H_2) d\theta_2} \ .$$

Construir esta média ponderada de razões de verosimilhança é o curso de ação tradicional dentro do paradigma Bayesiano baseado na teoria da decisão e guiado pelas metáforas da aposta-utilidade ou do cassino, como discutido em detalhe em Stern [61]. Um ponto forte desta abordagem é que um único cálculo credal, a saber, probabilidade, é utilizado para manejar todos os cômputos envolvendo quantidades incertas. Esta abordagem é muito eficaz em vários casos envolvendo hipóteses não precisas, mas entra em sérias dificuldades no caso de hipóteses precisas representando sub-modelos de dimemsão mais baixa em um dado modelo.

O paradoxo de Lindley e outros sintomas correlatos indicam que a atitude ingênua de atribuir de forma ad-hoc probabilidades não nulas a conjuntos de medida (no sentido de Lebesgue) nula, é problemático. Existe uma vasta e sempre crescente literatura propondo, com este propósito, métodos cada vez mais sofisticados para construir distribuições a priori para casos específicos. No entanto, a própria necessidade de soluções tão intrincadas parece indicar que pode valer a pena desenvolver abordagens diferentes para testas hipóteses precisas.

Existe também uma forte linha de pensamento dentro da escola Bayesiana baseada em teoria da decisão argumentando que hipóteses precisas fazem pouco sentido dentro deste arcabouço, e que as dificuldades aludidas no último parágrafo apenas fornecem justificativas e confirmações adicionais para considerar hipóteses precisas como enunciados mal-postos. Embora tentadora sob a perspectiva interna da teoria da decisão, esta posição não consegue suportar a forte demada feita pela comunidade científica por procedimentos adequados para testar hipóteses precisas. Hipóteses precisas são formuladas naturalmente em muitas áreas de pesquisa, especialmente em ciências exatas, como discutido em Stern [56-61].

6.1 A Teoria da Decisão Rejeita a Lógica Possibilística

O passo de otimização do FBST, $\theta^* = \arg\max_H p_n(\theta)$, pode ser interpretado como 'representando' H por seu melhor caso. Argumentamos na seção 5 que esta escolha para o valor de um parâmetro fixo mas desconhecido era muito razoável dentro do arcabouço freqüentista. Esta escolha segue princípios bem estabelecidos da lógica possibilística, como explicado em Darwiche [14], Darwiche e Ginsberg [15], e Borges e Stern [8]. Das considerações conceituais feitas nesta seção, todavia, fica claro que o mesmo passo de otimização cai completamente fora do tradicional arcaboço da estatística Bayesiana baseada na teoria da decisão. Denis Lindley [36] pessoalmente nos alertou que tomar este caminho, na sua opinião, é uma opção pouco sábia:

> *Várias propostas foram feitas sobre de que maneira se poderia calcular $ev(H)$. Duas delas são os p-valores e as probabilidades a posteriori, às quais vocês adicionaram uma terceira. Suponham que olhemos para $ev(H)$ como um conceito abstrato e perguntemos a nós mesmos que propriedades ele deveria ter. Por exemplo, se A e B são duas hipóteses... Eu sou compelido a acreditar que $ev(H)$ deveria satisfazer os axiomas SP_1 a SP_5 no capítulo 6 do livro de DeGroot... Aceitando estes cinco axiomas, DeGroot prova que $ev(H)$ deve ovedecer a todas as regras da probabilidade, em particular que $ev(A \text{ ou } B) = ev(A) + ev(B) - ev(A \text{ e } B)$. Como vocês mesmos apontaram... sua forma de $ev(H)$ não satifaz esta regra, mas sim $ev(A \text{ ou } B) = \max[ev(A),\ ev(B)]$.*
>
> *Aparte desta abordagem axiomática, muitas pessoas, incluindo eu mesmo, objetaram contra o uso do conceito de possibilidade baseados no fato de que, assim fazendo, podemos calcular $ev(A \text{ ou } B)$ a partir de $ev(A)$ e $ev(B)$, sem levar em consideração a inter-relação entre A e B... Probabilidade requer três números, adequados para descrever a inter-relação entre duas hipóteses; possibilidade requer apenas dois e, for esta razão, é geralmente considerada inadequada.*

John Skilling [53], alertou-nos ainda mais enfaticamente contra seguir a renegada lógica possibilística, pois esta nos leva à heresia da otimização:

> *Tome a máxima verossimilhança. A dificuldade reside no fato de que este único ponto ótimo pode ser altamente atípico na medida sendo acessada... Ao usar um ponto singular, θ^*, para representar H, o FBST falha em escapar da inadequação de representar um conjunto por um ponto... Eu não posso aceitar tal coisa!*

7 O Método de Lindley

O tratamento dado a hipóteses precisas pelos p-valores é muito menos problemático que o oferecido por fatores de Bayes, ao menos nos aspectos previamente discutidos concernentes às propriedades de conjuntos de medida nula. Otimização em sub-variedades algébricas definidas por restrições de igualdade e desigualdade é o problema padrão de programação matemática, requerendo apenas o uso de bons algoritmos numéricos, vide Andreani et al. [3]. Em contraste, o cômputo de integrais não nulas sobre sub-variedades algébricas próprias, requer que se façam várias escolhas ad-hoc bem como a definição de objetos especialmente desenhados (como prioris artificiais e outros oxímoros) para o modelo estatístico em estudo. Ademais, todo este esforço pode não ser eficaz, resultando em procedimentos estatísticos que são problemáticos no tratamento de hipótess precisas.

Esta situação estimulou a pesquisa em métodos de compromisso, que permitem o uso conjunto de operadores integrais 'probabilisticos' e operadores optimais 'possibilísticos' dentro de um arcabouço que é fundamentamente Bayesiano, conquanto faz uso de medidas de probabilidade no espaço paramétrico. Vários destes métodos de compromisso foram desenvolvidos. Em nossa opinião, muitos destes procedimentos híbridos probabilístico-possibilitícos (não o FBST) carecem de uma fundação sólida, tanto teórica quanto epistemológica. De qualquer forma, ao apresentar qualquer dos métodos prá-existentes para lidar com hipóteses precisas, vários autores Bayesianos manifestam explicitamente um *caveat emptor* aos usuários, alertando-os que tais procedimentos são ferramentas pragmáticas que devem ser utilizadas com muita cautela, vide por exemplo Williams [65, p.234].

p-valores Bayesianos olham para a *freqüência de amostras extremas sob parâmetros médios* localizados sobre a hipótese, isto é, eles medem a probabilidade freqüentista de observarmos um banco de dados mais extremo que o em verdade observado, considerando um parâmetro θ distribuido sobre H de acordo com uma probabilidade a posteriory conveniente, $g_n(\theta)$ em H. Embora interessantes, p-valores Bayesianos e outras soluções de compromisso, não estão diretamente ligadas às idéias levando ao FBST, e serão analisadas em artigos futuros.

Esta seção discute o método de Lindley, uma abordagem para testar uma hipótese pontual, $H : \theta = \theta^0$, baseado na cobertura (ou não) de θ^0 por um intervalo de credibilidade de tamanho prescrito, vide Lee [33, Sec4.3, p.123]. Hald [24] fornece um relato detalhado do uso de intervalos de confiança e credibilidade em estatística, desde os tempos de Laplace e Gauss; vide também Barnett [4, Sec.5.5]. No método de Lindley, HPDSs são o centro de atenção de todos os procedimentos computacionais e, neste sentido, poderia ser considerado um precursor direto do FBST. Denis Lindley em pessoa, considera tais métodos apenas como procedimentos práticos para testar hipóteses pontuais, tendo porém sérias reservas sobre os fundamentos teóricos subjacentes. Iremos introduzir este método pelas palavras de alguns autores que o utilizaram em aplicações importantes:

> *A distribuição normal tem a notável propriedade de permitir que enunciados equivalentes sejam feitos quer tendo X quer tendo Θ como o espaço relevante suportando as distribuições de probabilidade... Uma interpretação Bayesiana dos habituais testes F está assim disponível refraseando a noção de teoria amostral de que um valor nulo é significante se um intervalo de confiança não o inclui, trocando confiança por credibilidade... Essencialmente,*

ao rejeitar o valor nulo, estamos dizendo que ele não tem alta probabilidade (densidade) a posteriori em comparação com outros valores.
Embora esta idéias permitam que práticas ortodoxas sejam interpretadas em termos probabilisticos, disto não se conclúi que tais práticas devam ser adotadas... Agora voltaremos nossa atenção de conceitos de teoria amostral para uma análise Bayesiana honesta de um problema de decisão (e portanto, de um problema de inferência associado). D.V.Lindley [35, p.18,19].

Em modelos lineares Normais, regiões de mais alta densidade são sempre regiões conexas, ou intervalos, e estas provêm inferências intervalares e testes de hipótese através do uso de uma posteriori Normal, ou de uma distribuição T ou F. West e Harrison [64, Sec.17.3.5, p.643].
Um gráfico temporal de [valores] estimados com uma indicação da incerteza associada, assim como medida pela variância da posteriori, fornece uma clara e útil indicação visual da contribuição dos [parâmetros escalares de interesse]. West e Harrison [64, Sec.8.6.7, p.256,257].

Teste de uma Hipótese Nula Precisa Através de Intervalos de Confiança: Alguns Bayesianos são favoráveis a testar, digamos, $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, aceitando H_0 se θ_0 pertence a um dado conjunto de credibilidade. Isto é similar à relação entre intervalos de confiança e testes clássicos, exceto que ali os testes são invertidos para obter intervalos de confiança. Isto tudo precisa ser encarado como uma maneira muito informal de fazer um teste. Se realmente acreditarmos que uma hipótese nula precisa é uma teoria bem-formulada e merece ser testada, certamente haveriamos de querer atribuir-lhe uma probabilidade a posteriori. Isto não é possível nesta abordagem.
Devido ao fato de uma inferência baseadas em intervalos de credibilidade muitas vezes ter boas propriedades freqüntistas, um teste neles baseado também é similar a um teste clássico. Isto contrasta bastante com a inferência baseada em fatores de Bayes e probabilidades a posteriori. Ghosh et al. [23, Sec.2.7.3-4, p.48-50].

Outros autores influentes no campo da estatística Bayesiana não estão dispostos a ser tão complacentes com transgressões heterodoxas em relação à doutrina da teoria da decisão, vide por exemplo Berger e Delampady [7, Sec1.2, p.319; Sec.4.3, p.328]:

Opinião 3: Testar é algo Irrelevante; Deveríamos concentrar-nos em conjuntos de confiança, testando os mesmo se necessário. *Esta opinião é errada, porque ignora a suposta natureza especial de θ_0. Um ponto pode estar fora de um conjunto de confiança de 95% e, ainda assim, não entrar em forte contradição com os dados. Apenas calculando um fator de Bayes (ou medidas condicionais relacionadas) podemos julgar quão bem os dados suportam um ponto distinguido, θ_0... O fator de Bayes comunica a evidência nos dados contra θ_0, e [uma região de confiança] C a magnitude da possível discrepância.*

Como já haviamos afirmado, no método de Lindley, HPDSs são o centro de atenção de todos os procedimentos computacionais e, neste sentido, o mesmo pode ser considerado como um precursor do FBST. No entanto, o método de Lindley e o FBST são em essência duas abordagens muito diferentes com respeito ao conceito de significância de hipótese. A lista seguinte de pontos contrastantes deve tornar clara estas diferenças. Esta lista tem também o objetivo de expor alguns obstáculos que, em nossa opinião, obstruíram as pesquisas nesta área, ao menos no caminho historicamente percorrido no desenvolvimento da corrente principal da estatística Bayesiana.

(a) Interpretações como Intervalo de Cobertura: A interpretação original de intervalo de cobertura do método de Lindley freqüentemente acarreta fortes requisitos topológicos como a de ter HPDSs simplesmente conexos. Estes requisitos topológicos, por seu turno, engendram condições analíticas como unimodalidade e monotonicidade sobre as densidades a posteriori subjacentes no modelo estatístico em estudo.

(b) Eliminação de Parâmetros Molestos: O método de Lindley pode ser coherentemente estendido de forma a lidar com uma hipótese composta, H, por redução a uma hipótese pontual, $\mathcal{D}(H) = \{\delta^0\}$, através de um procedimento aceitável de eliminação de parâmetros molestos. Como discutido nas seções 4 e 5, no arcabouço da estatística Bayesiana baseada em teoria da decisão, procedimentos aceitáveis para este propósito são baseados em marginalização e outros operadores de integração, mas nunca em operações possibilísticas de otimização.

(c) Não-invariância: Como discutido na seção 4, o método de Lindley é essencialmente não invariante, um motivo consagrado para questionar os fundamentos teóricos de qualquer método baseado em HPDSs.

(d) Função de Perda: Antes da formulação de uma função de perda apropriada por Madruga et al. [37], procedimentos de teste para hipóteses precisas baseados em HPDSs não tinham sido devidamente derivados dentro do arcabouço de teoria da decisão, mesmo considerando que funções de perda bastante parecidas foram discutidas em O'Hagan [40, Sec.2.5, p.54-59]. A falta desta fundamentação foi uma constante fonte de queixas em relação ao caráter heterodoxo destes procedimentos.

(e) Status Ontológico de Hipóteses Precisas: Mesmo depois do trabalho de Madruga et al. [36, 36], sob a perspectiva de teoria da decisão, hipóteses precisas são percebidas como sendo problemas mal-postos. Em contraste, o arcaborço epistemológico do construtivismo cognitivo suporta plenamente hipóteses precisas, vide Stern [56-61].

Pode-se argumentar que, dentro do arcabouço de teoria da decisão, os pontos (c), (d) e (e) tornaram o método de Lindley mal-e-mal aceitável como procedimento pragmático que *deve ser pensado como uma maneira muito informal de fazer um teste.* Ademais, os pontos (a), (b) e (c) parecem ter restringido o escopo de interesse das aplicações práticas do método de Lindley quase que exclusivamente a modelos lineares Gaussianos, como em Box e Tiao [10] ou West e

Harrison [64]. No entanto, fomos recentemente informados do trabalho de Sanjib Basu [6], onde o autor se afasta de interpretações estritamente baseadas em intervalos de cobertura, e rompe com os tradicionais requisitos de ter HPDSs simplesmente conexos. Assim fazendo, mesmo que não invariante e limitado a hipóteses pontuais, este trabalho pode ser visto como um precursor ainda mais próximo do FBST.

Paradoxalmente (da perspectiva do FBST e do construtivismo cognitivo), mas também muito interessante (da perspectiva histórica), o paradigma de inferêrncia por intervalos de cobertura parece ter percolado (regressivamente), vindo de (por vezes muito) antigas fontes na literatura da estatística matemática para a literatura de possibilidade ou conjuntos fuzzy. Para desenvolvimentos recentes neste tipo de inferência possibilista baseada em intervalos de cobertura, vide Castineira et al [13] e Salicone [49]. A próxima citação, de Dubois et al. [17, Sec,3,p.282], explica a motivação dos autores neste curso de investigação mesmo se, como reconhecido pelos mesmos na Observação 3.2, p.282,283, as definições decorrentes só fazem sentido para distribuições unimodais e monotônicas.

> *Uma expressão fechada da distribuição de possibilidade induzida por intervalos de confiança entorno da moda, x^m, é obtida para densidades de probabilidade contínuas e unimodais estritamente crescentes à esquerda e decrescentes à direita de x^m... Neste artigo, utilizamos a terminologia 'intervalo de confiança' para substitutos intervalares confiáveis para distribuições de probabilidade. Isto não corresponde à terminologia tradicional... Nossa noção de intervalo de confiança está muito mais próxima do intervalo fiducial de Fisher.*

8 Comenários Finais e Pesquisa Futura

8.1 Testes de Significância Preditivos vs. Epistêmicos

Na citação de Ghosh et al. [23] feita na última seção, os autores elogiam as boas qualidades freqüentistas da inferência Bayesiana baseada em intervalos de credibilidade. Leonard e Hsu [34, p.142-143] fazem comenários similares. Estas boas propriedades freqüentistas podem explicar a excelente performance do FBST em ensaios de bancada comparativos baseados na freqüencia de erros do tipo-I e tipo-II, utilizados para avaliar o desempenho do FBST em várias aplicações já publicadas. No entanto, Skilling [53] vê estas medidas de desempenho com grande suspeição. Opiniões contraditórias sobre este assunto são comuns na literatura de estatística Bayesiana: As boas propriedades freqüentistas de intervalos de cobertura são hora exaltadas como uma qualidade maravilhosa, hora desprezadas como um aspecto irrelevante.

Acreditamos que apreciações contraditórias destas propriedades freqüentistas e opiniões conflitantes sobre medidas de desempenho estão profundamente entrelaçadas com dois conceitos muito distintos de significância estatística, a saber, a noção de poder preditivo de uma teoria versus a idéia de verificação epistêmica da mesma teoria. O último parágrafo da citação de Berger e Delampady [7] feita na última seção já nos indicou a inportância de fazer esta mesma distinção. Cremos também que estes dois conceitos, embora interligados, requerem testes de significância essencialmente distintos, que devem ser acompanhados de medidas de desempenho compatíveis e, portanto, também distintas. Este assunto merece o benefício de pesquisa futura.

8.2 Transformações de Probabilidade-Possibilidade Alternativas

Dubois e Prade [19, p.177,180] definem duas transformações de probabilidade-possibilidade alternativas, especialmente adequadas para densidades contínuas,

$$\kappa(\varphi) = \int_{\Theta} \min\left[p\left(\theta\right) p\left(\varphi\right)\right] p(\theta) d\theta \ ; \quad \xi(\varphi) = \frac{p(\varphi)}{\widehat{p}} \ , \ \widehat{p} = \sup_{\Theta} p(\theta) \ .$$

Sob condições razoáveis de regularidade, ambas as transformações são facilmente estendidas para medidas de possibilidade quando computadas no argumento $\theta^* = \arg\max_H p(\theta)$.

Estas transformações alternativas têm várias interpretações interessantes. Para a primeira transformação alternativa, a identidade $\pi(\widehat{\theta}) = \kappa(\widehat{\theta}) = 1$ pode ser interpretada como duas maneiras alternativas de calcular o volume total sob o gráfico de $p(\theta)$, integrando sobre 'fatias' verticais ou horizontais. A equivalência destas duas maneiras de calcular a probabilidade total é uma conseqüência do teorema de Fubini que, por seu turno, pode ser interpretado pelo princípio de Cavalieri, expressando uma noção que precede as formalizações do cálculo tanto de Newton quanto de Leibniz, vide Fubini [22] e Palmieri [41].

Analiticamente, a transformação alternativa $\kappa(\varphi)$ pode ser mais fácil de lidar que a transformação de possibilidade padrão, $\pi(\varphi)$, porque o integrando, $\mathbf{1}\left[p(\theta) \leq p(\varphi)\right] p(\theta)$, que é (necessariamente) descontínuo, é substituído pela função, (possivelmente) contínua, $\min\left[p(\theta), p(\varphi)\right]$. Ademais, qualquer resultado enunciado para uma destas medidas pode ser imediatamente traduzido para a outra, pois, tomando $v = p(\theta)$, $\delta(\theta) = \kappa(\theta) - \pi(\theta) = v * \mu(\overline{T}(v))$, onde μ é a medida de Lebesgue adequada.

A segunda medida alternativa, $\xi(H)$, é ainda mais fácil de lidar analiticamente, embora também se afaste ainda mais da medida padrão. Várias outras transformações de probabilidade-possibilidade alternativas tem sido propostas na literatura, vide por exemplo Dubois e Prade [19], Dubois et al. [18, 21], Castineira et al. [13], Dhar [17], Jumarie [27, 28], Mauris et al. [39], Salicone [49], Yamada [62] e Wonneberger [66]. Em pesquisas futuras, temos a intenção de estudar generalizações do e-valor baseadas em algumas destas transformações de probabilidade-possibilidade alternativas.

8.3 Agradecimentos

Os autores são gratos pelo apoio recebido dos departamentos de Matemática Aplicada e Estatística do IME-USP - Instituto de Matemática e Estatística da Universidade de São Paulo, FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, e CNPq - Conselho Nacional de Desenvolvimento Científico e

Tecnológico (bolsas PQ-306318-2008-3 e PQ-302046-2009-7). Parte do material deste artigo foi inicialmente apresentado no EBL-2011 - XVI Conferência Brasileira de Lógica, de 9 a 13 de maio de 2011 no LNCC - Laboratório Nacional de Computação Cienífica, Petrópolis, e também no COBAL-2011 - III Congresso Bayesiano da América Latina, de 23 a 27 de outubro de 2011 na UFRO - Universidad de La Frontera, Pucón, Araucanía, Chile. A versão em língua inglesa deste artigo encontra-se submetida à publicação.

References

[1] S.I.Amari (2007). *Methods of Information Geometry.* Providence, RI: AMS.

[2] S.I.Amari, O.E.Barndorff-Nielsen, R.E.Kass, S.L.Lauritzen, C.R.Rao (1987). *Differential Geometry in Statistical Inference.* IMS Lecture Notes Monograph, v.10. Hayward, CA: IMS.

[3] R.Andreani, E.Birgin, J.Martinez, M.Schuverdt (2007). On Augmented Lagrangian Methods w.General Lower-level Constraints. *SIAM J.Optimization,* 18, 1286-1309.

[4] V.Barnett (1999). *Comparative Statistical Inference.* Chichester John Wiley.

[5] D.Basu (1988). Statistical Information and Likelihood. Edited by J.K.Ghosh. *Lect. Notes in Statistics*, 45.

[6] S.Basu (1996). A New Look at Bayesian Point Null Hypothesis Testing. *Sankhya,* 58,A,2, 292-310.

[7] J.O.Berger, M.Delampady (1987). Testing Precise Hypothesis. Statistical Science, 2,3,317-335. Comments of D.R.Cox, p.335-336, M.Eaton, p.337-338; A.Zellner, p.339-341; M.J.Bayarry, p.342-344; G.Casella and R.L.Berger, p.344-347; J.B. Kadane 347-348; and Rejoinder of J.O.Berger and M.Delampady p.348-352.

[8] W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-values. *Logic Journal of the IGPL*, 15, 5-6, 401-420.

[9] G.E.P.Box, G.C.Tiao (1965). Multiparameter Problems From a Bayesian Point of View. *Ann. Math. Statist.,* 36,5,1468-1482.

[10] G.E.P.Box, G.C.,Tiao (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

[11] E.Castillo, A.Cobo J.A.Gutierrez, R.E.Pruneda (1998). *Functional Networks with Applications: A Neural-Based Paradigm.* NY: Springer

[12] E.Castillo, A.Iglesias, R.Ruíz-Cobo (2005). *Functional Equations in Applied Sciences.*

[13] E.Castineira, S.Cubillo, E.Trillas (2007). On the Coherence Between Probability and Possibility Measures. *Information Theories & Applications,* 14, 303-310.

[14] A.Y.Darwiche (1993). *A Symbolic Generalization of Probability Theory.* Ph.D. Thesis, Stanford Univ.

[15] A.Y.Darwiche, M.L.Ginsberg (1992). A Symbolic Generalization of Probability Theory. AAAI-92. 10-th Conf. American Association for Artificial Intelligence.

[16] M.Diniz, C.A.B.Pereira, J.M.Stern (2011). Unit Roots: Bayesian Significance Test. *Communications in Statistics - Theory and Methods,* 40,23,4200-4213.

[17] M.Dhar (2012). Probability- Possibility Transformations: A Brief Revisit. To appear in *Annals of Fuzzy Mathematics and Informatics.*

[18] D.Dubois, L.Foulloy, G.Mauris, H.Prade (2004) Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. *Reliable Computing,* 10, 273-297.

[19] D.Dudois H.Prade (1982). On Several Representations of an Uncertain Body of Evidence. p. 167-181 in M.M.Gupta, E.Sanchez (eds.) *Fuzzy Information and Decision Processes,* North-Holland.

[20] D.Dudois H.Prade (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* NY: Plenum.

[21] D.Dubois, H.Prade, S.Sandri, (1993). On possibility-probability transformations. In *Fuzzy Logic,* R. Lowen, M. Roubens, eds. p.103-112.

[22] G. Fubini (1958). Sugli integrali multipli. *Opere scelte,* 2,243-249. Cremonese.

[23] J.K.Ghosh, M.Delampady, T.Samanta (2006). *An Introduction to Bayesian Analysis: Theory and Methods.* NY: Springer.

[24] A.Hald (2007). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935.* NY: Springer. Ch.8 Credibility and Confidence Intervals by Laplace and Gauss.

[25] T.Z.Irony, M.Lauretto, C.A.B.Pereira, and J.M.Stern (2002). A Weibull Wearout Test: Full Bayesian Approach. In: Y.Hayakawa, T.Irony, M.Xie, edit. Systems and Bayesian Reliability, 287–300. *Quality, Reliability & Engineering Statistics*, 5, Singapore: World Scientific.

[26] R.Johnson, D.Chakrabarty, E.O'Sullivan, S.Raychaudhury (2009). Comparing X-ray and Dynamical Mass Profiles in The Early-Type Galaxy NGC 4636. *The Astrophysical Journal,* 706.

[27] G.Jumarie (1995a). Possibility, probability and relative information: a unified approach via geometric programming. *Kybernetes,* 24,1,18-33.

[28] G.Jumarie (1995b). Further results on possibility-probability conversion, relative information and informational invariance. *Cybernetics and Systems,* 26,1,111-128.

[29] R.E.Kass, A.E.Raftery (1995). Bayes Factors. J.of the American Statistical Association, 90, 430, 773-795.

[30] O.Kempthorne, L.Folks (1971). *Probability, Statistics and Data Analysis.* Ames: Iowa State Univ. Press.

[31] G.J.Klir, T.A.Folger (1988). *Fuzzy Sets, Uncertainty and Information.* NY: Prentice Hall.

[32] M.Lauretto, C.A.B.Pereira, J.M.Stern, S.Zacks (2003). Full Bayesian Significance Test Applied to Multivariate Normal Structure Models. *Brazilian Journal of Probability and Statistics,* 17, 147-168.

[33] P.M.Lee (2004). *Bayesian Statistics.* Chichester: Wiley.

[34] T.Leonard, j.s.j.Hsu (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary researchers.* Sec.3.2.E, p.109, (Bayesian Intervals) and Sec.3.2.F, p.109-110, (Investigating Hypotheses).

[35] D.V.Lindley (1972). *Bayesian Statistics, A Review.* Montpelier,VM: SIAM.

[36] D.V.Lindley (2006). Personal letter to Prof. Carlos Pereira concerning the FBST. `www.ime.usp.br/∼jstern/miscellanea/citacoes/Lindley06p1.pdf`

[37] M.R.Madruga, L.G.Esteves, S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.

[38] M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference,* 117, 185–198.

[39] G.Mauris, V.Lasserre, L.Foulloy (1997). A Simple Probability-Possibility Transformation for Measurement Error Representation: A truncated triangular transformation. World Congress of International Fuzzy Systems Assoc., IFSA, Prague, Czech Republic, 476-481,

[40] A.O'Hagan (1994). *Bayesian Inference.* NY: Halsted Press.

[41] P.Palmieri (2009). Superposition: on Cavalieri's Practice of Mathematics. *Archive for History of Exact Sciences,* 63,5, 471-495.

[42] C.A.B.Pereira, D.V.Lindley (1987). Examples Questioning the use of Partial Likelihood. *The Statistician*, 36, 15–20.

[43] C.A.B.Pereira, J.M.Stern, (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69–80.

[44] C.A.B.Pereira, J.M.Stern, S.Wechsler (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Anal.* 3,1,79-100.

[45] C.A.B.Pereira, S.Wechsler (1993). On the Concept of p-value. *Brazilian Journal of Probability and Statistics*, 7, 159–177.

[46] Pickett Inc. (1965). N525 Stat-Rule, A Multi-Purpose Sliderule for General and Statistical Use (Instruction manual). Santa Barbara, CA, USA.

[47] L.L.R.Rifo, S.Torres (2009). Full Bayesian Analysis for a Class of Jump-Diffusion Models. *Communications in Statistics - Theory and Methods,* 38, 1262-1271.

[48] J.Rodrigues (2006). Full Bayesian Significance Test for Zero-Inflated Distributions. *Communications in Statistics - Theory and Methods,* 35, 299-307.

[49] S.Salicone (2007). *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence.* NY: Springer.

[50] L.Schwartz (1966). *Mathematics for the Physical Sciences.* NY: Addison-Wesley.

[51] G.L.S. Shackle (1961). *Decision Order and Time in Human Affairs.* Cambridge University Press.

[52] G.Shafer (1975). *A Mathematical Theory of Evidence.* Princeton University Press.

[53] J.Skilling (2011). Open review for paper 1146, MaxEnt 2010.

[54] J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. Laptec-2003, *Frontiers in Artificial Intelligence and its Applications,* 101, 139–147.

[55] J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *Lecture Notes Artificial Intelligence,* 3171, 134–143.

[56] J.M.Stern (2007a). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing*, 14, 1, 9-36.

[57] J.M.Stern (2007b). Language and the Self-Reference Paradox. *Cybernetics and Human Knowing*, 14, 4, 71-92.

[58] J.M.Stern (2008a). Decoupling, Sparsity, Randomization, and Objective Bayesian Inference. *Cybernetics and Human Knowing*, 15, 2, 49-68.

[59] J.M.Stern (2008b). *Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses.* Tutorial book for MaxEnt 2008, The 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. July 6-11 of 2008, Boracéia, São Paulo, Brazil.

[60] J.M.Stern (2011a). Symmetry, Invariance and Ontology in Physics and Statistics. *Symmetry,* 3, 3, 611-635.

[61] J.M.Stern (2011b). Constructive Verification, Empirical Induction, and Falibilist Deduction: A Threefold Contrast. *Information,* 2, 635-650.

[62] K.Yamada (2001) Probability-Possibility Transformation Based on Evidence Theory. Joint 9th IFSA World Congress and 20th NAFIPS Int.Conference, 70-75.

[63] S.Wechsler, C.A.B.Pereira, P.C.Marques (2008). Birnbaum's Theorem Redux. *AIP Conference Proceedings,* 1073, 96-100.

[64] M.West, J.Harrison (1997). *Bayesian Forecasting and Dynamic Models,* 2nd.ed. NY: Springer.

[65] D.Williams (2001) *Weighing the Odds.* Cambridge Univ. Press.

[66] S.Wonneberger (1994). Generalization of an invertible mapping between probability and possibility. *Fuzzy Sets and Systems,* 64, 229-240.

[67] L.A.Zadeh (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems,* 1,1,3-28.