**DESENVOLVIMENTO E IMPLICAÇÕES
DO PRINCÍPIO DA VEROSSIMILHANÇA**

Lurdes Yoshiko Tani Inoue

Dissertação apresentada ao Instituto de Matemática e Estatística da
Universidade de São Paulo para a obtenção do grau de mestre em
Estatística

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

- São Paulo, Janeiro de 1995 -

**DESENVOLVIMENTO E IMPLICAÇÕES DO PRINCÍPIO DA
VEROSSIMILHANÇA**

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e apresentada por
Lurdes Yoshiko Tani Inoue
e aprovada pela Comissão Julgadora.

São Paulo, 12 de janeiro de 1995.

Banca Examinadora

. Prof. Dr. Sergio Wechsler (Orientador)	- IME-USP
. Prof. Dr. Carlos Alberto de B. Pereira	- IME-USP
. Profa. Dra. Telba Zalkind Irony	- George Washington University

ABSTRACT

The Likelihood Principle is obtained from more primitive concepts as Conditionality and Sufficiency. Some of its implications, like the Stopping Rule Principle, are studied in detail. The difficulties caused by the violation of the Likelihood Principle are discussed through some examples on clinical trials. In addition, the problem of assessing bioequivalence of formulations of a given drug is approached from a strict Bayesian point of view as an example of implementation of the Likelihood Principle.

SUMÁRIO

A demonstração do Princípio da Verossimilhança é feita partindo-se dos conceitos mais primitivos de Condicionalidade e Suficiência. Algumas de suas implicações, como o Princípio da Regra de Parada, são estudadas em detalhe. As dificuldades causadas pela violação do Princípio da Verossimilhança são abordadas através de alguns exemplos voltados para a área de ensaios clínicos. Além disso, o problema da bioequivalência entre formulações de uma dada droga é abordado sob um ponto de vista estritamente bayesiano como um exemplo de implementação do Princípio da Verossimilhança.

AGRADECIMENTOS

Ao CNPq pelo apoio financeiro.

À Profa. Lisbeth K. Cordani e Oscar E.D. Paschoa pelo incentivo nas valorosas discussões em torno do tema bioequivalência.

Aos meus amigos pelo extremo carinho com que me apoiaram nos momentos mais difíceis.

Aos meus irmãos e à família Tanni pelo incentivo e pela constante inspiração.

Ao Prof. Dr. Sergio Wechsler pela esplêndida orientação. Sua presença amiga, paciência, atenção, otimismo e bom-humor foram fundamentais no desenvolvimento desse trabalho.

*Dedico este trabalho à memória
de meus pais: Kasuo e Satie*

ÍNDICE

Capítulo 1 - Introdução	1
Capítulo 2 - Desenvolvimento axiomático do Princípio da Verossimilhança	3
2.1. Introdução	3
2.2. O Princípio da Verossimilhança	4
2.3. Discussão crítica do Princípio da Verossimilhança	14
Capítulo 3 - Ensaios Clínicos e o Princípio da Verossimilhança	21
3.1. Introdução	21
3.2. Implicações do Princípio da Verossimilhança	26
3.2.1. A "quase" inevitabilidade da inferência bayesiana	26
3.2.2. A incompatibilidade com a inferência clássica	28
3.2.3. A irrelevância das regras de parada	32
3.2.4. A irrelevância dos mecanismos de censura	36
3.2.5. Aleatorização	41
3.3. Exemplos	43
3.4. Discussão	63
Capítulo 4 - Bioequivalência: Uma proposta Neo-Bayesiana de análise	66
4.1. Introdução	66
4.2. Métodos comuns de análise	68
4.3. Método alternativo de análise	70
4.4. Aplicações	71
Capítulo 5 - Conclusão	75
Referências Bibliográficas	78

CAPÍTULO 1 - INTRODUÇÃO

Declarando essencialmente que qualquer conclusão em um experimento estatístico a respeito de uma quantidade desconhecida está contida na função de verossimilhança (dessa quantidade) gerada pelos dados observados, o **Princípio da Verossimilhança** possui implicações profundas na Inferência Estatística. O Princípio não diz como proceder uma análise estatística, mas insiste em que o método de inferência deva se basear na função de verossimilhança gerada pelos dados.

O objetivo deste trabalho é apresentar o Princípio da Verossimilhança e algumas de suas implicações, além de uma discussão sobre as conseqüências de sua violação na área de ensaios clínicos. Apresenta ainda uma solução para o problema da bioequivalência implementando o Princípio da Verossimilhança com a utilização da abordagem bayesiana subjetivista.

A apresentação do Princípio da Verossimilhança é feita no capítulo 2. Na seção 2.2. apresentamos (formalmente) o Princípio da Verossimilhança com a definição do contexto estatístico, o enunciado de princípios mais primitivos e a conseqüente demonstração de Birnbaum. Na seção 2.3., apresentamos uma discussão crítica do desenvolvimento axiomático do Princípio da Verossimilhança.

No capítulo 3 algumas de suas implicações na inferência estatística são exploradas. Apresentamos o contexto de ensaios clínicos e algumas de suas particularidades e instâncias de interesse estatístico na seção 3.1. Na seção 3.2., apresentamos algumas implicações do Princípio da Verossimilhança sobre as instâncias de interesse estatístico abordadas na seção anterior. Os resultados apresentados são gerais e não limitados somente à área de ensaios clínicos. Na seção 3.3. apresentamos alguns exemplos que ilustram as dificuldades causadas

pela violação do Princípio da Verossimilhança. A discussão geral do capítulo é feita na seção 3.4. onde avaliamos alguns aspectos éticos.

No capítulo 4, apresentamos uma solução para o problema da bioequivalência sob uma abordagem preditivista. Na seção 4.1., apresentamos uma caracterização da bioequivalência e da metodologia de análise. Na seção 4.2., apresentamos os aspectos gerais da metodologia usualmente utilizada para a análise estatística da bioequivalência. Propomos, na seção 4.3., um método alternativo que obedece o Princípio da Verossimilhança com a adoção de um ponto de vista bayesiano subjetivista-preditivista para o problema (como preconizado por Bruno de Finetti). A aplicação desse método é apresentada em dois exemplos da seção 4.4.

No capítulo 5 apresentamos uma conclusão geral do trabalho.

CAPÍTULO 2 - DESENVOLVIMENTO AXIOMÁTICO DO PRINCÍPIO DA VEROSSIMILHANÇA

2.1. INTRODUÇÃO

O Princípio da Verossimilhança foi em grande parte impulsionado por idéias de R. A. Fisher e G. Barnard em vários artigos (ver, por exemplo, Barnard(1949) e Fisher(1959)). Segundo alguns autores, como Berger e Wolpert (1988), Fisher esteve próximo de declarar o princípio quando afirmou:

"the whole of the information supplied by a sample is comprised in the likelihood"

(Fisher (1959), p.70).

O Princípio da Verossimilhança, contudo, só se tornou proeminente quando A. Birnbaum(1962) proviu-o de um desenvolvimento axiomático partindo de conceitos mais simples de suficiência e condicionalidade (ver também Bjørnstad(1992), que apresenta uma introdução e discussão crítica de Birnbaum(1962)).

Segundo Joshi(1983), o desenvolvimento do Princípio da Verossimilhança está ligado ao desenvolvimento de uma metodologia de inferência estatística (a "Likelihood Inference" - inferência da verossimilhança) que é fundamentada inteiramente nesse princípio e que pretendia substituir as metodologias clássica e bayesiana. Observemos, contudo, que o princípio é compatível com a metodologia bayesiana. Mais ainda, Berger e Wolpert (1988) comentam que a submissão à análise bayesiana é a forma mais lógica e realística de implementação do Princípio da Verossimilhança. Além disso, a adoção desse princípio é um grande ponto de divergência entre bayesianos e freqüentistas (adeptos da metodologia clássica), tendo a sua violação levado vários destes últimos a abandonar a visão freqüentista em favor do ponto de vista bayesiano.

De modo informal, o Princípio da Verossimilhança enunciado por Berger e Wolpert (1988, p.19) diz o seguinte:

" Toda a informação fornecida por um experimento a respeito de uma quantidade desconhecida θ está contida na função de verossimilhança para θ gerada pelos dados. Duas funções de verossimilhança para θ (provenientes do mesmo ou até mesmo de experimentos distintos) contêm a mesma informação a respeito de θ se elas são proporcionais entre si. "

Apresentamos o desenvolvimento axiomático (devido a Birnbaum (1962)) do Princípio da Verossimilhança na seção 2.2. e na seção 2.3., apresentamos algumas críticas feitas à sua demonstração.

2.2. O PRINCÍPIO DA VEROSSIMILHANÇA

Seja θ uma quantidade de interesse de valor desconhecido com $\theta \in \Theta$: espaço paramétrico ($\Theta \subset \mathbb{R}^k$).

Com o intuito de obtermos informação a respeito de θ , realizamos um experimento estatístico E , que consiste em observar um vetor aleatório X com valores possíveis em um espaço amostral $\Omega_E \subset \mathbb{R}^m$. Seja $P_\theta(\cdot)$ a função de distribuição condicional de probabilidades de X em Ω_E . Admitamos que X possua função de densidade de probabilidade (ou, no caso discreto, função de probabilidade) condicional para todo $\theta \in \Theta$, a qual será representada por $f_\theta(\cdot)$. A coleção $\{f_\theta(\cdot): \theta \in \Theta\}$ será representada por $\{f_\theta\}$. Nesse contexto, representamos o experimento estatístico E pelo trio $(\Omega_E, \Theta, \{f_\theta\})$.

DEFINIÇÃO: Para um dado $x \in \Omega_E$ (um valor possível de X), a função de verossimilhança para θ gerada por x é definida, para $\forall \theta \in \Theta$, por $l_x(\theta) = f_\theta(x)$. Como é usual, representaremos a função de verossimilhança por $l_x(\cdot)$.

Após a realização do experimento E e da conseqüente observação do valor $X=x$, uma declaração a respeito de θ é feita. Não impomos restrição alguma sobre a natureza ou sobre as características desta "inferência". Ela pode ser qualquer tipo de declaração a respeito de θ . Por exemplo: uma estimativa, um intervalo de confiança, uma densidade a posteriori para θ , etc. Portanto, estamos considerando um conjunto abstrato D_Θ de (coleções finitas de) "declarações" a respeito de θ .

Um "método de inferência" deve produzir (pelo menos) uma declaração a respeito de θ para cada valor possível x do espaço amostral Ω_E do experimento E .

Usaremos a notação MI (método de inferência) seguindo a linha de raciocínio proposta por Dawid (1977) com o seu "padrão de inferência". Birnbaum (1962) (e também Berger e Wolpert (1988)) utilizam o termo "evidência" a respeito de θ que é fornecida por E e x (notação: $Ev(E,x)$). A evidência de Birnbaum (1962) é um conceito não definido e, portanto, alvo de críticas (LeCam, p.182-185.2 de Berger e Wolpert(1988)). Berger e Wolpert (1988) salientando que $Ev(E,x)$ possui uma natureza completamente arbitrária - podendo ser uma medida qualquer, ou uma coleção de medidas quaisquer - tentam se esquivar dessas críticas. Basu (1975) utiliza a notação $Inf(E,x)$ para a "informação" em uma transcrição que ele próprio qualifica como pseudo-matemática da expressão "*the whole of the information*" a respeito de θ fornecida pela observação de E e x . Para Basu o conceito de informação é abstrato ("*this is what the subject is all about*" (Basu (1975), p.1), e só faz sentido no contexto onde há uma "informação a priori" que devemos ter no início do processo de investigação de θ e um particular método inferencial que provê informação adicional.

Um método de inferência deve ser aplicável a uma larga variedade de experimentos realizáveis. Dado um método de inferência, a **comparação** das inferências feitas em alguns

desses diferentes experimentos é um aspecto essencial dos princípios de inferência. Partindo dessa idéia, Dawid (1977) introduz a noção de família de experimentos Ξ aos quais o método MI é aplicável. Experimentos pertinentes à classe Ξ são, portanto, comparáveis com relação ao resultado do método de inferência.

Birnbaum (1962) e Berger e Wolpert (1988) não fazem nenhuma referência à existência dessa classe de experimentos comparáveis. Para eles, Ξ é a classe de **todos** os experimentos realizáveis. Veremos adiante que a noção da classe Ξ é crucial para o desenvolvimento axiomático do Princípio da Verossimilhança.

No contexto acima descrito surge a seguinte definição:

DEFINIÇÃO: Seja $\mathcal{E}_\theta$ a classe de todos os experimentos estatísticos E onde $\Theta_E = \Theta$ (em outras palavras, $\mathcal{E}_\theta$ é o conjunto de todos os experimentos a "respeito de θ ").

Seja Ξ um subconjunto de $\mathcal{E}_\theta$. Um Método de Inferência a respeito de θ é uma função:

$$MI: \bigcup_{E \in \Xi} [\{E\} \times \Omega_E] \rightarrow D_\Theta$$

Em outras palavras, um "método de inferência" é uma função que, para todo E em uma classe Ξ de experimentos associa um número finito de declarações a respeito de θ a cada $x \in \Omega_E$.

Apresentamos a seguir o desenvolvimento axiomático devido a Birnbaum(1962) do Princípio da Verossimilhança resultante dos conceitos mais primitivos de condicionalidade e suficiência. (Ver também Basu(1975), Dawid(1977), Berger e Wolpert(1988) e Bjørnstad(1992))

Estamos supondo que Ω é discreto.

PRINCÍPIO DA CONDICIONALIDADE FRACA - Sejam $E_1 = (\Omega_1, \Theta, \{f_\theta^1\})$ e $E_2 = (\Omega_2, \Theta, \{f_\theta^2\})$ dois experimentos tais que E_1 e $E_2 \in \Xi$ e onde apenas o parâmetro desconhecido θ necessita ser comum aos dois experimentos. Consideremos o experimento "misto" E^* , no qual $J=1$ ou 2 é observado com probabilidades respectivamente iguais a α_1 e α_2 , sendo o experimento E_J então realizado. ($\alpha_1, \alpha_2 \in [0,1]$; $\alpha_1 + \alpha_2 = 1$, independentemente de θ , X_1 ou X_2 - onde X_1 e X_2 , respectivamente, são os vetores aleatórios de E_1 e E_2). Formalmente, $E^* = (\Omega^*, \Theta, \{f_\theta^*\})$, onde $\Omega^* = [\{1\} \times \Omega_1] \cup [\{2\} \times \Omega_2]$ e $X^* = (J, X_J)$ é observado com densidade $f_\theta^*((j, x_j)) = (\alpha_j \cdot f_\theta^j(x_j))$. Então é necessário que:

$$i) E^* \in \Xi ;$$

$$ii) MI(E^*, (j, x_j)) = MI(E_j, x_j),$$

Em outras palavras, Ξ deve ser fechada sob misturas e a inferência resultante de E^* deve ser meramente a inferência resultante do experimento E_j de fato realizado.

O seguinte exemplo esclarece a idéia do Princípio da Condicionalidade Fraca: Suponha que uma droga possa ser avaliada em dois laboratórios (por exemplo, localizados em São Paulo e no Rio de Janeiro) igualmente adequados para a análise. Uma moeda é lançada para a escolha do laboratório (com resultado cara indicando que o laboratório escolhido é o do Rio de Janeiro).

Temos, portanto, dois potenciais experimentos: um onde a droga é analisada em São Paulo (E_1) e outro onde a análise ocorre no Rio de Janeiro (E_2). No experimento misto, o resultado obtido com o lançamento da moeda (cara ou coroa) tem probabilidade $1/2$ (independentemente de θ , dos laboratórios e da análise). Se ocorre coroa ($j=1$) a droga é analisada no laboratório de São Paulo; se ocorre cara ($j=2$) a análise é no Rio de Janeiro.

Suponhamos que o resultado do lançamento da moeda seja coroa e, portanto, o laboratório escolhido seja o de São Paulo.

O bom senso nos diz que a inferência obtida nesse experimento misto deverá ser exclusivamente a inferência obtida no experimento efetivamente realizado, isto é, a inferência obtida a respeito de θ (digamos o pH da solução) resultará exclusivamente da análise feita em São Paulo, não devendo ser levada em conta a possibilidade de que a análise **pudesse** ter sido feita no Rio de Janeiro.

DEFINIÇÃO: Dizemos que uma estatística $T(X)$ é suficiente para o parâmetro θ no experimento E quando para cada valor possível t de T , a distribuição condicional de X , dados $T(X)=t$ e θ , não depende do valor θ .

Interpretemos essa definição: Se a estatística T for suficiente, $P(X|T=t, \theta) = P(X|T=t)$. Então, uma vez conhecido o valor de T , os dados não mais fornecerão informação a respeito de θ , pois ficará desaparecida a conexão entre X e θ .

PRINCÍPIO DA SUFICIÊNCIA FRACA - Seja $E=(\Omega_E, \Theta, \{f_\theta\})$ um experimento e suponhamos que $T(X)$ seja uma estatística suficiente para θ em E . Então, se $T(x_1)=T(x_2)$, é necessário que:

$$MI(E, x_1) = MI(E, x_2),$$

i.e., o Princípio da Suficiência Fraca requer que não faça a inferência distinção alguma entre amostras que tenham o mesmo valor da estatística suficiente.

PRINCÍPIO DA VEROSSIMILHANÇA (FORMAL) - Consideremos dois experimentos $E_1=(\Omega_1, \Theta, \{f_\theta^1\})$ e $E_2=(\Omega_2, \Theta, \{f_\theta^2\})$, onde θ é o mesmo objeto em cada um dos experimentos. Suponhamos que E_1 e $E_2 \in \Xi$, e que para as realizações x_1^* e x_2^* de E_1 e E_2 , respectivamente, tenhamos $l_{x_1^*}(\theta) \equiv c \cdot l_{x_2^*}(\theta)$ para alguma constante positiva c (i.e. $f_\theta^1(x_1^*)=c \cdot f_\theta^2(x_2^*)$, $\forall \theta \in \Theta$). Então, é necessário que:

$$MI(E_1, x_1^*) = MI(E_2, x_2^*).$$

COROLÁRIO 1 DO PRINCÍPIO DA VEROSSIMILHANÇA - Se $E=(\Omega_E, \Theta, \{f_\theta\})$ é um experimento (e se MI satisfaz o Princípio da Verossimilhança), então $MI(E, x)$ depende de E e x através de $l_x(\cdot)$ exclusivamente.

DEMONSTRAÇÃO:

Para $\forall x^* \in \Omega_E$, definimos o experimento E^* que consiste em observar a quantidade $Y_{x^*} = 1(X=x^*)$, ou seja, uma quantidade aleatória com distribuição de Bernoulli de parâmetro $f_\theta(x^*)$. A função de verossimilhança para θ (gerada por $X=x^*$) é $l_{x^*}(\cdot)$. Formalmente, esse experimento é assim descrito: $E^* = (\{0,1\}; \Theta; \{g_\theta\})$, com $g_\theta(1) = 1 - g_\theta(0) = l_{x^*}(\theta)$ para todo $\theta \in \Theta$. Podemos também escrever $E^* = (\{0,1\}; \Theta; \{G_\theta\})$, onde $\{G_\theta\}$ é a família das distribuições de Bernoulli ($l_{x^*}(\theta)$), indexada em Θ .

Devemos mostrar que $\exists$ uma função $\Psi(\cdot)$ tal que $MI(E, x^*) = \Psi(l_{x^*}(\cdot))$.

Ora, dada uma função de verossimilhança $l(\cdot)$, mesmo desconhecendo o seu ponto amostral gerador x^* , podemos definir perfeitamente o experimento $E^* = (\{0,1\}; \Theta; G_\theta)$. Como a função de verossimilhança $l_1(\theta)$ gerada por $y=1$ neste experimento satisfaz a identidade (em θ):

$$l_1(\theta) = \text{Prob}(Y_{x^*}=1|\theta) = \text{por construção de } Y_{x^*} = f_\theta(x^*) = l_{x^*}(\theta),$$

o Princípio da Verossimilhança garante (ver observação abaixo) que:

$$MI(E^*, 1) = MI(E, x^*)$$

e, portanto,

$$\begin{aligned} MI(E, x^*) &= MI(\{0, 1\}; \Theta; \{G_\theta\}; 1) \\ &= \Psi(I_{x^*}(.)) \end{aligned}$$

□

Observação: O leitor terá percebido que é crucial para o argumento acima que imponhamos a comparabilidade entre E^* e E com relação a MI, isto é, que E^* e $E \in \Xi$.

Segundo Joshi(1983), o enunciado desse corolário é incompleto, pois não diz **como** o "significado da evidência" é determinado a partir da função de verossimilhança de θ . Ele propõe $\Psi(I_{x^*}(.)) = I_{x^*}(.)$ (i.e. $\Psi(.) \equiv .$) e sugere que o corolário (com Ψ = identidade) é o "Princípio da Verossimilhança Total". Para Hill(1987), a própria função de verossimilhança é, de fato, a única definição possível de evidência sob uma perspectiva bayesiana-subjetivista.

Apresentamos agora o clássico teorema de Birnbaum (Ver Birnbaum(1962) e também Bjørnstad(1992)). A demonstração segue a de Berger e Wolpert(1988) estando no entanto adaptada à construção de método de inferência que fizemos, por sua vez inspirada em Dawid (1977).

TEOREMA : O Princípio da Condicionalidade Fraca e o Princípio da Suficiência Fraca implicam, conjuntamente, o Princípio da Verossimilhança Formal. Reciprocamente, o Princípio da Verossimilhança Formal implica o Princípio da Condicionalidade Fraca e o Princípio da Suficiência Fraca.

DEMONSTRAÇÃO:

($\Rightarrow$) Consideremos os experimentos E_1 e E_2 a respeito de uma mesma quantidade θ e o experimento misto E^* (como definido no Princípio da Condicionalidade Fraca acima).

Decorre do Princípio da Condicionalidade Fraca que:

$$MI(E^*, (j, x_j)) = MI(E_j, x_j) \quad (1)$$

Sejam $x_1^* \in \Omega_1$ e $x_2^* \in \Omega_2$ pontos amostrais gerando funções de verossimilhança proporcionais entre si e suponhamos que E_1 e $E_2 \in \Xi$. Então, definindo a estatística T em Ω^* da seguinte forma:

$$T(J, X_J) = \begin{cases} (1, x_1^*) & \text{se } J=2, X_2 = x_2^* \\ (J, X_J) & \text{em caso contrário} \end{cases},$$

percebemos que T é suficiente para θ em E^* , pois

i) $P_\theta(X^* = (j, x_j) \mid T = t \neq (1, x_1^*)) = I((j, x_j) = t)$ independentemente de θ , e

ii) $P_\theta(X^* = (1, x_1^*) \mid T = (1, x_1^*)) = 1 - P_\theta(X^* = (2, x_2^*) \mid T = (1, x_1^*)) =$

$$= \frac{f_\theta^*((1, x_1^*))}{f_\theta^*((1, x_1^*)) + f_\theta^*((2, x_2^*))} =$$

$$= \frac{\alpha_1 \cdot f_\theta^1(x_1^*)}{\alpha_1 \cdot f_\theta^1(x_1^*) + \alpha_2 \cdot f_\theta^2(x_2^*)} =$$

$$= (\text{Como } f_\theta^1(x_1^*) \equiv c \cdot f_\theta^2(x_2^*) \text{ para algum } c > 0) =$$

$$= \frac{c \cdot \alpha_1 \cdot f_\theta^2(x_2^*)}{c \cdot \alpha_1 \cdot f_\theta^2(x_2^*) + \alpha_2 \cdot f_\theta^2(x_2^*)} =$$

$$= \frac{c \cdot \alpha_1}{c \cdot \alpha_1 + \alpha_2}, \text{ independentemente de } \theta \text{ novamente.}$$

Como $T(1, x_1^*) = T(2, x_2^*)$ (por definição da estatística T), segue-se do Princípio da Suficiência Fraca que:

$$MI(E^*, (1, x_1^*)) = MI(E^*, (2, x_2^*)). \quad (2)$$

Combinando (1) e (2), obtemos finalmente: $MI(E_1, x_1^*) = MI(E_2, x_2^*)$.

($\Leftrightarrow$) 1- Vamos agora provar que o Princípio da Verossimilhança implica o Princípio da Condicionalidade Fraca.

No experimento misto E^* , temos $f_0^*(j, x_j) = (\alpha_j \cdot f_0^j(x_j))$ e, então,

$$l_{(j, x_j)}(\theta) = \alpha_j \cdot l_{x_j}(\theta) \text{ (com } \alpha_j > 0)$$

Portanto, a verossimilhança para θ gerada no experimento E^* é proporcional à verossimilhança gerada no experimento E_j decorrendo do Princípio da Verossimilhança que:

$$MI(E^*, (j, x_j)) = MI(E_j, x_j)$$

No argumento acima, é necessário supor que a classe Ξ de experimentos associada ao método MI seja fechada sob misturas.

2- Provemos finalmente que o Princípio da Verossimilhança implica o Princípio da Suficiência Fraca.

Suponhamos que $T(X)$ seja uma estatística suficiente para θ . Pelo Critério da Fatoração de Fisher-Neyman temos, para todo $\theta \in \Theta$ e para todo $x \in \Omega$, $f_\theta(x) = h(x) \cdot g(T(x), \theta)$ (para alguma função $h(\cdot) > 0$ e $g(\cdot) > 0$). Para x_1 e $x_2 \in \Omega$ tais que $T(x_1) = T(x_2)$ temos, portanto,

$$f_{\theta}(x_1) = \frac{h(x_1)}{h(x_2)} \cdot f_{\theta}(x_2) \text{ e, então:}$$

$$l_{x_1}(\theta) \equiv \frac{h(x_1)}{h(x_2)} l_{x_2}(\theta)$$

seguindo pelo Princípio da Verossimilhança que: $MI(E, x_1) = MI(E, x_2)$.

□

Como ressaltado por Berger e Wolpert (1988), os resultados acima afirmam que a violação do Princípio da Verossimilhança por parte de algum método de inferência deixa este na insustentável posição de estar rejeitando ou o Princípio da Suficiência Fraca ou o Princípio da Condicionalidade Fraca. Esta é a posição da grande maioria dos métodos freqüentistas.

O desenvolvimento axiomático do Princípio da Verossimilhança devido a Birnbaum (1962) e aqui apresentado é restrito a experimentos onde o espaço amostral é discreto. Para Basu (1975) essa qualificação não restringe realmente a aplicabilidade do Princípio da Verossimilhança já que, devido à nossa impossibilidade de medir com precisão infinita, dados são sempre percebidos de forma discreta. Observando que, contudo, distribuições contínuas são boas aproximações para os modelos discretos, Berger e Wolpert (1988) apresentam uma versão geral do Princípio da Verossimilhança que também decorre de conceitos adaptados de condicionalidade e suficiência. A demonstração dessa versão geral do Princípio da Verossimilhança é feita em um nível matemático não elementar e será aqui omitida sem perda de continuidade.

2.3. DISCUSSÃO CRÍTICA DO PRINCÍPIO DA VEROSSIMILHANÇA

Como observado por David Lane (na discussão da monografia de Berger e Wolpert (1988)) as formas de evadirmos o Princípio da Verossimilhança são, basicamente, as seguintes:

- 1) rejeição do paradigma $(\Omega, \Theta, \{P_\theta\})$;
- 2) modificação dos princípios pré-fundamentais (isto é, do Princípio da Suficiência Fraca e/ou do Princípio da Condicionalidade Fraca);
- 3) exigência que outros princípios sejam satisfeitos.

Apresentamos a seguir essas formas de "escapes" ao Princípio da Verossimilhança.

1) Rejeição do paradigma $(\Omega, \Theta, \{P_\theta\})$.

Como apontado por Lane, a rejeição do "cenário" dos experimentos pode ser feita pela não requisição de estrutura alguma, tratando a análise dos dados exclusivamente de forma exploratória (descritiva), ou por uma requisição de uma estrutura maior (como fazem os estruturalistas, por exemplo).

Nesse sentido, Fraser (1985) apresenta alguns "contra-exemplos" ao Princípio da Verossimilhança, refutando-o simplesmente porque, seguindo a teoria da inferência estrutural (ver, por exemplo, Fraser (1968, 1983) e Barnett (1982) para detalhes da abordagem estrutural), o Princípio da Verossimilhança não é satisfeito.

As situações apresentadas por Fraser como contra-exemplos ao Princípio da Verossimilhança são restritas à sua visão particular de descrição de um experimento. Ou seja, ele modifica a descrição usual de experimentos (isto é, de $(\Omega, \Theta, \{P_\theta\})$) ao incluir uma estrutura nos modelos. Ora, dentro de um novo "cenário" não faz sentido propor contra-exemplos ao Princípio da Verossimilhança tal como este foi enunciado! De um ponto de

vista não restritivo, essas situações simplesmente revelam a violação do Princípio da Verossimilhança pela inferência estrutural.

O paradigma $(\Omega, \Theta, \{P_\theta\})$ também é rejeitado por uma linha totalmente distinta da estruturalista: os bayesianos preditivistas não reconhecem sentido algum nos "parâmetros" θ e tratam o problema da inferência estatística como o de construção coerente da distribuição $f(y|x)$ para observações y desconhecidas e operacionalmente definidas após o aprendizado x (ver também Wechsler (1993) e seção 3.2.1.).

Devemos observar adicionalmente que o contexto da discussão de David Lane é preditivista e não o representado pelo modelo $(\Omega, \Theta, \{P_\theta\})$.

Mesmo fora da linha preditivista há críticas à descrição $(\Omega, \Theta, \{P_\theta\})$ no que diz respeito ao seu significado. Lane atribui três possíveis significados aos elementos de Θ :

- a) θ é a distribuição P_θ ;
- b) Θ é um conjunto abstrato que indexa P_θ ;
- c) θ é um valor possível para algum parâmetro físico, e P_θ é a distribuição de X

quando θ é o valor verdadeiro daquele parâmetro.

Ele observa que a interpretação dessa "quantidade desconhecida" é fundamental, pois ela pode, por exemplo, induzir a uma restrição do Princípio da Verossimilhança. Vamos, portanto, avaliar cada interpretação citada acima.

A interpretação (a), de fato, expõe o Princípio da Verossimilhança a algumas discussões.

Uma das críticas diz respeito a adoção de um modelo paramétrico com uma densidade para X . A argumentação é de que modelos quase nunca são conhecidos exatamente, então

o Princípio da Verossimilhança tem uma aplicabilidade restrita. Em problemas não paramétricos (como representados na interpretação (a)), onde não há uma função da verossimilhança claramente definida, os argumentos de Birnbaum, seguindo essa linha de raciocínio, tornam-se questionáveis. Agora vejamos:

EXEMPLO: Suponhamos que X tenha valores possíveis no conjunto $\Omega = \{x_1, x_2, \dots\}$ e seja $\theta = (\theta_1, \theta_2, \theta_3, \dots)$ tal que $\Theta = \{\theta: 0 \leq \theta_i \leq 1 \text{ e } \sum \theta_i = 1\}$ e definamos: $f_\theta(x_i) = \theta_i$.

Θ representa vários modelos ($\{f_\theta\}$ é a família de densidades em Ω). Nesta situação não-paramétrica existe uma função de verossimilhança (ela é igual a $l_x(\theta) = \theta_i$) e o Princípio da Verossimilhança é aplicável.

O Princípio da Verossimilhança nos diz que a informação trazida pelos dados a respeito de uma quantidade desconhecida está contida na função de verossimilhança definida pelo modelo assumido para a distribuição dos dados. Ele é aplicável ainda que não se conheça o modelo, como observado no exemplo acima.

Em resumo, o Princípio da Verossimilhança não se restringe a modelos paramétricos. Berger e Wolpert (1988) observam que a implementação do Princípio da Verossimilhança pode ser mais difícil, contudo, em situações não-paramétricas, pois a função da verossimilhança torna-se difícil de ser utilizada a menos que restringida por uma distribuição a priori.

É interessante observar que essa crítica, ainda que dirigida ao Princípio da Verossimilhança, tem implicações profundas na teoria estatística já que questiona o uso de modelos estatísticos quando muitos procedimentos clássicos são extremamente dependentes do modelo paramétrico assumido.

Outra crítica deve-se a Lane. Ele diz que sob a interpretação (a) não se tem a aplicabilidade do Princípio da Condicionalidade Fraca sobre uma mistura, pois somente experimentos com mesma distribuição apresentam "o mesmo θ ". Em outras palavras, somente misturas de distribuições idênticas fazem sentido. A conseqüente existência de um único experimento torna o Princípio da Verossimilhança indistinguível do Princípio da Suficiência Fraca (como já observado por Basu (1975)). Dessa forma, o Princípio da Verossimilhança torna-se restrito e não nos permite comparar dois pontos amostrais procedentes de dois experimentos distintos com relação a qualquer método de inferência (essa forma restrita do Princípio da Verossimilhança é denominada Princípio Fortemente Restrito da Verossimilhança por Basu(1975)). Lane, por outro lado, observa que com a interpretação (b), essa restrição desaparece.

A interpretação (c), segundo Lane, é a mais problemática, pois trata parâmetros como elementos físicos, reais. Ele observa que a apresentação do Princípio da Verossimilhança por Berger e Wolpert segue a interpretação (c) ao propor que " θ seja o mesmo objeto em cada experimento" (o que Lane interpreta como fisicamente ou conceitualmente a mesma quantidade). Por outro lado, o conjunto Θ meramente serve para indexar a família de distribuições $\{P_\theta\}$ (seguindo, dessa forma, a interpretação (b)).

Berger e Wolpert (1988) observam que o Princípio da Verossimilhança pode ser aplicado nos casos em que a interpretação de θ segue ao exposto em (a) ou (c). Apesar de apresentar uma descrição matemática mais precisa, ressaltam que a interpretação (b) não se enquadra adequadamente ao contexto do Princípio da Verossimilhança, pois a existência de um conjunto índice Θ comum aos experimentos da mistura não garante que o parâmetro θ seja (também) comum.

Para Berger e Wolpert, o fato de que o Princípio da Verossimilhança seja aplicável (até mesmo) em situações em que θ é um "parâmetro físico real" é uma justificativa para o

"descuido" com a definição de θ . Com isto eles tentaram evitar questões filosóficas, pois: *"Convincing a practicing statistician, who routinely uses models, to base his analysis on the observed likelihood function is a significant practical step. Informing him that his model parameters really have no meaning is unlikely to cause much improvement in his statistical practice"* (Berger e Wolpert, p. 194).

2) Modificação dos princípios pré-fundamentais.

2.1) Modificação do Princípio da Condicionalidade Fraca:

Durbin(1970) propôs uma versão enfraquecida do Princípio da Condicionalidade Fraca, a qual se aplica somente a variáveis condicionantes que dependem unicamente da estatística suficiente mínima.

Sob essa versão, o Princípio da Verossimilhança não pode ser demonstrado. Para chegar a esta conclusão, ele recorreu ao seguinte argumento: No experimento misto E^* (descrito no Princípio da Condicionalidade Fraca), a estatística $T(J, X_J)$ é suficiente (como demonstrado na seção 2.2). Contudo, após ter observado $J=j$, o valor conhecido de J não pode fazer parte da estatística suficiente mínima. Desse modo, a demonstração do Princípio da Verossimilhança partindo-se da adoção do Princípio da Condicionalidade proposto por Durbin fica comprometida, pois com J não fazendo parte da estatística suficiente mínima, o condicionamento em $J=j$ não é admitido.

2.2) Modificação do Princípio da Suficiência Fraca

Kalbfleisch(1975) propôs uma versão enfraquecida do Princípio da Suficiência Fraca que não se aplica a misturas de experimentos. O Princípio da Verossimilhança não pode ser demonstrado partindo-se do Princípio da Condicionalidade Fraca e dessa versão enfraquecida do Princípio da Suficiência Fraca proposta por Kalbfleisch já que o

desenvolvimento axiomático proposto por Birnbaum(1962) depende da construção de um experimento E^* (mistura de experimentos E_1 e E_2).

Joshi (1990) apontava uma falha na demonstração do Princípio da Verossimilhança. Ele observou que pelo Princípio da Condicionalidade Fraca somente o experimento realmente realizado é relevante para inferências a respeito da quantidade desconhecida θ . Assim, não se poderia aplicar o Princípio da Suficiência Fraca a (J, X_J) . Como observado por Berger (1990) a demonstração do Princípio da Verossimilhança (seção 2.2) não é incorreta. O Princípio da Suficiência Fraca aplica-se a qualquer experimento e , portanto, também ao experimento misto. A falha na demonstração do Princípio da Verossimilhança apontada por Joshi, como podemos perceber, está relacionado com a versão enfraquecida do Princípio da Suficiência Fraca acima exposto.

Como podemos observar, as modificações apresentadas por Durbin(1970) e Kalbfleisch(1975), respectivamente, dos princípios fundamentais de condicionalidade e suficiência são extremamente restritivas. De fato, as versões dos princípios de condicionalidade e suficiência não se aplicam a situações gerais de experimentos, o que torna injusticável a adoção dessas versões como princípios fundamentais na opinião de Berger e Wolpert(1988).

3) Adoção de outros princípios como fundamento.

Vamos limitar a nossa discussão à adoção do Princípio da Confiança (como fazem os adeptos da escola freqüentista). Informalmente, o Princípio da Confiança diz que "qualquer estatístico que usa uma metodologia em que se faz afirmações ou obtém-se conclusões com uma acurácia especificada, deve garantir que, em repetições, a sua real acurácia deve ser

pelo menos aquela especificada (prometida)" (Berger e Wolpert(1988), p.66-67). Conceitos como coeficiente de confiança e erro do tipo I obedecem o Princípio da Confiança. Este, por sua vez, é fortemente baseado no conceito de probabilidade freqüentista.

Esse princípio é conflitante com o Princípio da Verossimilhança. Berger e Wolpert (1988) mostram a inconsistência de se violar o Princípio da Verossimilhança utilizando o conceito de coerência (ver também Loschi (1992)). Observam que pelo Princípio da Verossimilhança as observações x_1 e x_2 , respectivamente, provenientes de experimentos E_1 e E_2 , ao gerarem funções de verossimilhança proporcionais trazem a mesma "informação" a respeito de θ . Em um cenário de jogos, ao estabelecer "prêmios" (por exemplo, coeficientes de confiança) diferentes a um mesmo bilhete (no exemplo citado acima, um intervalo de confiança) o *bookmaker* viola o Princípio da Verossimilhança. Com efeito, Berger e Wolpert(1988) enunciam um teorema onde fica claro que agindo assim o *bookmaker* é incoerente, isto é, é um perdedor com certeza (ver também Berger e Wolpert (1988) para a avaliação da violação do Princípio da Verossimilhança no contexto de Teoria da Decisão utilizando o conceito de admissibilidade).

CAPÍTULO 3 - ENSAIOS CLÍNICOS E O PRINCÍPIO DA VEROSSIMILHANÇA

3.1. INTRODUÇÃO

Em uma fase preliminar de estudo, novas drogas são utilizadas em animais com o intuito de se avaliar efeitos farmacológicos e toxicidade. No caso de apresentarem resultados iniciais satisfatórios, esses ensaios iniciais dão também informações das doses iniciais em seres humanos.

A apreciação dos efeitos de uma nova droga em indivíduos humanos é tipicamente realizada em quatro fases. A **fase I** constitui-se de um estudo piloto que avalia efeitos farmacológicos da droga e dose de tolerância. Envolve sujeitos voluntários "normais" com o propósito de se reduzir o risco de problemas tóxicos e evitar confundimento entre doença e efeito farmacológico. O monitoramento dos participantes desses ensaios (geralmente variando de 4 a 20 indivíduos) é intenso. A avaliação de bioequivalência também é efetuada nos ensaios clínicos da fase I (para maiores detalhes, ver capítulo 4).

Quando os níveis de dose seguros são estabelecidos, a droga é então administrada em pacientes - **fase II** - para determinar se há evidência suficiente de eficácia preliminar para justificar estudos adicionais. Os pacientes selecionados, em número maior do que o observado na fase I, são cautelosamente monitorados. Eles vêm de uma população relativamente restrita. Apresentam a doença sob investigação, mas são relativamente livres de outras condições de confundimento (por exemplo, terapias concomitantes). A população pode ser gradativamente estendida (mas usualmente não se estende àquela população consumidora do produto). Nessa fase, a avaliação da droga consiste de estudos clínicos farmacológicos (fase IIA) e ensaios clínicos controlados (fase IIB). Os estudos farmacológicos são planejados com o intuito de se estimar os limites efetivos de dosagem,

incluindo caracterização da curva dose-resposta e determinação de dose mínima efetiva. A fase IIB também pode incluir estudos farmacocinéticos em pacientes onde se avaliam a absorção, a distribuição, o metabolismo e a eliminação do produto. Os estudos da fase IIB provêm uma demonstração inicial da segurança e eficácia da droga em uma dose ou doses avaliadas nos estudos clínicos farmacológicos.

A fase III é uma extensão da fase II englobando os ensaios clínicos controlados comparativos planejados com o objetivo de se obter informações mais precisas sobre efeitos adversos e eficácia. A população de pacientes pode ser menos restrita que aquela determinada na fase II, e o número de pacientes pode variar de centenas a mais de mil. Os pacientes podem ser menos intensivamente observados do que nas fases I e II.

Em uma fase adicional, fase IV, que consiste de ensaios pós-aprovação da droga por órgão regulador competente, os estudos prosseguem com diferentes propósitos: avaliar a incidência de efeitos adversos raros, explorar um específico efeito farmacológico, determinar efeitos da droga sobre morbidade e mortalidade, etc. Pode se constituir de ensaios clínicos em uma população de pacientes não estudada anteriormente (por exemplo, crianças e jovens); ou para investigar outros possíveis usos da droga.

Fundamental para que uma nova droga seja utilizada na prática médica, é a sua demonstração de eficácia quando comparada com uma terapia usual. Estudos controlados, onde alguns pacientes são tratados como experimentais por utilização da droga experimental e outros como controles por omissão (ou substituição) da nova droga, permitem um padrão de comparabilidade. Esses controles podem ser concorrentes (quando o mesmo protocolo é

seguido para os grupos controle e experimental) ou históricos ⁽¹⁾. Apesar de reduzirem o custo da pesquisa e até liberar o médico para aplicar um tratamento melhor aos pacientes experimentais, os controles históricos apresentam desvantagens já que o tratamento controle de pacientes históricos é fixo, e os grupos nem sempre são muito comparáveis.

Focalizando os ensaios clínicos com controles concorrentes, o que se justifica pela sua larga aceitação na comunidade científica, dois aspectos são merecedores de alguma atenção: a aleatorização, e o planejamento.

A aleatorização é tida como um fundamento básico nos ensaios clínicos, tendo como propósitos evitar vícios na atribuição de tratamentos e o confundimento de variáveis desconhecidas. Por exemplo, suponhamos que os médicos atribuam tratamentos baseados (unicamente) no exame do paciente. Então, um vício pode ter sido introduzido: os prognósticos podem estar confundidos com tratamento.

A aleatorização é também crucial para os propósitos de inferências estatísticas (ver seção 3.3.) freqüentistas. A suposição de que os pacientes do estudo possam ser considerados como uma amostra aleatória de alguma população é fundamental, contudo, a aleatorização não garante essa representatividade. Suponhamos, por exemplo, que um terço da população de pacientes sejam mulheres. Uma particular aleatorização dos pacientes daquela população a dois grupos de tratamentos não necessariamente (e a menos de uma estratificação da população) produz grupos com iguais proporções de mulheres (e assim por diante com as outras variáveis). De fato, a estratificação da população é recomendada por

(1) Há dois tipos de controles históricos:

1- controles tratados na mesma instituição e pelos mesmos médicos dos pacientes experimentais e,
2- controles tratados em diferentes instituições (aqui se incluem controles apresentados na literatura).

estatísticos bayesianos de modo a adiar a aleatorização o tanto quanto for possível (Berry (1990)).

Ainda que os tratamentos sejam aleatorizados, a possibilidade de que um paciente responda de um modo diferente de acordo com o seu conhecimento do tratamento recebido, ou de que o médico atribua tratamento diferenciado ou meça diferentemente a resposta de pacientes, que ele sabe, serem experimentais, pode interferir no estudo. Uma solução é o mascaramento "double-blind" onde nem o paciente, nem o médico são informados do tratamento atribuído. Observemos que esta solução, embora largamente reconhecida como ideal, nem sempre é praticável já que nem sempre é possível esconder o tratamento para o médico (por exemplo, quando envolvem procedimentos cirúrgicos, ou quando um tratamento tem um efeito "auto-revelador").

De modo generalizado podemos classificar os planejamentos dos ensaios clínicos em dois grandes grupos:

1) Planejamentos não seqüenciais: caracterizados pela inclusão de um número fixo de pacientes para os quais o tratamento é atribuído aleatoriamente. Aqui se incluem, por exemplo, os ensaios "cross-over" e estudos paralelos (Rodda et al.(1988) e Berry(1990) descrevem e discutem estes e outros planejamentos mais comumente utilizados nos ensaios clínicos farmacológicos).

2) Planejamentos seqüenciais: o monitoramento dos dados pode ser feito continuamente, ou periodicamente. Esses planejamentos permitem que os resultados provisórios afetem a atribuição dos tratamentos para os próximos pacientes, ou que a parada do experimento ocorra mais cedo.

A mudança do esquema usual de aleatorização 50-50 (isto é, esquema que atribui probabilidades iguais para cada tratamento) face aos resultados provisórios apresenta vantagens nos planejamentos "adaptativos". Suponha, por exemplo, que uma droga

experimental apresente, interinamente, resultados de melhor eficiência. Se os "novos" pacientes forem aleatorizados aos grupos tratamento e controle com probabilidades modificadas (no exemplo, a probabilidade de atribuir um novo paciente à droga experimental é aumentada) pode-se reduzir o número de pacientes que recebem tratamento desfavorável e, comparando-se com o esquema de aleatorização 50-50, há uma melhora das chances de que os participantes recebam o melhor tratamento. Por outro lado, a parada precoce do experimento sugerida pelos dados provisórios pode ser crucial. Continuar um ensaio além do ponto onde há evidência de que uma droga é efetivamente melhor do que outra, torna a aleatorização não ética.

Para utilizar a **inferência clássica** (no uso de intervalos de confiança, "p-values", etc), em geral, não é suficiente conhecer os dados. É preciso conhecer e considerar **todos** os detalhes do planejamento. Nos estudos seqüenciais deve haver também uma completa especificação dos dados provisórios para se requisitar a parada. Como resultado dessas limitações, poucos ensaios clínicos patrocinados por companhias farmacêuticas permitem uma parada precoce (embora, como Berry (1990) observa, os dados sejam monitorados periodicamente nos ensaios clínicos da fase I e fase II, por exemplo).

O Princípio da Verossimilhança apresenta fortes implicações na inferência estatística e, em particular, as conseqüências práticas em áreas como a de ensaios clínicos são profundas. Discutimos nas seções 3.2.1. e 3.2.2. o relacionamento do Princípio da Verossimilhança com duas escolas de inferência: a bayesiana e a clássica. As implicações teóricas do Princípio da Verossimilhança são apresentadas nas seções 3.2.3. a 3.2.5. Apresentamos e discutimos na seção 3.3. alguns exemplos que ilustram a situação absurda de se utilizar métodos que violam o Princípio da Verossimilhança. Na seção 3.4. apresentamos uma discussão sobre o papel do Princípio da Verossimilhança em ensaios clínicos.

3.2. IMPLICAÇÕES DO PRINCÍPIO DA VEROSSIMILHANÇA

3.2.1. A "QUASE" INEVITABILIDADE DA INFERÊNCIA BAYESIANA

Na visão bayesiana a única descrição sensata (coerente) de quantidades desconhecidas (e, portanto, aleatórias) é através de probabilidades (Lindley(1990) e Loschi(1992)). Em um experimento estatístico E, além dos dados, o parâmetro é também um objeto aleatório em uma representação matemática de incerteza (Piccinato(1992)). Nesse contexto, o foco de atenção da inferência bayesiana é a obtenção da distribuição de probabilidades em Θ , dadas as informações disponíveis. Representando toda a nossa incerteza sobre θ (anterior à obtenção dos dados) por uma distribuição a priori em Θ , digamos $\pi(\cdot)$, após a observação dos dados obtemos (pelo Teorema de Bayes) a "atualização" da nossa incerteza na distribuição de probabilidades a posteriori $\pi(\cdot|x)$.

TEOREMA - O Princípio da Verossimilhança é implicado pela inferência bayesiana.

DEMONSTRAÇÃO

Seja $\pi(\cdot)$ a densidade a priori para θ ⁽²⁾.

A densidade a posteriori de θ é obtida pelo Teorema de Bayes:

$$\pi(\theta | x) = \frac{\pi(\theta).f(x|\theta)}{\sum_{\Theta} \pi(\theta).f(x|\theta)} = \frac{\pi(\theta).l_x(\theta)}{g(x)} \quad \text{onde } g(x) = \sum_{\Theta} \pi(\theta).l_x(\theta)$$

A densidade a posteriori depende do experimento somente através da função de verossimilhança para θ gerada pela observação x . Como toda a inferência bayesiana segue

⁽²⁾ Supomos que $\pi(\cdot)$ reflete uma opinião sobre θ que é independente de E e anterior à observação de X.

da obtenção de $\pi(\theta|x)$, o Princípio da Verossimilhança é uma consequência imediata do enfoque bayesiano.

□

Apesar de implicado pela inferência bayesiana, o Princípio da Verossimilhança não é considerado um princípio bayesiano pelos bayesianos (Wechsler(1989)). Nas palavras de de Finetti(1975) o Princípio da Verossimilhança *"inevitably appears to be rather obvious, and certainly not worth getting excited about"* (em outras palavras: o Princípio da Verossimilhança é tão óbvio que não devemos ficar muito entusiasmados com ele).

O Princípio da Verossimilhança não implica, contudo, a inferência bayesiana. Como citado no capítulo 2, a inferência da verossimilhança também obedece o Princípio da Verossimilhança (para maiores detalhes da inferência da verossimilhança veja, por exemplo, Edwards (1972)). Daí, a "quase" inevitabilidade da inferência bayesiana.

Além disso, o Princípio da Verossimilhança é enunciado em um contexto (não-preditivista ⁽³⁾) Fisheriano de modelos indexados por "parâmetros". De fato, como ressaltado por Wechsler (1989), os bayesianos subjetivistas criticam a ausência de opiniões a priori sobre as quantidades desconhecidas (e portanto aleatórias) na formulação do Princípio da Verossimilhança.

⁽³⁾ Em um contexto neo-bayesiano (ou predictivista) "PROBABILIDADE NÃO EXISTE" (deFinetti (1972)) e tampouco parâmetro. Sob este enfoque se rejeita a arbitrariedade da consideração da existência de parâmetros. Eles **podem** ser induzidos naturalmente através de julgamentos de invariância (e.g. permutabilidade) sobre os dados (quantidades aleatórias observáveis).

Para maiores detalhes do argumento predictivista veja Wechsler(1993).

O BAYESIANO SUBJETIVISTA versus O BAYESIANO OBJETIVISTA:

Um bayesiano subjetivista interpreta uma distribuição de probabilidade como uma opinião e, portanto, ela pode variar de pessoa a pessoa, pois a experiência (acumulada) que cada um apresenta a respeito de uma mesma quantidade pode ser diferente. Essa subjetividade é, contudo, alvo de críticas.

A idéia de utilizar uma priori não-informativa, que representa "ignorância total" e "deixa os dados falarem por si mesmos o mais possível" seduz os bayesianos objetivistas. Sob essa visão bayesiana alternativa, a distribuição a priori é vista como sendo "objetiva", uma medida lógica da incerteza total. (Para maiores detalhes ver, por exemplo, Barnett(1982), Dawid(1983) e Bernardo e Smith (1993)). É interessante observar que bayesianos objetivistas violam o Princípio da Verossimilhança (ver, exemplo 1, seção 3.3.). Além disso, para os bayesianos subjetivistas, a posição "objetivista" é, antes que uma alternativa na escolha da priori, absurda. Os subjetivistas rejeitam a noção de "objetividade" física e definem "objetividade" como unanimidade de opiniões (sinceras!). A unanimidade forçada tampouco é tolerada.

3.2.2. A INCOMPATIBILIDADE COM A INFERÊNCIA CLÁSSICA

Sob o enfoque clássico ou freqüentista, probabilidade é definida em termos de limites de freqüências em repetições hipotéticas de um ensaio. Qualquer procedimento inferencial decorre do comportamento de uma seqüência infinita de repetições do mesmo experimento (sob mesmas condições) com um mesmo valor "verdadeiro" do parâmetro desconhecido. Piccinato (1992) enfatiza que tais ensaios são somente hipotéticos, e que esse procedimento deveria ser desenvolvido em um campo lógico, independentemente da disponibilidade prática de uma ou mais amostras aleatórias.

Como uma consequência do desenvolvimento de procedimentos dependentes do conhecimento do espaço amostral, a inferência clássica viola o Princípio da Verossimilhança ao considerar dados não observados na inferência.

Na literatura há uma série de artigos que apresentam exemplos onde constatamos a violação do Princípio da Verossimilhança no uso de procedimentos clássicos. (Veja por exemplo: Berger e Berry(1988), Berry(1987 - a) e a seção 3.3., adiante.).

É interessante observar que a estimação de máxima verossimilhança não viola o Princípio da Verossimilhança. Contudo, provido de propriedades assintóticas, os estimadores de máxima verossimilhança perdem essa característica inicial.

TESTES DE HIPÓTESES versus TESTES DE SIGNIFICÂNCIA

R.A. Fisher, E.S. Pearson e J. Neyman são considerados os criadores da teoria clássica de testes de hipóteses e, os responsáveis pela existência de pelo menos duas escolas distintas de abordagem deste problema. A distinção usualmente difundida é a de que os testes de hipóteses de Neyman e Pearson são realizados com um nível de significância fixado, enquanto que na teoria de testes de significância de Fisher o objetivo é a apresentação de um nível descritivo de significância ("p-value") que pode ou não ser acompanhado de um "pronunciamento" de significância do resultado⁽⁴⁾. Esta é a visão mais superficial das diferenças existentes entre as escolas Fisheriana e Neyman-Pearsoniana.

(4) Barnett(1982) apresenta ainda uma distinção para testes de significância. **Teste de Significância Pura** corresponde ao processo determinar (unicamente) o "p-value". Este teste tem caráter meramente inferencial já que provém um meio de delimitar o conhecimento de θ (sobre a inconsistência do dado observado) quando θ pertence ao espaço paramétrico definido pela hipótese nula. Um **Teste de Significância de nível α** corresponde a um processo adicional de escolher um nível de significância α e rejeitar a hipótese nula ao nível α se p ("p-value") $\leq \alpha$, ou equivalentemente escolher um valor crítico $t_0(\alpha)$ para a estatística de teste $t(X)$ e rejeitar a hipótese nula se $t(x) > t_0(\alpha)$.

Na abordagem de Neyman e Pearson, a otimização da escolha de um particular teste de hipóteses (um procedimento inferencial) necessita da especificação da hipótese alternativa, bem como da hipótese nula. Estas definem dois tipos de erros: a falsa rejeição da hipótese nula (erro do tipo I) e a sua falsa aceitação (erro do tipo II). O teste escolhido é aquele que minimiza a probabilidade do erro do tipo II sujeito a uma probabilidade máxima fixada do erro do tipo I. Sob este enfoque, um teste é um algoritmo, um procedimento que permite rejeitar ou aceitar hipóteses, isto é, um procedimento indutivo.

Para Fisher, a inferência tem uma característica indutiva, um processo que permite obter conhecimentos gerais partindo do particular e ele via, com horror, a doutrina (de Neyman e Pearson) em que o "raciocínio" não pode ser aplicado aos dados empíricos para obter inferências. Sob a sua lógica, a diferença entre as duas escolas reside entre extrair conclusões (no enfoque de Fisher) e tomar decisões (no enfoque de Neyman e Pearson). Fisher, em suas críticas, argumenta que a idéia de aceitar uma hipótese (simplesmente porque os dados não a contradizem) era absurda e que, embora conceitualmente a noção de erro do tipo II possa ser importante, ela não pode ser avaliada quantitativamente.

Outra diferença entre Fisher e Neyman-Pearson reside no conceito de condicionalidade. O princípio básico da inferência Fisheriana é calcular probabilidades condicionais no valor observado de uma estatística ancilar adequada (Dawid(1983)). Suponhamos, por exemplo, que uma quantidade aleatória, condicionalmente dado μ , tenha distribuição normal com média μ e variância 1 se o resultado do lançamento de uma moeda honesta for cara, e de variância 4 em caso contrário. O lançamento da moeda resultou em cara, e portanto, a quantidade aleatória tem, condicionalmente dado μ , distribuição Normal(μ , 1). Para Fisher, o cálculo de probabilidade é condicional no resultado do lançamento da moeda (isto é, baseia-se na distribuição Normal (μ , 1)); enquanto que para Neyman-Pearson a possibilidade de que o resultado fosse coroa deve ser considerada. Em

outras palavras, sob este único aspecto, a abordagem de Fisher respeita o Princípio da Condicionalidade e a de Neyman-Pearson o viola. (Ressaltemos, contudo, que ao calcular o nível descritivo, ainda que condicional no valor do lançamento da moeda, a utilização de dados "mais extremos" do que o observado na abordagem Fisheriana viola o Princípio da Condicionalidade.)

Para maiores detalhes a respeito das diferenças de abordagens do problema de teste de hipóteses entre Fisher e Neyman-Pearson veja Dawid (1983), Lehmann(1993) e Inman(1994).

Pereira e Wechsler (1993) mostram as dificuldades causadas pela não consideração da hipótese alternativa na definição dos testes de significância através de alguns exemplos. Eles definem um "P-value", isto é, um "p-value" que considera a hipótese alternativa e que evita essas inconsistências. O cálculo do "P-value" constitui um procedimento que viola o Princípio da Verossimilhança (como observado pelos próprios autores). Contudo, a adoção desse conceito nos testes de significância soluciona a diferença existente entre as escolas Fisheriana e Neyman-Pearsoniana no aspecto da consideração de hipóteses alternativas.

3.2.3. A IRRELEVÂNCIA DAS REGRAS DE PARADA

Seja $(E_1, E_2, \dots)$ uma seqüência de experimentos estatísticos que consistem em observar X_i em Ω_i , isto é, $E_i = (\Omega_i, \Theta, \{P_{i,\theta}\})$, $(i=1, 2, \dots)$. Para cada instante n ($n=1, 2, \dots$),

podemos definir o experimento composto $E^n = (\Omega^n, \Theta, \{P_\theta^n\})$, onde $\Omega^n = \prod_{i=1}^n \Omega_i$ e $\{P_\theta^n\}$ é a

classe de distribuições de probabilidade sobre Ω^n indexada em Θ associada à distribuição de um processo estocástico $\{P_{1,\theta}, P_{1,\theta}, P_{3,\theta}, \dots\}$ (se os experimentos forem independentes, P_θ^n é a medida produto das distribuições de probabilidade, indexadas em Θ , sobre Ω_i ($i=1, 2, \dots, n$)).

DEFINIÇÃO: Um ensaio seqüencial é qualquer forma de coleção de dados em que a decisão de continuar ou não a coleta de dados depende, de alguma forma, da informação até então obtida.

DEFINIÇÃO: Uma regra de parada, τ , de um experimento seqüencial é uma seqüência $(\tau_0, \tau_1, \tau_2, \dots)$ em que $\tau_0 \in [0, 1]$ é uma constante e, para $m \geq 1$, $\tau_m: \Omega^m \rightarrow [0, 1]$ é uma função (mensurável) sobre Ω^m tendo a seguinte interpretação:

- τ_0 é a probabilidade de que o experimento encerre-se sem que nenhuma observação tenha sido obtida (por exemplo, $\tau_0 = 1$ quando o experimento é inviabilizado pelo alto custo da amostragem);
- $\tau_m(x^m)$ representa a probabilidade condicional de encerrar-se a amostragem após m observações, dado que estas m observações são $x^m = (x_1, \dots, x_m)$.

A regra de parada é chamada determinística ou não-aleatorizada se $\tau_m(x^m) \in \{0, 1\}$ para todo $m \geq 0$ e todo $x^m \in \Omega^m$. Caso contrário, é chamada uma regra de parada aleatorizada.

DEFINIÇÃO: O tempo de parada, N_τ , correspondente a uma regra de parada τ é o tamanho da amostra obtida no experimento seqüencial.

Uma regra de parada τ é própria se $P_\theta(N_\tau < \infty) = 1, \forall \theta \in \Theta$.

EXEMPLO: Sejam $X_1, X_2, \dots$ quantidades aleatórias que são condicionalmente, dado θ , independentes e identicamente distribuídas segundo Bernoulli (θ).

Situação 1: Seja $E^{(\tau^1)}$ um experimento com tamanho amostral n fixado. A regra de parada τ^1 é definida por:

$$\tau_0^1 = 0; \quad \tau_m^1(x^m) = \begin{cases} 0 & \text{se } m < n \\ 1 & \text{se } m = n \end{cases}$$

Situação 2: Seja $E^{(\tau^2)}$ um experimento seqüencial em que a parada ocorre ao se observar o k -ésimo sucesso. A regra de parada τ^2 é definida por :

$$\tau_0^2 = 0; \quad \tau_m^2(x^m) = \begin{cases} 0 & \text{se } \sum_{i=1}^m x_i < k \\ 1 & \text{se } \sum_{i=1}^m x_i = k \end{cases}$$

DEFINIÇÃO: Um experimento seqüencial $E^{(\tau)}$ com regra de parada τ própria é assim definido: $E^{(\tau)} = (\Omega^{(\tau)}, \Theta, \{P_\theta^{N_\tau}\})$ onde:

$$\Omega^{(\tau)} = \bigcup_{N_\tau=1}^{\infty} [\{N_\tau\} \times \Omega^{N_\tau}] \text{ com } N_\tau: \text{tempo de parada associado à regra } \tau.$$

Vamos supor que Ω_i seja discreto (para todo i) e que, para todo j exista $f(x^j|\theta)$, a densidade conjunta de $\mathbf{X}^j = (X_1, \dots, X_j)$.

Enunciaremos agora o Princípio da Regra de Parada.

PRINCÍPIO DA REGRA DE PARADA - Sejam $E^{(\tau)} = (\Omega^{(\tau)}, \Theta, \{P_\theta^{N_\tau}\})$ um experimento seqüencial com regra de parada τ , própria, $(E^{(\tau)} \in \Xi)$, e $\{E^i, i \in \mathbb{N}\}$ a seqüência de experimentos compostos, onde $E^i = (\Omega^i, \Theta, \{P_\theta^i\})$ ($E^i \in \Xi$, para todo $i \in \mathbb{N}$). Suponhamos que após a realização de $E^{(\tau)}$ tenhamos observado $\mathbf{x}^n = (x_1, \dots, x_n)$. Então é necessário que:

$$MI(E^{(\tau)}, (n, \mathbf{x}^n)) = MI(E^n, \mathbf{x}^n).$$

Em outras palavras, a inferência resultante de $E^{(\tau)}$ deve ser meramente a inferência resultante do experimento E^n com tamanho amostral fixado, ou seja, após a realização do experimento seqüencial, a regra de parada deve ser irrelevante para inferências a respeito de θ .

Como é esperado, o Princípio da Regra de Parada vigora quando o Princípio da Verossimilhança é válido, como veremos no seguinte corolário.

COROLÁRIO 2 : O Princípio da Regra de Parada decorre do Princípio da Verossimilhança.

DEMONSTRAÇÃO:

Considere uma regra de parada τ própria fixada.

Seja $\mathbf{x}^n$ a amostra obtida ao final do experimento seqüencial $E^{(\tau)}$. Observemos que para todo $\theta \in \Theta$,

$$f((n, \mathbf{x}^n) | \theta) = \left[\tau_n(\mathbf{x}^n) \cdot \prod_{i=1}^{(n-1)} [1 - \tau_i(\mathbf{x}^i)] \right] \cdot [1 - \tau_0] \cdot f(\mathbf{x}^n | \theta).$$

Logo a verossimilhança para θ gerada por $(n, \mathbf{x}^n)$ será:

$$l_{(n, \mathbf{x}^n)}^{(\tau)}(\theta) = \left[\tau_n(\mathbf{x}^n) \cdot \prod_{i=1}^{(n-1)} [1 - \tau_i(\mathbf{x}^i)] \right] \cdot [1 - \tau_0] \cdot l_{\mathbf{x}^n}(\theta)$$

Como $\left[\tau_n(\mathbf{x}^n) \cdot \prod_{i=1}^{(n-1)} [1 - \tau_i(\mathbf{x}^i)] \right] \cdot [1 - \tau_0]$ é uma constante (independe de θ) segue que:

$$l_{(n, \mathbf{x}^n)}^{(\tau)}(\theta) \propto l_{\mathbf{x}^n}(\theta)$$

Observemos que $l_{\mathbf{x}^n}(\theta)$ é (também) a função de verossimilhança gerada pela observação de $\mathbf{x}^n$ em E^n , o experimento composto de tamanho amostral n fixado. Da proporcionalidade das funções de verossimilhanças geradas por $(n, \mathbf{x}^n)$ em $E^{(\tau)}$ e $\mathbf{x}^n$ em E^n , para todo $\theta \in \Theta$, e **desde que** $E^{(\tau)}$ e $E^n \in \Xi$ (ver observação abaixo), decorre do Princípio da Verossimilhança que:

$$MI(E^{(\tau)}, (n, \mathbf{x}^n)) = MI(E^n, \mathbf{x}^n).$$

□

Observação: Novamente o leitor deverá ter percebido que é crucial para a demonstração acima que imponhamos a comparabilidade entre $E^{(\tau)}$ e E^n com relação a MI, isto é, é necessário que $E^{(\tau)}$ e $E^n \in \Xi$.

3.2.4. A IRRELEVÂNCIA DOS MECANISMOS DE CENSURA

Seja $E = (\Omega, \Theta, \{P_\theta\})$ um experimento estatístico não-seqüencial. Uma censura fixa ocorre quando, ao invés de X , observamos $Y=g(X)$, onde g é alguma função conhecida (usualmente, g não é injetiva). Formalmente temos a seguinte definição:

DEFINIÇÃO: Uma censura fixa é uma função g

$$g: \Omega \rightarrow \mathcal{Y}.$$

O respectivo experimento censurado é $E^g = (\mathcal{Y}, \Theta, \{P_\theta \circ g^{-1}\})$, onde $g^{-1}(A)$ é a imagem inversa $\{x \in \Omega: g(x) \in A\}$.

Quando a censura envolve um truncamento e a ocorrência deste também é registrada, temos o caso particular:

$$\mathcal{Y} = \Omega \times \{0, 1\},$$

onde $g(x)=(y, 1)$ indica que houve truncamento, enquanto

$g(x)=(x, 0)$ indica que não houve truncamento.

EXEMPLO 1: Suponhamos que $\mathbf{X} = (X_1, X_2, \dots, X_n)$, sejam os tempos de sobrevida de n pacientes diagnosticados como HIV-positivos e acompanhados em um ensaio clínico por até 10 anos. Para o i -ésimo paciente, na realidade observamos:

$$Y_i = (Y_i^1, Y_i^2) = (\min\{X_i, 10\}, I[X_i > 10]),$$

ou seja, o tempo de sobrevida censurado (possivelmente truncado em 10) e o indicador de truncamento.

Neste exemplo, $Y_i = g(X_i) = (\min\{X_i, 10\}, I[X_i > 10])$, e

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_n) = g(X_1, X_2, \dots, X_n) = g(\mathbf{X}).$$

Este tipo de truncamento é conhecido como "censura do tipo I" (à direita).

EXEMPLO 2: Suponhamos agora que $X = (X_1, X_2, \dots, X_n)$, mas que somente as r menores observações serão registradas, enquanto que as $(n-r)$ maiores serão truncadas no valor da r -ésima delas. Neste caso,

$$Y_i = g(X_i) = (\min\{X_i, X_{(r)}\}, I[X_{(r)} < X_i])$$

Este tipo de truncamento é conhecido como "censura do tipo II" (à direita).

Dawid (1977) considera a situação em que não há valor de truncamento, isto é, a situação em que:

$$\begin{aligned} g: \Omega &\rightarrow Y = A \cup \{\varphi\} \\ x &\rightarrow g(x) = \begin{cases} x & \text{se } x \in A \\ \varphi & \text{se } x \notin A \end{cases} \end{aligned}$$

onde: A é um subconjunto de Ω e

φ é um "símbolo", i.e., $\varphi \notin \Omega$.

Se permitirmos $\varphi \in \Omega$, (ou, em particular, $\varphi \in A$), teremos o valor de truncamento na observação Y_i . Com a inclusão do indicador de truncamento, retornaremos ao caso anterior onde $\mathcal{Y} = \Omega \times \{0, 1\}$, ou $\mathcal{Y} = A \times \{0, 1\}$.

O leitor já estará antevendo que o Princípio da Verossimilhança não permitirá que o mecanismo de censura exerça influência na inferência quando o valor da observação for não-truncado. Para fazer esta construção, precisamos da seguinte definição:

DEFINIÇÃO : Seja $E=(\Omega, \Theta, \{P_\theta\})$ um experimento (não censurado) e consideremos dois mecanismos de censura fixa, g_1 e g_2 . Seja A um subconjunto de Ω . Dizemos que g_1 e g_2 são equivalentes sobre A quando:

$$g_1^{-1}(g_1(x)) = g_2^{-1}(g_2(x)) \text{ , qualquer que seja } x \in A.$$

A definição diz que, para todo x em A , a "recuperação" obtida a partir do valor registrado $y=g(x)$ é a mesma, seja com censura g_1 ou com censura g_2 .

A não-censura sobre A é o caso em que $g^{-1}(g(x)) = \{x\}$ para todo $x \in A$. Se $A=\Omega$ temos o caso particular da não-censura.

EXEMPLO 3: Suponhamos que $\mathbf{X} = (X_1, X_2, \dots, X_n)$ e que g_1 seja a censura do tipo I à direita em 10, enquanto que g_2 seja a censura do tipo II à direita em $X_{(r)}$.

Podemos então verificar que g_1 e g_2 são equivalentes sobre o subconjunto $A = \{\mathbf{x} = (x_1, \dots, x_n) \in \Omega : x_{(r)} = 10\}$.

De fato, se $x \in A$, $g_1^{-1}(g_1(x)) = g_2^{-1}(g_2(x)) = \{x\}$ sempre que $x_{(n)} < 10$ e $g_1^{-1}(g_1(x)) = g_2^{-1}(g_2(x)) = \{w = (w_1, \dots, w_n) : w_i = x_i \text{ se } x_i < 10, w_i \in [10, +\infty) \text{ se } x_i = 10\}$. (Como curiosidade, deve-se registrar que a Inferência freqüentista tende a tratar as duas censuras acima de modo distinto).

Enunciaremos agora o Princípio da Censura apresentado em Dawid (1977).

PRINCÍPIO DA CENSURA - Seja $E=(\Omega, \Theta, \{P_\theta\})$ um experimento estatístico não-censurado e g uma censura do tipo $g:\Omega \rightarrow A \cup \{\varphi\}$, onde $A \subset \Omega$ e $g(x) = x$ se $x \in A$. Então é necessário que:

$$i) E \in \Xi \Rightarrow E^g \in \Xi;$$

$$ii) x \in A \Rightarrow MI(E^g, g(x)) = MI(E, x) .$$

Ou seja, o Princípio da Censura diz que a classe Ξ deve ser fechada sob censuras e que, no subconjunto de Ω onde não há censura, o mecanismo de censura deve ser ignorado na inferência.

O Princípio da Censura é um caso particular do seguinte princípio:

PRINCÍPIO DAS CENSURAS EQUIVALENTES - (Berger e Wolpert (1988), p. 94) - Seja $E = (\Omega, \Theta, \{P_\theta\})$ um experimento não-censurado e seja $A \subset \Omega$. Se E^{g_1} e E^{g_2} são experimentos associados a E definidos por censuras fixas g_1 e g_2 equivalentes sobre A , então é necessário que:

$$i) E \in \Xi \Rightarrow E^{g_i} \in \Xi, i=1,2.$$

$$ii) MI(E^{g_1}, g_1(x)) = MI(E^{g_2}, g_2(x)) \text{ sempre que } x \in A.$$

De fato, se tomarmos :

$$g_1:\Omega \rightarrow A \cup \{\varphi\}$$

$$x \rightarrow g(x) = \begin{cases} x & \text{se } x \in A \\ \varphi & \text{se } x \notin A \end{cases}$$

e $g_2 =$ função identidade em Ω , poderemos verificar facilmente que g_1 e g_2 são equivalentes sobre A , pois se $x \in A, g_1(x)=x, g_2(x)=x$ e $g_1^{-1}(g_1(x))=g_2^{-1}(g_2(x))=\{x\}$. Neste caso, o

Princípio das Censuras Equivalentes diz que $MI(E^{g_1}, g_1(x)) = MI(E, x)$ sempre que $x \in A$ (uma vez que $E^{g_2} = E$) e este é justamente o Princípio da Censura anterior (caso particular do Princípio das Censuras Equivalentes).

Resta finalmente provarmos que o Princípio acima decorre do Princípio da Verossimilhança.

COROLÁRIO 3 - O Princípio das Censuras Equivalentes decorre do Princípio da Verossimilhança.

DEMONSTRAÇÃO:

Como g_1 e g_2 são equivalentes sobre $A \subset \Omega$, temos, para todo $x \in A$ e para todo $\theta \in \Theta$,

$$\begin{aligned} l_{g_1(x)}(\theta) &= P_\theta[g_1(X) = g_1(x)] = P_\theta[X \in g_1^{-1}(g_1(x))] = \text{pela equivalência} = \\ &= P_\theta[X \in g_2^{-1}(g_2(x))] = l_{g_2(x)}(\theta) \end{aligned}$$

Portanto, sempre que $x \in A$, o Princípio da Verossimilhança garante que

$$MI(E^{g_1}, g_1(x)) = MI(E^{g_2}, g_2(x)), \forall x \in A.$$

□

Observação: Mais uma vez, foi crucial para o argumento a comparabilidade entre os experimentos censurados (e entre cada experimento censurado e E) com relação a MI, isto é, é necessário impor que $E^{g_i} \in \Xi$, para $i=1,2$.

3.2.5. ALEATORIZAÇÃO

Seja $\mathcal{P} = \{1, 2, \dots, N\}$ uma população finita de objetos distinguíveis, as unidades amostrais, com tamanho N conhecido. Para cada unidade i está associado um número real (ou um vetor) X_i ($i=1, \dots, N$). A quantidade $\mathbf{X} = (X_1, X_2, \dots, X_N) \in \Omega^N$ é desconhecida (o espaço amostral Ω_i é conhecido, e X_i é desconhecida até que a unidade i seja inspecionada).

É de interesse inferir algo a respeito da quantidade θ utilizando a informação provida por uma amostra. O parâmetro θ é uma função $\theta = \theta(X_1, X_2, \dots, X_N)$. (Ver também, Zuazola et al. (1994)).

A amostragem $\mathcal{A}$ produz um subconjunto $\mathbf{i} = (i_1, i_2, \dots, i_n)$ de $\mathcal{P}$ (onde $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq N$). As unidades de $\mathbf{i}$ são inspecionadas e seus valores x_i são revelados. A observação amostral é representada da seguinte forma:

$$\mathbf{z} = (\mathbf{i}, \mathbf{y}) = \{ \mathbf{i} = (i_1, \dots, i_n), \mathbf{y} = (y_1, \dots, y_n) = (x_{i_1}, \dots, x_{i_n}) \text{ onde } i_j \in \mathcal{P} \text{ para } j = 1, \dots, n \}$$

Dado um plano amostral $\mathcal{A}$, para cada possível $\mathbf{i}$, existe um número $p(\mathbf{i})$, a probabilidade de seleção do subconjunto $\mathbf{i}$ de $\mathcal{P}$. Usualmente, $p(\mathbf{i})$ não depende de θ . Ela pode depender de θ somente através dos valores previamente observados (isto é, de $\mathbf{y}$), quando o plano amostral é seqüencial (Ver, Basu(1969)).

EXEMPLO: (Basu (1969))

Considere o seguinte plano amostral seqüencial: Escolha a unidade 1 e observe X_1 . Se $X_1 > K$ então escolha a unidade 2, caso contrário escolha a unidade N . Para este plano $\mathbf{i} = (1, 2)$ ou $\mathbf{i} = (1, N)$ e $p(\mathbf{i})$ depende de $\theta = (X_1, \dots, X_N)$ através de x_1 .

Planos amostrais do tipo acima descrito apresentam $p(i)$ dependente de θ somente através de y .

Após a observação de z , podemos definir Θ_z o conjunto dos pontos paramétricos que são consistentes com a amostra z , isto é:

$$\Theta_z = \{\theta: \theta \in \Theta \text{ e } X_{i_k} = y_k \text{ para } k = 1, 2, \dots, n\}$$

Seja $P_\theta(z)$ a probabilidade de que a amostra gerada seja $z = (i, y)$. Temos que: $P_\theta(z) = P_\theta(i, y) = P_\theta(i) \cdot P_\theta(y|i) = p(i) \cdot P_\theta(y|i) = p(i) \cdot I_{(\theta \in \Theta_z)}$. Assim, a função de verossimilhança gerada por z é: $l_z(\theta) = P_\theta(z) = p(i) \cdot I_{(\theta \in \Theta_z)} \propto I_{(\theta \in \Theta_z)}$, isto é, a verossimilhança é plana no conjunto Θ_z e, mais do que isso, independe do mecanismo de aleatorização (aqui a aleatorização é utilizada para a seleção da amostra).

Embora o contexto descrito acima seja de amostragem, de forma análoga, verifica-se que também na situação de planejamento de experimentos (alocação aleatória de tratamentos) a função de verossimilhança gerada pelos dados independe do mecanismo de aleatorização.

A consideração do mecanismo de aleatorização na análise estatística, após a obtenção dos dados, viola o Princípio da Verossimilhança. Tal situação é ilustrada através do exemplo 5 da seção 3.3.

3.3. EXEMPLOS

Apresentamos a seguir alguns exemplos onde se verifica a violação do Princípio da Verossimilhança. Os exemplos 1, 2 e 3 são versões dos exemplos originalmente descritos por Berry (1987 - a) (ver também Inoue e Wechsler (1994)). O exemplo 4 é inspirado no exemplo de Pratt descrito na discussão do artigo de Birnbaum (1962). Finalmente, o exemplo 5 é essencialmente devido a Basu (1980).

EXEMPLO 1:

Suponha que os pacientes envolvidos em um ensaio clínico recebam (cada um deles) as drogas A e B, e que um pesquisador anota a droga preferível para cada paciente (suponha adicionalmente que "empates" na preferência sejam impossíveis).

Seja θ a proporção (na população de pacientes) de preferências pela droga A e consideremos as observações condicionalmente independentes com distribuição comum Bernoulli de parâmetro θ , dado θ .

Seja $\mathbf{x} = (B, B, B, A, B, B, A, B, B, B, B, A)$, por exemplo, o vetor de observações. Consideremos as seguintes "regras de parada" que poderiam definir o experimento:

E_1 : O pesquisador havia decidido tomar exatamente 12 observações

E_2 : O pesquisador havia decidido prosseguir a amostragem até obter 3 preferências pela droga A.

As funções de verossimilhança para θ geradas em E_1 e em E_2 são:

$$l_{\mathbf{x}}^1(\theta) = \binom{12}{3} \theta^3 (1-\theta)^9 \text{ e,}$$

$$l_{\mathbf{x}}^2(\theta) = \binom{11}{2} \theta^3 (1-\theta)^9, \text{ respectivamente.}$$

As funções de verossimilhanças são proporcionais e a informação sobre θ é a mesma tanto em E_1 quanto em E_2 .

Os estimadores não viciados para θ , respectivamente em E_1 e E_2 , são:

$$\bar{\theta}^{(1)} = \bar{X}_n = \sum_{i=1}^n X_i / n \quad \text{e,}$$

$$\bar{\theta}^{(2)} = (k-1)/(N-1) \quad \text{onde } k: \text{ número de preferências pela droga A,}$$

N: número de observações necessárias até que se obtenha k preferências pela droga A.

Obtemos $\bar{\theta}^{(1)} = 3/12=0.250$ e, $\bar{\theta}^{(2)} = 2/11=0.182$, isto é, estimativas (não-viciadas) de θ diferentes. Elas evidenciam a dependência sobre a regra de parada do experimento e a violação do Princípio da Verossimilhança.

Suponhamos agora que desejamos testar: $H_0: \theta=1/2$ "versus" $H_1: \theta < 1/2$. Calculemos o nível descritivo p (nível de significância observado ou "p-value") que é a probabilidade, sob a hipótese nula, de ocorrência de resultado igual ou mais extremo que o observado.

Em E_1 , como o pesquisador havia decidido tomar exatamente 12 observações devemos considerar a estatística suficiente:

Y : variável aleatória que representa o número de preferências pela droga A.

Sob a hipótese nula, $Y \sim \text{Binomial}(n, 1/2)$ e então $p^{(1)} = P_{H_0}(Y \leq 3) \cong 0.073$.

Na situação E_2 , como o pesquisador havia decidido conduzir o experimento até obter 3 preferências pela droga A devemos considerar a estatística suficiente:

Z : variável aleatória que representa o número de ensaios necessários até a obtenção de 3 preferências pela droga A.

Sob a hipótese nula, $Z \sim \text{Binomial Negativa}(3; 1/2)$ e $p^{(2)} = P_{H_0}(Y \geq 12) \cong 0.033$.

Fixando-se um nível de significância de 5% em E_1 não rejeitaríamos a hipótese nula, mas em E_2 sim. Isto é, estaríamos chegando a diferentes conclusões para um mesmo conjunto de dados de um experimento com planejamentos distintos.

Observemos também a consideração de dados não obtidos no cálculo do nível descritivo decorrente simplesmente de sua própria definição (ver Pereira e Wechsler (1993)). Além disso, a especificação de resultados mais extremos depende da regra de parada.

A violação do Princípio da Verossimilhança fica evidente tanto na estimação não-viciada quanto no teste de hipóteses: um mesmo conjunto de dados leva a diferentes estimativas/conclusões dependendo da regra de parada.

A análise freqüentista, portanto, não é consistente com o Princípio da Verossimilhança, pois ao considerar as regras de parada após a observação dos dados, pode chegar a diferentes conclusões. Além disso, os critérios de otimalidade envolvem quase sempre integrações sobre o espaço amostral, isto é, consideram pontos amostrais não realizados.

Muitos experimentos são realizados sem preconceberem uma regra de parada. A decisão de parar pode ocorrer por vários motivos tais como o esgotamento de recursos financeiros para a pesquisa ou de tempo hábil (com o pesquisador prestes a entregar a tese!), ou até mesmo por cansaço. Nessas situações nenhum procedimento de inferência estatística clássica pode, teoricamente, ser considerado.

Avaliemos esse exemplo sob a ótica de um bayesiano "objetivista". Adotando-se a priori de Jeffreys obtém-se, $\pi_1(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$ e $\pi_2(\theta) \propto \theta^{-1/2}(1-\theta)^{-1}$, respectivamente prioris para θ no experimento E_1 e em E_2 . Observemos que as prioris são diferentes não porque revelam opiniões diferentes a respeito do parâmetro θ , mas porque os experimentos são distintos (possuem regras de parada distintas).

Como as funções de verossimilhança geradas em E_1 e em E_2 são, respectivamente, iguais a: $l_x^1(\theta) = \binom{12}{3} \theta^3(1-\theta)^9$ e, $l_x^2(\theta) = \binom{11}{2} \theta^3(1-\theta)^9$, obtemos: $\pi_1(\theta|x) \propto \theta^{5/2}(1-\theta)^{17/2}$ e, $\pi_2(\theta|x) \propto \theta^{5/2}(1-\theta)^8$, isto é, a distribuição a posteriori para θ em E_1 é Beta(7/2, 19/2) e em E_2 é Beta(7/2, 9). As distribuições a posteriori para θ são diferentes revelando a violação do Princípio da Verossimilhança (veja seção 4.1. também).

Observemos que, na possibilidade de dois experimentos (E_1 e E_2), uma "mistura" de duas prioris está sendo considerada. Como apontado por Hill(1974), a utilização de prioris diferentes (segundo o plano amostral adotado) revela a situação (absurda!) em que o "lançamento de uma moeda" dita a priori que deve ser usada.

Vale lembrar que um verdadeiro bayesiano (subjetivista) já viola o Princípio da Verossimilhança, pois a sua priori é uma opinião (de fato) e independe de aspectos do planejamento do experimento - ela é anterior e independe do planejamento! O bayesiano subjetivista só tem, portanto, uma opinião a priori para θ mesmo diante de dois (ou mais) protocolos experimentais passíveis de execução.

EXEMPLO 2:

Um cientista consulta um estatístico da escola freqüentista e traz 100 observações que são supostas condicionalmente, dado θ , independentes e com distribuição comum $\text{Normal}(\theta, 1)$. Ele deseja testar as hipóteses: $H_0: \theta = 0$ versus $H_1: \theta \neq 0$. Como a média das observações é $\bar{x}_{100} = 0.2$ temos que o valor da estatística do teste é: $Z = \sqrt{100} (\bar{x}_{100} - \theta) =$ (sob H_0) $= 10 \times 0.2 = 2$. O nível descritivo do teste bilateral é $p = 2P_{H_0}(\bar{X}_{100} \geq 0.2) = 2P_{H_0}(Z \geq 2) = 0.0456$.

Com um nível de significância de 5%, há "forte" evidência contra a hipótese nula e o estatístico pode decidir pela significância estatística dos dados. No entanto, o estatístico cautelosamente pergunta ao cientista: "Por que você parou após ter obtido 100 observações?". Se o cientista declara que resolveu de antemão tomar 100 observações e parar o ensaio, qualquer que fosse a decisão, então não há problemas com relação à decisão de significância estatística.

Contudo, se o cientista declara que caso as 100 observações não apresentassem um resultado significativo tomaria um outro lote de 100 observações, temos então um caso em que a regra de parada não é aquela considerada na obtenção do nível descritivo acima.

A rigor, o procedimento adotado pelo cientista seria o representado abaixo:

- Se $\sqrt{100}|\bar{x}_{100}| \geq k$, rejeita-se H_0 e encerra-se o experimento;
- Se $\sqrt{100}|\bar{x}_{100}| < k$, toma-se outras 100 observações. Se $\sqrt{200}|\bar{x}_{200}| \geq k$, rejeita-se H_0 . Caso contrário, encerra-se o experimento e aceita-se H_0 .

Para um nível $\alpha=0.05$, k deve ser igual a 2.18 (Pocock (1977)) e não 1.96! No caso temos: $\sqrt{100}|\bar{x}_{100}| = 2 < 2.18$.

Dessa forma, não podemos encerrar o experimento, contrariando a conclusão que seria tomada na situação que foi anteriormente exposta. O cientista deverá então tomar outras 100 observações.

Vamos supor que o cientista conduza a segunda amostragem obtendo as 100 novas observações. Consideremos as seguintes possibilidades:

$$\text{I)} \sqrt{200} |\bar{x}_{200}| = 2.20 > 2.18.$$

Aqui temos um resultado estatisticamente significativo. Nesse caso, o estatístico pode perguntar: "O que seria feito se os resultados não revelassem significância?". Vamos supor então que o cientista declare que, se houver renovação de sua bolsa de apoio à pesquisa, tomaria 100 novas observações. Caso contrário, sem dinheiro, encerraria o experimento. Nessas condições, o estatístico dirá que se houver renovação da bolsa, não é possível tirar conclusões até a obtenção das próximas observações. Se a bolsa não for renovada, então "proclama-se" a significância contra a hipótese nula ao nível $\alpha = 0.05$.

O que é interessante notar nesta situação é a forte dependência, na inferência clássica, da regra de parada, e que as conclusões sobre θ podem depender de fatores divorciados de θ , tais como uma renovação de bolsa de estudo.

$$\text{II)} \sqrt{200} |\bar{x}_{200}| = 2.10 < 2.18.$$

Então os resultados não mostram significância. Observe-se, entretanto, que se o cientista não tivesse feito a pausa no experimento, teríamos $\sqrt{200} |\bar{x}_{200}| = 2.10 > 1.96$, ou seja, os dados revelariam forte evidência contra a hipótese nula para um nível $\alpha = 0.05$.

Percebemos então que a conclusão do experimento na inferência clássica não depende apenas dos dados obtidos, mas também das intenções do cientista a respeito de continuar ou

não a amostragem a cada etapa. Um fato que torna ainda mais "dramático" este problema é que a interrupção da amostragem pode ocorrer involuntariamente: A situação no exemplo é equivalente à possibilidade de, podendo o cientista obter 200 observações de uma única vez, a sua confissão de ter examinado as primeiras 100 observações tornou $\bar{x}_{200}=2.10$ não-significante. No entanto, se o cientista houvesse se esquecido de examinar o valor de $\bar{x}_{100}$, o mesmo valor de $\bar{x}_{200}=2.10$ passaria a ser significativo! Ora, os dados são rigorosamente os mesmos, mas as conclusões são opostas! Novamente, temos a violação do Princípio da Verossimilhança. Alterações no protocolo do experimento ocorrem em situações práticas e têm importante impacto na análise clássica.

Por outro lado, na análise bayesiana a distribuição dos parâmetros desconhecidos é atualizada a todo momento. Assim, nenhuma incoerência surge quando lidamos com experimentos realizados por etapas. A continuação ou não do experimento pode ser decidida em qualquer momento com base nos custos e benefícios que ponderam a função de perda pela distribuição do parâmetro desconhecido.

EXEMPLO 3:

Consideremos um experimento que é conduzido com a finalidade de comparar "drogas" para o tratamento de leucemia aguda: A: droga 6-mercaptopurina e B: placebo (controle). Suponhamos que os pacientes sejam tratados em pares homogêneos. A escolha de droga ou placebo é aleatorizado para cada paciente que compõe o par.

O ensaio modela as observações como sendo de Bernoulli indicando qual droga é mais efetiva para cada par. Supõe-se que empates na comparação sejam impossíveis.

Temos então: $X_i = \begin{cases} 1 & \text{... Droga A revela - se melhor} \\ 0 & \text{... "Droga" B (placebo) melhor} \end{cases}$, no i -ésimo par ($i=1,2,\dots,n$).

Consideremos θ = chance de que a droga A tenha melhor desempenho do que a "droga" B num dado par.

Assim, $X_i \sim$ condicionalmente independentes e identicamente distribuídas Bernoulli (θ), dado θ .

Estamos interessados em testar: $H_0: \theta=1/2$ "versus" $H_1: \theta \neq 1/2$.

Considera-se um planejamento que é seqüencial, pois os dados são acumulados e analisados continuamente até que o número de preferências por uma das drogas menos as preferências pela outra droga se torne significativo ou, até que fique evidenciado que não há base para rejeição da hipótese nula (H_0). Deste modo, o número de pares utilizado é aleatório.

Freireich et al. (1963) usaram o projeto seqüencial sugerido por Armitage(1961): Planejamento seqüencial restrito. Os resultados são apresentados na tabela 1 e ilustrados na figura 1.

FIGURA 1: Projeto seqüencial de um ensaio clínico comparativo em que a amostragem pára quando o excesso de preferências pelo tratamento A ou tratamento B torna-se suficientemente grande (página 120 de Berry (1987 - a)).

A figura 1 ilustra os limites para decisão. O ensaio pára e H_0 é rejeitada quando o experimento alcança ou o limite inferior ou o limite superior. Se a trajetória do experimento se estende ao limite mais à direita, o ensaio atinge a região não conclusiva e não é estabelecida preferência entre as drogas. A característica do plano é que a probabilidade de atingir o limite superior ou o limite inferior quando $\theta=1/2$ é 0.025. Deste modo, o nível de significância é $2 \times 0.025 = 0.05$ ($\alpha = 0.05$).

Pela tabela 1 verificamos que o ensaio é encerrado após 18 pares (i.e. quando $n_A = 15$ e $n_B = 3$).

Na coluna referente a "p-value" bilateral temos os níveis descritivos de significância ("p-value") em cada ponto da coleção de dados, quando a distribuição binomial com $\theta = 1/2$ e $n = 1, 2, \dots, 21$ é usada.

Seja $p = P(H_0)$ e $f(\theta|H_1)$ a densidade a priori condicional (dada H_1) de θ . Na tabela 1 temos em (1) as probabilidades $P(H_0|n_A, n_B)$ calculadas para $p = 1/2$ e $f(\theta|H_1)$ uniforme em $(0, 1)$ ⁽⁵⁾. Em (2) temos que $P(\theta < 1/2 | n_A, n_B, H_1) = P(\theta < 1/2 | n_A, n_B)$ (pois $p = 0$ e $f(\theta|H_1)$ uniforme em $(0, 1)$) ⁽⁶⁾.

Nesse experimento, tivemos 3 pares adicionais após a parada do experimento. Na análise clássica, esses pares não podem ser usados numa inferência formal, isto é, eles precisam ser ignorados.

$$(5) \quad P(\theta = 1/2 | n_A, n_B) = \frac{p(1/2)^{n_A + n_B}}{p(1/2)^{n_A + n_B} + (1-p) \frac{\Gamma(n_A + 1)\Gamma(n_B + 1)}{\Gamma(n_A + n_B + 2)}}$$

$$(6) \quad P(\theta < 1/2 | n_A, n_B) = \frac{\Gamma(n_A + n_B + 2)}{\Gamma(n_A + 1)\Gamma(n_B + 1)} \int_0^{1/2} \theta^{n_A} (1 - \theta)^{n_B} d\theta$$

TABELA 1: Análise de um ensaio seqüencial envolvendo pacientes com leucemia aguda.
(página 120 de Berry (1987 - a))

par	tratamento preferível	$n_A - n_B$	nível de significância bilateral	"p-value" bilateral	Probabilidades Correntes	
					$\theta=1/2$ (1)	"B" é melhor (2)
1	A	1	-	1.0000	0.5000	0.2500
2	B	0	-	1.0000	0.6000	0.5000
3	A	1	-	1.0000	0.6000	0.3100
4	A	2	-	0.6300	0.5600	0.1900
5	A	3	-	0.3800	0.4800	0.1100
6	B	2	-	0.6900	0.6200	0.2300
7	A	3	-	0.4500	0.5800	0.1400
8	A	4	-	0.2900	0.5000	0.0900
9	A	5	-	0.1800	0.4100	0.0550
10	A	6	-	0.1100	0.3300	0.0330
11	A	7	-	0.0650	0.2400	0.0190
12	A	8	-	0.0390	0.1700	0.0110
13	A	9	-	0.0220	0.1200	0.0065
14	B	8	-	0.0570	0.2500	0.0180
15	A	9	-	0.0350	0.1800	0.0110
16	A	10	-	0.0210	0.1300	0.0064
17	A	11	-	0.0130	0.0850	0.0038
18	A	12	0.05	0.0075	0.0560	0.0022
19	A	13	-	0.0044	0.0340	0.0013
20	A	14	-	0.0026	0.0220	0.0008
21	A	15	-	0.0015	0.0140	0.0005

(1): Priori uniforme para $\theta \neq 1/2$, $p=1/2$.

(2): Priori uniforme para θ e $p=0$.

n_A : número de preferências pela droga A.

n_B : número de preferências por placebo.

O efeito de "olhadelas" repetidas durante o experimento é bem pronunciado, pois o planejamento é alterado. Consideremos $n=18$ pares: O nível descritivo de significância para a regra de parada binomial é 0.0075, que lança muito mais suspeita contra a hipótese nula do que aquele nível no plano seqüencial, de valor 0.05.

Observa-se, então, a violação do Princípio da Verossimilhança: Para um mesmo conjunto de dados a decisão de rejeitar ou não a hipótese nula pode ser diferente e depender do planejamento. Isto é mais evidente, por exemplo, quando tomamos o 12º par. Nesse caso, o nível descritivo de significância bilateral se o tamanho de amostra $n=12$ fosse fixado de antemão é 0.039 e rejeitaríamos H_0 ao nível $\alpha = 0.05$, (ao parar o ensaio). Pelo projeto seqüencial não poderíamos sequer encerrar o ensaio.

Uma análise bayesiana deste ensaio clínico envolve cálculos de probabilidades posteriores e a parada ocorre quando a evidência fornecida pelos dados parece ser conclusiva, sem a interferência da regra de parada.

Tomando-se para "nível de conclusão" probabilidades correntes menores ou iguais a 0.05, poderemos constatar na tabela que uma decisão é atingida aós a 10ª observação, quando a probabilidade de que B seja melhor é 0.033. Decidimos então que a droga A é melhor e atingimos uma decisão antes do procedimento clássico seqüencial que pára somente após a observação do 18º par.

Um dos principais aspectos que devemos considerar neste exemplo é a questão ética envolvida. Um médico que acredita que um tratamento seja mais efetivo que o outro não pode submeter pacientes a tratamentos inferiores. Têm-se a imposição ética de usar os tratamentos superiores tão logo esta superioridade seja revelada.

Dessa forma, é natural que em um ensaio seqüencial deva-se interromper o experimento assim que alguma diferença significativa seja detectada.

Em uma análise clássica formal não temos a possibilidade de mudar o planejamento interrompendo o ensaio, não obstante as evidências fornecidas pelo conjunto de dados. Sob este aspecto, a análise bayesiana é bastante flexível e conseqüentemente mais ética.

Apresentamos até este ponto resultados da análise bayesiana envolvendo uma distribuição uniforme(0,1) a priori para θ . Apresentamos na tabela 2 as probabilidades correntes de que a droga B seja melhor ($P(\theta < 1/2 | \text{dados})$) adotando outras densidades a priori.

Através da tabela 2, verificamos que o procedimento bayesiano é vantajoso mesmo quando a opinião inicial do médico não é uniforme em (0,1). Em geral, consegue-se obter resultados conclusivos (com $P(\theta < 1/2 | \text{dados}) \leq 0.05$) e, por isso, encerrar-se o ensaio antes do projeto seqüencial clássico, com outras opiniões a priori.

Mesmo com opiniões a priori que apresentam valor esperado inferior a 1/2, sendo portanto "desfavoráveis" ao tratamento com a droga A, consegue-se, em geral, obter resultados conclusivos antes do projeto seqüencial clássico.

Há um ponto relevante que deve ser esclarecido sobre a análise bayesiana: verificamos que adotando prioris distintas podemos chegar a resultados conclusivos diferentes em "instantes" diferentes. Por exemplo: se adotamos uma priori Beta(3, 1) obtemos $P(\theta < 1/2 | \text{dados}) \leq 0.05$ com o 5º par, enquanto que se tivermos uma priori Beta(4,2), o resultado conclusivo ocorre com o 8º par.

No entanto, isto não representa uma violação do Princípio da Verossimilhança! A análise bayesiana não dependeu de aspectos do planejamento do experimento, mas da opinião a priori do médico a respeito da eficácia dos tratamentos. Os dados atualizam essa informação, confirmando ou não, a opinião inicial do médico. Em suma, a análise bayesiana jamais viola o Princípio da Verossimilhança.

TABELA 2: Probabilidades correntes de que a "droga B" seja melhor para outras prioris

par	tratamento preferível	$n_A - n_B$	Beta(3,1)	Beta(4,2)	Probabilidades Correntes		
					Beta(2,2)	Beta(1,3)	Beta(2,3)
1	A	1	0.0625	0.1094	0.3125	0.6875	0.5000
2	B	0	0.1875	0.2266	0.5000	0.8125	0.6563
3	A	1	0.1094	0.1445	0.3437	0.6563	0.5000
4	A	2	0.0625	0.0898	0.2266	0.5000	0.3632
5	A	3	0.0352	0.0547	0.1445	0.3632	0.2539
6	B	2	0.0898	0.1132	0.2539	0.5000	0.3769
7	A	3	0.0547	0.0730	0.1718	0.3769	0.2743
8	A	4	0.0327	0.0462	0.1132	0.2743	0.1921
9	A	5	0.0193	0.0295	0.0730	0.1921	0.1321
10	A	6	0.0113	0.0185	0.0462	0.1321	0.0893
11	A	7	0.0065	0.0116	0.0295	0.0893	0.0596
12	A	8	0.0042	0.0072	0.0185	0.0596	0.0394
13	A	9	0.0025	0.0044	0.0116	0.0394	0.0258
14	B	8	0.0072	0.0109	0.0258	0.0719	0.0492
15	A	9	0.0044	0.0070	0.0168	0.0492	0.0334
16	A	10	0.0027	0.0044	0.0109	0.0034	0.0225
17	A	11	0.0017	0.0028	0.0070	0.0225	0.0151
18	A	12	0.0010	0.0018	0.0044	0.0151	0.0100
19	A	13	0.0006	0.0011	0.0028	0.0100	0.0066
20	A	14	0.0004	0.0007	0.0018	0.0066	0.0043
21	A	15	0.0002	0.0004	0.0011	0.0043	0.0028

EXEMPLO 4:

Um fabricante afirmou que os termômetros produzidos pela sua empresa geram observações com distribuição normal de variabilidade igual a $0,01^\circ\text{C}$ e valor esperado θ , a temperatura do paciente. Um indivíduo hipocondríaco dispõe de um desses termômetros com o qual avalia a sua temperatura. Leu no termômetro que estava com $36,5^\circ\text{C}$ e informou este resultado ao vizinho, um estatístico clássico. Este, por sua vez, calculou um intervalo de confiança (com nível de confiança igual a 95%) para a sua temperatura (θ) obtendo: $IC(\theta, 95\%) = [36,3^\circ\text{C} ; 36,7^\circ\text{C}]$. O estatístico tranquilizou o hipocondríaco dizendo que o intervalo não continha valores iguais ou superiores a 37°C (limite de detecção de estado febril). O hipocondríaco voltou satisfeito para casa.

Após receber o vizinho, o estatístico foi ler um jornal onde encontrou um informe do fabricante do termômetro. O fabricante avisava que alguns termômetros que ele tinha lançado no mercado apresentavam defeito de fabricação e que, por isto, eram incapazes de registrar temperaturas superiores a 37°C . O estatístico foi, então, à casa do hipocondríaco informar que a observação no termômetro é censurada no valor 37 e que a distribuição da leitura da temperatura é uma normal truncada (com a mesma densidade para temperaturas menores do que 37 e com probabilidade igual a

$$P(37|\theta, \sigma = 0,1) = \int_{37}^{\infty} \left\{ \frac{1}{(2\pi)^{1/2} \cdot (0,1)} \exp\left\{ \frac{-1}{0,02} (x - \theta)^2 \right\} \right\} dx \text{ quando a temperatura registrada é}$$

igual a 37°C) o que iria conduzir a um resultado diferente, pois o IC seria outro.

O hipocondríaco ficou achando tudo muito esquisito. O que tinha a ver uma nota de jornal com a avaliação da temperatura? O termômetro tinha registrado temperatura menor do que 37°C , então por que a condição do aparelho iria interferir na análise? Ele acabou chegando à conclusão de que ele era normal, e o estatístico, psicopata !

Este exemplo mostra como é intuitivamente óbvio o Princípio da Censura. A informação a respeito de θ obtida de uma observação não truncada (no exemplo o valor 36,5) de um experimento em que a censura é possível deve ser a mesma daquela que é obtida com uma mesma observação em uma versão não censurada do experimento.

A explicação de que a condição do aparelho deve ser considerada porque esta interfere na observação da temperatura é inadequada já que a observação 36,5 não sofreu prejuízo (isto é, não foi censurada). A insistente preocupação com a possibilidade de censura, revela a preocupação com dados não observados. Isto revela a violação do Princípio da Verossimilhança!

EXEMPLO 5:

Um cientista deseja testar se uma nova droga (tratamento) é superior a uma droga convencional (controle) quanto ao ganho de peso. Para isto, dispõe de 15 pares de indivíduos homogêneos (mesmo sexo, mesma faixa etária, etc.). Em cada par, o cientista seleciona um indivíduo para o tratamento e outro para o controle. Sejam t_i e c_i as respostas dos sujeitos tratamento e controle, respectivamente, do i -ésimo par. Suponhamos que $(t_1, c_1), \dots, (t_{15}, c_{15})$ sejam condicionalmente, dados $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, independentes com distribuição Normal bi-variada com função densidade de probabilidade comum e igual a:

$$f(t, c | \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) =$$

$$= \frac{1}{2\pi\sqrt{\sigma_1\sigma_2(1-\rho^2)}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{t-\mu_1}{\sqrt{\sigma_1}} \right)^2 + \left(\frac{c-\mu_2}{\sqrt{\sigma_2}} \right)^2 - 2\rho \left(\frac{t-\mu_1}{\sqrt{\sigma_1}} \right) \left(\frac{c-\mu_2}{\sqrt{\sigma_2}} \right) \right] \right\}$$

O cientista observa que: i) $T = \sum t_i - \sum c_i \gg 0$ e ii) $d_i = t_i - c_i > 0$, para todo $i = 1, \dots, 15$ e conclui que há forte evidência a favor da nova droga. Decide então procurar um estatístico para quantificar essa evidência.

O estatístico decide por um teste unilateral da hipótese nula (as drogas têm mesmo efeito sobre o ganho de peso) com a estatística de teste: $T = \sum d_i$. A diferença d_i entre as respostas dos indivíduos do i -ésimo par pode ser explicada em termos de uma possível diferença entre tratamentos e outros fatores "nuisance" como diferenças entre os indivíduos (não totalmente controladas pelo cientista), infecção, perda de apetite, etc., que podem agir de modo diferenciado entre os indivíduos do par. Se a hipótese nula é verdadeira, então não há diferença entre tratamentos; assim a resposta diferença d_i pode ser causada por esses outros fatores. A aleatorização permite ao estatístico eliminar esses fatores. Vamos verificar como isto ocorre.

Suponhamos que o cientista tivesse utilizado uma moeda honesta para aleatorizar a droga tratamento a cada par (isto é, os indivíduos de cada par têm igual probabilidade de receber a droga experimental) e registrado a diferença d_i . O cientista pode especular a respeito de uma hipotética execução do experimento em que todos os fatores experimentais (controlados e não controlados) são supostamente mantidos fixos e iguais ao que ele obteve no experimento original. Só é permitido que a alocação dos tratamentos seja aleatória. Se a hipótese nula é verdadeira, então a diferença de respostas, d_i , para o i -ésimo par deve ter sido causada por fatores "nuisance". Na execução hipotética desse experimento, eles são mantidos fixos. Portanto, no novo experimento as diferenças de respostas para o i -ésimo par podem levar somente a valores d_i e $-d_i$ (para todo i), dependendo da alocação do i -ésimo par ser a mesma do experimento prévio (original) ou diferente. Se denotamos as diferenças de respostas por $(d_1^*, \dots, d_{15}^*)$, então é claro que sob a hipótese nula, o espaço amostral (para as

diferenças de respostas) é um conjunto de 2^{15} vetores do tipo $(\pm d_1, \dots, \pm d_{15})$

equiprováveis. Seja $T^* = \sum_{i=1}^{15} d_i^*$. O nível descritivo dos dados é: $p = P(T^* \geq T \mid H_0)$.

Observemos que $d_i > 0$ para todo i . Portanto, $T^* \geq T$ se, e somente se, $d_i^* = d_i$ para todo i . Segue que: $p = P(T^* \geq T \mid H_0) = 2^{-15}$. Esta é a análise clássica em que o mecanismo de aleatorização envolvido na alocação de tratamentos às unidades experimentais é utilizado no cálculo de medidas de evidência.

O argumento utilizado no teste condicional da permutação (não-paramétrico) independe de qualquer suposição probabilística sobre uma população de indivíduos. Além disso, os indivíduos que participam do estudo não são considerados amostra aleatória dessa população.

Todas essas "elucubrações" foram feitas pelo estatístico que apresentou, então, a análise para o cientista. O cientista confessa estar bastante intrigado com a análise. Ele não entende, por exemplo, a relevância das probabilidades da aleatorização nessa análise. O que teria acontecido se a moeda utilizada não fosse honesta? O estatístico explica que o cenário do experimento hipotético continuaria o mesmo. A única alteração é de que a moeda é viciada tal que com probabilidade igual a, digamos, $1/4$, a droga experimental é atribuída a um indivíduo do par. Então $P(d_i^* = d_i) = 1/4$ ou $3/4$ dependendo da alocação ser (t_i, c_i) ou (c_i, t_i) de modo que, $p = P(T^* \geq T \mid H_0) = \left(\frac{1}{4}\right)^m \cdot \left(\frac{3}{4}\right)^{(15-m)}$ onde m é o número de alocações do tipo (t, c) no experimento original. O cientista acha isto um absurdo. Como o nível descritivo podia depender tanto da alocação dos tratamentos? O modo de dispor as "etiquetas" nos pares não parece relevante para o cientista.

Sentindo um desconforto imenso, já que o estatístico insiste na grandiosidade de se ter um experimento aleatorizado, o cientista confessa que, de fato, não tinha aleatorizado a alocação dos tratamentos no par. Angustiado, ele explica que, como cientista, ele foi treinado a manter sob controle e balancear todos os fatores "nuisance" tanto quanto isto fosse possível. Após a formação dos pares, ele tinha observado diferenças em relação a peso, altura, etc. Ele tentou balancear essas diferenças de modo a formar dois grupos com peso médio, altura média, etc. similares e a alocação então foi atribuída ao grupo (e não ao par) através de um processo similar ao de lançar uma moeda honesta. Ele queria saber o que acontecia com a análise.

O estatístico explica que o espaço amostral para as diferenças $(d_1^*, \dots, d_{15}^*)$, nesse novo contexto, é composto por $(d_1, \dots, d_{15}) \in (-d_1, \dots, -d_{15})$ de modo que: $p = P(T^* \geq T \mid H_0) = 1/2$, e portanto, os dados não revelam mais significância.

Não pretendemos discutir o valor da aleatorização em si mesma, para a alocação dos tratamentos e a escolha de uma amostra, isto é, no planejamento do experimento. De fato, ela pode ser útil. Entretanto, a utilização de métodos de inferência que se baseiam nas probabilidades da aleatorização é questionada, já que os detalhes do mecanismo de aleatorização devem ser irrelevantes para inferências, em observância ao Princípio da Verossimilhança.

Como Basu (1980) observa, a aleatorização injeta um elemento de incerteza a respeito da droga que vai ser atribuída a um indivíduo. Esta incerteza é removida no momento em que se torna conhecida a droga que o indivíduo vai receber. Observemos também que não há probabilidade genuína nesse exemplo, já que a incerteza a respeito do efeito do tratamento jamais é expressa por probabilidades (somente a da aleatorização).

A consideração do mecanismo de aleatorização resulta na violação do Princípio da Verossimilhança, como podemos constatar nesse exemplo.

Suponhamos que a aleatorização do tratamento tenha sido obtida com o lançamento de uma moeda honesta. Observemos que os dados são os mesmos. Estes geram a função de

verossimilhança $l_{(t,c)}^1(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = 2^{-15} \prod_{i=1}^{15} f(t_i, c_i | \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ quando a aleatorização do tratamento ocorre em cada par e,

$l_{(t,c)}^2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = 2^{-1} \prod_{i=1}^{15} f(t_i, c_i | \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ quando a aleatorização ocorre no grupo.

As funções de verossimilhança são (claramente) proporcionais, contudo, os níveis descritivos obtidos para cada situação são diferentes revelando a violação do Princípio da Verossimilhança.

3.4. DISCUSSÃO

A comunidade científica, em sua grande maioria, considera que os ensaios clínicos aleatorizados são éticos e constituem o padrão "ouro" da pesquisa científica. Um argumento comum é que somente no conhecimento pleno dos efeitos da droga o ensaio não deveria ser executado. O fato de que eles sejam aleatorizados é justificado pela eliminação de vícios na atribuição de tratamentos, o que removeria a responsabilidade do clínico.

Quanto ao argumento esboçado acima devemos primeiro observar que não existe uma situação sobre a qual tenhamos um conhecimento completo. Vivemos sob incerteza e o que existe são graus de incerteza (sua única medida coerente é a probabilidade).

A justificativa de que a aleatorização deve ser feita porque ela elimina vícios na atribuição de tratamentos e, portanto, a responsabilidade do clínico gera controvérsias. A aleatorização não garante a representatividade da população e muito menos que uma droga experimental não seja sistematicamente atribuída aos pacientes, digamos, mais severamente doentes.

Os ensaios clínicos aleatorizados talvez fossem éticos quando os pesquisadores não conseguem fazer distinção alguma entre as terapias. Entretanto, se o médico acha que a droga experimental é superior à droga controle, com alta probabilidade, então o procedimento de aleatorizar não é ético. Além disso, um pesquisador pode estar indiferente somente no início do ensaio já que o acúmulo de dados atualiza a sua opinião no que diz respeito à eficácia da droga.

Os procedimentos planejados especificamente para o monitoramento sequencial dos dados são, contudo, infreqüentes na prática devido à dificuldade encontrada pelos estatísticos (clássicos) na análise dos dados. O'Brien (1990) justifica essa posição dizendo que o encerramento do ensaio não permite obter respostas satisfatórias relacionadas com eficácia e segurança da droga. Agora vejamos: Um médico que se recusa a examinar um

paciente é inocente quando este paciente morre? Esta atitude pode ser considerada como negligência. Do mesmo modo, não seria negligência ignorar os dados provisórios e somente avaliá-los ao final do ensaio? Sob este ponto de vista, há fortes motivações para monitorar os dados sequencialmente. A necessidade ética de encerrar um ensaio clínico quando a saúde do paciente está em jogo e quando há uma superioridade de uma terapia sobre a outra é inquestionável. A oportunidade de encerrar mais cedo um estudo resulta em diminuição dos custos que em alguns estudos é crucial. Além disso, uma resposta mais rápida para a questão científica pode facilitar a iniciação de estudos adicionais ou indicar novas áreas de estudo e liberar os pacientes do estudo.

O dilema ético no monitoramento dos dados de um ensaio clínico reside basicamente em balancear o interesse de cada paciente dentro do ensaio (isto é, a ética individual de aleatorizar o próximo paciente) e o interesse a longo prazo em obter conclusões confiáveis sobre os dados (isto é, a ética coletiva de adotar condutas apropriadas para os pacientes futuros). Em outras palavras, consideremos um paciente futuro. Este paciente iria preferir o tratamento que melhor se adequasse à sua condição clínica. Mas se a todo paciente do ensaio fosse dado o tratamento que vem apresentando até então melhores resultados no próprio ensaio, uma das duas terapias seria provavelmente atribuída a uma esmagadora maioria de pacientes. Esse desbalanceamento no ensaio clínico pode limitar as conclusões comparativas a respeito do melhor tratamento (ver, por exemplo, Ware(1989) e Pocock(1990)). Se cada paciente é tratado tão bem quanto possível, os pacientes futuros em sua totalidade podem não ser: Eis aí o conflito entre a ética individual e a ética coletiva. Para Berry (1993), o único ensaio verdadeiramente ético provê informação completa sobre eficácia e efeito aos pacientes envolvidos, e permite aos pacientes escolherem sua própria terapia. Assim, aleatorizar só é ético se os pacientes estão completamente informados e escolhem ser aleatorizados.

Devemos observar que ao mesmo tempo que o monitoramento ou não de um ensaio clínico tem implicações éticas, muitas das objeções feitas a um experimento em que não há aleatorização, ou que não segue um protocolo experimental, devem-se às restrições causadas pela utilização de métodos clássicos extremamente dependentes do planejamento dos ensaios clínicos. Sob o ponto de vista do Princípio da Verossimilhança, tais restrições não têm fundamento. Instâncias relacionadas especificamente com o planejamento e, portanto, a aleatorização, a possibilidade de censura, a regra de parada, etc. não devem ser consideradas na inferência. Como Berger e Wolpert (1988) observam, na prática a visão condicional do experimento termina por ser adotada sempre que um protocolo experimental é alterado, seja porque alguém resolveu olhar os dados, ou porque acabou o papel na impressora, ou algo parecido. Esta é, de fato, uma posição pouco coerente, como expresso por Cornfield:

" The biostatistician's responsibility for providing biomedical scientists with a satisfactory explication of inference cannot, in my opinion, be satisfied by applying certain principles when he agrees with their consequences and by disregarding them when he doesn't. " (Cornfield (1966), p.19 e 20).

CAPÍTULO 4 - BIOEQUIVALÊNCIA: UMA PROPOSTA NEO-BAYESIANA DE ANÁLISE

4.1. INTRODUÇÃO

O planejamento, a execução e a avaliação de estudos de bioequivalência têm recebido especial atenção das indústrias farmacêuticas e autoridades da saúde na última década. Apesar dos esforços, nenhum consenso universal foi alcançado (Balant et al. (1991)). Por esta razão, anterior a qualquer descrição estatística, pelo menos dois pontos controversos relacionados com aspectos médicos são dignos de consideração aqui, isto é, o conceito de bioequivalência e sua significância terapêutica e, sua assim chamada equivalência terapêutica.

Bioequivalência é um conceito originalmente definido para a comparação da biodisponibilidade de formulações (cápsula, tablete, líquido intravenoso, etc.) de uma droga. Sob condições idênticas, espera-se que a taxa e a extensão da absorção do componente ativo ou terapêutico das formulações difiram somente dentro de limites toleráveis (Westlake (1979); Longer et al. (1992)). A equivalência "farmacêutica" das duas formulações, entretanto, não necessariamente assegura uma equivalência "biológica ou terapêutica".

Devido às restrições práticas e éticas, na prática, supondo-se que há relação direta entre níveis plasmáticos e efeitos terapêuticos, voluntários saudáveis são selecionados no lugar de pacientes e dados farmacocinéticos são avaliados ao invés de parâmetros clínicos. Desse modo, em geral, a avaliação da equivalência farmacêutica é considerada para assegurar também a equivalência terapêutica e consequentemente, a eficácia clínica. De um ponto de vista clínico, seria mais apropriado abordar bioequivalência em base terapêutica.

Especificamente considerando a análise estatística, muitos estudos relatados modelam o problema de um ponto de vista clássico ou frequentista. Sob tal enfoque, um parâmetro de

"bioequivalência é usualmente considerado e é conceitualmente uma razão de medidas médias de biodisponibilidade em uma população numerosa e fictícia". Veja, por exemplo, Metzler (1974) ou Mandallaz e Mau (1981).

Variações de até 20% nos parâmetros farmacocinéticos são arbitrariamente supostas não ter significado clínico. Assim, uma formulação teste (T) e uma de referência (R) são consideradas bioequivalentes se os parâmetros de locação (médias ou medianas) não diferem por mais do que 20% daquela formulação de referência (i.e. $|\mu_T - \mu_R| < 0.20 \mu_R$). Veja Westlake (1979) ou Steinijans et al. (1991).

Devemos observar que estes modelos, contudo, não podem garantir que um indivíduo seja indiferente a uma troca de formulações "terapeuticamente equivalentes". Além do mais, o problema conceitual com a modelagem estatística pode ser vista através de um exemplo extremo e direto: se para a metade da população P a formulação A apresenta uma biodisponibilidade claramente maior do que para a formulação B, enquanto que para a outra metade de P isto acontece com a formulação B no lugar de A, a bioequivalência tradicional (i.e. baseada em médias) das formulações A e B é declarada mesmo que nenhum paciente seja indiferente a elas.

Enfatizamos que as abordagens bayesianas (usuais) à bioequivalência modelam o problema da mesma maneira, i.e, eles consideram medidas médias de bioequivalência como parâmetros de interesse. (Selwyn et al. (1981) ou Fluehler et al. (1980)).

Sugerimos que uma abordagem genuinamente bayesiana deva considerar somente parâmetros definidos operacionalmente. Além do mais, um tratamento bayesiano deveria ser preditivo e não deve admitir parametrizações arbitrárias (Wechsler (1993), Zuazola (1993)). Estas idéias são perfeitamente concordantes com a noção intuitiva de que a bioequivalência para um par de formulações tem de ser vista como uma propriedade do

indivíduo, não do par. Cuidadasas medidas regulatórias precisam considerar que fração da população relevante se beneficiará (por ser terapeuticamente indiferente) à mudança de formulações. Esta noção foi claramente apresentada por Anderson e Hauck (1990). Entretanto, o trabalho deles é apresentado em um contexto freqüentista.

A violação do Princípio da Verossimilhança pela abordagem freqüentista, ou pelo uso de prioris não-informativas (ver capítulo 3) como fazem os bayesianos objetivistas (ver, por exemplo, Selwyn et al. (1981)) é evitada com a nossa proposta bayesiana subjetivista e completamente preditivista para o problema de avaliação de proporções de pessoas bioequivalentes na população.

4.2. MÉTODOS COMUNS DE ANÁLISE

O seguinte caso é apresentado na literatura: duas formulações de uma dada droga são comparadas. Uma formulação é usualmente de referência ou padrão (R), enquanto que a outra formulação (T) é nova ou de teste. A substituição de R por T é de interesse e pode ocorrer se as duas formulações forem julgadas bioequivalentes. Tais comparações são realizadas baseadas na análise de medidas como área sob a curva (AUC - "area under curve") ou concentração máxima da curva de concentração ($C_{\max}$ - "maximal concentration"), que indiretamente medem a biodisponibilidade das formulações. As respostas são usualmente obtidas após um experimento "crossover" incluindo n pessoas. O modelo estatístico é especificado por:

$$X_{ijk} = \mu + \gamma_i + \beta_j + \phi_k + \varepsilon_{ijk} \quad , \text{ onde:}$$

X_{ijk} : valor observado de AUC (ou $C_{\max}$) do i-ésimo indivíduo no j-ésimo período com a k-ésima formulação;

μ : média geral;

γ_i : efeito do i-ésimo indivíduo;

β_j : efeito do j-ésimo período;

ϕ_k : efeito da k-ésima formulação;

ε_{ijk} : erro (suposto ser normalmente distribuído),

com as restrições usuais dos modelos de análise de variância.

Sob o modelo acima, a rejeição da hipótese nula $H_0: \phi_R = \phi_T$ (igualdade dos efeitos das formulações) não tem, necessariamente, relevância clínica (nem farmacêutica), já que o principal interesse está na avaliação da diferença entre as formulações. Portanto, bioequivalência é declarada se um teste de hipóteses de que a razão $\theta = (\mu + \phi_T) / (\mu + \phi_R)$ entre as respostas médias das formulações de referência e de teste esteja entre certos limites é aceita. Os limites r_1 e r_2 são definidos por um médico ou autoridade da saúde e são usualmente próximos de 1. Usualmente, a hipótese de bioequivalência é testada verificando-se se (r_1, r_2) contém um intervalo de confiança para θ .

Por outro lado, a abordagem padrão bayesiana à bioequivalência reduz-se ao cálculo da probabilidade posterior de que a hipótese de bioequivalência seja verdadeira quando uma densidade a priori não informativa para θ é usada (Fluehler et al.(1980), Fluehler et al. (1983)). A bioequivalência é declarada quando esta probabilidade a posteriori é alta.

4.3. MÉTODO ALTERNATIVO DE ANÁLISE

Como descrito na seção 2, os enfoques tradicionais de avaliação da bioequivalência são baseados no conceito de bioequivalência média, i.e, sobre (a razão de) médias de respostas sobre uma população fictícia numerosa. Conseqüentemente, não se pode assegurar que em tal população "bioequivalente" exista um indivíduo para o qual as formulações produzam respostas similares (veja exemplo da seção 1). Isto nos motiva a tratar a bioequivalência como uma característica individual de todo indivíduo.

Neste sentido, a bioequivalência é declarada para todo indivíduo que apresenta a razão de respostas próxima a 1. Em outras palavras, se diz que o indivíduo i é bioequivalente contra um par de formulações R e T se a razão dele $\theta_i = X_{iT} / X_{iR}$ pertence a um intervalo $[r_1, r_2]$.

O modelo agora manipula um conjunto de variáveis indicadoras, $Y_i = I$ (i -ésimo indivíduo é bioequivalente), para $i=1, 2, \dots, n$. Seja π a proporção de indivíduos bioequivalentes na população relevante. Supomos que as variáveis Y_i são condicionalmente independentes e identicamente distribuídas como Bernoulli(π), dado π . Este modelo é uma consequência do julgamento de permutabilidade dos indivíduos com respeito a bioequivalência e também de uma suposição de ausência de efeitos de "carry-over" e de período. Esta última suposição, que pode ser válida quando o período de "wash-out" é suficientemente longo, é também necessária para Anderson e Hauck(1990) com o teste frequentista TIER ("test of individual equivalence ratios").

Tendo uma Beta(a, b) como densidade a priori para π , a densidade a posteriori para π ,

dado $(y_1, \dots, y_n)$, é uma Beta($a + \sum_{i=1}^n y_i, b + n - \sum_{i=1}^n y_i$). A bioequivalência entre as

formulações é então declarada caso a densidade posterior para π tem alta probabilidade. Isto corresponde ao caso em que, posterior à observação dos dados, há indícios de que uma

grande proporção de indivíduos são indiferentes a uma troca de formulações. A especificação de quão grande deveria ser tal proporção pode depender também de uma função utilidade específica de um médico ou de uma autoridade da saúde. Isto é uma questão regulatória, como apontado por Anderson e Hauck(1990). A função utilidade seria integrada com respeito a densidade a posteriori de π . Uma referência para teoria de decisão é DeGroot (1970). Veja também Berry, et al.(1992) para aplicações.

4.4. APLICAÇÕES

EXEMPLO 1:

Consideramos os dados apresentados por Fluehler et al (1983) em um estudo "crossover" comparativo de biodisponibilidades realizado com 12 indivíduos. A resposta aqui é a AUC, a área sob a curva de concentração sangüínea. A análise usual foi realizada com $r_1=0.8$ e $r_2= 1.2$. Estes limites estão realmente de acordo com as linhas gerais do FDA ("Food and Drug Administration"). A probabilidade posterior é $P(0.8 \leq \theta \leq 1.2 | \text{dados}) = 0.846$. Portanto, uma análise bayesiana usual produz fraca evidência de bioequivalência entre as formulações teste e de referência. A tabela 1 apresenta a razão de respostas para cada indivíduo.

Tabela 1: Razão de respostas de AUC em um estudo de biodisponibilidade (de Fluehler et al. (1983)).

SUJEITO	X_R	X_T	X_T/X_R
1	144.57	115.21	0.80
2	98.17	106.60	1.09
3	121.87	129.70	1.06
4	30.20	52.85	1.75
5	131.51	59.42	0.45
6	104.17	152.76	1.47
7	71.54	31.24	0.44
8	71.98	108.22	1.50
9	78.83	82.05	1.04
10	140.48	101.10	0.72
11	75.27	58.72	0.78
12	111.56	83.27	0.75

Usando uma Beta (1,1) (i.e. a uniforme em (0,1)) a priori para π obtemos uma Beta(5,9) a posteriori para π , pois como podemos ver na tabela 1 há somente 4 pessoas bioequivalentes entre os 12 indivíduos incluídos no experimento. Um exame da função distribuição Beta(5,9) produz para o exemplo:

$$\Pr(\pi > 0.4|\text{dados}) = 0.353$$

$$\Pr(\pi > 0.5|\text{dados}) = 0.133$$

$$\Pr(\pi > 0.6|\text{dados}) = 0.032$$

$$\Pr(\pi > 0.7|\text{dados}) = 0.004$$

Como a probabilidade a posteriori de que a proporção de indivíduos bioequivalente seja grande é muito pequena, ficamos inclinados a concluir que as formulações não deveriam ser trocadas.

Mencionemos que a mesma conclusão seria alcançada pela aplicação do TIER de Anderson e Hauck(1990): com confiança 95%, pelo menos 18.1% da população é bioequivalente. Este valor é muito pequeno para a conclusão a favor da bioequivalência.

EXEMPLO 2:

Examinaremos os dados de Clayton e Leslie (1981) que alegam a bioequivalência entre as duas formulações de erythromycin quando tomados imediatamente após a alimentação. Eles realizaram um estudo "crossover" com 18 indivíduos. Os valores e razões de AUC sob as duas formulações são apresentadas na tabela 2.

Tomando-se os limites $r_1=0.8$ e $r_2=1.2$, Anderson e Hauck (1990) dizem que os dados não apoiam a bioequivalência entre as formulações: O limite inferior, com 95% de confiança para a proporção de indivíduos bioequivalentes é de somente 12%.

Usando uma densidade uniforme em $(0,1)$, obtemos uma densidade a posteriori $\text{Beta}(5,15)$ para π , pois há 4 sujeitos bioequivalentes na amostra. Como consequência temos:

$$\Pr(\pi > 0.4|\text{dados}) = 0.0696$$

$$\Pr(\pi > 0.5|\text{dados}) = 0.0096$$

$$\Pr(\pi > 0.6|\text{dados}) = 0.0006$$

$$\Pr(\pi > 0.7|\text{dados}) \approx 0.0000$$

Tabela 2: Razão de respostas de AUC em um estudo de biodisponibilidade (de Clayton e Leslie (1981)).

SUJEITO	X_R	X_T	X_T/X_R
1	2.52	5.47	0.46
2	8.87	4.84	1.83
3	0.79	2.25	0.35
4	1.68	1.82	0.92
5	6.95	7.87	0.88
6	1.05	3.25	0.32
7	0.99	12.39	0.08
8	5.60	4.77	1.17
9	3.16	1.88	1.68
10	3.19	4.98	0.64
11	9.83	7.14	1.38
12	2.91	1.81	1.61
13	4.58	7.34	0.62
14	7.05	4.25	1.66
15	3.41	6.66	0.51
16	2.49	4.76	0.52
17	6.18	7.16	0.86
18	2.85	5.52	0.52

Como a probabilidade a posteriori, dado os dados, de uma grande proporção de casos bioequivalentes na população é pequena, concluímos que as formulações não são bioequivalentes, em oposição ao alegado por Clayton e Leslie(1981).

CAPÍTULO 5 - CONCLUSÃO

Fundamental para a demonstração do Princípio da Verossimilhança é a consideração da classe Ξ de experimentos comparáveis, como foi feito no capítulo 2. A rigor, a violação do Princípio da Verossimilhança por algum método de inferência ocorre ou não dentro dessa respectiva classe de experimentos. O método $\underline{e}$ sua classe violam (ou não) o Princípio da Verossimilhança.

Admitindo, como Dawid (1977), que um princípio é um teste pelo qual um método deve passar, o método de inferência deve ser investigado em todos os possíveis experimentos a respeito de θ na classe Ξ . Se esta classe de experimentos não inclui todos os experimentos a respeito da quantidade desconhecida θ , então até mesmo um procedimento clássico pode estar não violando o Princípio da Verossimilhança bastando, por exemplo, que esta classe contenha somente o experimento realizado. Esta é uma forma de "escapar" da violação ao Princípio da Verossimilhança que é, contudo, muito restritiva, pois limita o alcance da discussão a, digamos, um único tipo de experimento. De fato, algumas formas de evasão discutidas no capítulo 2 se enquadram no contexto descrito acima, pois, por exemplo, ao rejeitar a descrição usual dos experimentos estatísticos por requisição de uma estrutura, a classe Ξ fica reduzida.

Por outro lado, a modificação dos princípios pré-fundamentais (isto é, de condicionalidade e suficiência) propostas por Durbin(1970) e Kalbfleisch(1975) revela-se como uma tentativa para justificar a não obediência ao Princípio da Verossimilhança. Isto porque, adotando-se os princípios propostos por Durbin e Kalbfleisch, o Princípio da Verossimilhança não pode ser demonstrado e, portanto, perde o seu apelo lógico.

Outra forma de evasão discutida no capítulo 2 justifica a não obediência ao Princípio da Verossimilhança pela adoção do Princípio da Confiança. Esta posição também é pouco sedutora, já que revela-se incoerente.

Implicações do Princípio da Verossimilhança como a irrelevância da regra de parada, do mecanismo de censura e da aleatorização são bastante sedutoras ao senso prático conforme constatamos no capítulo 3 através de uma série de exemplos. Elas depõem contra os métodos clássicos de inferência, extrema e permanentemente dependentes das características de planejamento do experimento.

Embora o Princípio da Verossimilhança seja um princípio de inferência, ele possui algumas implicações éticas. Ele não diz como conduzir um experimento (isto é, se os tratamentos devem ou não ser aleatorizados, quando o experimento deve encerrar-se ou não, etc.), mas o seu enfoque condicional permite que uma pessoa observe os dados provisórios, e até altere o protocolo experimental. A justificativa clássica para, no entanto, seguir "cegamente" o protocolo experimental é de que somente assim resultados confiáveis podem ser obtidos. É nesse ponto que se inserem algumas questões éticas. É ético continuar executando um experimento após obtermos informações de que um tratamento expõe os pacientes a um risco desnecessário? É ético ignorarmos os dados provisórios para só analisá-los ao final do experimento? Essas questões de fato inserem outra fonte de grandes discussões entre os cientistas: a ética individual e a ética coletiva.

A adoção do ponto de vista condicional no Princípio da Verossimilhança permite que o experimento encerre-se a qualquer momento e que dados provisórios sejam avaliados, sem, no entanto, por fim à dialética entre ética coletiva e individual.

Já no capítulo 3, vimos que somente o bayesiano subjetivista não viola o Princípio da Verossimilhança. Esse fato definiu a proposta de um método bayesiano subjetivista para o problema da bioequivalência em obediência ao Princípio da Verossimilhança no capítulo 4.

Nesse capítulo, a bioequivalência foi operacionalmente tratada como uma característica individual (e não populacional, coletiva) e a proporção de indivíduos como a quantidade de interesse (ao invés da razão entre médias de respostas aos tratamentos). Fica claro que a noção de bioequivalência individual deveria ser preferível à noção de bioequivalência populacional, já que uma droga bioequivalente à outra na população nem sempre apresenta resposta similar no indivíduo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anderson, S. e Hauck, W.W. (1990). Consideration of Individual Bioequivalence. **Journal of Pharmacokinetics and Biopharmaceutics**, **18**, No.3, 259-273.
- Armitage, P. (1961). **Sequential Medical Trials**. Oxford: Blackwell Scientific Publications.
- Balant, L.P. , Benet, L.Z., Blume, H., Bozler, G., Breimer, D.D., Eichelbaum, M., Gundert-Remy, U., Hirtz, J.L., Mutschler, E., Midha, K.K., Rauws, A.G., Ritschel, W.A., Sansom, L.N. Skelly, J.P., Vollmer, K-O. (1991). Is there a need for more precise definitions of bioavailability? Conclusions of a Consensus Workshop. **European Journal of Clinical Pharmacology**, **40**, 123-126.
- Barnard, G.A.(1949). Statistical inference (with discussion). **J. Roy. Statist. Soc. B**, **11**, 115-139.
- Barnett, V. (1982). **Comparative Statistical Inference**. 2nd. edition. Chichester: John Wiley & Sons.
- Basu, D. (1969). Role of the Sufficiency and Likelihood Principles in Survey Theory. **Sankhyā Ser. A**, **31**, 441-453.
- Basu, D. (1975). Statistical information and likelihood (with discussion). **Sankhyā Ser. A**, **37**, 1-71.
- Basu, D. (1980). Randomization analysis of experimental data: the Fisher Randomization Test (with discussion). **J. Amer. Statist. Assoc.**, **75**, 575-595.
- Berger, J.O. (1980). **Statistical Decision Theory - Foundations, Concepts, and Methods**. New York: Springer-Verlag.
- Berger, J.O. (1985). In defense of the Likelihood Principle: Axiomatics and Coherency. In: **Bayesian Statistics II**, J.M. Bernardo, M.H. DeGroot, D. Lindley, and A. Smith (eds.). Valencia: University Press.

- Berger, J.O. (1990). Birnbaum's theorem is correct: a reply to a claim by Joshi. In: **Journal of Statistical Planning and Inference**, **26**, 112-113.
- Berger, J.O. e Berry, D. A. (1988). The Relevance of Stopping Rules in Statistical Inference. In. **Statistical Decision Theory and Related Topics IV**. vol.1. S.S. Gupta and J.O. Berger (eds.). New York:Springer-Verlag.
- Berger, J.O. e Wolpert, R.L. (1988). **The Likelihood Principle**. IMS Lectures Notes, Monograph Series, vol.6, 2nd. Edition. Hayward: Institute of Mathematical Statistics.
- Bernardo, J.M. e Smith, A.F. (1993). **Bayesian Theory**. New York: John Wiley & Sons.
- Berry, D. A.(1987 - a). Interim analysis in clinical trials - The role of the Likelihood Principle. **The American Statistician**, **41**, 117-122.
- Berry, D. A.(1987 - b). Statistical inference, designing, clinical trials, and pharmaceutical company decisions. **The Statistician**, **36**, 181-189.
- Berry, D. A. (1990). Basic Principles in Designing and Analyzing Clinical Studies. In: **Statistical methodology in the pharmaceutical sciences**. D.A. Berry (ed.) .New York: Marcel Dekker, Inc.
- Berry, D. A. (1993). A Case for Bayesianism in Clinical Trials. **Statistics in Medicine**, **12**, 1377-1393.
- Berry, D.A., Wolff, M.C. and Sack, D. (1992). Public Health Decision Making: a Sequential Vaccine Trial. In. **Bayesian Statistics**, vol. 4. Oxford: Clarendon Press.
- Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). **J. Amer. Statist. Assoc.** , **57**, 269-306.
- Birnbaum, A. (1972). More on concepts of statistical evidence. **J. Amer. Statist. Assoc.**, **67**, 858-861.

- Bjørnstad, J. F.(1992) Introduction to Birnbaum (1962) On the Foundations of Statistical Inference. In: **Breakthroughs in Statistics - Foundations and Basic Theory** - vol.1 . S. Kotz and N. L. Johnson (eds) New York: Springer-Verlag.
- Clayton, D., Leslie, A. (1981). The Bioavailability of Erythromycin Stearate Versus Entericcoated Erythromycin Base When Taken Immediately Before and After Food. **The Journal of International Medical Research**, 9, 470-477.
- Cornfield, J. (1966). Sequential Trials, Sequential Analysis and the Likelihood Principle. **The American Statistician**, 18-23.
- Dawid, A.P. (1977). Conformity of Inference Patterns. In: **Recent Developments in Statistics**. J.R. Barra, et al. (eds). Amsterdam: North-Holland Publishing Company.
- Dawid, A.P. (1983). Inference, Statistical: I. In: **Encyclopedia of Statistical Sciences**. vol. 4. New York: John Wiley, 89-105.
- de Finetti, B. (1972). **Probability, Induction and Statistics**. New York: John Wiley & Sons.
- de Finetti, B. (1975). **Theory of Probability**, vol. 2. Chichester: John Wiley & Sons.
- DeGroot, M.H. (1970). **Optimal Statistical Decisions**. MacGraw-Hill.
- DeGroot, M.H. (1975). **Probability and Statistics**. Massachusetts: Addison-Wesley Publishing Company, Inc.
- Durbin, J. (1970). On Birnbaum's Theorem on the Relation between Sufficiency, Conditionality and Likelihood (with discussion). **Journal of the American Statistical Association**, 65, 395-403.
- Edwards, A.W.F. (1972). **Likelihood**. Cambridge: University Press.
- Evans, M.J., Fraser, D.A.S. e Monette, G. (1986). On principles and arguments to likelihood (with discussion). **The Canadian Journal of Statistics**, 14, 181-199.

- Fisher, R. A. (1959). **Statistical Methods and Scientific Inference**. Edinburgh: Oliver and Boyd (2nd. edition).
- Fluehler, H., Hirtz, J., Moser, H.A.(1980). An aid to Decision-Making in Bioequivalence Assessment. **Journal of Pharmacokinetics and Biopharmaceutics**, 9, No.2, 235-243.
- Fluehler, H., Grieve, A.P., Mandallaz, D., Mau, J., Moser, H.A. (1983). Bayesian Approach to Bioequivalence Assessment: An Example. **Journal of Pharmaceutical Sciences**, 72, No.10, 1178-1181.
- Fraser, D.A.S. (1968). **The Structure of Inference**. New York: John Wiley & Sons.
- Fraser, D.A.S. (1983). Structural Inference. In: **Encyclopedia of Statistical Sciences**. vol. 9. New York: John Wiley, 20-27.
- Fraser, D.A.S. (1983). Structural Models. In: **Encyclopedia of Statistical Sciences**. vol. 9. New York: John Wiley, 27-32.
- Fraser, D.A.S. (1983). Structural Prediction. In: **Encyclopedia of Statistical Sciences**. vol. 9. New York: John Wiley, 32.
- Fraser, D.A.S. (1983). Structural Probability. In: **Encyclopedia of Statistical Sciences**. vol. 9. New York: John Wiley, 32.
- Fraser, D.A.S., Monette, G. e Ng, K.W. (1985). Marginalization, Likelihood and Structured Models. In: **Multivariate Analysis-VI**, P.R. Krishnaiah (ed.). New York: Academic Press.
- Freireich, E.J., Gehan, E., Frei, E., Schroeder, L.R., Wolman, I.J., Anbari, R., Burgert, E.O., Mills, S.D., Pinkel, D., Selawry, O.S., Moon, J.H., Gendel, B.R., Spurr, C.L., Storrs, R., Haurani, F., Hoogstraten, B., and Lee, S. (1963). The effect of 6-Mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. **Blood**, 21, 699-716.

- Goldstein, M. e Howard, J.V. (1991). A likelihood paradox (with discussion). **J. Roy. Statist. Soc. B**, **53**, 619-628.
- Good, I.J. (1990). In defense of 'the' Likelihood Principle. In: **Journal of Statistical Planning and Inference**, **26**, 122-123.
- Hill, B. M. (1974). Review of "Bayesian Inference in Statistical Analysis" by G.E.P.Box and G. Tiao. **Technometrics**, **16**, 478-479.
- Hill, B. M. (1981). On some statistical paradoxes and non-conglomerability (with discussion). In: **Bayesian Statistics**. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.) Valencia: University Press, 39-66.
- Hill, B.M. (1987). The validity of the Likelihood Principle. **The American Statistician**, **41**, 95-100.
- Inman, H.F. (1994). Karl Pearson and R.A. Fisher on Statistical Tests: A 1935 Exchange from *Nature*, **The American Statistician**, **48**, 2-11.
- Inoue, L.Y.T. and Wechsler, S. (1994). Inferência Comparada: A Violação do Princípio da Verossimilhança. **Revista de Matemática e Estatística**, **12**, 19-31.
- Joshi, V. M. (1983). Likelihood Principle. In: **Encyclopedia of Statistical Sciences**. vol. 4. New York: John Wiley & Sons, 644-647.
- Joshi, V.M. (1989). A Counter-example against the Likelihood Principle. **J. Roy. Statistic. Soc. B**, **51**, 215-216.
- Joshi, V.M. (1990). Fallacy in the proof of Birnbaum's theorem. In: **Journal of Statistical Planning and Inference**, **26**, 111-112.
- Kadane, J.B. (1986). Progress Toward a More Ethical Method for Clinical Trials. **The Journal of Medicine and Philosophy**, **11**, 385-404.
- Kadane, J.B. and Sedransk, N. (1979). Toward a More Ethical Clinical Trial. In: **Bayesian Statistics**. Bernardo et al. (eds).

- Kadane, J.B. and Seidenfeld, T. (1990). Randomization in a Bayesian Perspective. **Journal of Statistical Planning and Inference**, **25**, 329-345.
- Kalbfleisch, J.D. (1975). Sufficiency and Conditionality (with discussion). **Biometrika**, **62**, 251-269.
- Lehmann, E.L. (1993). The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? . **Journal of the American Statistical Association**, **88**, No. 424, 1242-1249.
- Lindley, D.V. (1983). **Lectures on Bayesian Statistics**. Publicações do Instituto de Matemática e Estatística da Universidade de São Paulo.
- Lindley, D.V. (1990). The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics. **Statistical Science**, **5**, No. 1, 44-89.
- Longer, M.A., Schaefer, H.G., Derendorf, H. (1992). Grundlagen der Beurteilung von Bioäquivalenzstudien. **Pharm. Ztg. Wiss.**, **1**, 15-22.
- Loschi, R.H. (1992). **Coerência, Probabilidade e Calibração**. São Paulo: IME-USP (tese de mestrado).
- Mandallaz, D. and Mau, J. (1981). Comparison of Different Methods for Decision-Making in Bioequivalence Assessment. **Biometrics**, **37**, 213-222.
- Metzler, C.M. (1974). Bioavailability - A Problem in Equivalence. **Biometrics**, **30**, 309-317.
- Meyer, M.C. (1991). Drug Product Selection - Part 2: Scientific Basis of Bioavailability and Bioequivalence Testing. **Am. Pharm. NS31(8)**, 47-53.
- O'Brien, P.C. (1990). Group Sequential Methods in Clinical Trials. In: **Statistical methodology in the pharmaceutical sciences**. D.A.Berry (ed.). New York: Marcel Dekker, Inc.
- Pereira, C.A.B. and Wechsler, S.(1993). On the Concept of P-Value. **REBRAPE**, **7**, 159-177.

- Piccinato, L. (1992). Critical Issues in different inferential paradigms. **J. Ital. Statist. Soc.**, **2**, 251-274.
- Pocock, S.J.(1977). Group sequential methods in the design and analysis of clinical trials. **Biometrika**, **64**, 191-199.
- Pocock, S.J.(1990). Statistical and Ethical Issues in Monitoring Clinical Trials. **Statistics in Medicine**, **12**, 1459-1469.
- Rodda, B.E., Tsianco, M.C., Bolognese, J.A. and Kersten, M.K. (1988). Clinical Development. In: **Biopharmaceutical Statistics for Drug Development**. K.E.Pearce (ed.) New York: Marcel Dekker, Inc.
- Schuirman, D. J. (1990). Design of Bioavailability/Bioequivalence Studies. **Drug Inf. J.** **24**, 315-323.
- Schwabe, U. (1988). Zur Frage der Bioaquivalenz von Generika. **Internist**, **29**, 147-159.
- Seidenfeld, T. (1992). R.A. Fisher's Fiducial Argument and Bayes' Theorem. **Statistical Science**, **7**, 358-368.
- Selwyn, M.R., Dempster, A.P., Hall, N.R. (1981). A Bayesian Approach to Bioequivalence for the 2x2 Changeover Design. **Biometrics**, **37**, 11-21.
- Shapiro, S.H. e Louis, T.A. (1983). **Clinical Trials - Issues and Approaches**. New York: Marcel Dekker, Inc.
- Steinijans, V.W. and Hauschke, D.(1990). Update on the statistical analysis of bioequivalence studies. **Int. J. Clin. Pharmacol. Ther. Toxicol.** **28(3)**, 105-110.
- Steinijans, V.W. , Hauschke, D., Jonkman, J.H.G. (1991). Controversies in Bioequivalence Studies. **Clin. Pharmacokinet.**, **22**, 1-7.
- Stone, M. (1983). Probability Fiducial. In: **Encyclopedia of Statistical Sciences**. vol. **3**. New York: John Wiley, 81-86..

- Ware, J. H. (1989). Investigating Therapies of Potentially Great Benefit: ECMO (with discussion). **Statistical Science**, 4, No. 4, 298-340.
- Wechsler, S. (1989). **On de Finettian Decision-Making**. Berkeley: University of California.
- Wechsler, S. (1993). Exchangeability and Predictivism. **Erkenntnis**, 38, 343-350.
- Westlake, W.J.(1979). Statistical Aspects of Comparative Bioavailability Trials. **Biometrics**, 35, 273-280.
- Zabell, S. L. (1992). R.A.Fisher and the Fiducial Argument. **Statistical Science** , 7, 369-387.
- Zuazola, P.L.I. (1993). Finite Forms of De Finetti's Theorem - the Predictivistic Approach to Statistical Inference in Finite Populations (In Portuguese). São Paulo: IME-USP.
- Zuazola, P.L.I. , Pereira, C.A.B and Wechsler, S. (1994). Predictivistic Statistical Inference in Finite Populations. IME-USP - Relatório Técnico No. 9406.