

UNIVERSIDADE TÉCNICA DE LISBOA  
INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

ENSAIO DE HIPÓTESES.  
PERSPECTIVAS CLÁSSICA E BAYESIANA.

Maria João de Almeida Fonseca Rabaça

Junho de 1996

X - 96 - 052028 - 6

E. Biblioteca  
1698-G. 44397



HA31.5  
R33  
1996

**UNIVERSIDADE TÉCNICA DE LISBOA**  
**INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO**

**ENSAIO DE HIPÓTESES.**

**PERSPECTIVAS CLÁSSICA E BAYESIANA.**

Dissertação apresentada como requisito parcial  
para a obtenção do grau de mestre em  
Matemática Aplicada à Economia e Gestão.

**Maria João de Almeida Fonseca Rabaça**

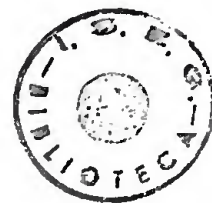
**Junho de 1996**



## Agradecimentos

Ao Professor Doutor Bento Murteira, pelo incentivo, disponibilidade e orientação prestados. O seu apoio foi fundamental à realização deste trabalho, sendo, no entanto, quaisquer erros ou omissões da minha inteira responsabilidade.

A todos os colegas, familiares e amigos que, de algum modo contribuíram para a realização deste trabalho.



## Ensaio de Hipóteses. Perspectivas Clássica e Bayesiana.

Maria João de Almeida Fonseca Rabaça.

Mestrado em: Matemática Aplicada à Economia e Gestão.

Orientador: Professor Doutor Bento José Ferreira Murteira.

Provas concluídas em:

### RESUMO

Estatísticos Clássicos e Bayesianos utilizam medidas de evidência diferentes no Ensaio de Hipóteses, preferindo os primeiros o valor- $p$  ou nível de significância e os segundos a probabilidade *a posteriori* de  $H_0$  ou o factor Bayes.

Nos testes de hipóteses unilaterais parece verificar-se uma certa conciliação entre a probabilidade *a posteriori* de  $H_0$  e o valor- $p$ . De facto, para diferentes classes de distribuições *a priori* do parâmetro  $\theta$  e determinadas características da função densidade de probabilidade  $f(x|\theta)$  da população em causa é possível mostrar que se verifica igualdade numérica entre o valor- $p$  e  $\inf P\{H_0|x\}$ ; para outras classes de distribuições *a priori* ou se  $f(x|\theta)$  não satisfizer determinados requisitos, mostra-se que  $\inf P\{H_0|x\} < p$ .

Estes resultados estão em marcado contraste com os obtidos para testes cuja hipótese nula é constituída apenas por um ponto (bilaterais ou não). Para este tipo de testes, parece não haver “harmonia” entre os resultados Clássicos e Bayesianos ( $\inf P\{H_0|x\} > p$ ). São utilizadas diferentes classes de distribuições *a priori* e modelos amostrais, o que confere alguma generalidade a este problema.

O presente trabalho apresenta a situação anteriormente descrita, fazendo referência a aspectos que, estando relacionados com as medidas de evidência ou o tipo de teste utilizado podem, segundo os autores estudados, vir a ser esclarecedoras dos factores que determinam a existência ou não de conciliação entre Clássicos e Bayesianos.

**Palavras chave:** valor- $p$ , probabilidades *a priori* e *a posteriori*, factor Bayes, testes unilaterais, testes bilaterais, paradoxo de Jeffreys-Lindley.

# ÍNDICE

<b>Capítulo 1- Introdução: O problema do ensaio de hipóteses</b>	
1.1- Perspectiva Clássica	1
1.1.1- Ensaio de significância	2
1.1.2- Ensaio de hipóteses	5
1.1.3- Lema de Neyman-Pearson	8
1.2- Perspectiva Bayesiana	9
<b>Capítulo 2- Conciliação entre Clássicos e Bayesianos</b>	
2.1- Introdução	13
2.2- Conciliação em testes de hipóteses unilaterais	14
2.3- Conciliação para diferentes modelos experimentais e diferentes e classes de distribuições <i>a priori</i>	18
2.4- Igualdade numérica entre o valor- <i>p</i> e a probabilidade <i>a posteriori</i> de $H_0$ .	23
2.5- Comentários	28
<b>Capítulo 3- Situações em que não há conciliação</b>	
3.1- Introdução	30
3.2- Perspectiva histórica da inexistência de conciliação	31
3.3- Paradoxo de Jeffreys-Lindley	32
3.3.1- Apresentação do paradoxo por Lindley	33
3.3.2- Apresentação do paradoxo por Jeffreys	38
3.4- Procura de generalidade para o problema de inexistência de conciliação	44
3.4.1- Limites inferiores em $G_A$	46
3.4.2- Limites inferiores em $G_S$	51
3.4.3- Limites inferiores em $G_{US}$	55
3.4.4- Limites inferiores em $G_{NOR}$	59
3.4.5- Uma situação não simétrica	63
3.5- Comentários	68
3.5.1- A hipótese nula é constituída apenas por um ponto	70
3.5.2- Objectividade da distribuição de credibilidades <i>a priori</i>	70
3.5.3- A definição do valor- <i>p</i>	71
3.5.4- Variação das medidas Bayesianas de evidência	73

# CAPÍTULO 1

## INTRODUÇÃO: O PROBLEMA DO ENSAIO DE HIPÓTESES

O presente estudo pretende abordar o tema do **ensaio de hipóteses**. O respectivo desenvolvimento é feito começando por apresentar as perspectivas Clássica e Bayesiana sem explorar qualquer relação que possa existir entre uma e outra.

Assim, inicialmente, a apresentação das duas correntes é conduzida em termos independentes. Seguidamente passam a tratar-se as situações em que, entre Clássicos e Bayesianos:

- existe conciliação (capítulo 2);
- não existe conciliação (capítulo 3).

Como vai ver-se, consegue-se alguma conciliação nas situações em que se ensaiam hipóteses unilaterais, isto é, do tipo,

$$H_0: \theta \geq 0 \text{ versus } H_1: \theta < 0 .$$

Por outro lado, nas situações em que ensaiam hipóteses bilaterais, isto é, do tipo,

$$H_0: \theta = \theta_0 \text{ versus } H_1: \theta \neq \theta_0 ,$$

essa conciliação não se verifica. Nos casos em que não existe conciliação dá-se particular destaque ao **paradoxo de Jeffreys-Lindley**.

### 1.1 - Perspectiva Clássica

Entre os procedimentos da inferência clássica, para além da estimação por pontos ou estimação por intervalos ou regiões, ainda podem considerar-se os ensaios de significância e os ensaios de hipóteses.

Entre os clássicos há já uma certa clivagem entre os que os que preferem os ensaios de significância de Fisher e os que defendem os ensaios de hipóteses na óptica de Neyman-Pearson.

### 1.1.1 - Ensaio de significância

Nos ensaios de significância não existe hipótese alternativa e o objectivo da análise estatística é verificar em que medida os dados recolhidos suportam a hipótese formulada, geralmente designada por hipótese da nulidade ou hipótese nula.

Considere-se uma população em que se observa uma característica  $X$  com função de distribuição de expressão analítica conhecida  $F(x|\theta)$  e com um (ou vários) parâmetro desconhecido  $\theta$ , assumindo valores em espaço próprio  $\theta \in \Theta$ , conhecido por espaço parâmetro. Em geral,  $F(x|\theta)$  pertence a uma família,

$$(1.1) \quad \mathcal{F} = \{F(x|\theta) : \theta \in \Theta\},$$

especificada pelo analista em função dos conhecimentos substantivos e da natureza da experiência de que decorre a observação de  $X$ . Em vez de  $F(x|\theta)$ , pode trabalhar-se com a função probabilidade (caso discreto) ou a função densidade de probabilidade (caso contínuo) representadas indiferentemente<sup>1</sup> por  $f(x|\theta)$ . Em geral, no caso paramétrico pretendem fazer-se inferências sobre o verdadeiro valor do parâmetro  $\theta$ , conhecendo apenas a expressão analítica da função de distribuição e uma amostra casual de valores de  $X$ ,

$$(1.2) \quad (X_1, X_2, \dots, X_N).$$

As  $N$  variáveis aleatórias independentes  $(X_1, X_2, \dots, X_N)$  são cópias de  $X$ , isto é, variáveis aleatórias i.i.d. (independentes e identicamente distribuídas).

As hipóteses a estabelecer em relação ao parâmetro são designadas hipóteses paramétricas e é delas que se ocupa exclusivamente o presente estudo.

---

<sup>1</sup> A diferença resulta claramente do contexto.



Em dada população, seja  $H_0: \theta = \theta_0$  a hipótese nula, isto é, a hipótese de que o verdadeiro valor do parâmetro  $\theta$  é igual a  $\theta_0$ . Considere-se uma realização das  $N$  variáveis aleatórias  $(X_1, X_2, \dots, X_N)$ , seja,

$$(1.3) \quad (x_1, x_2, \dots, x_N).$$

Como já foi dito, o objectivo dos ensaios de significância é avaliar em que medida os dados recolhidos estão ou não em consonância com a hipótese  $H_0$ .

Para isso: (1) utiliza-se uma medida  $T$ , ou seja uma função da amostra (estatística teste,  $T$ ) que traduz o desvio ou afastamento entre o resultado esperado quando  $H_0$  é verdadeira e o resultado observado; (2) calcula-se a probabilidade de se verificar um resultado tão bom ou pior do que o observado, admitindo que a hipótese é verdadeira.

Com,

$$(1.4) \quad T = T(x_1, x_2, \dots, x_N) = t_{obs},$$

tem-se,

$$(1.5) \quad P\{T \geq t_{obs} / H_0\} = P_{obs} = \text{valor-}p.$$

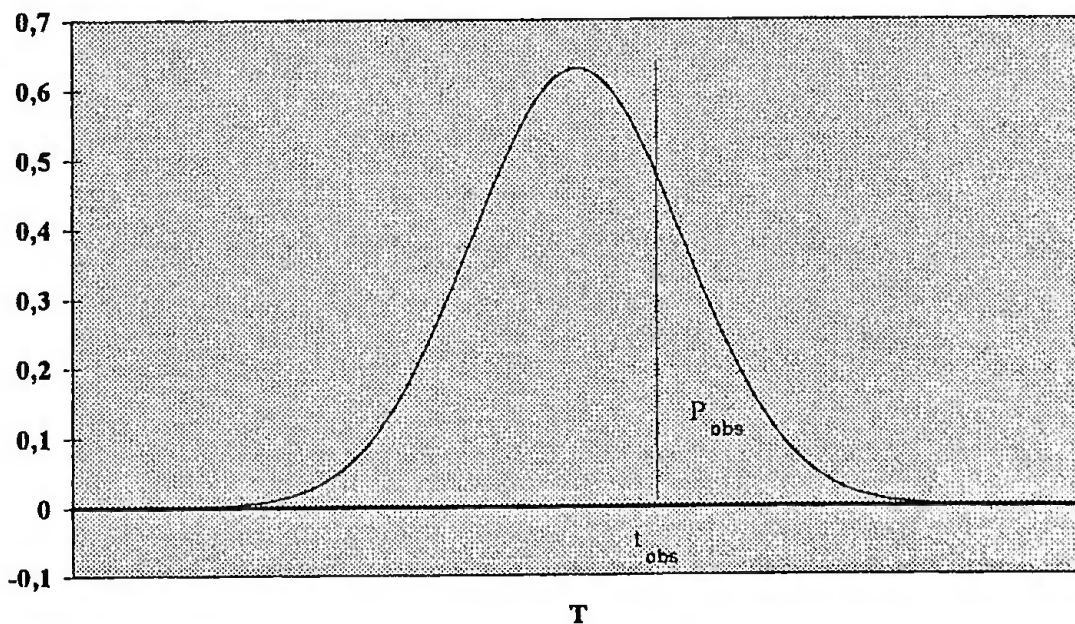


Fig.1.1

A probabilidade  $P_{obs}$  ou valor-p recebe a designação de **nível de significância** e é determinante para concluir se os dados recolhidos dão ou não consistência à hipótese  $H_0$ .

Quanto menor for esta probabilidade menor será a confirmação ou suporte que os dados fornecem à hipótese na medida em que:

- o acontecimento –  $T \geq t_{obs} | H_0$  – ou seja, o facto de se recolher um resultado tão bom ou pior do que o observado, tem uma probabilidade muito pequena;
- se um acontecimento tem uma probabilidade muito pequena isso significa, em termos práticos, que esse acontecimento raramente se verifica;
- mas o acontecimento ocorreu. Logo, o resultado não suporta a hipótese com a qual se fez o cálculo da probabilidade referida.

A questão que se poderá pôr é a seguinte: qual o nível de significância, acima do qual os dados confirmam a hipótese e abaixo do qual acontece o contrário?

Não existe um nível de significância nestas condições, mas, na prática, é costume estabelecer níveis de significância críticos tais como 0.05, 0.01 ou 0.001. Por exemplo, se na situação a estudar,  $0.001 < P_{obs} < 0.01$ , então pode dizer-se que a evidência contra  $H_0$  não é significativa ao nível de 0.1% mas é significativa ao nível de 1%, etc. Em qualquer caso, a escolha é subjectiva, compete ao investigador, não compete à teoria estatística.

Os seguidores de Fisher nunca falam em aceitação ou rejeição da hipótese  $H_0$ ; limitam-se a dizer que  $P_{obs}$  é a medida da evidência que a observação  $(x_1, x_2, \dots, x_n)$  dá da hipótese  $H_0$  sem entrar no problema da decisão. Aliás, Fisher faz a seguinte afirmação: “a perspectiva evidencial é a que interessa ao cientista, porquanto decidir pela aceitação ou rejeição com base em custos-benefícios ou perdas e lucros é vestir a pele do comerciante”

Há por outro lado clássicos que, distanciando-se da posição defendida por Fisher, preferem concluir com base na desigualdade já indicada a título de

exemplo,  $0.001 < P_{obs} < 0.01$ , que a hipótese é de rejeitar ao nível de 1% mas não é de rejeitar ao nível de 0.1%.

É de notar que, embora com o objectivo de determinar o verdadeiro valor do parâmetro, a probabilidade calculada nos ensaios de significância ( $P_{obs}$ ) é uma probabilidade com base nos dados, ou seja  $P\{T \geq t_{obs} | H_0\} = P_{obs}$ , em que  $T$  é uma estatística sendo portanto função da amostra.

### 1.1.2 Ensaio de hipóteses.

Os seguidores de Fisher pretendem, com os ensaios de significância, quantificar a evidência da hipótese, utilizando para isso os dados recolhidos (amostragem casual, por exemplo) sobre a população em causa.

Foi mencionado que alguns estatísticos clássicos se distanciam da posição defendida por Fisher e preferem dizer que determinada hipótese é de rejeitar ao nível de  $\alpha \times 100\%$  mas não é de rejeitar ao nível de  $\alpha' \times 100\%$ .

Na presente secção aborda-se a teoria de Neyman-Pearson sobre o ensaio de hipóteses que difere em vários aspectos da teoria de Fisher sobre os ensaios de significância.

A teoria de Neyman-Pearson considera importante a decisão a tomar pelo estatístico. Propõe que, face aos dados recolhidos, o estatístico decida pela rejeição ou aceitação da hipótese nula, posição substancialmente diferente da defendida por Fisher que se preocupa apenas com a medida que os dados dão da evidência da hipótese.

Tal como no caso dos ensaios de significância vão considerar-se apenas hipóteses paramétricas, isto é, parte-se do princípio que a função de distribuição  $F(x|\theta)$  ou a função de densidade  $f(x|\theta)$  da população considerada são conhecidas, permanecendo apenas desconhecido o verdadeiro valor do parâmetro  $\theta \in \Theta$ . Para facilitar a exposição considera-se no presente estudo que  $\Theta$  é unidimensional.

Seja a variável aleatória  $X$  e  $\mathcal{X}$  o espaço amostra quando se observa uma amostra casual de dimensão  $N$ ,  $(X_1, X_2, \dots, X_N)$ . Supõe-se conhecido o modelo,

ou seja  $\mathbb{F} = \{F(x/\theta) : \theta \in \Theta\}$  ou  $\mathbb{F} = \{f(x/\theta) : \theta \in \Theta\}$  e, portanto, a hipótese paramétrica a testar é do tipo

$$H_0: \theta \in \Theta_0 \text{ versus } H_1: \theta \in \Theta_1,$$

Nos ensaios de hipóteses há sempre uma hipótese nula e uma alternativa, podendo cada uma delas ser simples ou composta conforme seja constituída por um ou mais elementos. Sendo  $\Theta$  o espaço parâmetro, conjunto de todos os valores possíveis que o parâmetro pode tomar,  $\Theta_0$  e  $\Theta_1$  obedecem às condições seguintes:  $\Theta_0 \cup \Theta_1 = \Theta$  e  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Testar a hipótese  $H_0$  significa recolher uma amostra  $(x_1, x_2, \dots, x_N)$  ou seja, obter uma realização das variáveis aleatórias  $(X_1, X_2, \dots, X_N)$  e verificar qual o posicionamento dessa amostra no espaço amostra. O fundamento deste procedimento reside no facto de existirem regiões do espaço amostra  $\mathbf{X}$  em que a hipótese  $H_0$  é mais aceitável. Nesta ordem de ideias o espaço amostra  $\mathbf{X}$  é particionado em duas regiões:  $R$  (*região crítica ou região de rejeição*) e  $A = \mathbf{X} - R$  (*região de aceitação*) de tal forma que se:

- $(x_1, x_2, \dots, x_N) \in A \Rightarrow$  aceita-se  $H_0$ ;
- $(x_1, x_2, \dots, x_N) \in R \Rightarrow$  rejeita-se  $H_0$ .

A questão fundamental que se põe ao proceder ao ensaio de uma hipótese é a seguinte: qual a melhor forma de particionar o espaço amostra, ou seja, qual o teste óptimo? Há muitas formas possíveis de particionar o espaço amostra mas o que interessa é encontrar a partição óptima, ou seja, o teste óptimo o qual fica perfeitamente definido se for conhecida a região crítica óptima e consequentemente a correspondente região de aceitação.

Uma vez que toda a inferência estatística é incerta, pois, é um procedimento que permite passar do particular para o geral, os Clássicos defendem que o teste óptimo deve ser aquele que minimiza as probabilidades de cometer cada um dos erros possíveis.

Os erros possíveis no ensaio de hipóteses são:

- **Erro de 1ª. espécie** - rejeitar  $H_0$  quando  $H_0$  é verdadeira.

- Erro de 2ª. espécie - aceitar  $H_0$  quando  $H_0$  é falsa.

O Quadro 1.1 representa, não só as probabilidades de cometer os erros referidos como as probabilidades das decisões correctas.

Quadro 1.1

	$H_0$ Verdadeira	$H_0$ Falsa
Rejeita-se $H_0$	$\alpha = P\{\text{Erro de 1ª. Espécie}\}$	$1 - \beta = P\{\text{Decisão Correcta}\}$
Aceita-se $H_0$	$1 - \alpha = P\{\text{Decisão Correcta}\}$	$\beta = P\{\text{Erro de 2ª. Espécie}\}$

Assim,

$$\begin{aligned}
 \alpha &= P\{\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}\} \\
 (1.6) \quad &= P\{(X_1, X_2, \dots, X_N) \in R \mid H_0\} \\
 &= \int \dots \int_R \prod f(x_i / \theta_0) dx_1 \dots dx_N,
 \end{aligned}$$

é a dimensão do teste<sup>1</sup> ou probabilidade de cometer um erro de 1ª espécie;

$$\begin{aligned}
 \beta &= P\{\text{aceitar } H_0 \mid H_0 \text{ falsa}\} \\
 (1.7) \quad &= P\{(x_1, x_2, \dots, x_n) \in A \mid H_1\} \\
 &= \int \dots \int_A \prod f(x_i / \theta_1) dx_1 \dots dx_n,
 \end{aligned}$$

é a probabilidade de cometer um erro de 2ª espécie.

É através das probabilidades  $\alpha$  e  $\beta$  que se procura o “*melhor teste*” ou teste óptimo. À primeira vista, o teste óptimo parece ser aquele que permite a minimização destas duas probabilidades; mas, este propósito é impossível de concretizar, na medida em os dois objectivos conflituam: as probabilidades dos

<sup>1</sup> Também designado nível de significância.

dois tipos de erro variam em sentido contrário (quando uma diminui a outra aumenta e vice-versa).

No entanto, é possível a minimização de uma das probabilidades fixada a outra. Normalmente fixa-se a probabilidade de um erro de 1ª. espécie (faz-se  $\alpha=0,001, 0,01$  ou  $0,05$ ) e procura minimizar-se a probabilidade de um erro de 2ª. espécie. Se o valor obtido para esta probabilidade não for satisfatório o único caminho para a reduzir, sem alterar probabilidade de um erro de 1ª. espécie, consiste em aumentar a dimensão da amostra ( $N$ ). O aumento do número de elementos da amostra acarreta custos, o que faz com que em muitos casos se opte por uma solução de compromisso em que se avaliam as consequências que, para a situação em causa, resultam de cada um destes tipos de erro e se escolha naturalmente a mais vantajosa.

### 1.1.3 - Lema de Neyman-Pearson

O lema de Neyman-Pearson, que adiante se refere, contempla o caso particular em que se procede ao ensaio de uma hipótese simples  $H_0: \theta = \theta_0$  versus uma alternativa também simples  $H_1: \theta = \theta_1$ , caso em que o espaço parâmetro tem apenas dois elementos  $\Theta = \{\theta_0, \theta_1\}$ <sup>1</sup>. Em tal caso, o lema ensina um método para determinar a região crítica que para uma dada probabilidade de um erro de 1ª. espécie (seja  $\alpha$ ) minimiza a probabilidade de um erro de 2ª. espécie (seja  $\beta$ ). Uma tal região crítica óptima também se designa por **teste mais potente** ou **teste MP** (a potência é a probabilidade de não cometer um erro de 2ª. Espécie ou de rejeitar correctamente).

Para recordar o lema de Neyman-Pearson considere-se o ensaio  $H_0: \theta = \theta_0$  versus  $H_1: \theta = \theta_1$ ; seja uma amostra casual de uma população com função densidade  $f(x/\theta)$ ,  $\theta \in \{\theta_0, \theta_1\}$ , seja  $(x_1, x_2, \dots, x_N)$  ponto corrente do espaço

---

<sup>1</sup> Embora as suas aplicações transcendam muito este caso particular.

amostra e “c” um número real positivo. O teste mais potente é definido pela região crítica  $R$ , tal que:

$$(1.8) \quad R = \{ (x_1, x_2, \dots, x_N) \in \mathbf{X} : \prod f(x_i/\theta_1) / \prod f(x_i/\theta_0) \geq c \},$$

onde a constante “c” é determinada em função do valor fixado para  $\alpha$ .

O lema de Neyman-Pearson refere-se apenas a ensaios de uma hipótese simples contra uma alternativa simples mas pode também aplicar-se em ensaios unilaterais, com vista à obtenção em tais casos de testes uniformemente mais potentes (UMP).

Refere-se, finalmente, que tal como acontece com os ensaios de significância as decisões são tomadas com base no comportamento geral dos dados recolhidos apesar das hipóteses dizerem respeito ao verdadeiro valor do parâmetro.

## 1.2 - Perspectiva Bayesiana

No campo Bayesiano, a maneira de “trabalhar” o ensaio de hipóteses é mais directa. Calculam-se as probabilidades de cada uma das hipóteses (*a posteriori* e por vezes também *a priori*) que são condições que dizem respeito ao parâmetro e é com base na comparação dessas probabilidades que se toma uma decisão.

Como já foi dito os estatísticos Clássicos tratam os ensaios de hipóteses e os ensaios de significância de forma mais “indirecta” na medida em que, para tomarem uma decisão (ensaios de hipóteses) ou simplesmente referirem a medida da evidência que os dados conferem à hipótese (ensaios de significância), utilizam os dados, ou seja, calculam probabilidades com base nos dados e não no parâmetro cujo valor estão a testar.

A abordagem Bayesiana passa por uma extensão do modelo clássico, do qual divergem ao considerar o parâmetro  $\theta$  como uma variável aleatória ao contrário dos Clássicos que considerando  $\theta$  desconhecido mas fixo. Para os Bayesianos o que é desconhecido é incerto e toda a incerteza tem de ser

quantificada em termos de probabilidade. Se a experiência consiste na observação da variável aleatória  $X$  (ou do vector aleatório  $(X_1, X_2, \dots, X_N)$ ) com função densidade de probabilidade  $f(x|\theta)$ , há que pensar da distribuição conjunta do par  $(X, \theta)$ , que se escreve,

$$h(x, \theta) = f(x|\theta)h(\theta),$$

onde  $h(\theta)$  é a distribuição *a priori* de  $\theta$  que traduz formalmente a informação inicial<sup>1</sup> que o investigador tem acerca de  $\theta$  e que, na prática corrente, é de natureza subjectiva. A distribuição marginal de  $X$ , seja  $f(x)$ , que no contexto Bayesiano se designa por distribuição preditiva, obtém-se, como é sabido, com a integração,

$$(1.9) \quad f(x) = \int_{\Theta} h(x, \theta) d\theta = \int_{\Theta} f(x|\theta)h(\theta) d\theta$$

O recurso ao teorema de Bayes conduz à chamada distribuição *a posteriori* de  $\theta$ ,

$$(1.10) \quad h(\theta|x) = \frac{f(x|\theta)h(\theta)}{f(x)} = \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x|\theta)h(\theta) d\theta}.$$

A distribuição *a posteriori* – base de toda a inferência Bayesiana – formaliza a transformação das perspectivas do investigador acerca de  $\theta$  quando a informação inicial, donde a distribuição *a priori*, é suplementada com a informação experimental (observação de  $X=x$ ).

Assim, considerando o ensaio

$$H_0: \theta \in \Theta_0 \text{ versus } H_1: \theta \in \Theta_1$$

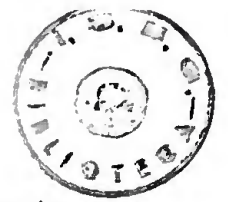
e admitindo por enquanto que, quando  $h(\theta)$  é contínua,  $\Theta_0$  não é conjunto singular ( $P\{\Theta_0\} > 0$ ), os Bayesianos calculam as probabilidades *a posteriori* das hipóteses  $H_0$  e  $H_1$  respectivamente,

$$(1.11) \quad P\{H_0|x\} = \int_{\Theta_0} h(\theta|x) d\theta, \quad P\{H_1|x\} = \int_{\Theta_1} h(\theta|x) d\theta,$$

---

<sup>1</sup> Que pode ser de qualquer natureza, inclusivamente palpites, intuições, inclinações, etc., do investigador ou decisor.





e ficam a dispor da credibilidade ou plausibilidade das duas hipóteses que podem utilizar em função dos objectivos que perseguem. Se,

$$(1.12) \quad P\{H_0|x\} > P\{H_1|x\} \Leftrightarrow P\{H_0|x\} > 1 - P\{H_0|x\} \Leftrightarrow P\{H_0|x\} > 1/2,$$

e tiverem de fazer uma opção é natural que se inclinem para  $H_0$ . Note-se  $P\{H_0|x\}$  e  $P\{H_1|x\}$  são medidas de precisão final enquanto na perspectiva clássica se recorre sobretudo à precisão inicial.

Os Bayesianos também costumam calcular o rácio das “vantagens” *a posteriori* ou seja,

$$\frac{P\{H_0|x\}}{P\{H_1|x\}},$$

por exemplo,  $P\{H_0|x\}/P\{H_1|x\} = 8$  significa que, *a posteriori*,  $H_0$  é oito vezes mais credível do que  $H_1$ . Aliás, consideram não menos interessante comparar o rácio das vantagens *a posteriori* com o rácio das vantagens *a priori*, surgindo assim o **factor Bayes**,

$$(1.13) \quad B = \frac{P\{H_0|x\} / P\{H_1|x\}}{P\{H_0\} / P\{H_1\}},$$

onde, é claro,

$$P\{H_0\} = \int_{\Theta_0} h(\theta) d\theta, \quad P\{H_1\} = \int_{\Theta_1} h(\theta) d\theta,$$

são as probabilidades *a priori* das hipóteses. O factor Bayes permite verificar em que medida os dados recolhidos suportam ou contrariam  $H_0$ . Assim, grosso modo,

- se  $B > 1$  - os dados suportam  $H_0$  ou seja, os dados não contrariam a hipótese nula;
- se  $B < 1$  - os dados contrariam a hipótese  $H_0$  que perde vantagem em relação a  $H_1$ .

É interessante verificar que, considerando o ensaio de uma hipótese simples contra uma hipótese simples e admitindo que a distribuição *a priori* de  $\theta$  é

discreta, o factor Bayes é a razão de verosimilhanças, elemento importante para os Clássicos na identificação da região crítica (veja-se (1.8)).

Com efeito, com  $H_0: \theta = \theta_0$  versus  $H_1: \theta = \theta_1$ , tem-se,

$$P\{H_0|x\} = \frac{f(x|\theta_0)P\{H_0\}}{f(x|\theta_0)P\{H_0\} + f(x|\theta_1)P\{H_1\}},$$

$$P\{H_1|x\} = \frac{f(x|\theta_1)P\{H_1\}}{f(x|\theta_0)P\{H_0\} + f(x|\theta_1)P\{H_1\}}.$$

donde se conclui que o factor Bayes,

$$\begin{aligned} B &= \frac{P\{H_0|x\} / P\{H_1|x\}}{P\{H_0\} / P\{H_1\}}, \\ &= \frac{f(x|\theta_0)}{f(x|\theta_1)}, \end{aligned}$$

é precisamente a razão de verosimilhanças. Esta situação, apesar de particular, é importante na medida em que traduz uma certa “aproximação” entre as perspectivas Clássica e Bayesiana.

# CAPÍTULO 2

## CONCILIAÇÃO ENTRE CLÁSSICOS E BAYESIANOS

### 2.1 - Introdução

Como foi referido, há situações de conciliação entre Clássicos e Bayesianos (testes unilaterais do tipo  $H_0: \theta \leq \theta_0$  versus  $H_1: \theta > \theta_0$ ) e outras em que esse “encontro” não parece ser possível (testes bilaterais do tipo  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ ).

No presente capítulo vão abordar-se algumas situações em que existe conciliação entre Clássicos e Bayesianos, ou seja, em que tanto uns como outros, face a resultados aproximadamente ou mesmo numericamente iguais, tomam a mesma posição quanto à hipótese nula em ensaio embora mantenham diferenças na interpretação..

Clássicos e Bayesianos têm formas diferentes de tratar os testes de hipóteses, usando, por conseguinte, “medidas” diferentes para tirarem conclusões sobre a hipótese nula.

Nos estudos realizados sobre esta problemática a medida de evidência clássica utilizada é o nível de significância  $\alpha$  ou valor- $p$ ; no estabelecimento do confronto são preferidos os ensaios de significância à filosofia de Neyman-Pearson sobre os ensaios de hipóteses, não sendo portanto utilizadas as probabilidades de cometer erros de 1ª e 2ª espécie.

Para os Bayesianos a medida utilizada é a probabilidade *a posteriori* de a hipótese  $H_0$  ser verdadeira.

O problema da conciliação entre Clássicos e Bayesianos passa então pela identificação de situações em que estas “medidas” se aproximam ou se distanciam.

Depois da introdução apresentam-se situações em que existe conciliação e, em seguida, abre-se o caminho para a terceira parte do trabalho, ou seja, para as situações em que não existe conciliação.

## 2.2 - Conciliação em testes de hipóteses unilaterais

Na presente secção vai apresentar-se um caso particular em que existe conciliação entre Clássicos e Bayesianos para testes unilaterais, ou seja, testes do tipo  $H_0: \theta \leq \theta_0$  versus  $H_1: \theta > \theta_0$ . Existem alguns estudos em que se mostra que a medida frequentista de evidência (valor- $p$ ) e a medida Bayesiana (probabilidade *a posteriori* da hipótese  $H_0$  nuns casos e factor Bayes noutros) podem ser aproximadamente iguais. Entre estes estudos podem mencionar-se os de Pratt (1965) e DeGroot (1973). Casella e Berger (1987) concluíram que a “evidência” pode ser reconciliada para o caso dos testes unilaterais e obtiveram, para algumas classes de distribuições *a priori* do parâmetro  $\theta$ , consideradas como imparciais, igualdade numérica entre o ínfimo – dentro dessas classes – da probabilidade *a posteriori* de  $H_0$  e o valor- $p$  frequentista. Noutros casos, ou seja, para outras classes de distribuições *a priori*, mostraram ser o ínfimo da probabilidade *a posteriori* de  $H_0$  um limite inferior do valor- $p$ , facto que os levou a afirmar que há significativos indícios de que o valor- $p$  está na fronteira ou no interior do intervalo correspondente às medidas Bayesianas de evidência.

Depois deste breve enquadramento apresenta-se um exemplo que expressa bem a conciliação entre Bayesianos e frequentistas no caso dos testes unilaterais.

Considere-se uma amostra casual constituída por  $N$  variáveis aleatórias independentes  $X_i$  ( $i = 1, 2, \dots, N$ ) com distribuição Normal,  $X_i \sim N(\theta, \sigma^2)$ , em que a variância  $\sigma^2$  é conhecida. Seja o ensaio da hipótese  $H_0: \theta \leq \theta_0$  contra a

alternativa  $H_1: \theta > \theta_0$  e  $\pi(\theta)=1$  a distribuição *a priori* não informativa de  $\theta$ . Obtida uma amostra com média  $\bar{x}$ , os frequentistas (perspectiva Fisheriana), calculam o nível de significância  $\alpha$  ou valor- $p$ <sup>1</sup>,

$$\begin{aligned}\alpha &= P\{\bar{X} \geq \bar{x} | \theta = \theta_0\} = P\left\{\frac{\sqrt{N}(\bar{X} - \theta_0)}{\sigma} \geq \frac{\sqrt{N}(\bar{x} - \theta_0)}{\sigma}\right\} \\ &= P\left\{U \geq \frac{\sqrt{N}(\bar{x} - \theta_0)}{\sigma}\right\} = 1 - \Phi\left[\frac{\sqrt{N}(\bar{x} - \theta_0)}{\sigma}\right].\end{aligned}$$

Os Bayesianos avaliam a hipótese  $H_0$  calculando a respectiva probabilidade *a posteriori*. Para isso, têm de começar por identificar a distribuição *a posteriori* de  $\theta$ , isto é,

$$h(\theta|\bar{x}) = \prod_{i=1}^N (\sigma\sqrt{2\pi})^{-1} \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\} / \int_{-\infty}^{+\infty} \prod_{i=1}^N (\sigma\sqrt{2\pi})^{-1} \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\} d\theta$$

ou seja,

$$h(\theta|\bar{x}) = \frac{\sqrt{N}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{N}{2\sigma^2}(\theta - \bar{x})^2\right\},$$

que permite concluir que  $h(\theta|\bar{x}) \equiv N(\bar{x}, \sigma^2 / N)$ .

Uma vez que a distribuição *a posteriori* de  $\theta$  está identificada, passa-se ao cálculo de  $P\{H_0|\bar{x}\}$ :

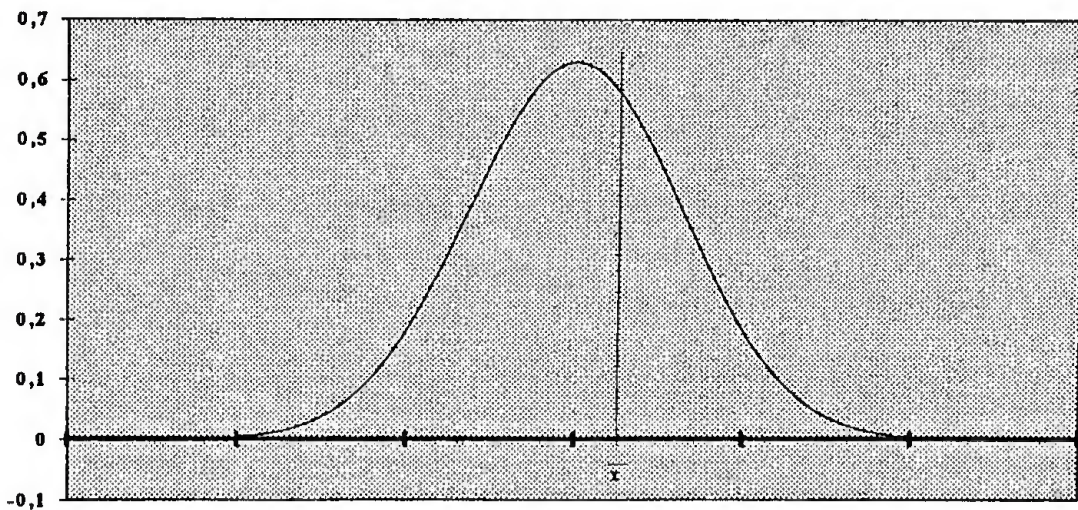
$$\begin{aligned}P\{H_0|\bar{x}\} &= P\{\theta \leq \theta_0|\bar{x}\} = P\left\{\frac{\sqrt{N}(\theta - \bar{x})}{\sigma} \leq \frac{\sqrt{N}(\theta_0 - \bar{x})}{\sigma}\right\} \\ &= P\left\{U \leq \frac{\sqrt{N}(\theta_0 - \bar{x})}{\sigma}\right\} = \Phi\left[-\frac{\sqrt{N}(\bar{x} - \theta_0)}{\sigma}\right] = 1 - \Phi\left[\frac{\sqrt{N}(\bar{x} - \theta_0)}{\sigma}\right].\end{aligned}$$

---

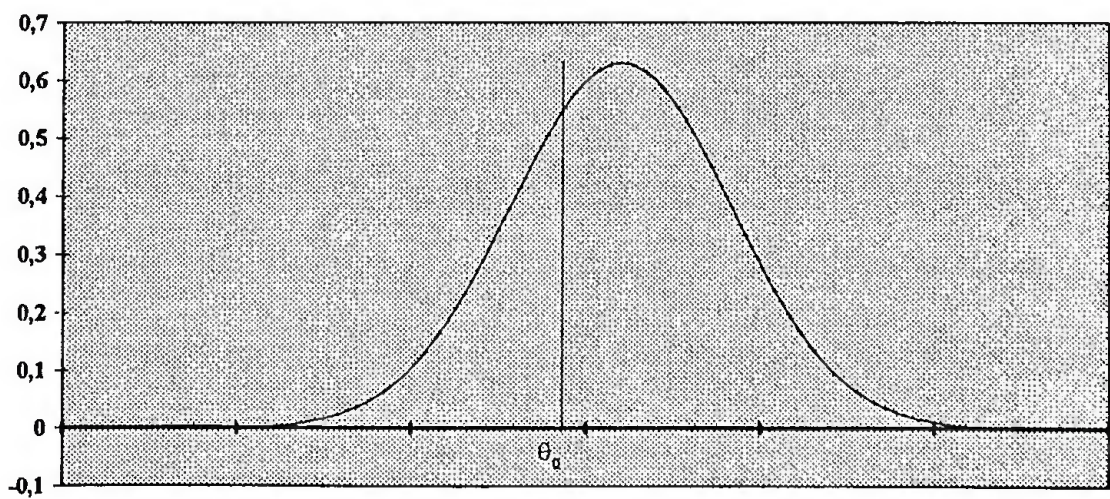
<sup>1</sup> Na expressão abaixo e em expressões futuras a variável aleatória  $U$  designa sempre uma Normal estandardizada. Adopta-se, também, a convenção habitual em que  $\phi$  é a respectiva função densidade de probabilidade e  $\Phi$  a respectiva função de distribuição.

Como pode verificar-se, os valores referentes ao nível de significância  $\alpha$  e a  $P\{H_0|\bar{x}\}$  são numericamente idênticos  $[P\{\bar{X} \geq \bar{x}|\theta = \theta_0\} = P\{\theta \leq \theta_0|\bar{x}\}]$  tendo, no entanto, interpretações diferentes.

**Fig.2.1**



**Fig.2.2**



As Fig.2.1 e 2.2 caracterizam as duas situações. A Fig.2.1 representa a distribuição por amostragem da média  $\bar{X}$ ; a região à direita do segmento de recta corresponde à probabilidade de ser  $\bar{X} \geq \bar{x}$ , ou seja, ao valor- $p$ . Repare-se que a distribuição tem média igual a  $\theta_0$ . A Fig.2.2 representa a distribuição *a posteriori* de  $\theta$  e a zona à esquerda do segmento de recta corresponde à probabilidade de ser  $\theta \leq \theta_0$  quando foi observado  $\bar{x}$ . Repare-se que a distribuição da Fig.2.2 tem média igual a  $\bar{x}$ . [Exemplo baseado em Hays e Winkler(1970) e reproduzido em Murteira(1988)].

Muitos autores consideram que a igualdade numérica que se verifica entre  $P(\bar{X} \geq \bar{x} | \theta = \theta_0)$  e  $P(\theta \leq \theta_0)$  representa a conciliação entre Clássicos e Bayesianos no caso dos testes unilaterais e referem que os Clássicos, perante estes resultados, deviam admitir a extensão das medidas Bayesianas de evidência ao campo clássico, através da interpretação do nível de significância em termos de graus de credibilidade.

O nível de significância é bastante criticado pelos Bayesianos que argumentam, por exemplo, que o mesmo é calculado com base em dados que deviam ter sido observados mas não foram. Por exemplo, em relação a esta questão Jeffreys afirma: *“a hypothesis which may be true may be rejected because it has not predicted observable results which have not occurred”*. Os Bayesianos referem ainda que este aspecto contradiz o princípio da verosimilhança segundo o qual a inferência deve basear-se só e somente nos dados efectivamente observados.

Reforçando a crítica acima os Bayesianos insistem no facto de o valor- $p$  ser uma probabilidade referente aos dados e não uma probabilidade referente ao parâmetro; a hipótese a ensaiar diz respeito ao parâmetro, o que leva os Bayesianos a considerar a sua forma de abordar o problema muito mais directa, na medida em que calculam a probabilidade *a posteriori* da hipótese nula, ou seja, a probabilidade do parâmetro pertencer a determinado conjunto.

### 2.3 - Conciliação para diferentes modelos experimentais e diferentes classes de distribuições *a priori*.

O problema que se põe, em seguida, é o de saber se a “conciliação”, traduzida pela igualdade numérica do valor-*p* e da probabilidade *a posteriori* da hipótese  $H_0$ , se verifica noutras situações e não apenas no caso particular em que o modelo experimental é a Normal e a distribuição *a priori* do parâmetro é não informativa. O facto de a distribuição *a priori* ser não informativa poderia restringir bastante o campo de obtenção da referida “conciliação” na medida em que, neste caso, contam apenas os dados amostrais, o que, de alguma forma, atenua as diferenças entre o tratamento Clássico e Bayesiano.

Casella e Berger (1987) mostram que não, ou seja, que esta “conciliação” se verifica em muitas outras situações, nomeadamente quando a distribuição *a priori* deixa de ser não informativa. Verificaram que: (1) para algumas classes de distribuições *a priori* existe igualdade numérica entre as medidas de evidência Clássica e Bayesiana; (2) para outras classes de distribuições *a priori*, a medida Bayesiana de evidência mostra ser um limite inferior do valor-*p*, o que os leva a concluir que a medida frequentista está certamente na fronteira ou no interior do intervalo das medidas Bayesianas de evidência.

A exposição que vai fazer-se segue de perto aqueles dois autores. O que pretende mostrar-se é que a igualdade numérica entre as medidas Clássica e Bayesiana de evidência se verifica em condições diferentes das do caso particular apresentado anteriormente (modelo experimental Normal e distribuição *a priori* não informativa).

As medidas de evidência Clássica e Bayesiana que vão utilizar-se são respectivamente o valor-*p* ou nível de significância  $\alpha$  e o ínfimo da probabilidade *a posteriori* de  $H_0$  obtido numa classe de distribuições *a priori* do parâmetro em relação ao qual, sem perda de generalidade, se considera o ensaio  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ .



A experiência consiste na observação da variável aleatória  $X$  com função densidade de probabilidade  $f(x-\theta)$  verificando as seguintes propriedades:

- (a)  $f(x-\theta)$  é simétrica em relação a zero;
- (b)  $f(x-\theta)$  tem razão de verossimilhança monótona (RVM).



Observado  $X=x$ , as medidas de evidência são:

Clássica:

$$\alpha = \text{valor} - p = p(x) = P\{X \geq x | \theta = 0\} = \int_x^{\infty} f(t) dt.$$

Bayesiana:

$$P\{H_0 | x\} = P\{\theta \leq 0 | x\} = \int_{-\infty}^0 f(x-\theta)\pi(\theta) d\theta / \int_{-\infty}^{+\infty} f(x-\theta)\pi(\theta) d\theta.$$

Como já foi referido, a busca de generalidade para a igualdade numérica das medidas de evidência Clássica e Bayesiana passa também pela utilização de outras classes de distribuições *a priori* do parâmetro  $\theta$ . Como o valor- $p$  é baseado num modelo frequentista objectivo, parece ser mais indicado utilizar distribuições *a priori* “imparciais”, ou seja, distribuições que atribuam igual peso às hipóteses nula e alternativa.

São três as classes de distribuições *a priori* consideradas:

$\Gamma_S = \{\text{distribuições simétricas em relação a zero}\};$

$\Gamma_{US} = \{\text{distribuições com densidades unimodais, simétricas em relação a zero}\};$

$\Gamma_{NOR} = \{\text{distribuições Normais } (0, \tau^2), 0 < \tau^2 < \infty\}.$

Para o teste de  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ , se  $f(x-\theta)$  é simétrica em relação a zero e tem RVM e se a distribuição *a priori* pertence à classe  $\Gamma_S$  pode concluir-se que,

$$p(x) \geq \inf_{\pi \in \Gamma_S} P\{H_0 | x\} .$$

Antes de entrar na demonstração deste resultado convém notar que apenas interessa considerar o caso em que o valor observado da variável aleatória  $X$  é positivo, ou seja,  $x > 0$ . Para  $x < 0$  tanto Bayesianos como Clássicos se inclinam obviamente para a hipótese  $H_0$ .

Começa então por provar-se que para o teste de hipóteses  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ , se  $f(x-\theta)$  é simétrica em relação a zero, tem RVM e  $x > 0$ , então:

$$(2.1) \quad \inf_{\pi \in \Gamma_S} P\{H_0|x\} = \inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} \leq p(x).$$

Utiliza-se na demonstração a classe de distribuições *a priori*,

$\Gamma_{2PS} = \{ \text{distribuições simétricas em relação a zero com massa concentrada apenas em dois pontos} \}.$

A igualdade (2.1) decorre da aplicação do resultado seguinte:

**Lema 2.1** - Seja  $\Gamma = \{ \pi_\alpha: \alpha \in A \}$ , a classe de distribuições *a priori* definidas na recta real e indexadas pelo conjunto  $A$ . Seja  $\Gamma_M$  o conjunto de todas as “misturas” de elementos de  $\Gamma$ , isto é,

$$(2.2) \quad \pi \in \Gamma_M \Leftrightarrow \pi(B) = \int_A \pi_\alpha(B) dP(\alpha).$$

para alguma medida de probabilidade  $P$  em  $A$  e para todo o conjunto mensurável  $B$ . Então,

$$(2.3) \quad \inf_{\pi \in \Gamma_M} P\{H_0|x\} = \inf_{\pi \in \Gamma} P\{H_0|x\}.$$

Dem. Uma vez que a igualdade indicada em (2.1) está justificada, porquanto  $\pi \in \Gamma_S$  é uma mistura de elementos de  $\Gamma_{2PS}$ , passa-se à demonstração da desigualdade,

$$\inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} \leq p(x).$$

Como  $\pi \in \Gamma_{2PS}$  dá probabilidade 1/2 a dois pontos  $\theta = \pm k$ , tem-se:

$$P\{H_0|x\} = P\{\theta \leq 0|x\} = \frac{\frac{1}{2} f(x+k)}{\frac{1}{2} f(x-k) + \frac{1}{2} f(x+k)} = \frac{f(x+k)}{f(x-k) + f(x+k)}.$$

Devido às propriedades de  $f(x-\theta)$  pode concluir-se que  $P\{H_0|x\}$  é decrescente em  $k$ , donde,

$$\inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} = \lim_{k \rightarrow \infty} \frac{f(x+k)}{f(x-k) + f(x+k)} = \lim_{k \rightarrow \infty} \left[ 1 + \frac{f(x-k)}{f(x+k)} \right]^{-1},$$

e, devido à simetria de  $f(x-\theta)$ , obtém-se:

$$\inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} = \lim_{k \rightarrow \infty} \left[ 1 + \frac{f(k-x)}{f(k+x)} \right]^{-1}.$$

Para o resto da demonstração admite-se que  $f'(t)$  existe para qualquer valor de  $t$  e que o suporte de  $f(x-\theta)$  é a recta real. Como  $f(x-\theta)$  tem RVM pode escrever-se  $f(t) = \exp\{-g(t)\}$ , onde  $g$  é convexa, sendo então  $f(t)$  log-côncava. Pode concluir-se então que:

$$\begin{aligned} \frac{f(k-x)}{f(k+x)} &= \frac{\exp\{-g(k-x)\}}{\exp\{-g(k+x)\}}, \\ &= \exp\{g(k+x) - g(k-x)\}, \\ &= \exp\{2xg'(c)\}, \\ &\geq 2xg'(k-x). \end{aligned}$$

Este último resultado é devido ao Teorema de Lagrange e ao facto de  $g(t)$  ser convexa, sendo, portanto,  $-g(t)$  côncava. Seja  $\ell = \lim_{t \rightarrow \infty} g'(t)$ , que deve existir devido ao facto de  $g'(t)$  ser crescente. Se  $\ell = \infty$  o resultado pretendido é trivialmente verdadeiro. Admita-se então que  $\ell < \infty$ . Substituindo  $g'(k-x)$  por  $\ell$  obtém-se:

$$(2.4) \quad \inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} \leq \frac{1}{1 + \exp\{2\ell x\}}.$$

Em seguida note-se que para  $t > 0$ , a razão,

$$\frac{f(t)}{\exp\{-\ell t\}} = \exp\{\ell t - g(t)\},$$

é crescente em  $t$ ,  $\ell \geq g'(t)$ . Portanto,

$$p(x) = \int_x^\infty f(t)dt = \frac{\int_x^\infty f(t)dt}{2 \int_0^\infty f(t)dt} \geq \frac{\int_x^\infty \exp\{-\ell t\}dt}{2 \int_0^\infty \exp\{-\ell t\}dt} = \frac{1}{2} \exp\{-\ell x\}.$$

Combinando esta desigualdade com (2.4) obtém-se,

$$(2.5) \quad p(x) \geq \frac{1}{2} \exp\{-\ell x\} > \frac{1}{1 + \exp\{2\ell x\}} \geq \inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\},$$

e fica assim provado que,

$$p(x) \geq \inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\} = \inf_{\pi \in \Gamma_S} P\{H_0|x\},$$

ou seja, que na classe de distribuições *a priori* do parâmetro que são simétricas em relação a zero, o  $\inf P\{H_0|x\}$  é um limite inferior de  $p(x)$ . É ainda de referir que para densidades  $f(x-\theta)$  cujo suporte é a recta real, sendo  $\ell \neq 0$  a desigualdade entre  $\inf_{\pi \in \Gamma_{2PS}} P\{H_0|x\}$  e  $p(x)$  é estrita. Se  $f(x-\theta)$  tem suporte limitado, então a igualdade pode ser atingida.

Com este resultado é possível considerar que o valor- $p$  está na proximidade do  $\inf P\{H_0|x\}$ , desde que para o teste de  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ ,  $f(x-\theta)$  seja simétrica em relação a zero, tenha RVM e  $\pi(\theta)$  seja uma distribuição *a priori* de  $\theta$  simétrica em relação a zero. Permite também antever que a proximidade, ou até mesmo a igualdade das medidas Clássica e Bayesiana, é certamente possível para um grande número de casos, o que vai sem dúvida ao encontro da generalização do resultado obtido no exemplo considerado com modelo experimental Normal e distribuição *a priori* não informativa.

## 2.4 - Igualdade numérica entre o valor- $p$ e o $\inf P\{H_0|x\}$

Em certas condições, ou seja, se forem consideradas determinadas classes de distribuições *a priori* do parâmetro  $\theta$  e se  $f(x-\theta)$  satisfizer certos requisitos é possível estabelecer a igualdade numérica entre as medidas de evidência Clássica e Bayesiana.

Considere-se o teste de  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$  quando  $X$  tem função densidade de probabilidade  $f(x-\theta)$ , simétrica em relação a zero e com RVM. Admita-se que se realizou a observação  $X=x>0$  e que a classe de distribuições a priori do parâmetro  $\theta$  é  $\Gamma_{US} = \{\text{distribuições com densidades unimodais, simétricas em relação a zero}\}$ . Nestas condições, considerando a classe  $US = \{\text{distribuições uniformes simétricas em relação a zero}\}$ , é possível provar que,

$$(2.6) \quad \inf_{\pi \in \Gamma_{US}} P\{H_0|x\} = \inf_{\pi \in US} P\{H_0|x\} = p(x),$$

isto é, que se verifica igualdade numérica entre as medidas de evidência Clássica e Bayesiana.

Omitindo a demonstração do resultado [veja-se Casella e Berger(1987)] apresenta-se um exemplo em que se verifica (2.6).

Considere-se que  $X$  tem distribuição logística com função densidade de probabilidade,

$$f(x-\theta) = \exp\left\{\frac{-(x-\theta)}{\beta}\right\} / \beta \left[1 + \exp\left\{\frac{-(x-\theta)}{\beta}\right\}\right]^2,$$

$$-\infty < x < +\infty, -\infty < \theta < +\infty, \beta > 0.$$

Considere-se o teste  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$  e uma observação  $X=x$  ( $x>0$ ). Está-se em condições de aplicar resultado (2.6) na medida em que a função densidade de probabilidade da distribuição logística é simétrica em relação a zero e tem razão de verossimilhança monótona (RVM),

$$\begin{aligned}
p(x) = P\{X \geq x | \theta = 0\} &= \int_x^{\infty} \frac{\exp\left\{\frac{-(t-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{\frac{-(t-\theta)}{\beta}\right\}\right]^2} dt \\
&= \left[ \frac{\exp\left\{\frac{-x}{\beta}\right\}}{1 + \exp\left\{\frac{-x}{\beta}\right\}} \right]_x^{\infty} = \left[ 1 + \exp\left\{\frac{x}{\beta}\right\} \right]^{-1}
\end{aligned}$$

Por outra parte, se  $\pi(\theta) \in US$ , ou seja, se é uma distribuição uniforme em  $(-k, +k)$ , tem-se,

$$P\{H_0 | x\} = \frac{\int_{-k}^0 \frac{\exp\left\{\frac{-(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{\frac{-(x-\theta)}{\beta}\right\}\right]^2} d\theta}{\int_{-k}^k \frac{\exp\left\{\frac{-(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{\frac{-(x-\theta)}{\beta}\right\}\right]^2} d\theta},$$

donde,

$$\inf_{\pi \in US} P\{H_0 | x\} = \lim_{k \rightarrow \infty} \frac{\int_{-k}^0 \frac{\exp\left\{\frac{-(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{\frac{-(x-\theta)}{\beta}\right\}\right]^2} d\theta}{\int_{-k}^k \frac{\exp\left\{\frac{-(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{\frac{-(x-\theta)}{\beta}\right\}\right]^2} d\theta}$$

isto é,

$$\inf_{\pi \in US} P\{H_0|x\} = \frac{\int_{-\infty}^0 \frac{\exp\left\{-\frac{(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{-\frac{(x-\theta)}{\beta}\right\}\right]^2} d\theta}{\int_{-\infty}^{+\infty} \frac{\exp\left\{-\frac{(x-\theta)}{\beta}\right\}}{\beta \left[1 + \exp\left\{-\frac{(x-\theta)}{\beta}\right\}\right]^2} d\theta}$$

e finalmente,

$$\inf_{\pi \in US} P\{H_0|x\} = \left[ \frac{\exp\left\{-\frac{(x-\theta)}{\beta}\right\}}{1 + \exp\left\{-\frac{(x-\theta)}{\beta}\right\}} \right]_{-\infty}^0 = \left[ 1 + \exp\left\{\frac{x}{\beta}\right\} \right]^{-1}.$$

Concluindo-se então ser,

$$p(x) = \inf_{\pi \in US} P\{H_0|x\}$$

Como pela aplicação do lema 2.1,  $\inf_{\pi \in US} P\{H_0|x\} = \inf_{\pi \in \Gamma_{US}} P\{H_0|x\}$ , tem-se,

$$p(x) = \inf_{\pi \in \Gamma_{US}} P\{H_0|x\}$$

Apresenta-se seguidamente uma outra situação em que se verifica igualdade numérica entre as medidas de evidência Cássica e Bayesiana. Neste caso a classe de distribuições a priori do parâmetro  $\theta$  não será  $\Gamma_{US}$  mas sim  $\Gamma_{NOR}$ .

Seja então  $X \sim N(\theta, \sigma^2)$  uma variável aleatória Normal que tem, como é bem sabido, função densidade de probabilidade  $f(x-\theta)$  simétrica em relação à origem e RVM. Considere-se mais uma vez o teste da hipóteses  $H_0: \theta \leq 0$  versus

$H_1: \theta > 0$  com base numa observação  $X=x$  ( $x>0$ ). Suponha-se que  $\pi(\theta)$  é um elemento da classe  $\Gamma_{NOR}$  ou seja  $\pi(\theta) \equiv N(0, \tau^2)$ . Nestas condições é bem sabido que  $\theta|x \sim N[\mu(x), \rho^{-1}]$ , onde,

$$\mu(x) = \frac{x\tau^2}{\sigma^2 + \tau^2} \quad \text{e} \quad \rho^{-1} = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2} \quad ,$$

e se tem,

$$p(x) = P\{X \geq x | \theta = 0\} = P\left\{U \geq \frac{x}{\sigma}\right\} = 1 - \Phi\left(\frac{x}{\sigma}\right),$$

$$P\{H_0|x\} = P\{\theta \leq 0 | X = x\} = P\left\{U \leq -\frac{x\tau}{\sigma\sqrt{\sigma^2 + \tau^2}}\right\} = 1 - \Phi\left(\frac{x\tau}{\sigma\sqrt{\sigma^2 + \tau^2}}\right),$$

$$\inf_{\pi \in \Gamma_{NOR}} P\{H_0|x\} = \lim_{\tau^2 \rightarrow \infty} \left[1 - \Phi\left(\frac{x\tau}{\sigma\sqrt{\sigma^2 + \tau^2}}\right)\right] = 1 - \Phi\left(\frac{x}{\sigma}\right),$$

concluindo-se também neste caso que,

$$p(x) = \inf_{\pi \in \Gamma_{NOR}} P\{H_0|x\}.$$

Foram apresentados dois exemplos, sendo num deles considerada a classe de distribuições  $\Gamma_{US}$  e no outro a classe  $\Gamma_{NOR}$ . Verificou-se, tanto num como no outro, que  $\inf P\{H_0|x\} = p(x)$ . Para estes dois casos considerou-se  $f(x-\theta)$  como sendo simétrica e com RVM.

Para terminar a presente secção apresenta-se um Teorema que estabelece a conciliação entre a evidência Clássica e Bayesiana para todas as classes de distribuições *a priori* do parâmetro, que tal como  $\Gamma_{NOR}$ , consistem em transformações de escala de uma função densidade limitada, simétrica e unimodal.

Sendo  $f(x-\theta)$  função densidade de probabilidade simétrica, com RVM, demonstra-se que, para todas as classes de distribuições *a priori* acima referidas, se verifica igualdade numérica entre o valor-*p* e o  $\inf P\{H_0|x\}$ , o que sem dúvida traduz já, uma significativa generalização do caso particular apresentado



inicialmente (modelo experimental Normal e distribuição *a priori* não informativa).

**Teorema 2.1:** Seja  $g(\theta)$  uma distribuição *a priori* do parâmetro  $\theta$ , limitada, simétrica e unimodal. Considere-se a classe,

$$\Gamma^\sigma(g) = \{\pi_\sigma : \pi_\sigma(\theta) = (1/\sigma)g(\theta/\sigma), \sigma > 0\}.$$

Para o teste da hipótese  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ , se  $f(x-\theta)$  é simétrica em relação a zero, tem RVM e se  $x > 0$ , então

$$(2.7) \quad \inf_{\pi_\sigma \in \Gamma^\sigma(g)} P\{H_0|x\} = p(x).$$

Dem. Como  $\Gamma^\sigma(g) \subset \Gamma_S$ , tem-se por (2.1),

$$(2.8) \quad \inf_{\pi_\sigma \in \Gamma^\sigma(g)} P\{H_0|x\} \geq p(x).$$

Para se justificar a igualdade (2.7) terá que se provar ser,

$$\inf_{\pi_\sigma \in \Gamma^\sigma(g)} P\{H_0|x\} \leq p(x).$$

Com efeito,

$$(2.9) \quad \begin{aligned} \inf_{\pi_\sigma \in \Gamma^\sigma(g)} P\{H_0|x\} &\leq \lim_{\sigma \rightarrow \infty} P\{H_0|x\} = \\ &= \lim_{\sigma \rightarrow \infty} \frac{\int_{-\infty}^0 f(x-\theta)g(\theta/\sigma)d\theta}{\int_{-\infty}^{+\infty} f(x-\theta)g(\theta/\sigma)d\theta}. \end{aligned}$$

O facto de  $g$  ser limitada permite escrever,

$$\lim_{\sigma \rightarrow \infty} \frac{\int_{-\infty}^0 f(x-\theta)g(\theta/\sigma)d\theta}{\int_{-\infty}^{+\infty} f(x-\theta)g(\theta/\sigma)d\theta} = \frac{\int_{-\infty}^0 f(x-\theta)[\lim_{\sigma \rightarrow \infty} g(\theta/\sigma)]d\theta}{\int_{-\infty}^{+\infty} f(x-\theta)[\lim_{\sigma \rightarrow \infty} g(\theta/\sigma)]d\theta}.$$

Além disso, como  $g$  é simétrica e unimodal então,  $\lim_{\sigma \rightarrow \infty} g(\theta / \sigma) = g_0$ , limite que existe e é positivo. Assim,

$$\lim_{\sigma \rightarrow \infty} P\{H_0|x\} = \frac{\int_{-\infty}^0 f(x-\theta)g_0 d\theta}{\int_{-\infty}^{+\infty} f(x-\theta)g_0 d\theta} = \frac{\int_{-\infty}^0 f(x-\theta) d\theta}{\int_{-\infty}^{+\infty} f(x-\theta) d\theta} = p(x).$$

Por (2.8) e (2.9) fica provado o teorema.

## 2.5 - Comentários

Considerando, sem perda de generalidade, o teste  $H_0: \theta \leq 0$  versus  $H_1: \theta > 0$ , verificou-se que para determinadas classes de distribuições *a priori*, sendo  $f(x-\theta)$  simétrica em relação a zero e possuindo RVM), se tem igualdade numérica entre o valor- $p$  e o  $\inf P\{H_0|x\}$ . Para outras classes de distribuições *a priori*, sendo  $f(x-\theta)$  assimétrica ou sem RVM, pode apenas concluir-se que o valor- $p$  é um limite superior do  $\inf P\{H_0|x\}$ .

Pode então afirmar-se que para testes do tipo  $H_0: \theta \leq \theta_0$  versus  $H_1: \theta > \theta_0$ , o valor- $p$  está muito próximo das medidas Bayesianas de evidência. Infelizmente esta afirmação não pode manter-se quando passam a considerar-se testes bilaterais. Assim, Dickey(1977) afirma que para estes testes o valor- $p$  “exagera” a evidência contra  $H_0$ , mesmo quando comparado com o ínfimo das medidas Bayesianas de evidência. Na mesma ordem de ideias Berger e Sellke(1987) mostraram que nos testes bilaterais em que a hipótese nula é constituída apenas por um valor (*point null hypothesis*), o  $\inf P\{H_0|x\}$  é muito maior que o valor- $p$  para diferentes classes de distribuições *a priori*. Esta conclusão levou-os a declarar que os níveis de significância “... *can be highly misleading measures of the evidence provided by the data against the null hypothesis*”.

Como acabou de ver-se há situações em que os resultados são substancialmente diferentes. Por exemplo, no que diz respeito à conclusão de Dickey (1977) pode referir-se que em testes do tipo  $H_0: \theta \leq \theta_0$  versus  $H_1: \theta > \theta_0$  o valor- $p$  não “exagera” a evidência contra  $H_0$  pois  $p(x) \geq \inf P\{H_0|x\}$ . A questão que se põe é a de saber quais os factores que são essenciais na diferenciação dos dois problemas.

Esta questão é abordada nas secções finais do presente trabalho. No entanto, convém desde já por referir, que muita coisa se tem escrito sobre esta questão e que a polémica vai desde a razoabilidade de utilização do valor- $p$  como medida de evidência, até à afirmação de que a disparidade dos resultados se deve ao facto de a hipótese nula ser constituída apenas por um valor e não à bilateralidade do teste. Segundo Casella e Berger (1987), o que justifica esta diferença é o facto de alguma “*prior mass*” estar concentrada num único ponto, sendo a restante distribuída por  $H_1$ . Parece assim poder concluir-se que, na perspectiva destes autores, a diferença de resultados seria devida ao facto de a hipótese nula ser constituída apenas por um valor e não à bilateralidade do teste de hipóteses. Como reforço para esta justificação referem que, considerando o teste de hipóteses  $H_0: \theta = 0$  versus  $H_1: \theta > 0$ , Berger e Sellke (1987) obtiveram valores bastante diferentes para o valor- $p$  e  $\inf P\{H_0|x\}$ , devendo-se este facto à natureza da hipótese nula (constituída apenas por um valor) e não ao tipo de teste, que neste caso é um teste unilateral.

# CAPÍTULO 3

## SITUAÇÕES EM QUE NÃO HÁ CONCILIAÇÃO

### 3.1- Introdução

No Capítulo 2 houve oportunidade de verificar que, para testes unilaterais, ou seja, testes do tipo  $H_0: \theta \leq \theta_0$  versus  $H_1: \theta > \theta_0$ , há uma certa conciliação entre os resultados Clássicos e Bayesianos. No presente capítulo passam a considerar-se testes do tipo  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$  ou  $H_0: \theta = \theta_0$  versus  $H_1: \theta > \theta_0$ , quer dizer, testes cuja hipótese nula é constituída apenas por um valor  $\theta_0$ .

No primeiro ponto, procura mostrar-se que nestas situações os resultados Clássicos diferem dos Bayesianos, parecendo poder concluir-se que não há conciliação.

No segundo ponto, apresenta-se um pequeno historial relativo ao problema, que expressa bem a atenção que a questão tem despertado em muitos estudiosos.

No terceiro ponto, refere-se a situação conhecida por “Paradoxo de Jeffreys” ou “Paradoxo de Lindley” que evidencia, embora em casos particulares, a disparidade de resultados obtidos por Clássicos e Bayesianos.

Por último, mostra-se que os resultados a que chegaram Jeffreys e Lindley têm uma certa generalidade. São obtidos assim resultados semelhantes para diferentes modelos amostrais e diferentes classes de distribuição *a priori* do parâmetro desconhecido<sup>1</sup>.

---

<sup>1</sup> Tratam-se apenas casos paramétricos.

### 3.2 - Perspectiva histórica da inexistência de conciliação

O problema da inexistência de conciliação entre Clássicos e Bayesianos, no que diz respeito ao teste de hipóteses simples despertou a atenção de muitos estudiosos, sendo inúmeros os trabalhos que versam esta questão.

De entre os vários autores que trataram o problema, Berger e Delampady (1987) dão referências completas dos seguintes: Jeffreys (1957, 1961), Good (1950, 1958, 1965, 1967, 1983, 1985, 1986), Lindley (1957, 1961, 1965, 1977) Raiffa and Schlaiffer (1961), Edwards, Lindman and Savage (1963), Smith (1965), Pratt (1965), Zellner (1971, 1984), Dickey (1971, 1973, 1974, 1980), Lempers (1971), DeGroot (1973), Dempster (1973), Dickey (1977), Leamer (1978), Smith and Shiegelhalter (1980), Zellner and Siow (1980), Bernardo (1980), Hill (1982), Good (1984), Shafer (1982), Diamond and Forrester (1983) e Gómez and de la Horra Navarro (1984). Grande parte destes trabalhos versa o estudo da relação entre a probabilidade *a posteriori* da hipótese nula ( $P\{H_0|x\}$ ) e o valor- $p$ , sendo utilizadas determinadas distribuições *a priori* para o parâmetro desconhecido.

Mais tarde, a questão da inexistência de conciliação foi tratada pela análise da relação entre o valor- $p$  e os limites inferiores de  $P\{H_0|x\}$  e de  $B$  (factor Bayes). Pode encontrar-se este tipo de abordagem em Edwards, Lindman and Savage (1963), Hildreth (1963), Good (1967, 1983, 1984), Dickey (1973, 1977), Berger (1986), Berger and Sellke (1987), Casella and Berger (1987), Delampady (1986a, 1986b, 1986c) e Delampady and Berger (1987).

Em relação a esta problemática destacam-se as situações seguintes:

- **Jeffreys e Lindley**, abordam o problema da conciliação entre Bayesianos e Clássicos calculando a probabilidade *a posteriori* de  $H_0$  e comparando-a com o nível de significância ou valor- $p$ . Para fazer o cálculo consideraram distribuições *a priori* do parâmetro  $\theta$ , específicas, tais como a distribuição uniforme ou a distribuição Normal com variância igual à da variável aleatória  $X$ , representativa de determinada característica da população em causa.

Lindley(1957) refere, no entanto, que o paradoxo no essencial não é novo. A diferença entre as duas aproximações foi também apresentada por Jeffreys que, no entanto, se preocupou mais em evidenciar as semelhanças entre os seus resultados e os de Fisher do que em assinalar as diferenças.

•**Edwards, Lindman e Savage (1963)** merecem menção especial na medida em que foram os primeiros a utilizar os limites inferiores para  $P\{H_0|x\}$  e  $B$ (factor Bayes) e a orientar a sua pesquisa no sentido de obter alguma generalidade para o conflito evidenciado pelo paradoxo de Jeffreys-Lindley.

• **Berger and Sellke (1987)** insistem no facto de que o conflito evidenciado pelo paradoxo de Jeffreys-Lindley se verificar para diferentes classes de distribuições *a priori*, não constituindo portanto um resultado pontual. Preocupam-se em utilizar distribuições *a priori* “objectivas”, tendo sido pioneiros na escolha da classe,

$$G = \{ \text{densidades simétricas em relação a } \theta_0 \text{ e não crescentes em } |\theta - \theta_0| \}.$$

### 3.3 - Paradoxo de Jeffreys-Lindley

Este paradoxo tem designação associada aos nomes de Jeffreys e de Lindley pois tanto Lindley (1957) como Jeffreys (1961) evidenciaram o conflito entre as medidas Clássicas e Bayesianas de evidência em testes do tipo  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ . Qualquer deles escolheu uma população com função de distribuição normal, tendo considerado determinadas distribuições *a priori* para o parâmetro  $\theta$ . O primeiro autor admitiu que o parâmetro  $\theta$  tinha, *a priori*, distribuição uniforme e o segundo que o parâmetro se distribuía segundo uma Normal cuja variância igualava a do modelo experimental ( $\tau^2 = \sigma^2$ ).

### 3.3.1 Apresentação do paradoxo por Lindley.

Seja  $(x_1, x_2, \dots, x_n)$  uma realização de  $(X_1, X_2, \dots, X_n)$ , sendo  $X_i$  variáveis aleatórias i.i.d. com distribuição normal com valor médio  $\theta$  e variância conhecida  $\sigma^2$ . Lindley (1957) considera o teste da hipótese  $H_0: \theta = \theta_0$  contra a hipótese  $H_1: \theta \neq \theta_0$  e utiliza a seguinte distribuição *a priori* para o parâmetro  $\theta$ ,

$$(3.1) \quad \begin{cases} h(\theta) = \pi_0 & \text{se } \theta = \theta_0, \\ h(\theta) = (1 - \pi_0)g(\theta) & \text{se } \theta \neq \theta_0, \end{cases}$$

sendo  $g(\theta)$  uma distribuição uniforme. Significa então que atribui a probabilidade  $\pi_0$  ao evento  $\theta = \theta_0$  e distribui o resto da probabilidade  $(1 - \pi_0)$  pelo conjunto  $I = \{\theta: \theta \neq \theta_0\}$ . Nestas condições, a probabilidade *a posteriori* de  $\theta = \theta_0$ , considerando os dados amostrais  $(x_1, x_2, \dots, x_n)$ , é a seguinte,

$$(3.2) \quad P\{H_0|\bar{x}\} = \frac{\pi_0 \exp\left\{-\frac{n(\bar{x} - \theta_0)^2}{2\sigma^2}\right\}}{\pi_0 \exp\left\{-\frac{n(\bar{x} - \theta_0)^2}{2\sigma^2}\right\} + (1 - \pi_0) \int_I \exp\left\{-\frac{n(\bar{x} - \theta)^2}{2\sigma^2}\right\} d\theta}.$$

A expressão  $P\{H_0|\bar{x}\}$  resulta de se tomar, como é habitual, a observação da estatística suficiente mínima  $\bar{x}$  em vez da amostra  $(x_1, x_2, \dots, x_n)$ . Para o integral,

$$\int_I \exp\left[-n(\bar{x} - \theta)^2 / 2\sigma^2\right] d\theta,$$

pode tomar-se o valor proposto por Lindley,

$$\int_I \exp \left[ -n(\bar{x} - \theta)^2 / 2\sigma^2 \right] d\theta = \frac{\sigma\sqrt{2\pi}}{\sqrt{n}} \int_I \frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}} \exp \left[ -n(\bar{x} - \theta)^2 / 2\sigma^2 \right] d\theta ,$$

$$= \frac{\sigma\sqrt{2\pi}}{\sqrt{n}} .$$

Sendo assim,

$$P\{H_0|\bar{x}\} = \frac{\pi_0 \exp \left\{ -\frac{n(\bar{x} - \theta_0)^2}{2\sigma^2} \right\}}{\pi_0 \exp \left\{ -\frac{n(\bar{x} - \theta_0)^2}{2\sigma^2} \right\} + (1 - \pi_0) \frac{\sigma\sqrt{2\pi}}{\sqrt{n}}} .$$

A medida de evidência clássica utilizada em Lindley foi o nível de significância  $\alpha$  ou valor- $p$ . Nestas condições, tem-se,

$$(3.3) \quad P\{\bar{X} \geq \bar{x}\} = P\left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \geq \frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}}\right) = \alpha \Rightarrow \frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}} = \lambda_\alpha ,$$

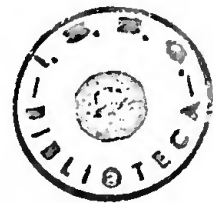
em que  $\lambda_\alpha$  é um número que depende de  $\alpha$  e pode ser encontrado nas tabelas da Normal [ $1 - \Phi(\lambda_\alpha) = \alpha$ ].

De (3.3) pode deduzir-se que:

$$\bar{x} = \theta_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}} \Leftrightarrow \bar{x} - \theta_0 = \lambda_\alpha \frac{\sigma}{\sqrt{n}}$$

Fazendo a substituição em (3.2) tem-se:





$$(3.4) \quad P\{H_0 | \bar{x}\} = \frac{\pi_0 \exp\left[-n \lambda_\alpha^2 \frac{\sigma^2}{n} / 2\sigma^2\right]}{\pi_0 \exp\left[-n \lambda_\alpha^2 \frac{\sigma^2}{n} / 2\sigma^2\right] + (1 - \pi_0) \frac{\sigma\sqrt{2\pi}}{\sqrt{n}}}$$
$$= \frac{\pi_0 \exp\left[-\frac{1}{2} \lambda_\alpha^2\right]}{\pi_0 \exp\left[-\frac{1}{2} \lambda_\alpha^2\right] + (1 - \pi_0) \frac{\sigma\sqrt{2\pi}}{\sqrt{n}}}$$

O nível de significância ou valor- $p$  é  $\alpha$  e isto significa, segundo os Clássicos, que  $\alpha$  é a medida da evidência que os dados  $(x_1, x_2, \dots, x_n)$ , através da estatística suficiente  $\bar{x}$ , dão da hipótese nula.

Se  $\alpha = 0.05$  significa que a probabilidade de obter um resultado tão mau ou pior que  $\bar{x}$ , é 0.05. Sendo  $\alpha$  pequeno, os Clássicos consideram que não têm condições para aceitar  $H_0$ . Por outro lado, os Bayesianos podem determinar um valor  $n$ , dependente de  $\pi_0$  e  $\alpha = 0.05$ , de tal forma que  $P\{H_0 | \bar{x}\} = 1 - \alpha = 0.95$ .

É assim evidente o conflito entre Bayesianos e Clássicos neste exemplo concreto. Quando  $\alpha = 0.05$  os Clássicos asseguram que existe forte evidência contra a hipótese nula  $H_0: \theta = \theta_0$ , enquanto os Bayesianos não têm motivo para rejeitar a hipótese  $H_0$  pois  $P\{H_0 | \bar{x}\} = 0.95$ .

De facto, repare-se no que acontece quando  $n \rightarrow \infty$ .

Se  $n \rightarrow \infty$  então, como pode verificar-se por (3.4) que  $P\{H_0 | \bar{x}\} \rightarrow 1$ , para  $\alpha$  e  $\pi_0$  fixos. Pode então concluir-se que, quando  $n \rightarrow \infty$ , a diferença entre o valor- $p$  e  $P\{H_0 | \bar{x}\}$  tem tendência a acentuar-se. É esta discrepância entre o valor- $p$  e  $P\{H_0 | \bar{x}\}$  quando  $n \rightarrow \infty$  que recebe a designação de paradoxo de Jeffreys - Lindley.

A questão que se põe seguidamente é a de saber se as condições admitidas à partida são razoáveis, isto é, se a estatística teste utilizada e a distribuição *a priori* do parâmetro  $\theta$  são aceitáveis.

No que diz respeito à estatística teste, existe um grande número de situações em que é assintoticamente Normal, com variância conhecida, o que faz com que o exemplo anteriormente considerado não seja despropositado.

Relativamente à distribuição *a priori* do parâmetro o conflito parece persistir com distribuições *a priori* mistas, isto é, distribuições *a priori* que permitam probabilidade positiva para o valor especificado pela hipótese ( $\theta = \theta_0$ ) e distribuam a restante probabilidade pelos valores da hipótese alternativa ( $H_1: \theta \neq \theta_0$ ). Lindley(1957) refere que, desde que a distribuição *a priori* tenha estas características e seja limitada,  $P\{H_0|\bar{x}\}$  tende para 1 quando  $n \rightarrow \infty$ .

Parece então que o exemplo considerado em Lindley (1957) não está demasiado particularizado pois segundo aquele autor podem obter-se resultados semelhantes para distribuições *a priori* distinta de (3.1), ou seja, neste artigo, encontra-se já a procura de uma certa generalidade para os resultados obtidos no caso concreto considerado.

Que fazer? Qual a medida da evidência mais indicada? Esta é a grande interrogação. Os Clássicos defendem a utilização do valor-*p* e os Bayesianos avançam com a probabilidade *a posteriori* da hipótese de nulidade mas, mesmo no seio destes dois grupos, o entendimento não é fácil. Há estatísticos que se dizem “Clássicos” e não consideram o valor-*p* como medida significativa de evidência e Bayesianos que consideram o valor-*p* como medida bastante útil.

Pela leitura dos comentários referentes aos resultados numéricos (3.3) e (3.4) parece poder concluir-se que o valor-*p* é mais aceitável que  $P\{H_0|\bar{x}\}$  na medida em que, para um valor fixo de  $\alpha$ , podem calcular-se diferentes valores para  $P\{H_0|\bar{x}\}$ .



Assim,  $P\{H_0|\bar{x}\}$  varia em função da dimensão da amostra o que causa uma certa “variabilidade” nas conclusões a tirar. Para um mesmo valor- $p$  que corresponde, por exemplo, à rejeição da hipótese nula, pode obter-se um determinado valor para  $P\{H_0|\bar{x}\}$ , numa pequena amostra, em determinado momento, e outro valor para  $P\{H_0|\bar{x}\}$  numa grande amostra, no momento seguinte. A não “variabilidade” do valor- $p$  associar-se-ia “confiança”, à “variabilidade” de  $P\{H_0|\bar{x}\}$  estaria associada “desconfiança”.

Perante esta aparente “fraqueza” da probabilidade *a posteriori* de  $H_0$ , os Bayesianos mostram ser (veja-se (3.4)),

$$P\{H_0|\bar{x}\} = \frac{\pi_0 \exp(-\frac{1}{2} \lambda_\alpha^2)}{\pi_0 \exp(-\frac{1}{2} \lambda_\alpha^2) + (1 - \pi_0) \sigma \sqrt{\frac{2\pi}{n}}},$$
$$= \frac{\pi_0 \exp(-\frac{1}{2} \lambda_\alpha^2) \sqrt{\frac{n}{2\pi\sigma^2}}}{\pi_0 \exp(-\frac{1}{2} \lambda_\alpha^2) \sqrt{\frac{n}{2\pi\sigma^2}} + (1 - \pi_0)}$$

donde, designando  $\exp(-\frac{1}{2} \lambda_\alpha^2) \sqrt{\frac{n}{2\pi\sigma^2}}$  por  $f_n$ , vem  $f_n \rightarrow \infty$  quando  $n \rightarrow \infty$ , sendo  $\lambda_\alpha$  fixo. Assim, para um nível de significância fixo, a função de verosimilhança da hipótese nula aumenta indefinidamente com o tamanho da amostra.

Uma vez que  $f_n$  quantifica a verosimilhança associada ao valor do parâmetro especificado pela hipótese da nulidade parece serem bem diferentes as situações em que se observa um determinado acontecimento e aquelas em que se consideram acontecimentos tão maus ou piores do que o acontecimento observado. Este fenómeno serve então de crítica ao valor- $p$ , na medida em que este é calculado, utilizando não só o acontecimento observado, mas todos aqueles que são piores do que ele. O que parecia constituir um “ponto fraco” de  $P\{H_0|\bar{x}\}$  é utilizado pelos Bayesianos para, de imediato, criticarem a própria essência do valor- $p$ . É no entanto de referir, relativamente ao desenvolvimento feito com base

na função de verosimilhança, que não foram considerados os valores alternativos de  $\theta$ , o que é importante numa análise deste tipo.

Analisada a apresentação do paradoxo de Jeffreys-Lindley, fundamentalmente baseada em Lindley (1957), segue-se a apresentação do mesmo paradoxo através de um exemplo em que se utiliza uma distribuição *a priori* tipo Jeffreys. A distribuição *a priori* neste exemplo está muito próxima da utilizada em Jeffreys (1961), embora presentemente Jeffreys aconselhe a utilização da Cauchy.

### 3.3.2 Apresentação do paradoxo por Jeffreys

Considere-se uma população em que se observa uma característica  $X$  com função densidade de probabilidade  $f(x|\theta)$ , Normal, sendo  $\theta$  o parâmetro desconhecido (neste caso o valor médio). Admita-se que a variância é conhecida. Pretende, mais uma vez, testar-se a hipótese  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ .

A distribuição *a priori* do parâmetro  $\theta$  a utilizar neste exemplo é a que foi considerada em Berger and Sellke (1987), sendo designada por estes como *priori* do tipo Jeffreys.

Para esta *priori*  $\pi_0 = 1/2$  mas, contrariamente ao exemplo utilizado por Lindley, o resto da probabilidade é distribuída no conjunto  $\{\theta: \theta \neq \theta_0\}$  segundo uma normal cujo valor médio é  $\theta_0$  mas de variância igual à variância da variável amostral, ou seja,  $g(\theta) \equiv N(\theta_0, \sigma^2)$ .

De uma forma geral,

$$(3.5) \quad P\{H_0 | x\} = \frac{\pi_0 f(x|\theta_0)}{m(x)},$$

onde  $m(x)$  é a distribuição preditiva ou densidade marginal de  $X$ ,

$$(3.6) \quad m(x) = f(x|\theta_0) \pi_0 + (1-\pi_0) m_g(x),$$

e

$$(3.7) \quad m_g(x) = \int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta.$$

Para a distribuição *a priori* considerada,

$$(3.8) \quad \begin{aligned} P\{H_0 | x\} &= \frac{\pi_0 f(x|\theta_0)}{f(x|\theta_0)\pi_0 + (1-\pi_0) \int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta}, \\ &= \left[ \frac{f(x|\theta_0)\pi_0 + (1-\pi_0) \int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta}{\pi_0 f(x|\theta_0)} \right]^{-1} \\ &= \left[ 1 + \frac{1-\pi_0}{\pi_0} \times \frac{m_g(x)}{f(x|\theta_0)} \right]^{-1} \end{aligned}$$

De interesse é também o factor Bayes que relaciona as vantagens *a posteriori* com as vantagens *a priori*. Por definição, o factor Bayes tem por expressão,

$$B_g(x) = \frac{\frac{P\{H_0|x\}}{1-P\{H_0|x\}}}{\frac{P\{H_0\}}{1-P\{H_0\}}}.$$

Considerando a distribuição *a priori* utilizada, tem-se,

$$\frac{P\{H_0|x\}}{1-P\{H_0|x\}} = \frac{\left[ 1 + \frac{1-\pi_0}{\pi_0} \times \frac{m_g(x)}{f(x|\theta_0)} \right]^{-1}}{1 - \left[ 1 + \frac{1-\pi_0}{\pi_0} \times \frac{m_g(x)}{f(x|\theta_0)} \right]^{-1}} = \frac{\pi_0}{1-\pi_0} \times \frac{f(x|\theta_0)}{m_g(x)},$$

donde,

$$(3.9) \quad B_g(x) = \frac{f(x|\theta_0)}{m_g(x)}.$$

Se a distribuição *a priori* for do tipo Jeffreys tem-se, considerando que se observa  $\bar{x}$ , tem-se,

$$B_g(\bar{x}) = \frac{f(\bar{x}|\theta_0)}{m_g(\bar{x})} = \frac{\left[2\pi \frac{\sigma^2}{n}\right]^{-1/2} \exp\left\{-\frac{n}{2}(\bar{x} - \theta_0)^2 / \sigma^2\right\}}{\left[2\pi\sigma^2(1+n^{-1})\right]^{-1/2} \exp\left\{-\frac{1}{2}(\bar{x} - \theta_0)^2 / [\sigma^2(1+n^{-1})]\right\}},$$

depois de notar que a distribuição marginal de  $\bar{X}$  é  $N(\theta_0, \frac{\sigma^2}{n} + \sigma^2)$ , ou seja,

$$m_g(\bar{x}) \equiv N(\theta_0, \frac{\sigma^2}{n} + \sigma^2).$$

Simplificando, vem,

$$(3.10) \quad B_g(\bar{x}) = (1+n)^{1/2} \exp\left\{-\frac{1}{2} \frac{(\bar{x} - \theta_0)^2}{\sigma^2/n} \frac{\sigma^2}{\sigma^2/n + \sigma^2}\right\},$$

$$= (1+n)^{1/2} \exp\left\{-\frac{1}{2} t^2 / (1+n^{-1})\right\},$$

com  $t = \sqrt{n}(\bar{x} - \theta_0) / \sigma$ , notação que se mantém em tudo o que segue.

Uma vez que  $B_g(\bar{x})$  é dado por (3.10), sendo  $\pi_0=1/2$ , tem-se,

$$(3.11) \quad P\{H_0|\bar{x}\} = \left[1 + \frac{1-\pi_0}{\pi_0} (B_g(\bar{x}))^{-1}\right]^{-1} = \left[1 + \frac{1}{B_g(\bar{x})}\right]^{-1}$$

$$= \left[1 + (1+n)^{-1/2} \exp\left\{\frac{1}{2} t^2 / (1+n^{-1})\right\}\right]^{-1}$$

São agora apresentados resultados numéricos obtidos a partir de (3.10) e (3.11). A tabela 3.1<sup>1</sup> representa alguns valores de  $P\{H_0|\bar{x}\}$  para valores de  $n$  e de  $t$ . Para cada valor de  $t$  é indicado o correspondente *valor-p*.

**Tabela 3.1**

		Dimensão da amostra ( $n$ )						
$t$	$p$	1	5	10	20	50	100	1000
1.645	0.10	0.42	0.44	0.47	0.56	0.65	0.72	0.89
1.960	0.05	0.35	0.33	0.37	0.42	0.52	0.60	0.82
2.576	0.01	0.21	0.13	0.14	0.16	0.22	0.27	0.53
3.291	0.001	0.086	0.026	0.024	0.026	0.034	0.045	0.124

Consultando a tabela pode verificar-se que o conflito entre o valor- $p$  e a probabilidade *a posteriori* da hipótese nula se verifica também neste caso. O paradoxo de Jeffreys é também evidente para este exemplo.

Para um determinado valor- $p$  pode verificar-se facilmente que, quando  $n \rightarrow \infty$ ,  $P\{H_0|\bar{x}\} \rightarrow 1$ ,

$$(3.12) \quad \lim_{n \rightarrow \infty} P\{H_0|\bar{x}\} = \lim_{n \rightarrow \infty} \left[ 1 + (1+n)^{-1/2} \exp\left\{\frac{1}{2}t^2 / (1+n^{-1})\right\} \right]^{-1} = 1.$$

Por exemplo, se  $n=100$  e  $t=1.645$  os Clássicos rejeitam a hipótese nula ao nível de 0.10 mas, por ser,  $P\{H_0|\bar{x}\} = 0.72$ , os Bayesianos não encontram justificação para rejeitar  $H_0$ . É assim visível que, para os mesmos dados amostrais, os Clássicos rejeitam a hipótese  $H_0$  não tendo os Bayesianos motivo para o fazer.

<sup>1</sup> Esta tabela foi recolhida de Berger e Sellke(1987).

Como Lindley (1957) já tinha referido, os resultados que obteve também se verificam noutras situações. O exemplo com uma *priori* do tipo Jeffreys (1961) insere-se no “quadro” previsto por Lindley na medida em que a variável aleatória  $X$  se distribui segundo uma Normal e a distribuição *a priori* tem a probabilidade  $\pi_0$  concentrada no ponto ( $\theta = \theta_0$ ) e a restante distribuída por  $\{\theta: \theta \neq \theta_0\}$  segundo a densidade de probabilidade  $N(\theta_0, \sigma^2)$ , que é limitada. Com um exemplo deste tipo, em Jeffreys (1961) confirmam-se os resultados obtidos por Lindley (1957).

Embora a discussão sobre estes resultados seja retomada posteriormente é desde já de referir que, em relação ao paradoxo de Jeffreys-Lindley, se discutem situações do tipo:

- A hipótese  $H_0: \theta = \theta_0$  é pouco realista pois raros são os casos em que se pretende ensaiar uma hipótese constituída por um só ponto. Se, muito raramente, se verifica uma situação deste tipo, não tem então justificação dar-lhe muita atenção no campo da Estatística. Opinião contrária têm os que, como Jeffreys (1961), referem que, na realidade, não são tão poucas as situações em que se pretendem ensaiar hipóteses do tipo  $H_0: |\theta - \theta_0| \leq \varepsilon$  que, para um valor de  $\varepsilon$  suficientemente pequeno, podem ser aproximados por  $H_0: \theta = \theta_0$ .

Afirmam que, apesar de serem consideradas por alguns como pouco realistas, são muito utilizadas na prática o que faz com que os estatísticos se devam preocupar com o seu estudo.

Relativamente a este assunto, Berger e Delampady (1987), depois de alguns cálculos, afirmam que a aproximação de uma hipótese do tipo  $H_0: |\theta - \theta_0| \leq \varepsilon$  por  $H_0: \theta = \theta_0$ , considerando uma população com distribuição Normal, é razoável, se a amplitude do intervalo for igual ou menor que metade do desvio padrão da amostra.

- O conflito expresso pelo paradoxo Jeffreys-Lindley é evidenciado pelo tamanho da amostra ( $\lim_{n \rightarrow \infty} P\{H_0 | \bar{x}\} = 1$ ).



Referem no entanto outros autores, nomeadamente Berger e Sellke (1987), que a discrepância entre as medidas Clássicas e Bayesianas é independente do tamanho da amostra.

- Sobre o facto de se ter considerado que a variância da distribuição *a priori* é igual à variância da população em causa. Berger (1985) refere que tem sentido questionar a escolha  $\tau = \sigma$ , mas que o problema pode ser ultrapassado calculando o ínfimo da probabilidade *a posteriori* de  $H_0$  sobre a classe das distribuições *a priori* Normais com variância  $\tau$ .

- A escolha de  $\pi_0 = \frac{1}{2}$  não traz grande polémica na medida em que se considera que uma escolha objectiva deverá dar igual peso às duas hipóteses, não favorecendo nenhuma delas. Existem no entanto defensores de que  $\pi_0$  deveria ser maior que  $\frac{1}{2}$  pois  $H_0$  corresponde muitas vezes à teoria estabelecida.

- O paradoxo agrava o conflito pois, quando  $n \rightarrow \infty$ ,  $P\{H_0|\bar{x}\} \rightarrow 1$ , mas, segundo Berger and Delampady (1987), não tem grande justificação o facto de esta questão gerar tanta polémica. Referem que, quando  $n \rightarrow \infty$ , não é possível aproximar a hipótese  $H_0^*: |\theta - \theta_0| \leq \varepsilon$  por  $H_0: \theta = \theta_0$  pois a variância da amostra tende para zero. Provam que, para que a aproximação fosse possível, a amplitude do intervalo referente à hipótese  $H_0^*$  deveria ser igual ou menor que metade do desvio padrão da amostra  $\left(\frac{\sigma}{\sqrt{n}}\right)$ , como aliás já foi dito. Se  $n \rightarrow \infty$  então  $\frac{\sigma}{\sqrt{n}} \rightarrow 0$ , o que invalida a aproximação. Assim, não sendo possível a aproximação considerar-se-ia a hipótese  $H_0^*: |\theta - \theta_0| \leq \varepsilon$  para a qual se verifica que,

$$P\{H_0|\bar{x}\} \rightarrow \alpha \text{ quando } n \rightarrow \infty.$$

Este último resultado está de acordo com os resultados a que chegaram, entre outros, Casella and Berger (1987) e que foram apresentados no capítulo 2 deste trabalho. Isto acontece porque, concentrando-se a função de verosimilhança

na extremidade do intervalo referente à hipótese nula (onde ela deve estar localizada para que o *valor-p* seja  $\alpha$ ), o intervalo nulo transforma-se num intervalo unilateral, verificando-se nestes casos uma certa conciliação entre os resultados Clássicos e Bayesianos.

### 3.4. - Procura de generalidade para o problema de inexistência de conciliação.

Em que situações mais gerais parece não haver conciliação entre as medidas Clássicas e Bayesianas de evidência? Mostra-se que o conflito referente ao paradoxo de Jeffreys-Lindley persiste, sendo independente da dimensão da amostra.

Os exemplos a considerar variam em função da distribuição de probabilidade da população em causa e da classe de distribuições *a priori* do parâmetro.

Para os casos que a seguir são apresentados considera-se o vector aleatório

$$(3.13) \quad (X_1, X_2, \dots, X_n),$$

em que  $X_i$  são i.i.d. com função densidade de probabilidade  $f(x|\theta)$ . O parâmetro  $\theta$  é um elemento desconhecido do intervalo  $\Theta = (-\infty, +\infty)$ .

Quando pretendem testar a hipótese  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$  e instituir uma distribuição *a priori* para o parâmetro  $\theta$  os Bayesianos deparam com o seguinte problema: se  $\Theta$  é escalar,  $\Theta = (-\infty, +\infty)$  e se a distribuição *a priori* do parâmetro  $\theta$  é contínua, então  $P\{H_0|x\} = P\{\theta = \theta_0|x\}$  é igual a zero  $\forall x$  em consequência de ser *a priori*  $P\{H_0\} = P\{\theta = \theta_0\} = 0$ . Nestas condições a hipótese  $H_0: \theta = \theta_0$  seria sempre rejeitada.

Com o objectivo de ultrapassar este obstáculo os Bayesianos referem que, uma vez que se considera a hipótese  $H_0: \theta = \theta_0$ , o valor  $\theta_0$  tem uma importância diferente dos outros valores de  $\theta$  e, sendo assim, consideram razoável atribuir *a priori* a probabilidade  $\pi_0 > 0$  a  $\theta = \theta_0$  e aos valores  $\theta \neq \theta_0$  a densidade de

probabilidade  $(1-\pi_0)g(\theta)$ , em que  $g(\theta)$  é uma função densidade de probabilidade própria. Ter-se-ia então uma distribuição *a priori* mista,

$$(3.14) \quad h(\theta) = \begin{cases} \pi_0 & \text{se } \theta = \theta_0, \\ (1-\pi_0)g(\theta) & \text{se } \theta \neq \theta_0. \end{cases}$$

Nos exemplos que a seguir são apresentados tem-se como objectivo a comparação das três medidas de evidência (*Valor-p*, *factor Bayes* ou *razão de verosimilhanças ponderada e probabilidade a posteriori de  $H_0$* ).

O factor Bayes ou razão de verosimilhança ponderada tem a sua expressão apresentada em (3.9). A utilização do factor Bayes como medida Bayesiana de evidência é bastante frequente, porquanto não só evita a escolha de  $\pi_0$  como permite de alguma forma estabelecer uma certa ligação com o campo clássico (pode ser visto como uma razão de verosimilhanças).

A probabilidade *a posteriori* de  $H_0$  tem a expressão matemática deduzida em (3.8). Por (3.9) pode concluir-se,

$$(3.15) \quad P\{H_0|x\} = \left[ 1 + \frac{1-\pi_0}{\pi_0} \times \frac{m_g(x)}{f(x|\theta_0)} \right]^{-1},$$

isto é,

$$(3.16) \quad P\{H_0|x\} = \left[ 1 + \frac{1-\pi_0}{\pi_0} \frac{1}{B_g(x)} \right]^{-1}.$$

Argumenta-se por vezes que a discrepância entre Clássicos e Bayesianos tem a sua explicação na distribuição *a priori* escolhida. Com o objectivo de evitar esta situação têm sido utilizados os limites inferiores de  $P\{H_0|x\}$  e  $B_g(x)$ , fazendo variar  $g(\theta)$ , a distribuição *a priori* quando  $H_1$  é verdadeira, numa ampla classe de distribuições.

Sendo assim, obtém-se o limite inferior do factor Bayes,

$$(3.17) \quad \underline{B}_g(x) = \inf_{g \in G} B_g(x) = \frac{f(x|\theta_*)}{\sup_{g \in G} m_g(x)}.$$

e o limite inferior da probabilidade *a posteriori*,

$$(3.18) \quad \underline{P}\{H_0|x\} = \inf_{g \in G} P\{H_0|x\} = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \cdot \frac{1}{\underline{B}_g(x)} \right]^{-1}.$$

Como já foi referido anteriormente, a procura de generalidade para a situação de conflito evidenciada pelo paradoxo de Jeffreys-Lindley passa pela verificação de que se podem obter resultados semelhantes utilizando diferentes modelos experimentais e diferentes classes de distribuições *a priori*.

As classes de distribuição *a priori* usadas nos exemplos que a seguir se apresentam são essencialmente as seguintes,

$G_A = \{ \text{todas as funções densidade de probabilidade} \};$

$G_{US} = \{ \text{densidades de probabilidade simétricas e não crescentes} \\ \text{em } |\theta - \theta_0| \};$

$G_{NOR} = \{ \text{densidades de probabilidade } N(\theta_0, \tau^2) \text{ em que } 0 < \tau^2 < \infty \};$

$C = \{ \text{densidades conjugadas da Binomial com valor médio } \theta_0 \};$

$M_e = \{ \text{densidades com média } \theta_0 \};$

$M_e M_o = \{ \text{densidades com média e mediana } \theta_0 \}.$

### 3.4.1 Limites inferiores para $G_A$

Considere-se, mais uma vez, que o vector aleatório  $(X_1, X_2, \dots, X_n)$  mencionado em (3.13) é uma amostra casual de uma população Normal. Pretende



ensaiar-se a hipótese  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ , sendo  $h(\theta)$ , tal como está definida em (3.14) com  $g(\theta) \in G_A$ , e comparar as medidas de evidência no sentido de saber o que fariam Clássicos e Bayesianos se tivessem que decidir entre as hipóteses  $H_0$  e  $H_1$ .

Assim, para uma dada concretização do amostra casual  $(X_1, X_2, \dots, X_n)$ , da qual se utiliza a estatística suficiente  $\bar{X}$  a que corresponde um determinado valor- $p$ , são calculados os limites inferiores de  $P\{H_0|\bar{x}\}$  e  $B_g(\bar{x})$  e finalmente comparados os resultados obtidos. Se,

$$P\{H_0|\bar{x}\} = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \times \frac{m_g(\bar{x})}{f(\bar{x}|\theta_0)} \right]^{-1},$$

então,

$$(3.19) \quad P\{H_0|\bar{x}\} \geq \left[ 1 + \frac{1 - \pi_0}{\pi_0} \frac{s(\bar{x})}{f(\bar{x}|\theta_0)} \right]^{-1},$$

em que,

$$(3.20) \quad s(\bar{x}) = \sup_{g \in G_A} m_g(\bar{x}) = \sup_{\theta \neq \theta_0} f(\bar{x}|\theta) = f(\bar{x}|\hat{\theta}).$$

onde  $\hat{\theta}$  é o estimador de máxima verosimilhança.

De facto,

$$(3.21) \quad \begin{aligned} m_g(\bar{x}) &= \int_{\theta \neq \theta_0} f(\bar{x}|\theta) g(\theta) d\theta \\ &\leq \int_{\theta \neq \theta_0} \left[ \sup_{\theta \neq \theta_0} f(\bar{x}|\theta) \right] g(\theta) d\theta \\ &= f(\bar{x}|\hat{\theta}) \int_{\theta \neq \theta_0} g(\theta) d\theta = f(\bar{x}|\hat{\theta}) \end{aligned}$$

porque,

$$\int_{\theta \neq \theta_0} g(\theta) d\theta = 1,$$

não exercendo o ponto  $\theta = \theta_0$  influência na integração na medida em que  $g(\theta)$  é uma densidade contínua.

Consequentemente por (3.20) e (3.21),

$$(3.22) \quad s(\bar{x}) = f(\bar{x}|\bar{x}) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} \exp\left\{\frac{\bar{x} - \bar{x}}{2\sigma^2/n}\right\} = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} = \{2\pi\sigma^2/n\}^{-1/2}.$$

Substituindo (3.22) em (3.19), tem-se,

$$(3.23) \quad P\{H_0|\bar{x}\} \geq \left[ 1 + \frac{1 - \pi_0}{\pi_0} \frac{\{2\pi\sigma^2/n\}^{-1/2}}{\{2\pi\sigma^2/n\}^{-1/2} \exp\{-(\bar{x} - \theta_0)^2 / (2\sigma^2/n)\}} \right]^{-1}$$

$$= \left[ 1 + \frac{1 - \pi_0}{\pi_0} \exp\left\{\frac{1}{2}t^2\right\} \right]^{-1}$$

e, ainda,

$$(3.24) \quad B_g(\bar{x}) \geq \frac{f(\bar{x}|\theta_0)}{\sup m_g(\bar{x})} = \frac{f(\bar{x}|\theta_0)}{s(\bar{x})}.$$

Por (3.23) e (3.24),

$$B_g(\bar{x}) \geq \frac{\{2\pi\sigma^2/n\}^{-1/2} \exp\{-\frac{(\bar{x}-\theta_0)^2}{2\sigma^2/n}\}}{\{2\pi\sigma^2/n\}^{-1/2}}$$

Consequentemente,

$$(3.25) \quad \underline{B}_g(\bar{x}) = \frac{1}{\exp\left\{\frac{1}{2}t^2\right\}}$$

Para  $\pi_0 = \frac{1}{2}$ , o que é visto pela maioria dos Bayesianos como uma solução objectiva na medida em que não favorece nenhuma das hipóteses, obtêm-se os valores da tabela 3.2<sup>1</sup> que a seguir se apresenta e onde se adopta notação mais conveniente para efeitos comparativos.

**Tabela 3.2**

$t$	$p$	$\underline{P}\{H_0 \bar{x}, G_A\}$	$\underline{B}\{\bar{x}, G_A\}$	$\underline{P}\{H_0 \bar{x}, G_A\}/pt$
1.645	0.10	0.205	1/3.87	1.25
1.960	0.05	0.127	1/6.83	1.30
2.576	0.01	0.035	1/27.6	1.36
3.291	0.001	0.0044	1/224.83	1.35

Por exemplo, para  $t=1.960$  podem ler-se na Tabela 3.2,

$$\underline{P}\{H_0|\bar{x}, G_A\} = \left[1 + \frac{1-1/2}{1/2} \exp\left\{\frac{1}{2}(1.960)^2\right\}\right]^{-1} = 0.127,$$

$$\underline{B}\{\bar{x}, G_A\} = \frac{1}{\exp\left\{\frac{1}{2}t^2\right\}} = \frac{1}{\exp\left\{\frac{1}{2}(1.960)^2\right\}} \cong \frac{1}{6.83},$$

<sup>1</sup> Os valores desta tabela foram recolhidos de Berger(1985) e Berger e Sellke(1987).

$$\underline{P}\{H_0|\bar{x}, G_A\} / pt = \frac{0.127}{0.05 \times 1.96} \cong 1.30.$$

Por observação da tabela 3.2 pode concluir-se,

- Os valores de  $\underline{P}\{H_0|\bar{x}, G_A\}$  são menores do que os da tabela 3.1 ( $P\{H_0|\bar{x}\}$  para uma densidade *a priori* tipo Jeffreys (1961)), sendo no entanto limites inferiores, o que não acontece com a tabela 3.1.

- Os valores de  $\underline{P}\{H_0|\bar{x}, G_A\}$  e  $\underline{B}\{\bar{x}, G_A\}$  (com a agravante de serem limites inferiores) são bastante maiores do que o respectivo valor-*p* o que significa que, também para este caso, se verifica discrepância entre as medidas Clássica e Bayesiana.

- A diferença entre o valor-*p* e  $\underline{P}\{H_0|\bar{x}, G_A\}$  e  $\underline{B}\{\bar{x}, G_A\}$  verifica-se, independentemente da dimensão da amostra.

- Para  $t = 1.960$  obtém-se  $p = 0.05$ . Um clássico rejeitaria  $H_0$  o que não se verificaria com um Bayesiano, para o qual,  $\underline{P}\{H_0|\bar{x}, G_A\} = 0.127$  não constitui evidência suficientemente forte contra  $H_0$ . Para a decisão na óptica Bayesiana contaria também, o facto, de se tratar de um limite inferior e de  $G_A$  ser o conjunto de todas as densidades *a priori*, uma classe demasiado ampla, da qual se podem obter resultados tendenciosos ou seja, demasiado dirigidos a  $H_1$ .

- Considerando o limite inferior do factor Bayes a discrepância é bem evidente pois para o mesmo  $t = 1.96$  obtém-se  $\underline{B}\{\bar{x}, G_A\} = 1/6.83$  o que significa que os dados transformam  $H_1$  numa hipótese apenas 6.83 vezes mais favorável do que  $H_0$ .

- Os valores da última coluna são mais estáveis verificando-se (conforme refere Berger(1985)), para  $t > 1.68$ ,

$$(3.26) \quad \underline{P}\{H_0|\bar{x}, G_A\} \geq 1.25pt.$$



O interesse deste resultado é poder concluir-se que  $\underline{P}\{H_0|\bar{x}, G_A\}$  é pelo menos igual a  $1.25pt$ , para qualquer densidade *a priori*, uma vez que este resultado se obtém considerando  $G_A$ .

### 3.4.2.- Limites inferiores em $G_S$

Na secção anterior admitiu-se que a densidade de probabilidade *a priori* do parâmetro  $\theta$  para  $H_1$  verdadeira podia variar no conjunto  $G_A = \{\text{todas as densidade de probabilidade}\}$ . Este facto pode dirigir os resultados parcialmente contra  $H_0$  na medida em que, sendo uma classe demasiado ampla pode permitir a minimização de  $P\{H_0|\bar{x}, G_A\}$  ou  $B_g\{\bar{x}, G_A\}$  para uma distribuição  $g(\theta)$  demasiado desfavorável a  $H_0$ . O facto de os resultados obtidos para as medidas condicionais de evidência em  $G_A$  serem muito menores do que os que se obtiveram para uma densidade *a priori* tipo Jeffreys permite sugerir a utilização de outras classes de distribuição *a priori*, menos amplas e, por conseguinte, menos tendenciosas em relação a  $H_1$ .

Uma vez que se continua a considerar uma população com função densidade Normal parece ter algum sentido considerar a classe  $G_S$ . São então calculadas e comparadas as medidas de evidência Clássica e Bayesiana, considerando esta classe de densidades *a priori*.

Nos cálculos a realizar é utilizada a classe  $G_{2PS} = \{\text{todas as distribuições simétricas constituídas por dois pontos}\}$  pois, devido ao lema 2.1 tem-se,

$$(3.27) \quad \underline{P}\{H_0|\bar{x}, G_{2PS}\} = \underline{P}\{H_0|\bar{x}, G_S\},$$

e

$$(3.28) \quad \underline{B}\{\bar{x}, G_{2PS}\} = \underline{B}\{\bar{x}, G_S\}.$$

Seja então  $(X_1, X_2, \dots, X_n)$  a variável amostral em que  $X_i \sim N(\theta, \sigma^2)$ , com  $X_i$  i.i.d.. A distribuição *a priori* do parâmetro é, como nos casos anteriores, uma distribuição mista do tipo da apresentada em (3.14) sendo  $g(\theta) \in G_S$ .

Pretende mais uma vez testar-se a hipótese  $H_0: \theta = \theta_0$  versus  $\theta \neq \theta_0$ . São seguidamente deduzidas as expressões para  $\underline{P}\{H_0|\bar{x}, G_S\}$  e  $\underline{B}\{\bar{x}, G_S\}$  e, tal como no exemplo anterior, comparados estes valores com o valor- $p$ .

Por (3.17), vem,

$$(3.29) \quad \underline{B}\{\bar{x}, G_{2PS}\} = \inf_{g \in G_{2PS}} B_g(\bar{x}) = \frac{f(\bar{x}|\theta_0)}{\sup_{g \in G_{2PS}} m_g(\bar{x})}.$$

Se  $t \leq 1$ , a distribuição *a priori*  $g \in G_{2PS}$  que maximiza  $m_g(\bar{x})$  é degenerada e atribui a totalidade da probabilidade a  $\theta_0$ . Sendo assim, tem-se,

$$(3.30) \quad \underline{B}\{\bar{x}, G_{2PS}\} = \frac{f(\bar{x}|\theta_0)}{f(\bar{x}|\theta_0)g(\theta_0)} = 1,$$

porque  $g(\theta_0) = 1$ , e, por (3.18), vem,

$$(3.31) \quad \underline{P}\{H_0|\bar{x}, G_{2PS}\} = \left[1 + \frac{1 - \pi_0}{\pi_0}\right]^{-1} = \left[\frac{1}{\pi_0}\right]^{-1} = \pi_0.$$

Por (3.30) e (3.31) pode concluir-se que, para  $t \leq 1$ , os Bayesianos aceitariam sempre a hipótese  $H_0$ . Tem então interesse analisar o que acontece quando  $t > 1$ .

Neste caso,  $m_g(\bar{x})$  é maximizada por uma distribuição  $g(\theta) \in G_{2PS}$  não degenerada. Esta distribuição atribui probabilidade  $\frac{1}{2}$  a cada um dos valores  $\hat{\theta}(x)$  e  $2\theta_0 - \hat{\theta}(x)$ .

Assim sendo,

$$\begin{aligned}
\underline{B}\{\bar{x}, G_{2PS}\} &= \frac{f(\bar{x}|\theta_0)}{\sup_{g \in G_{2PS}} m_g(\bar{x})} \\
(3.32) \quad &= \frac{\frac{1}{\sqrt{2\pi\sigma/\sqrt{n}}} \exp\left\{-\frac{(\bar{x}-\theta_0)^2}{2\sigma^2/n}\right\}}{\frac{1}{\sqrt{2\pi\sigma/\sqrt{n}}} \exp\left\{-\frac{(\bar{x}-\bar{x})^2}{2\sigma^2/n}\right\} + \frac{1}{\sqrt{2\pi\sigma/\sqrt{n}}} \exp\left\{-\frac{(\bar{x}-2\theta_0+\bar{x})^2}{2\sigma^2/n}\right\}} \\
&= \frac{\varphi(t)}{\frac{1}{2}\varphi(0) + \frac{1}{2}\varphi(2t)}
\end{aligned}$$

Por (3.28), vem,

$$(3.33) \quad \underline{B}\{\bar{x}, G_s\} = \frac{\varphi(t)}{\frac{1}{2}\varphi(0) + \frac{1}{2}\varphi(2t)}.$$

Por (3.18) e (3.27), tem-se,

$$(3.34) \quad \underline{P}\{H_0|\bar{x}, G_s\} = \left[ 1 + \frac{1-\pi_0}{\pi_0} \frac{\frac{1}{2}\varphi(0) + \frac{1}{2}\varphi(2t)}{\varphi(t)} \right]^{-1}.$$

Na tabela 3.3<sup>1</sup>, podem encontrar-se, para determinado valor de  $t$  observado os respectivos, valor- $p$ ,  $\underline{P}\{H_0|\bar{x}, G_s\}$  e  $\underline{B}\{\bar{x}, G_s\}$ .

---

<sup>1</sup> Alguns dados desta tabela foram recolhidos de Berger e Sellke(1987)

Tabela 3.3

$t$	$p$	$\underline{P}\{H_0 \bar{x}, G_S\}$	$\underline{B}\{\bar{x}, G_S\}$
1.645	0.10	0.340	0.5146
1.960	0.05	0.227	0.2928
2.576	0.01	0.068	0.0725
3.291	0.001	0.0088	0.0089

Apresentem-se, como exemplo, os cálculos elaborados admitindo que se fez a observação  $t = 3.291$  e se considerou  $\pi_0 = \frac{1}{2}$ ,

$$\underline{P}\{H_0|\bar{x}, G_S\} = \left[ 1 + \frac{1/2 \frac{1}{2} \varphi(0) + \frac{1}{2} \varphi(6.582)}{1/2 \varphi(3.291)} \right]^{-1} \cong 0.0088$$

$$\underline{B}\{\bar{x}, G_S\} = \frac{\varphi(t)}{\frac{1}{2} \varphi(0) + \frac{1}{2} \varphi(2t)} = \frac{\varphi(3.291)}{\frac{1}{2} \varphi(0) + \frac{1}{2} \varphi(6.582)} \cong 0.00889$$

A observação da tabela 3. 3 permite concluir,

- A discrepância entre o valor- $p$  e as medidas Bayesianas de evidência ( $\underline{P}\{H_0|\bar{x}, G_S\}$  e  $\underline{B}\{\bar{x}, G_S\}$ ) continua a verificar-se ou seja, para a mesma observação  $t$ , as medidas condicionais de evidência são maiores que o valor- $p$ .
- Os valores de  $\underline{P}\{H_0|\bar{x}, G_S\}$  e  $\underline{B}\{\bar{x}, G_S\}$  são maiores do que  $\underline{P}\{H_0|\bar{x}, G_A\}$  e  $\underline{B}\{\bar{x}, G_A\}$  e isto, tal como se previa, devido ao facto de a classe de distribuição a priori  $G_S$  ser menos ampla. Nestas condições a maximização de  $m_g(\bar{x})$  não é tão parcialmente dirigida a favor de  $H_1$ .

### 3.4.3 - Limites inferiores em $G_{US}$ .

Com o objectivo de minimizar a inclinação em relação a  $H_1$ , os Bayesianos tentam definir da forma mais objectiva possível a distribuição *a priori* das credibilidades atribuídas aos diferentes valores do parâmetro.

Embora alguns estatísticos considerem que é muito difícil estabelecer objectivamente uma distribuição *a priori* pois ela é, em essência, uma distribuição subjectiva de credibilidades, Berger e Sellke(1987) foram pioneiros em relação ao facto de considerar que a classe  $G_{US} = \{\text{densidades unimodais simétricas}\}$  é bastante razoável em termos de objectividade ou seja, é uma classe onde não são favorecidos, como acontecia em classes anteriores, os valores alternativos a  $\theta = \theta_0$ .

Considere-se, mais uma vez, que se observa uma amostra casual de uma população  $N(\theta, \sigma^2)$  e se tem como objectivo ensaiar  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ .

A distribuição das credibilidades *a priori* é do tipo da apresentada em (3.14), com  $g(\theta) \in G_{US}$ , o que significa que a função densidade de probabilidade quando  $H_1$  é verdadeira tem que ser simétrica e unimodal, o mesmo é dizer simétrica e não crescente em  $|\theta - \theta_0|$ .

Tal como nas situações anteriores são calculados, para um dado valor  $t$  observado, o valor- $p$  e as medidas condicionais de evidência, com vista a serem comparadas para concluir se, também para este caso, se verifica falta de “harmonia” entre Clássicos e Bayesianos.

Se  $t \leq I$ , tem-se, por ser  $G_{US} \subset G_S$ ,

$$\underline{B}\{\bar{x}, G_{US}\} = 1 \text{ e } \underline{P}\{H_0 | \bar{x}, G_{US}\} = \pi_0 .$$

Logo, e tal como já tinha sido visto no caso anterior ( $G_S$ ), se  $t \leq I$  os Bayesianos são sempre conduzidos a concluir que os dados não contrariam a hipótese  $H_0$  ou seja, não rejeitam esta hipótese qualquer que seja a distribuição  $g(\theta) \in G_{US}$ .

Se  $t > I$  e uma vez que,

$$\underline{B}\{\bar{x}, US\} = \frac{f(\bar{x}|\theta_0)}{\sup_{g \in US} m_g(\bar{x})},$$

é importante conhecer-se a distribuição  $g(\theta) \in US$  que maximiza  $m_g(\bar{x})$ , onde  $US$  é a classe das distribuições uniformes simétricas. Considere-se que  $g(\theta)$  é a distribuição uniforme no intervalo  $(\theta_0 - K\sigma/\sqrt{n}, \theta_0 + K\sigma/\sqrt{n})$  para algum  $K > 0$ . A escolha da classe  $US$  deve-se a que  $\underline{B}\{\bar{x}, US\} = \underline{B}\{\bar{x}, G_{US}\}$  devido ao lema 2.1;

$$\begin{aligned} m_g(\bar{x}) &= \int_{\theta_0 - K\sigma/\sqrt{n}}^{\theta_0 + K\sigma/\sqrt{n}} f(\bar{x}|\theta)g(\theta)d\theta \\ &= \frac{1}{2K\sigma/\sqrt{n}} \int_{\theta_0 - K\sigma/\sqrt{n}}^{\theta_0 + K\sigma/\sqrt{n}} f(\bar{x}|\theta)d\theta \\ (3.35) \quad &= \frac{\sqrt{n}}{2K\sigma} P\left\{\theta_0 - \frac{K\sigma}{\sqrt{n}} < \theta < \theta_0 + \frac{K\sigma}{\sqrt{n}}\right\}, \\ &= \frac{\sqrt{n}}{\sigma} \frac{1}{2K} [\Phi(K-t) - \Phi(-(K+t))]. \end{aligned}$$

O valor de  $K$  que maximiza  $m_g(\bar{x})$  verifica a equação,

$$\frac{dm_g(\bar{x})}{dK} = \frac{\sqrt{n}}{\sigma} \frac{[\varphi(K-t) + \varphi(-(K+t))]2K - 2[\Phi(K-t) - \Phi(-(K+t))]}{(2K)^2} = 0.$$

Tem-se, então,

$$(3.36) \quad K = \frac{\Phi(K-t) - \Phi(-(K+t))}{\varphi(K-t) + \varphi(-(K+t))}.$$

e repare-se que,

$$f(x|\theta_0) = \frac{\sqrt{n}}{\sigma} \varphi\left(\frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}}\right) = \frac{\sqrt{n}}{\sigma} \varphi(t).$$

Assim, se  $t > 1$  e  $K$  maximiza  $m_g(\bar{x})$ , vem,

$$(3.37) \quad \underline{B}\{\bar{x}, US\} = \frac{f(\bar{x}|\theta_0)}{\sup_{g \in US} m_g(\bar{x})} = \frac{(\sqrt{n}/\sigma)\varphi(t)}{(\sqrt{n}/\sigma) \frac{1}{2t} K [\varphi(K+t) + \varphi(K-t)]}$$

$$= \frac{2\varphi(t)}{\varphi(K+t) + \varphi(K-t)}.$$

em que  $K > 0$  satisfaz (3.36). Enfim, recorrendo mais uma vez ao lema 2.1, tem-se que  $\underline{B}(\bar{x}, G_{US})$  também verifica (3.37).

Fazendo a substituição de (3.37) na expressão (3.18) vem,

$$(3.38) \quad \underline{P}\{H_0|\bar{x}, G_{US}\} = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \times \frac{\varphi(k+t) + \varphi(k-t)}{2\varphi(t)} \right]^{-1},$$

em que  $K > 0$  satisfaz (3.36).

Para  $t \geq 1.645$  pode obter-se uma boa aproximação de  $K$  utilizando a fórmula iterativa,

$$K_{i+1} = t + \left\{ 2 \log \left[ \frac{K_i}{\varphi(K_i - t)} \right] - 1.838 \right\}^{1/2}.$$

O processo iterativo deve iniciar-se com  $K_0 = t$  e ao fim de duas ou três iterações tem tendência a estabilizar.

Na tabela 3.4<sup>1</sup> indicam-se os valores referentes ao valor- $p$  e medidas condicionais de evidência para os valores de  $t$  que têm sido considerados neste capítulo. Considera-se que  $\pi_0 = \frac{1}{2}$ .

<sup>1</sup> Esta tabela foi recolhida de Berger(1985)



Tabela 3.4

$t$	$p$	$\underline{P}\{H_0 \bar{x}, G_{US}\}$	$\underline{B}\{x, G_{US}\}$
1.645	0.10	0.390	1/1.56
1.960	0.05	0.290	1/2.45
2.576	0.01	0.109	1/8.17
3.291	0.001	0.018	1/54.56

Como exemplo considere-se  $t=1.645$ . Uma boa aproximação do valor para  $K$  que maximiza  $m_g(\bar{x})$  obtem-se em três iterações:

$$K_0 = 1.645$$

$$K_1 = 1.645 + \left[ 2 \log(1.645 / \Phi(0)) - 1.838 \right]^{1/2} \cong 2.3824$$

$$K_2 = 1.645 + \left[ 2 \log(2.3824 / \Phi(2.3824 - 1.645)) - 1.838 \right]^{1/2} \cong 2.2930$$

$$K_3 = 1.645 + \left[ 2 \log(2.2930 / \Phi(2.2930 - 1.645)) - 1.838 \right]^{1/2} \cong 2.2915$$

Para  $K \cong 2.29$ , tem-se,

$$\underline{P}\{H_0|\bar{x}, G_{US}\} = \left[ 1 + \frac{1/2}{1/2} \frac{\varphi(2.29 + 1.645) + \varphi(2.29 - 1.645)}{2\varphi(1.645)} \right]^{-1} \cong 0.390$$

$$\begin{aligned} \underline{B}\{H_0|\bar{x}, G_{US}\} &= \frac{2\varphi(t)}{\varphi(k+t) + \varphi(k-t)} = \frac{2\varphi(1.645)}{\varphi(2.29 + 1.645) + \varphi(2.29 - 1.645)} = \\ &\cong 0.639 \cong \frac{1}{1.56} \end{aligned}$$

Analisando os resultados obtidos pode concluir-se que continua a existir falta de “harmonia” entre Clássicos e Bayesianos. Como pode verificar-se, para cada um dos valores de  $t$  observados o valor- $p$  correspondente continua a ser bastante inferior às medidas condicionais de evidência.

Com o objectivo de minimizar a inclinação da distribuição das credibilidades *a priori* em relação a  $H_1$  escolheu-se  $G_{US}$ , o que fez com que os limites inferiores das medidas Bayesianas fossem ligeiramente maiores do que os calculados



para as classes  $G_A$  e  $G_S$ . Assim, a utilização de uma classe de distribuição *a priori* mais “objectiva” veio realçar a discrepância que tem vindo a estudar-se.

#### 3.4.4 - Limites inferiores em $G_{NOR}$ .

Considere-se mais uma vez uma variável amostral  $N(\theta, \sigma^2)$ , tendo-se *a priori* uma distribuição mista com  $g(\theta) \in G_{NOR}$ . São agora comparados o valor- $p$  e as medidas condicionais de evidência quando minimizadas em  $G_{NOR} = \{ \text{densidades } N(\theta_0, \tau^2) : 0 < \tau^2 < +\infty \}$ , respectivamente  $\underline{P}\{H_0|\bar{x}, G_{NOR}\}$  e  $\underline{B}\{\bar{x}, G_{NOR}\}$ . Por último são comparados estes valores com os obtidos para as classes de distribuição *a priori* consideradas anteriormente.

A utilização da classe  $G_{US}$  teve como objectivo a tentativa de minimização da “inclinação” a favor de  $H_1$  que classes como  $G_A$  ou  $G_S$  podem proporcionar. Para os cálculos de  $\underline{P}\{H_0|\bar{x}, G_{US}\}$  e  $\underline{B}\{\bar{x}, G_{US}\}$  utilizou-se a classe  $US$  com base no lema 2.1. É agora utilizada a classe  $G_{NOR}$  pois as credibilidades *a priori* são certamente mais frequentemente distribuídas segundo uma Normal do que uniformemente.

Se  $t \leq 1$ ,

$$G_{NOR} \subset G_{US} \subset G_S \Rightarrow \underline{B}\{\bar{x}, G_{NOR}\} = 1, \underline{P}\{H_0|\bar{x}, G_{NOR}\} = \pi_0,$$

Se  $t > 1$ ,

$$\underline{B}\{\bar{x}, G_{NOR}\} = \frac{f(\bar{x}|\theta_0)}{\sup_{g \in G_{NOR}} m_g(\bar{x})},$$

isto é,

(3.39)

$$\underline{B}\{\bar{x}, G_{NOR}\} = \frac{\left[ 2\pi\sigma^2 / n \right]^{-1/2} \exp\left\{ -\frac{n}{2}(\bar{x} - \theta_0)^2 / \sigma^2 \right\}}{\sup_{g \in G_{NOR}} \left\{ \left[ 2\pi \left( \frac{\sigma^2}{n} + \tau^2 \right) \right]^{-1/2} \exp\left\{ -\frac{1}{2}(\bar{x} - \theta_0)^2 / \left( \frac{\sigma^2}{n} + \tau^2 \right) \right\} \right\}},$$

porque  $\bar{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right)$  e  $m_g(\bar{x}) \equiv N\left(\theta_0, \frac{\sigma^2}{n} + \tau^2\right)$ .

Para determinar o valor de  $\tau$  que maximiza  $m_g(\bar{x})$  tem que resolver-se a equação  $\frac{\partial}{\partial \tau} m_g(\bar{x}) = 0$ ,

$$\frac{\partial}{\partial \tau} m_g(\bar{x}) = \frac{-2\pi\tau}{\sqrt{(2\pi)^3 \left(\frac{\sigma^2}{n} + \tau^2\right)^3}} + \frac{1}{\sqrt{2\pi \left(\frac{\sigma^2}{n} + \tau^2\right) \left(\frac{\sigma^2}{n} + \tau^2\right)^2}} \frac{(\bar{x} - \theta_0)^2 \tau}{\left(\frac{\sigma^2}{n} + \tau^2\right)^2} = 0$$

$$\Rightarrow -2\pi\tau \left[ \left(\frac{\sigma^2}{n} + \tau^2\right) - (\bar{x} - \theta_0)^2 \right] = 0 \quad \Rightarrow \tau \begin{cases} = 0 \\ = \sqrt{(\bar{x} - \theta_0)^2 - \frac{\sigma^2}{n}} \end{cases}$$

Substituindo em (3.39) vem,

(3.40)

$$\underline{B}\{\bar{x}, G_{NOR}\} = \frac{[2\pi\sigma^2/n]^{-1/2} \exp\left\{-\frac{n}{2}(\bar{x} - \theta_0)^2 / \sigma^2\right\}}{\left[2\pi\left(\frac{\sigma^2}{n} + (\bar{x} - \theta_0)^2 - \frac{\sigma^2}{n}\right)\right]^{-1/2} \exp\left\{-\frac{1}{2}(\bar{x} - \theta_0)^2 / (\bar{x} - \theta_0)^2\right\}}$$

$$= \frac{\bar{x} - \theta_0}{\sigma / \sqrt{n}} e^{1/2} e^{-\frac{t^2}{2}} = \sqrt{e} t e^{-t^2/2}$$

e

$$(3.41) \quad \underline{P}\{H_0 | \bar{x}, G_{NOR}\} = \left[1 + \frac{1 - \pi_0}{\pi_0} \frac{1}{\sqrt{e} t e^{-t^2/2}}\right]^{-1} = \left[1 + \frac{1 - \pi_0}{\pi_0} \frac{e^{t^2/2}}{\sqrt{e} t}\right]^{-1}$$

Na tabela 3.5<sup>1</sup> estão indicados o valor- $p$ ,  $\underline{P}\{H_0 | \bar{x}, G_{NOR}\}$  e  $\underline{B}\{\bar{x}, G_{NOR}\}$  para cada um dos valores  $t$ , observados. Considera-se que  $\pi_0 = \frac{1}{2}$ .

<sup>1</sup> Alguns valores desta tabela foram recolhidos de Berger e Sellke(1987).

Tabela 3.5

$t$	$p$	$\underline{P}\{H_0 \bar{x}, G_{NOR}\}$	$\underline{B}\{\bar{x}, G_{NOR}\}$
1.645	0.10	0.412	0.7010
1.960	0.05	0.321	0.4734
2.576	0.01	0.133	0.1539
3.291	0.001	0.0235	0.02413

Como exemplo, para  $t = 1.960$ , vem,

$$\underline{P}\{H_0|\bar{x}, G_{NOR}\} = \left[ 1 + \frac{1/2}{1/2} \frac{e^{-\frac{(1.960)^2}{2}}}{1.960\sqrt{e}} \right]^{-1} \cong 0.321,$$

$$\underline{B}\{\bar{x}, G_{NOR}\} = \sqrt{e} t e^{-t^2/2} = 1.960\sqrt{e} e^{-\frac{(1.960)^2}{2}} \cong 0.4734.$$

Analisando os resultados da tabela 3.5 e comparando-os com os exemplos anteriores pode concluir-se que estes se aproximam muito dos limites inferiores obtidos para a classe  $G_{US}$ . Por outro lado, continua a verificar-se grande discrepância entre o valor- $p$  e as medidas Bayesianas de evidência.

Em todos os exemplos considerados supôs-se que  $(X_1, \dots, X_n)$  era uma amostra casual de uma população Normal. Pretendia ensaiar-se a hipótese  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ , tendo as credibilidades *a priori* uma distribuição mista, que atribuía a probabilidade  $\pi_0$  a  $\theta = \theta_0$  e distribuía a restante por  $\{\theta: \theta \neq \theta_0\}$ , de acordo com a função densidade de probabilidade  $(1-\pi_0)g(\theta)$ . As diferentes situações consideradas devem-se à escolha de diferentes classes de distribuição *a priori*.

Como já foi mencionado anteriormente os casos (Lindley (1957) e Jeffreys (1961)) que evidenciam o conflito, designado por paradoxo de Jeffreys-Lindley, agravam-no, na medida em que, quando,  $n \rightarrow \infty, P\{H_0|\bar{x}\} \rightarrow 1$ , o que significa que, quando  $n$  aumenta, a discrepância entre o valor- $p$  e a probabilidade *a posteriori* de  $H_0$  vai sendo cada vez maior.

Os quatro exemplos que se seguiram serviram para conferir alguma generalidade ao paradoxo de Jeffreys-Lindley e para mostrar que a falta de harmonia não depende exclusivamente da dimensão da amostra.

O cálculo da probabilidade *a posteriori* da hipótese nula deu lugar ao cálculo do limite inferior do factor Bayes pois, com este, para além de não ser necessário definir  $\pi_0$  ainda é possível estabelecer a “ponte” pois o factor Bayes pode ser visto como uma razão de verosimilhanças, neste caso, ponderada pela função densidade *a priori* de  $\theta$  quando  $H_1$  é verdadeira.

Surge assim a minimização das medidas Bayesianas em diferentes classes de distribuições a priori  $G_A$ ,  $G_S$ ,  $G_{US}$  e  $G_{NOR}$ , continuando a verificar-se em todas elas discrepância entre o valor- $p$  e as medidas condicionais de evidência.

Calcularam-se  $\underline{P}\{H_0|\bar{x}, G_A\}$  e  $\underline{B}\{\bar{x}, G_A\}$  e verificou-se que estes valores (embora maiores que o valor- $p$ ) poderiam resultar de uma escolha pouco objectiva para a distribuição das credibilidades *a priori*. A classe  $G_A$  é muito ampla, podendo portanto incluir distribuições  $g(\theta)$  demasiado “inclinadas” em relação a  $H_1$ , o que influenciaria certamente os resultados.

Com vista à escolha, o mais imparcial possível, da distribuição das credibilidades a priori, foi-se restringindo a classe  $G_A$  de modo a considerar  $G_S$ ,  $G_{US}$  e  $G_{NOR}$  não se verificando entre as duas últimas grande diferença a nível dos resultados obtidos.

É ainda de referir que a utilização de  $G_{NOR}$  tem vantagem em relação a  $G_{US}$  pois, é mais frequente uma distribuição de credibilidades Normal do que Uniforme – a distribuição usada para o cálculo de  $\underline{P}\{H_0|\bar{x}, G_{US}\}$  e  $\underline{B}\{\bar{x}, G_{US}\}$ . Por outro lado as expressões para  $\underline{P}\{H_0|\bar{x}, G_{NOR}\}$  e  $\underline{B}\{\bar{x}, G_{NOR}\}$  são de mais fácil utilização.

Berger e Sellke(1987) apresentam um gráfico em que comparam os limites inferiores do factor Bayes para as quatro classes de distribuição *a priori* utilizadas ( $G_A$ ,  $G_S$ ,  $G_{US}$  e  $G_{NOR}$ ), sendo visível a proximidade dos resultados obtidos para as classes  $G_{US}$  e  $G_{NOR}$ .

### 3.4.5 -Uma situação não simétrica

Na procura de generalidade para a situação designada por paradoxo de Jeffreys--Lindley consideraram-se as classes de distribuições *a priori* unimodais simétricas como as mais “objectivas,” nomeadamente a classe  $G_{NOR}$  por ser a mais “realista”. A escolha destas classes de densidades unimodais simétricas está também relacionada com o facto de a variável aleatória  $X$  se referir a uma população Normal que tem distribuição simétrica. Se o problema a testar não for simétrico em  $\theta$  não tem muito sentido utilizar este tipo de classes.

Na presente secção mencionam-se algumas situações em que a variável aleatória  $X$  se refere a uma população que se distribui segundo uma binomial, não havendo portanto simetria natural. Neste caso, as classes utilizadas nos exemplos anteriores parece não serem indicadas.

Berger e Delampady (1987) abordaram esta questão e propuseram o cálculo de limites inferiores para as medidas Bayesianas de evidência utilizando determinadas classes de distribuições *a priori*.

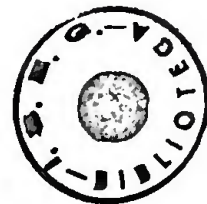
Embora a classe das densidades unimodais e simétricas em relação a  $\theta_0$  tenha dado muito “bons resultados” no caso da população Normal, parece não poder ser utilizada nesta situação devido à falta de simetria da distribuição de  $X$ . É então considerada uma variável aleatória  $\mu(\theta)$ ,

$$\mu(\theta) = \frac{\theta - \theta_0}{\sqrt{\theta(1 - \theta)}},$$

que depende de  $\theta$  e que tem função densidade de probabilidade simétrica em relação à origem;  $\mu(\theta)$  varia em  $(-\infty, +\infty)$  apesar de  $\theta \in (0,1)$ .

Considera-se então para classe de distribuições *a priori*, a classe  $H$  cujos elementos  $h$  são funções densidade de probabilidades unimodais e simétricas de  $\mu$  em relação à origem.

Transformando esta função densidade  $h$  numa função de  $\theta$  vem,



$$(3.42) \quad g(\theta) = h[\mu(\theta)] \frac{d\mu(\theta)}{d\theta},$$

em que,

$$\left| \frac{d\mu(\theta)}{d\theta} \right| = \frac{d\mu(\theta)}{d\theta},$$

pelo facto de  $\mu(\theta)$  ser crescente.

Seja  $(X_1, X_2, \dots, X_n)$  uma amostra casual com  $X_i \sim B(n; \theta)$  e  $G$  a classe das distribuições *a priori*  $g(\theta)$  da forma considerada em (3.42). Note-se que  $\theta_0$  é a mediana de  $g(\theta)$ ,

$$\int_0^{\theta_0} g(\theta) d\theta = \int_{-\infty}^0 h(\mu) d\mu = \frac{1}{2}$$

Pretende-se, tal como até aqui, ensaiar a hipótese  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$ , sendo a distribuição *a priori* de  $\theta$  do tipo da apresentada em (3.14) com  $g(\theta) \in G$ . Por (3.18), vem,

$$(3.43) \quad \inf_{g \in G} P\{H_0 | n, x\} = \left( 1 + \frac{1 - \pi_0}{\pi_0} \frac{\sup_{g \in G} \int_0^1 \binom{n}{x} \theta^x (1 - \theta)^{n-x} g(\theta) d\theta}{\binom{n}{x} \theta_0^x (1 - \theta_0)^{n-x}} \right)^{-1},$$

em que,

$$\begin{aligned} \sup_{g \in G} \int_0^1 \binom{n}{x} \theta^x (1 - \theta)^{n-x} g(\theta) d\theta &= \sup_{h \in H} \int_{-\infty}^{+\infty} \binom{n}{x} [\theta(\mu)]^x [1 - \theta(\mu)]^{n-x} h(\mu) d\mu \\ &= \sup_k \frac{1}{2k} \int_{-k}^{+k} \binom{n}{x} [\theta(\mu)]^x [1 - \theta(\mu)]^{n-x} d\mu \end{aligned}$$

Este último resultado obtém-se por aplicação da proposição:

**Teorema 3.1** - O supremo, sobre uma classe de distribuições *a priori* unimodais e simétricas em relação a  $\theta_0$ , com  $m_g(x) = \int f(x|\theta)g(\theta)d\theta$ , é atingido para uma distribuição uniforme, numa esfera de determinado raio  $k$ . Simbolicamente,

$$\sup_{g \in G_{US}} m_g(x) = \sup_k \left[ \frac{1}{v(k)} \int_{|\theta - \theta_0| \leq k} f(x|\theta) d\theta \right],$$

onde  $v(k)$  é o volume da esfera de raio  $k$ .

Por substituição deste último resultado em (3.43), vem,

$$(3.44)$$

$$\inf_{g \in G} P\{H_0|n, x\} = \left( 1 + \frac{1 - \pi_0}{\pi_0} \frac{\sup_k \frac{1}{2k} \int_{-k}^k \binom{n}{x} [\theta(\mu)]^x [1 - \theta(\mu)]^{n-x} d\mu}{\binom{n}{x} \theta_0^x (1 - \theta_0)^{n-x}} \right)^{-1},$$

onde  $\theta(\mu)$  é a função inversa de  $\mu(\theta)$ , dada por,

$$\theta(\mu) = \frac{\theta_0 + (\mu^2/2) + \mu \sqrt{(\mu^2/4) + \theta_0(1 - \theta_0)}}{1 + \mu^2}.$$

Uma vez encontrada a expressão  $\underline{P}\{H_0|n, x, G\}$ , Berger e Delampady (1987) calcularam, para  $\pi_0 = \frac{1}{2}$ , e determinados  $n$ ,  $x$  e  $\theta_0$ , vários valores para  $\underline{P}\{H_0|n, x, G\}$ , tendo chegado à conclusão que não eram muito diferentes dos obtidos numa população Normal para a classe de distribuições *a priori*  $G_{US}$ .

Parece então, poder concluir-se, que a “discrepância” entre as medidas Clássicas e Bayesianas não é exclusiva das populações Normais. Com este resultado a generalidade procurada para o paradoxo de Jeffreys-Lindley sai reforçada.

Considere-se agora que  $g(\theta)$  (distribuição das credibilidades *a priori* quando a hipótese  $H_1$  é verdadeira) é elemento da família beta ou seja, da família conjugada natural da binomial.

Tem-se ,

$$(3.45) \quad \inf_{g(\theta) \in B} P\{H_0 | n, x\} = \left[ 1 + \frac{1 - \pi_0}{\pi_0} \frac{\sup_{g \in B} \int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta}{f(x|\theta_0)} \right]^{-1}$$

em que  $B$  designa a família beta de parâmetros  $a > 0$  e  $b > 0$ . Se  $g(\theta) \in B$ , então,

$$g(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, \quad 0 < \theta < 1, \quad a > 0, \quad b > 0.$$

O valor médio de  $\theta$  é  $E\{\theta\} = \frac{a}{a+b} = \theta_0$ .

Tem-se então que, para algum  $c > 0$ ,

$$a = (a+b)\theta_0 \Rightarrow a = c\theta_0,$$

$$b = \frac{a(1-\theta_0)}{\theta_0} \Rightarrow b = c(1-\theta_0).$$

Seja então  $C \subset B$  o subconjunto de todas as funções densidade de probabilidade beta de parâmetros  $c\theta_0$  e  $c(1-\theta_0)$ . Consequentemente, para  $\{\theta: \theta \neq \theta_0\}$ ,

$$g(\theta) = \frac{\Gamma(c)}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))} \theta^{c\theta_0-1} (1-\theta)^{c(1-\theta_0)-1}$$

Voltando a (3.45),

$$(3.46) \quad \sup_{g \in C, \theta \neq \theta_0} \int f(x|\theta) g(\theta) d\theta =$$

$$= \sup_{c > 0} \left\{ \int_{\theta \neq \theta_0} \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(c)}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))} \theta^{c\theta_0-1} (1-\theta)^{c(1-\theta_0)-1} d\theta \right\}$$



ou seja,

$$\begin{aligned} & \sup_{g \in C} \int_{\theta \neq \theta_0} f(x|\theta)g(\theta)d\theta = \\ & = \sup_{c>0} \left\{ \frac{\Gamma(c)}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))} \binom{n}{x} \int_0^1 \theta^{x+c\theta_0-1} (1-\theta)^{n-x+c(1-\theta_0)-1} d\theta \right\}, \\ & = \sup_{c>0} \left\{ \frac{\Gamma(c)}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))} \binom{n}{x} \frac{\Gamma(x+c\theta_0)\Gamma(n-x+c(1-\theta_0))}{\Gamma(n+c)} \right\}. \end{aligned}$$

Substituindo a expressão referente a (3.46) em (3.45) vem,

(3.47)

$$\begin{aligned} \inf_{g \in C} P\{H_0|n, x\} &= \left[ 1 + \frac{1-\pi_0}{\pi_0} \frac{\sup_{c>0} \left\{ \frac{\binom{n}{x} \Gamma(c)\Gamma(x+c\theta_0)\Gamma(n-x+c(1-\theta_0))}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))\Gamma(n+c)} \right\}}{\binom{n}{x} \theta_0^x (1-\theta_0)^{n-x}} \right]^{-1}, \\ &= \left[ 1 + \frac{1-\pi_0}{\pi_0} \frac{\sup_{c>0} \frac{\Gamma(c)\Gamma(x+c\theta_0)\Gamma(n-x+c(1-\theta_0))}{\Gamma(c\theta_0)\Gamma(c(1-\theta_0))\Gamma(n+c)} \theta_0^x (1-\theta_0)^{n-x}}{\binom{n}{x} \theta_0^x (1-\theta_0)^{n-x}} \right]^{-1}. \end{aligned}$$

Berger e Delampady (1987) calcularam  $\underline{P}\{H_0|n, x, C\}$  para  $\pi_0 = \frac{1}{2}$  e vários valores de  $n$ ,  $x$  e  $\theta_0$ . O cálculo destes limites inferiores permite verificar que mais uma vez tem lugar a discrepância entre o valor- $p$  e  $\underline{P}\{H_0|n, x, C\}$ .

Na tabela 3.6<sup>1</sup> podem encontrar-se valores para o limite inferior da probabilidade *a posteriori* de  $H_0$  quando  $g(\theta)$  varia nas diferentes classes de distribuição *a priori* consideradas em Berger e Delampady (1987).

Os elementos das classes  $M_e$  e  $M_e M_o$  são funções densidade de probabilidade com mediana  $\theta_0(M_e)$  e média e mediana  $\theta_0(M_e M_o)$ .

**Tabela 3.6**

$p$	$\underline{P}\{H_0 n,x,C\}$	$\underline{P}\{H_0 n,x,G\}$	$\underline{P}\{H_0 n,x,M_e M_o\}$	$\underline{P}\{H_0 n,x,M_e\}$
0.01	0.09–0.22	0.06–0.13	0.07–0.16	0.04–0.07
0.05	0.20–0.42	0.22–0.29	0.21–0.25	0.14–0.20
0.10	0.32–0.48	0.28–0.37	0.27–0.38	0.20–0.33

### 3.5 - Comentários

Como pode constatar-se do presente estudo há situações de conciliação entre os resultados Clássicos e Bayesianos e outras em que esse “encontro” parece não se verificar.

Nos testes unilaterais (Capítulo2) do tipo  $H_0:\theta \leq \theta_0$  versus  $H_1:\theta > \theta_0$  parece haver “harmonia” entre os resultados Clássicos e Bayesianos, levando Casella e Berger (1987) a afirmar: *“Since the p value in the cases considered, is an upper bound on the infimum of  $Pr(H_0|x)$  it lies within or at the boundary of range of Bayesian measures of evidence demonstrating the existent to wich the Bayesian terminology can be attached”*.

Refira-se no entanto que esta “harmonia” se verifica apenas em determinadas condições (a função densidade  $f(x|\theta)$  referente à população em causa deve ser simé-

<sup>1</sup> Os valores desta tabela foram retirados de Berger e Delampady(1987).

trica em relação a  $\theta_0$  e ter razão de verosimilhança monótona). Caso estas condições não se verifiquem, os resultados obtidos podem ser completamente diferentes, o que permite a conclusão de que, a conciliação entre Clássicos e Bayesianos não é uma situação de absoluta generalidade, embora se verifique para diferentes e amplas classes de distribuição *a priori*. A este respeito Casella e Berger (1987) afirmam: “*Clearly there are classes of priors for which there are large discrepancies between  $\inf P_r(H_0|x)$  e  $p(x)$ ; the fact remains however, that reconciliation of measures of evidence is possible between the Bayesian and frequentist approaches.*”

Há no entanto outras situações (testes cuja hipótese nula é constituída por um ponto) em que este “encontro” não se verifica ou seja, em que, considerando os mesmos dados amostrais, Clássicos e Bayesianos tomam decisões opostas em relação à hipótese nula.

Repare-se que, embora o estudo do Capítulo 3 se refira a testes do tipo  $H_0:\theta = \theta_0$  versus  $H_1:\theta \neq \theta_0$  verifica-se certamente o mesmo “desencontro” caso se considerem testes do tipo  $H_0:\theta = \theta_0$  versus  $H_1:\theta > \theta_0$ . Casella e Berger(1987) referem que Berger e Sellke verificaram para um caso concreto que, considerando um teste do tipo  $H_0:\theta = \theta_0$  versus  $H_1:\theta \neq \theta_0$ , o valor- $p$  e o ínfimo da probabilidade  $\alpha$  *posteriori* da hipótese de nulidade são bastante diferentes.

Perante esta falta de conciliação relativamente aos testes (bilaterais ou não) cuja hipótese nula é constituída apenas por um ponto põe-se a questão de identificar os factores que poderão contribuir para tal desentendimento.

Clássicos e Bayesianos “responsabilizam-se” mutuamente referindo, tanto uns como outros, situações que podem suscitar alguma polémica. Os grandes “ataques” são dirigidos ao valor- $p$ , mas há também situações que, segundo os Clássicos, parece não serem muito favoráveis às medidas condicionais de evidência.

Seguidamente são apresentadas algumas destas situações mas, antes disso, refere-se que, apesar de se apresentarem frequentemente em oposição, a fronteira

entre uns e outros não é muito nítida na medida em que, como já foi dito, há Clássicos que não aceitam o valor- $p$  e Bayesianos que o consideram uma medida bastante razoável.

### 3.5.1. - A hipótese nula é constituída apenas por um ponto.

Devido à natureza escalar de  $\Theta = (-\infty, +\infty)$  é de esperar que a distribuição *a priori* de  $\theta$  seja contínua sendo  $P\{H_0\} = 0$  e conseqüentemente  $P\{H_0|x\} = 0$  para qualquer  $x$  observado. Os Bayesianos, insista-se, ultrapassam este obstáculo defendendo que, uma vez que se considera o valor particular  $\theta_0$ , tem importância especial, faz portanto sentido atribuir-lhe a probabilidade *a priori*  $\pi_0$  e distribuir a restante  $(1 - \pi_0)$  pelos valores  $\theta \neq \theta_0$ .

Os Clássicos, por outro lado, consideram que esta forma de “dar a volta ao problema” é bastante polémica.

O facto de  $\theta = \theta_0$  ser um ponto de massa discreta é de alguma importância na medida em que muitos a consideram como possível elemento perturbador da conciliação entre Clássicos e Bayesianos (como foi visto anteriormente verifica-se “desencontro” em situações em que, apesar da unilateralidade do teste, a hipótese nula é constituída apenas por um ponto).

### 3.5.2.- Objectividade da distribuição de credibilidades *a priori*.

Os defensores do valor- $p$  argumentam que, ao contrário do que acontece com este, uma medida objectiva de evidência, os Bayesianos não dispõem de uma medida com estas características pois a definição destas está sempre em função da especificação da distribuição das credibilidades *a priori* que é subjectiva na sua própria essência.



Em oposição a esta opinião situam-se os defensores das medidas condicionais de evidência que referem ter, num grande número de problemas, a possibilidade de considerar uma distribuição *a priori* não informativa. Reconhecem no entanto, que no caso dos testes em que a hipótese nula é constituída apenas por um ponto essa alternativa não é muito viável, tendo obrigatoriamente que escolher  $\pi_0$  e  $g(\theta)$ .

Admitem que  $\pi_0 = 1/2$  corresponde a uma escolha objectiva embora, até mesmo em relação a este aspecto, haja uma certa disparidade de opiniões. Argumentam alguns autores que, uma vez que se considera o particular valor  $\theta_0$ , este deve ter probabilidade superior a  $1/2$ , o que tem ainda a vantagem de facilitar a aproximação entre o valor- $p$  e as medidas de evidência Bayesianas.

A utilização do factor Bayes permite a não especificação de  $\pi_0$  mas  $g(\theta)$  tem que ser sempre definida. Enquanto uns defendem que não há escolha objectiva possível para  $g(\theta)$  outros consideram que o facto de considerar que esta distribuição de probabilidade é simétrica em relação a  $\theta_0$  e não crescente em  $|\theta - \theta_0|$  é uma especificação bastante satisfatória em termos de objectividade pois evita que se trate algum valor  $\theta \neq \theta_0$  como especial.

### 3.5.3. - A definição do valor- $p$

A grande crítica dirigida ao valor- $p$  tem a ver com a sua própria definição como probabilidade de observar um resultado tão bom ou pior que  $x$ , quando se observou apenas  $x$ . Para os Bayesianos este facto altera significativamente o resultado final. Em relação a esta questão Berger and Sellke (1987) apresentam a situação seguinte:

Uma vez que, para o cálculo do valor- $p$  não se utiliza apenas o valor observado mas todos os que são piores do que ele, considere-se, em vez de  $x$ , o conjunto,

$$Y = \{y: |T(y)| \geq |T(x)|\}.$$

Tem-se então que,

$$(3.48) \quad \underline{B}\{Y, G_A\} = \frac{f(x|\theta_0)}{\sup_{g \in G_A} m_g(Y)} = \frac{P_{\theta_0}(Y)}{\sup_{g \in G_A} m_g(Y)} = \frac{p}{\sup_{g \in G_A} m_g(Y)},$$

porque  $P_{\theta_0}(Y) = p$ .

Quando  $\pi_0 = \frac{1}{2}$  tem-se,

$$(3.49) \quad \underline{P}\{H_0|Y, G_A\} = \left[ 1 + \frac{\sup_{g \in G_A} m_g(Y)}{p} \right]^{-1}$$

sendo (veja-se Berger e Sellke(1987)),

$$(3.50) \quad \sup_{g \in G_A} m_g(Y) = 1.$$

Substituindo (3.50) em (3.49) vem,

$$\underline{P}\{H_0|Y, G_A\} = \left[ 1 + \frac{1}{p} \right]^{-1} = \frac{p}{p+1},$$

e  $\frac{p}{p+1} \cong p$  para pequenos valores de  $p$ .

Segundo Berger e Sellke (1987) o facto de se ter substituído o valor  $x$  observado pelo conjunto  $Y$  vem alterar significativamente o resultado (para pequenos

valores de  $p$ ) transformando a discrepância entre o valor- $p$  e  $P\{H_0|x, G_A\}$  numa situação de conciliação entre Clássicos e Bayesianos na medida em que,  $P\{H_0|Y, G_A\} \cong p$ . Considerou-se que  $f(x|\theta)$  é a função densidade de probabilidade da Normal e que  $G_A$  é a classe de todas as funções densidade de probabilidade.

A respeito desta situação Jeffreys(1980) escreveu: "*I have always considered the arguments for the use of  $P$  absurd. They amount to saying that a hypothesis that may or may not be true is rejected because a greater departure from the trial value was improbable; that is, that it has not predicted something that has not happened*".

#### 3.5.4.- Variação das medidas Bayesianas de evidência.

Como foi verificado nos dois exemplos apresentados no Capítulo 3 que se referem ao paradoxo de Jeffreys-Lindley a probabilidade *a posteriori* de  $H_0$  varia com a dimensão da amostra fazendo com que  $P\{H_0|\bar{x}\} \rightarrow 1$ , quando  $n \rightarrow \infty$ , constituindo este facto a exacerbação do conflito entre Clássicos e Bayesianos nos testes do tipo já referido.

Verificou-se posteriormente, considerando os limites inferiores da probabilidade *a posteriori* de  $H_0$  e do factor Bayes que a divergência entre Clássicos e Bayesianos não depende exclusivamente da dimensão da amostra.

Os resultados obtidos no Capítulo 3 (considerando sempre  $\pi_0 = 1/2$ ) permitem concluir que os valores numéricos referentes às medidas Bayesianas de evidência variam em função da classe de distribuições *a priori* escolhida, como era aliás de esperar. Segundo Berger e Delampady (1987) não é a forma funcional de  $g(\theta)$  (com excepção de situações extremas) que tem grande interferência na diferenciação dos resultados. No caso em que  $f(x|\theta)$  é a função densidade de probabilidade da Normal e  $g(\theta)$  pertence à classe das distribuições  $N(\theta_0, \tau^2)$ , esta classe podia ser

substituída pela da Cauchy, sem alteração significativa dos resultados. O que, segundo os autores deste artigo influencia de forma determinante o resultado é a escolha de  $\tau$ .

Seguidamente mostra-se como os valores de  $\tau$  podem influenciar as medidas Bayesianas de evidência.

Seja, mais uma vez, o teste de  $H_0: \theta = \theta_0$  versus  $H_1: \theta \neq \theta_0$  e uma população Normal sendo a função densidade de probabilidade  $f(x|\theta)$ . A distribuição de credibilidades a priori quando  $H_1$  é verdadeira é neste caso  $g(\theta) \equiv N(\mu, \tau^2)$  em que  $0 < \tau^2 < \infty$ .

O rácio das vantagens *a posteriori* tem por expressão,

$$\frac{P\{H_0|x\}}{P\{H_1|x\}} = \frac{\pi_0 f(x|\theta_0)}{(1-\pi_0) \int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta}$$

sendo,

$$\int_{\theta \neq \theta_0} f(x|\theta) g(\theta) d\theta \equiv N(\mu, \sigma^2 + \tau^2).$$

Consequentemente,

$$\begin{aligned} \frac{P\{H_0|x\}}{P\{H_1|x\}} &= \frac{\pi_0 \sigma^{-1} \exp\left\{-\frac{(x-\theta_0)^2}{2\sigma^2}\right\}}{(1-\pi_0) (\tau^2 + \sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x-\mu)^2}{2(\tau^2 + \sigma^2)}\right\}}, \\ &= \frac{\pi_0}{1-\pi_0} \left(1 + \frac{\tau^2}{\sigma^2}\right)^{1/2} \frac{\exp\left\{-\frac{(x-\theta_0)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x-\theta_0 + \theta_0 - \mu)^2}{2(\tau^2 + \sigma^2)}\right\}}, \\ &= \frac{\pi_0}{1-\pi_0} \left(1 + \frac{\tau^2}{\sigma^2}\right)^{1/2} \exp\left\{-\frac{\left[\frac{(x-\theta_0) - (\theta_0 - \mu) \frac{\sigma^2}{\tau^2}}{2\sigma^2 \left(1 + \frac{\sigma^2}{\tau^2}\right)}\right]^2}{2\sigma^2 \left(1 + \frac{\sigma^2}{\tau^2}\right)}\right\} \exp\left\{\frac{(\theta_0 - \mu)^2 \frac{\sigma^2}{\tau^2}}{2\sigma^2}\right\}, \end{aligned}$$



e finalmente

$$\frac{P\{H_0|x\}}{P\{H_1|x\}} = \frac{\pi_0}{1-\pi_0} \left(1 + \frac{\tau^2}{\sigma^2}\right)^{1/2} \exp\{-X\} \exp Y$$

em que,

$$X = \left\{x - \left[\theta_0 + (\theta_0 - \mu)\sigma^2 / \tau^2\right]\right\}^2 / 2\sigma^2 \left(1 + \frac{\sigma^2}{\tau^2}\right)$$

$$Y = (\theta_0 - \mu)^2 / 2\tau^2$$

As medidas Bayesianas de evidência são então,

$$B_g(x) = \left(1 + \frac{\tau^2}{\sigma^2}\right) \exp\{-X\} \exp Y,$$

e

$$P\{H_0|x\} = \left[1 + \frac{1-\pi_0}{\pi_0} \frac{\sigma^2}{\sigma^2 + \tau^2} \exp\{X\} \exp\{-Y\}\right]^{-1}.$$

Pode concluir-se das expressões para o factor Bayes e  $P\{H_0|x\}$  que, os seus valores numéricos dependem de  $\tau^2$ . Se  $\tau^2 \rightarrow \infty$ , então  $g(\theta)$  é não informativa e  $P\{H_0|x\} \rightarrow 1$ , obtendo-se uma situação semelhante à apresentada em Lindley(1957).

Sendo assim, caso as credibilidades *a priori* se distribuam segundo uma Normal, o que é bastante frequente, os Bayesianos não podem fugir à especificação de  $\tau^2$  o que, para os “não simpatizantes” das medidas Bayesianas de evidência constitui uma situação de desvantagem.

O processo a seguir consiste então no cálculo do infimo das medidas Bayesianas de evidência sobre a classe  $G_{NOR}$ .

O facto de ser necessária a especificação da distribuição das credibilidades *a priori* é sentido, inclusivamente pelos Bayesianos, como um inconveniente. Em muitas situações podem utilizar-se distribuições *a priori* não informativas mas,

segundo alguns autores este procedimento não é muito indicado quando se trata de situações referentes ao ensaio de hipóteses.

Como foi referido anteriormente Lindley(1957) utiliza uma *priori* não informativa mas Jeffreys(1961) aconselha já a utilização de distribuições próprias para as credibilidades *a priori*.

Mais recentemente Berger e Pericchi(1996) afirmam: “*The reason is that Bayes factor in hypothesis testing and model selection typically depend rather strongly on the prior distributions, much more than, in, say, estimation.(For instance, as the sample size grows, the influence of the prior distribution disappears in estimation, but does not in hypothesis testing or model selection)*”.

Para ultrapassar esta questão e admitindo que é em muitos casos difícil definir uma distribuição *a priori* própria, estes autores propõem para medida de evidência o factor Bayes intrínseco (IBF's). Esta medida obtém-se a partir do factor Bayes utilizando uma amostra de treino (training sample).

Nestas condições, se  $x_t$  for a amostra de treino, então ela é parte da amostra inicial  $x$ . A ideia consiste em usar  $x_t$  para converter  $h(\theta)$  (distribuição *a priori* imprópria) numa distribuição *a posteriori* própria  $h(\theta|x_t)$  de tal forma que,

$$h(\theta|x_t) = \frac{f(x_t|\theta)h(\theta)}{\int_{\Theta} f(x_t|\theta)h(\theta)d\theta}$$

e calcular seguidamente o factor Bayes intrínseco (the intrinsic Bayes factor), utilizando para isso os restantes dados amostrais  $x_{n-t}$  e  $h(\theta|x_t)$  como distribuição *a priori*.

Para um teste  $H_0:\theta \in \Theta_0$  versus  $H_1:\theta \in \Theta_1$ , tem-se,

$$B_1 = \frac{\int_{\Theta_0} f(x_{n-t}|\theta)h(\theta|x_t)d\theta}{\int_{\Theta_1} f(x_{n-t}|\theta)h(\theta|x_t)d\theta}$$

designando  $B_j$  o factor Bayes intrínseco.

O método utilizado em Aitkin (1991) pode também ser considerado como um processo para obter o factor Bayes em que se utiliza uma amostra de treino. Considera a amostra  $x$ , na sua totalidade, como amostra treino e utiliza o processo anteriormente descrito para determinar o factor Bayes.

Berger e Pericchi (1996) criticam esta utilização dupla dos dados amostrais referindo que não é consistente com a lógica usual Bayesiana e que interfere na obtenção de resultados assintóticos.

No ponto 3.5 foram apresentadas algumas questões que têm gerado polémica, ora em relação ao valor- $p$ , ora em relação às medidas condicionais de evidência. No desenrolar deste trabalho foram apresentadas algumas mais, não constituindo no entanto objectivo do presente estudo fazer enumeração desses aspectos mas sim de constatar que é certamente difícil identificar, sem quaisquer dúvidas, qual a medida indicada a utilizar, quando pretende ensaiar-se uma hipótese simples.

Verificou-se que há autores que defendem a utilização do valor- $p$  (e não apenas Clássicos) e outros (não necessariamente Bayesianos) que, com a mesma veemência aconselham o uso das medidas Bayesianas. Constatou-se que ao ensaiar testes cuja hipótese nula é constituída apenas por um ponto e considerando os mesmos dados amostrais, uns aceitam-na e outros rejeitam-na. Quem é que está a proceder correctamente?

Não se pode argumentar que a falta de conciliação entre Clássicos e Bayesianos é um fenómeno pontual. Como se teve oportunidade de verificar no decorrer do estudo, muitos foram os estudiosos que se debruçaram sobre o problema definindo uns, as situações em que há conciliação (testes unilaterais) e mostrando outros que a inexistência de conciliação é um fenómeno com alguma generalidade pois, para além de acontecer em populações com diferentes distribuições de probabilidade (a Normal e a Binomial foram os casos

considerados), verificam-se também para diversas e amplas classes de distribuição das credibilidades *a priori*.

O facto de ser possível conciliar as medidas Clássicas e Bayesianas nos testes unilaterais não deve constituir situação de exagerado optimismo na medida em que estes resultados têm lugar apenas se se verificarem determinadas condições ( $f(x|\theta)$  simétrica em relação a  $\theta_0$  e com RVM). Caso estas condições não se verifiquem os resultados são diferentes, não sendo no entanto razoável ser exageradamente pessimista pois é certo que a igualdade ou proximidade numérica de resultados Clássicos e Bayesianos se verifica em determinadas situações.

Refira-se ainda que para alguns autores o problema da inexistência de conciliação não tem muito sentido na medida em que na prática raramente tem interesse testar hipóteses constituídas por um único ponto. Consideram que é pouco realista pretender saber se determinado medicamento faz ou não efeito no tratamento de determinada doença; na prática é mais comum pretender saber se o medicamento faz ou não algum efeito. Opinião contrária têm os que consideram que os testes cujas hipóteses nulas são constituídas apenas por um ponto são muito boas aproximações de testes em que  $|\theta - \theta_0| \leq \varepsilon$  quando  $\varepsilon$  é suficientemente pequeno. A este respeito Berger e Sellke(1987) referem: *“And even if testing of a point null hypothesis were disreputable the reality is that people do it all the time”... “and we should do our best to see that it is done well”*.

Para finalizar parece ser sensato afirmar que é importante não prescindir de nenhuma destas medidas de evidência como objecto de estudo, no sentido de esclarecer, não só o modo como se relacionam mas também a forma como cada uma delas está definida. Casella e Berger(1987), respondendo aos comentários feitos ao seu artigo, afirmam: *“We agree with Professor Good that p values and Bayes factors (or posterior probabilities of  $H_0$ ) are here to stay”*. E acrescentam: *“Bayesians and frequentists may never agree on the appropriate way to analyze data and interpret results, but there is no reason why they cannot learn from one another. Whether or not measures of evidence can be reconciled is probably a*

*minor consideration; understanding what affects a measure of evidence is a major consideration”*

## **REFERÊNCIAS**



Aitkin, Murray (1991): "Posterior Bayes Factors", *Journal of the Royal Statistical Society, B*, Vol.53, No1,111-142.

Berger, James O. (1985): *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, Nova Iorque.

Berger, James O. e Delampady, Mohan (1987): "Testing Precise Hypotheses", *Statistical Science*, Vol.2, No 3, 317-352.

Berger, James O. e Pericchi, Luis R. (1996): "The Intrinsic Bayes Factor for Model Selection and Prediction", *Journal of the American Statistical Association*, Vol.91, No 433, 109-122.

Berger, James O. e Sellke, Thomas (1987): "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence", *Journal of American Statistical Association*, Vol. 82, No 397, 112-139.

Casella, George e Berger, Roger L. (1987): "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem", *Journal of American Statistical Association*, Vol. 82, No 397, 106-111.

Casella, George e Berger, Roger L. (1990): *Statistical Inference*, Wadsworth & Brooks, California.

DeGroot, M.H. (1973): "Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio", *Journal of the American Statistical Association*, Vol. 68, No 344, 966-969.

Delampady, Mohan (1989): "Lower Bounds on Bayes Factors for Interval Null Hypotheses", *Journal of American Statistical Association*, Vol. 84, No 405, 120-124.

Edwards, W., Lindman, H. e Savage, L. J. (1963): "Bayesian Statistical Inference For Psychological Research", *Psychological Review*, Vol. 70, No 3, 193-242.



Good, I.J.(1988): "The Interface Between Statistics and Philosophy of Science", *Statistical Science*, Vol.3, No 4, 386-412.

Good, I.J. (1992): "The Bayes/Non-Bayes Compromise: A Brief Review", *Journal of the American Statistical Association*, Vol. 87, No 419, 597-606.

Lindley, Dennis V. (1957): "A Statistical Paradox", *Biometrika*, 44, 187-192.

Lindley, Dennis V.(1990): "The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics", *Statistical Science*, Vol.5, No 1, 44-89.

Mello, F.G. (1973): *Introdução aos Métodos Estatísticos*, Vol.II, Livraria Escolar Editora, Lisboa.

Murteira, Bento J.F. (1988): *Estatística: Inferência e Decisão*, Imprensa Nacional-Casa da Moeda, Lisboa.

Murteira, Bento J.F. (1990): *Probabilidades e Estatística*, Vol.II, 2ª Ed., McGraw-Hill de Portugal Lda., Lisboa.

Oliveira, J.T. (1991): *Probabilidades e Estatística-Conceitos Métodos e Aplicações*, Vol.II, McGraw-Hill de Portugal Lda., Lisboa.

Pratt, John W.(1965): "Bayesian Interpretation of Standard Inference Statements", *Journal of the Royal Statistical Society*, Ser.B, 27, 169-203.