

The Internacional
Biometric Society



51^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria

24 a 26 de maio de 2006

Departamento de Bioestatística
Instituto de Biociências
UNESP – Botucatu

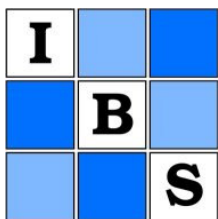
Minicurso

ESTATÍSTICA APLICADA: UMA VISÃO GENUINAMENTE BAYESIANA

Carlos Alberto de Bragança Pereira
IME/USP

Apoio





The International
Biometric Society



51^a Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria

24 a 26 de maio de 2006

Departamento de Bioestatística
Instituto de Biociências
UNESP – Botucatu

Minicurso

ESTATÍSTICA APLICADA: UMA VISÃO GENUINAMENTE BAYESIANA

Carlos Alberto de Bragança Pereira
IME/USP

Apoio



COMISSÃO ORGANIZADORA

José Eduardo Corrente - UNESP/Bot

Luzia A. Trinca - UNESP/Bot

Lídia R. de Carvalho - UNESP/Bot

Liciana V. A. Silveira - UNESP/Bot

Helenice de O. F. Silva - UNESP/Bot

Paulo F. A. Mancera - UNESP/Bot

COMISSÃO CIENTÍFICA

José Eduardo Corrente - UNESP/Bot

Joel Augusto Muniz - UFLA

Clarice Garcia B. Demétrio - ESALQ

Lídia R. de Carvalho – UNESP/Bot

Luzia A. Trinca - UNESP/Bot

COMISSÃO JULGADORA

Adair Regazzi - UFV

Ademar Sanches - UNESP/Jab

Antonio Carlos Pião - UNESP/RC

Arminda Lucia Siqueira - UFMG

Carlos R. Padovani - UNESP/Bot

Carlos T. dos S. Dias - ESALQ

Cecília Candolo - UFSCar

Cláudia Pio Ferreira - UNESP/Bot

Dilermando Perecin - UNESP/Jab

Dinara W. X. Fernandes – UFRS

Edwin M. M. Ortega - ESALQ

Enrico A. Colósimo - UFMG

Helenice de O. F. Silva - UNESP/Bot

João Riboldi - UFRS

José Eduardo Corrente - UNESP/Bot

José Silvio Govone - UNESP/RC

Júlia M. Pavan Soler - USP

Júlio S. de S. Bueno Filho - UFLA

Liciana V. A. Silveira - UNESP/Bot

Lídia R. de Carvalho - UNESP/Bot

Lindolfo Storck - UESM

Luzia A. Trinca - UNESP/Bot

Maria Cecília M. Barreto - UFSCar

Maria Sílvia de A. Moura - UFSCar

Paulo F. A. Mancera - UNESP/Bot

Thelma Sáfadi - UFLA

Carlos Alberto de Bragança Pereira

IME/USP

**ESTATÍSTICA APLICADA:
UMA VISÃO GENUINAMENTE
BAYESIANA**

Botucatu
2006

Patrocinadores



FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO
DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
Bibliotecária responsável: Selma Maria de Jesus

Pereira, Carlos Alberto de Bragança.

Estatística aplicada: uma visão genuinamente bayesiana / Carlos Alberto de Bragança Pereira. – Botucatu : Instituto de Biociências, Universidade Estadual Paulista, 2006.

Vi, 86 p. : 28cm.

Trabalho apresentado na 51ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, 24 a 26 de maio de 2006.

1. Bioestatística 2. Estatística aplicada 3. Biometria

CDD 574.0212

*Há um tempo em que é preciso
abandonar as roupas usadas,
que já têm a forma de nosso corpo,
e esquecer os nossos caminhos,
que nos levam sempre aos mesmos lugares.
É o tempo da travessia:
e, se não ousarmos fazê-la,
teremos ficado, para sempre,
à margem de nós mesmos.*
FERNANDO PESSOA

Estatística Aplicada: Uma Visão Genuinamente Bayesiana.

Prefácio

Fui surpreendido com um novo e honroso convite da Professora Luzia Trinca para ministrar um curso na reunião de Biometria. Novamente, trabalhos recentes são usados para poder aproveitar as idéias que estão mais presentes em minha memória e em meus interesses atuais.

O título escolhido é fruto do trabalho que venho desenvolvendo, com meu amigo e colega Julio da Motta Singer, dentro da disciplina Estatística Aplicada, oferecida aos alunos do último ano de Bacharelado de Estatística no IME-USP. Esta disciplina tem o objetivo de treinar o aluno para o trabalho de consultoria estatística. Recebemos profissionais e pesquisadores, atuantes nos diversos campos do conhecimento, que procuram auxílio para analisar informações e dados obtidos em trabalhos e pesquisas nas respectivas áreas. Desses trabalhos surgem questões tão interessantes que provocam desafios em nosso próprio trabalho estatístico. Não foram poucas as vezes que Julio e eu nos vimos envolvidos em discussões calorosas sobre a melhor forma de responder às questões estatísticas levantadas. Dessas nossas discussões surgiram problemas cujas soluções foram temas de publicações ou de teses e dissertações. Os desafios que vinham do Julio para mim eram sobre a possibilidade de encontrarmos soluções Bayesianas adequadas para determinados problemas. Meus desafios para ele, e que na grande maioria das vezes produziram soluções elegantes e adequadas, eram sobre a forma frequentista de responder às perguntas colocadas. Essa minha experiência pode ser traduzida com a seguinte afirmação: Embora participando como professor, *Estatística Aplicada* é a melhor disciplina que *cursei* em minha vida de estatístico. Entenda-se por melhor aquela disciplina que oferece ao participante o maior aprendizado ou desenvolvimento pessoal.

Ao aceitar o desafio de ministrar o curso e produzir o material didático correspondente, passei a escolher os assuntos a serem desenvolvidos. Considerei então a necessidade de apresentar, ao participante, as razões que poderiam produzir a vontade de considerar métodos alternativos de análise em seu trabalho cotidiano. O primeiro capítulo, sobre minhas reflexões, apresenta motivos para a necessidade de análises estatísticas sob novos paradigmas. O material discutido deve surpreender aqueles acostumados a usar apenas os métodos mais comuns da estatística frequentista.

O Capítulo 2 é sobre avaliação de testes diagnósticos. Ali é apresentada uma discussão de como livrar-se dos espaços amostrais no momento da análise estatística, pois estes podem não estar bem definidos. O capítulo é fruto de uma parceria com meu irmão, Pereira & Pereira (2005). Esse artigo é consequência do nosso entendimento de que verossimilhanças contêm todas as informações que os dados podem produzir sobre parâmetros desconhecidos e de interesse. Ao trabalhar sob tal paradigma, o estatístico inicia sua trajetória definitiva para o Bayesianismo.

É indispensável para o estatístico de sucesso a preocupação com os conceitos que utiliza. Embora informação seja a palavra mais comum da linguagem estatística, muitos não são capazes de apresentar uma definição precisa deste conceito. O Capítulo 3 é dedicado a uma discussão de como conceituar apropriadamente *informação*. Na verdade, o capítulo foi escrito em memória ao Professor Dev Basu, um mestre dos fundamentos da estatística. Como o leitor irá comprovar, serão apresentadas e discutidas as idéias de três professores renomados e que nunca se ausentam da memória do autor. Tive a grata satisfação de conhecer, assistir palestras e, principalmente, ler publicações dos professores Dev Basu, David Blackwell e Morris DeGroot. Devo a eles o pouco que consegui em minha carreira como professor de estatística. Recomendo aos leitores os trabalhos indispensáveis desses mestres: Basu (1975), Blackwell (1951) e DeGroot (1962).

Finalmente, usando toda a argumentação apresentada nos capítulos iniciais, o Capítulo 4 é dedicado aos testes de significância naquela que considero ser a visão genuinamente Bayesiana destes. Lembre-se que testes de significância são procedimentos usados para avaliar a consistência das observações em relação a uma hipótese sobre o parâmetro de interesse. Muitas vezes tais hipóteses são precisas: o espaço de definição da hipótese tem dimensão inferior à dimensão do espaço paramétrico original. Este é um aspecto que causa muitas dificuldades tanto aos frequentistas quanto aos Bayesianos. Ao tratar este problema como no Capítulo 4, temos a convicção que as dificuldades foram em grande parte eliminadas. O meu colaborador nesse capítulo é o Ricardo Vêncio. Nossas discussões sobre a melhor forma de como apresentar o FBST motivou o Ricardo a produzir as ilustrações auto-explicativas do Capítulo 4. Na verdade tenho tido a honra de ser co-autor do Ricardo em diversos artigos sobre biologia molecular. Embora seja oficialmente meu aluno de doutorado, na verdade atua muito mais como motivador dos diversos trabalhos que publicamos nos dois últimos anos.

O Capítulo 5 forma o principal objetivo de nosso curso: o uso da metodologia Bayesiana para resolver os problemas da prática da estatística aos dados reais. O primeiro problema, sobre preferência sexual em um experimento com *Drosophila melanogaster*, foi o tema de minha dissertação de mestrado em 1971. Visito novamente esse problema com o objetivo de ser genuinamente Bayesiano, o que não acontecia no meu trabalho original. Os dois outros exemplos foram discutidos no Centro de Estatística Aplicada no segundo semestre de 2005. Tivemos a sorte de encontrar as soluções Bayesianas que resolveram apropriadamente os problemas de prevalência apresentados. Em conjunto com os acadêmicos discutidos ao longo do texto, creio atingir com esses três exemplos reais os objetivos de um curso. É importante registrarmos que sem o apoio do povo e da administração municipal do EMBU o terceiro exemplo não existiria.

O material apresentado não é fruto de pesquisas solitárias do autor. É, na verdade, resultado do trabalho desenvolvido por um grupo de professores que usam seus pontos de vista diferenciados para poderem criar um clima de conflito constante e, conseqüentemente, o clima ideal de pesquisa original. Sou profundamente grato aos meus colegas e amigos do IME-USP, Julio Stern, Wagner Borges e Sergio Wechsler, que em conjunto com nossos ex-alunos, Luis Gustavo e Regina Madruga Tavares, e alunos, Marcelo Lauretto e Fábio Nakano, formamos o grupo de estudos e de pesquisas intitulado “The Full Bayesian Significance Test (FBST) Research Group”.

Ao leitor gostaria ainda de informar que essa monografia tem apenas o objetivo de divulgar idéias que vieram de grandes mestres. Fui mesmo um privilegiado, pois recebi aulas de Dev Basu, David Blackwell, Oscar Kempthorne, Frank Proschan, Dick Barlow, Morri DeGroot IJ Good e Dennis Lindley. O fato de tê-los conhecido e aproveitado do convívio por algum tempo é o maior trunfo que comigo carrego na vida acadêmica. Se o leitor estiver mesmo empenhado em conhecer estatística, não perca tempo com o presente trabalho e vá direto aos escritos desses grandes pensadores. Além desses mestres, tive também o privilégio de conviver intensamente com os Professores Euclides Custódio de Lima Filho e Flávio Wagner Rodrigues. Trago vivo na memória nossas conversas e os ensinamentos que recebi. Quando escrevia essas notas tive sempre comigo a lembrança desses amigos. O Flávio me pedindo para simplificar os argumentos e ser dramático nos contra-exemplos. O Euclides me pedindo para não perder o foco das aplicações práticas. Como se pode perceber, sou mesmo um privilegiado.

Inspirado por Alexandre Graham Bell, termino este prefácio com a frase que orienta o FBST-Research Group, *“se você seguir apenas por caminhos traçados, chegará onde todos poderiam chegar”*.

Carlos Alberto de Bragança Pereira
BioInfo – Núcleo de Bioinformática
MAE – Departamento de Estatística
IME – Instituto de Matemática e Estatística
USP – Universidade de São Paulo
São Paulo, 1 de maio de 2006

CONTEÚDO

| | |
|--|----|
| 1. Discutindo Conceitos e Métodos Através de Exemplos. | 01 |
| 1.1. Introdução | 01 |
| 1.2. Independência: Probabilística e Estatística | 01 |
| 1.3. Informação e Independência Condicional | 03 |
| 1.4. Confiança e Credibilidade | 04 |
| 1.5. Significância e Evidência | 06 |
| 1.6. Ilustração Gráfica | 10 |
| 1.7. Considerações Finais | 15 |
| 2. Testes de Diagnósticos em Medicina Clínica: | |
| Analisando Apenas as Verossimilhanças (em Inglês) | 17 |
| 2.1. Introduction | 17 |
| 2.2. Statistical Tests – A Brief History | 17 |
| 2.3. Likelihood Approach | 19 |
| 2.4. Ladder of Uncertainty and Controversies | 22 |
| 2.5. Diagnostic Tests and Statistical Verdicts | 24 |
| 2.6. Diagnosability | 26 |
| 2.7. Likelihood Analysis of a Diagnostic Test and Likelihood Ratio Plots | 28 |
| 2.8. Final Remarks | 30 |
| 3. Informação e Análise Pré Posteriori. | 33 |
| 3.1. Introdução | 33 |
| 3.2. Informação Segundo Basu | 33 |
| 3.3. Informação Segundo DeGroot | 35 |
| 3.4. Informação Segundo Blackwell: Suficiência de Blackwell | 36 |
| 3.5. Tamanho da Amostra | 39 |
| 4. FBST por gráficos. | 47 |
| 4.1. Introdução | 47 |
| 4.2. Teoria do FBST | 47 |
| 4.3. FBST: Gráficos Ilustrativos | 52 |
| 5. Estudo de casos submetidos ao CEA-USP. | 57 |
| 5.1. Preferência Sexual em <i>Drosophila melanogaster</i> | 57 |
| 5.2. Padrões de <i>Ectoparasitismo</i> em aves de cerrado | 61 |
| 5.3. Prevalência e Diagnóstico da Fibromialgia no EMBU | 69 |
| 6. Referências Bibliográficas | 83 |

1. Discutindo Conceitos e Métodos Através de Exemplos

1.1. INTRODUÇÃO

Conceitos estatísticos, fundamentais aos métodos que são usados no dia-a-dia do profissional, constituem o tema principal deste capítulo. A mensagem ao profissional é que faça uma maior reflexão sobre seu instrumento de trabalho, os métodos estatísticos. Devemos sempre questionar métodos que estão sendo aceitos universalmente para não permitir que sejam utilizados equivocadamente. Esse questionamento deve ser sobre a lógica do método e não sobre seu uso imediato. Isto porque, ao entender completamente um método, o profissional saberá quando e como usá-lo. Imaginem as consequências da necessidade do uso imediato de um método quando o usuário não conhece a lógica de sua criação.

Este capítulo é objetivo, pois toda a discussão é feita através exemplos que mostram a inadequação de métodos de uso comum do estatístico. Evidentemente a inadequação diz respeito à situação específica do exemplo. O uso inquestionável de um determinado método pode trazer sérios problemas ao usuário. Muitas vezes a resposta a um problema pode não ser adequada. Conceitos imprecisos também aparecem como causa de soluções inadequadas. De fato, este texto questiona a falta de precisão de linguagem no uso dos métodos estatísticos. Não só a forma de apresentação, mas também os próprios conceitos precisam ser revistos.

Discutem-se na seqüência os conceitos de independência probabilística e estatística; os conceitos de dependência condicional e informação; os conceitos de intervalos de confiança e de credibilidade; e finalmente os conceitos de significância e evidência. Ressaltamos que toda a discussão deste capítulo é feita por meio de exemplos especiais em contextos muito simples.

1.2. INDEPENDÊNCIA: PROBABILÍSTICA E ESTATÍSTICA

Em estatística, o conceito de dependência está intimamente ligado ao conceito mais usual de associação. No entendimento comum, dois eventos estão associados quando um influencia a ocorrência do outro. Por exemplo, a ocorrência de um pode aumentar a chance de o outro ocorrer e neste caso dizemos que a associação é positiva. Muitas vezes a ocorrência de um fenômeno pode diminuir ou mesmo eliminar a chance de o outro ocorrer e neste caso dizemos que a associação é negativa. Assim, a dependência entre duas quantidades aleatórias pode ocorrer em diferentes níveis, tanto positivamente quanto negativamente. Por outro lado, é sempre bom lembrar que os conceitos equivalentes de associação e dependência têm uma característica simétrica. Isto é, se x é dependente de y , com um determinado nível, então y é dependente de x com o mesmo nível.

A negação de dependência é a independência e este sim é um conceito preciso. Não há diferentes níveis de independência! Afirma-se que duas quantidades aleatórias são independentes quando, ao conhecer-se o valor de uma, a distribuição da outra não se altera. Em outras palavras, se x e y são duas quantidades aleatórias independentes então $p(y)=p(y|x)$; isto é, o fato de x ter sido observado não altera a distribuição de

y. Um bom exercício para o leitor é provar que as seguintes três igualdades são equivalentes:

- i) $p(y)=p(y|x)$;
- ii) $p(x)=p(x|y)$ e
- iii) $p(x,y)=p(y)p(x)$.

O conceito de independência probabilística é naturalmente estendido para o de independência condicional quando outra quantidade z entra no contexto. As quantidades x e y são ditas condicionalmente independentes, dado z , se $p(x,y|z)=p(y|z)p(x|z)$. Outro bom exercício para o leitor é mostrar a equivalência das seguintes igualdades:

- iv) $p(y|z)=p(y|x,z)$;
- v) $p(x|y)=p(x|y,z)$ e
- vi) $p(x,y|z)=p(y|z)p(x|z)$.

Independência estatística é na verdade independência condicional. Em estatística o parâmetro, desconhecido e de interesse, exerce o mesmo papel da quantidade z . Os Exemplos 1.2(1, 2 e 3) ilustram os conceitos de independência probabilística e estatística.

1.2.1. Exemplo: Normal Padrão

Considerem-se duas variáveis aleatórias normais independentes, x e y , com mesma distribuição normal com média zero e variância um. Inicialmente, sabe-se que embora os valores das duas quantidades possam ocorrer em toda a reta, muito provavelmente não estarão fora do intervalo $[-6;+6]$. Além disso, a informação de que $x=3$, por exemplo, não irá alterar a distribuição de y . Isto é, $p(y)=p(y|x)$. Novamente y pertence com grande chance ao intervalo $[-6;+6]$. O conhecimento da ocorrência de $x=3$ não alterou a distribuição de probabilidades inicialmente proposta para y . No presente caso, a independência é a probabilística no sentido usual.

1.2.2. Exemplo: Normal com Média Desconhecida

Sejam duas variáveis aleatórias normais independentes, x e y , com a mesma distribuição normal com média z , desconhecida, e variância um. Como o valor da média, z , é desconhecido, tanto x como y podem variar em toda a reta, sem preferências. A observação $x=3$ agora é bastante informativa com respeito à distribuição de y . Evidentemente, existe uma grande chance de que $x=3$ esteja no intervalo $[z-6;z+6]$. Assim, $-3 < z < 9$ e, portanto, y muito provavelmente irá pertencer ao intervalo $[-9;15]$. Note-se que, antes da observação de x , o valor de y era completamente desconhecido e poderia assumir qualquer valor em toda a reta. No entanto, após observar-se $x=3$, sabe-se que quase certamente y pertence ao intervalo finito $[-9;15]$. Isto prova que x e y não podem ser independentes, visto que a observação do valor de x modifica de fato a distribuição de y .

No exemplo acima o correto seria dizer: caso z fosse conhecido, x e y seriam independentes. Esse é o caso da independência estatística. Na linguagem probabilística, ao considerar-se z como variável aleatória, x e y seriam condicionalmente (em relação a z) independentes. Na perspectiva freqüentista da estatística, o parâmetro, representado aqui por z , não é considerado como variável aleatória. Conseqüentemente, sem uma estrutura probabilística para z , o uso do conceito de independência condi-

cional estaria prejudicado. Contudo, o estatístico poderia sempre afirmar que, no caso do valor de z ser conhecido, fosse ele qual fosse, x e y seriam independentes. Isto é, $p(x,y|z) = p(y|z)p(x|z)$.

Para, de forma definitiva, sensibilizar os professores de estatística, o Exemplo 3 é mais dramático do que o anterior.

1.2.3 Exemplo: Uniforme com Média Desconhecida

Considerem-se duas variáveis aleatórias independentes, x e y , com mesma distribuição uniforme no intervalo $[z-1; z+1]$, onde a média z é desconhecida. No início, pode-se dizer que $-\infty < y < +\infty$. No entanto, observando-se $x=3$, conclui-se que $2 < z < 4$ e, assim, y certamente pertencerá ao intervalo $[1;5]$. Isto é, o conhecimento de que $x = 3$ faz com que, novamente, apresente-se o caso de independência condicional (estatística) entre x e y , dado z .

Esse terceiro exemplo mostra que, embora ocorra a independência estatística – correspondendo, no caso de estrutura probabilística para z , à independência condicional de x e y , dado z –, na verdade tem-se uma forte dependência (não condicional). Note-se que ambas as variáveis possuem um elemento desconhecido em comum nas suas distribuições e esse elemento, o parâmetro z , é o que provoca a forte associação entre x e y .

1.3. INFORMAÇÃO E DEPENDÊNCIA CONDICIONAL

O exemplo discutido nesta seção é de igual importância, pois ressalta mais uma vez a diferença entre o trabalho estatístico e o probabilístico. O leitor deve entender, no entanto, que um não exclui o outro. Na verdade, o probabilístico complementa o estatístico, e vice-versa!

Apresenta-se apenas um exemplo, suficiente para a discussão, por ser crucial para o ensino da Estatística. Lembre-se de que a função de verossimilhança carrega toda a informação, contida nos dados x , sobre os parâmetros z . Diferenciem-se bem os dois elementos do modelo estatístico: a) a distribuição amostral é caracterizada pela função de probabilidade de x para cada valor fixado, z_0 , do parâmetro z , isto é, $p(x|z_0)$, e b) a função de verossimilhança para a observação x_0 de x é a distribuição amostral avaliada em x_0 , como função do parâmetro z ; isto é, $L(z|x_0) = p(x_0|z)$. Assim, a distribuição amostral pode ser caracterizada pela classe dos conjuntos $\mathcal{I}_z = \{f(x|z); x \in X\}$ e a verossimilhança da observação x pelo conjunto $\mathcal{I}_x = \{f(x|z); z \in Z\}$ onde X e Z são, respectivamente, os espaços amostral e paramétrico.

1.3.1. Exemplo: Um Par de Filhos

A escola de dança do bairro compete todos os anos com um grupo de 10 meninas no concurso nacional de artes. A professora Lily toma conhecimento de que as duas irmãs que fazem parte do grupo vão acompanhar os pais na imigração para o Chile. Ela é informada de que a nova família, que vai morar na casa onde moram atualmente as duas bailarinas, possui um par de crianças. Lily então pondera que se

forem duas meninas poderão substituir as duas que estão saindo do grupo. Dessa forma, terá tempo de treinar as novas componentes para a próxima competição. Acha então que tem probabilidade $\frac{1}{4}$ de o par ser formado por duas meninas. Seu espaço de probabilidades equiprováveis é $\{(m,m);(m,h);(h,m);(h,h)\}$. Conversando então com seu irmão, Jony, corretor de imóveis, a professora é informada de que o par é formado por pelo menos uma menina. Essa informação Jony obteve quando ouviu por telefone a mãe dizendo "fique quieta minha filha" para uma criança que fazia algum ruído perto do telefone. Lily fica alegre já que sua probabilidade aumentou para $\frac{1}{3}$; desse modo seu espaço de probabilidades equiprováveis é agora $\{(m,m);(m,h);(h,m)\}$. Mary, a esposa de Jony, entra na sala e diz que uma das crianças da nova moradora é uma menina que está no carro. Lily sai e vai ao carro para ver a criança. Fica mais feliz, pois entende que sua probabilidade aumenta para $\frac{1}{2}$ após ver a menina e se perguntar "qual a probabilidade de a outra criança ser menina?". Seu espaço de probabilidades equiprováveis, depois de ver a criança, é $\{m;h\}$, visto que a única fonte de incerteza é o sexo da outra criança. Aceitar passivamente o que foi apresentado na dramatização acima implica aceitar-se que a afirmação "pelo menos uma menina no par" possui diferentes pesos quando recebida por diferentes canais: audição e visão.

O olhar de um estatístico permite incorporar outras informações ocorridas no processo de coleta de dados. Para entender a racionalidade estatística, é preciso definir apropriadamente todas as entidades do processo. Seja o parâmetro $z = 1$ no caso de um par de meninas e $z = 0$ no caso de um casal. A probabilidade *a priori* de z é definida por $p(1) = 1/3 = 1 - p(0)$. Como variável observacional, para definir a verossimilhança, considere-se $x = 1$ se a criança que acompanha Mary é uma menina e $x = 0$ se menino. A função de verossimilhança, $p(x|z) = L(z|x)$, no caso presente de $x = 1$ é $L(z|1)$. No caso de $z = 1$ (par de meninas), $L(1|1) = 1$ e, se $z = 0$ (um casal), $L(0|1) = q$, onde q é um valor do intervalo $[0;1]$. Equivalentemente, pode-se escrever $p(x=1|z=0) = q$ e $p(x=1|z=1) = 1$. Com um cálculo simples e após lembrar que sua observação foi $x=1$, o leitor deve ser capaz de mostrar que a probabilidade *a posteriori* de $z = 1$ é então $p(z=1|x=1) = [1+2q]^{-1}$.

A princípio Lily pensou que Mary teria probabilidades iguais ($q = 1/2$) de trazer menino ou menina, no caso de o par ser um casal de crianças ($z=0$). Neste caso, Lily estaria certa em se alegrar com o aumento da sua probabilidade para um par de meninas, pois, se $q = 1/2$, a probabilidade *a posteriori* de $z = 1$ seria $1/2$. Mas Lily logo percebeu seu erro! Mary levou a criança, a pedido da mãe ocupada, para provar um vestidinho na nova loja de roupas femininas infantis, ali em frente. Assim Lily teve de considerar $q = 1$ e nesse caso *a posteriori* seria $1/3$, o mesmo valor da probabilidade *a priori* após a informação transmitida por Jony. Nesse caso a verossimilhança é constante e assim não informativa!

Como conclusão desta seção, o leitor deve entender que no momento da definição do modelo de probabilidades, é necessário sempre levar em consideração o modo como a ocorrência de eventos é observada.

1.4. CONFIANÇA E CREDIBILIDADE

O exemplo discutido nesta seção diz respeito à estimação por intervalo. O conceito de intervalo de confiança é o que mais dificulta a vida do estatístico. Muitas vezes este precisa explicar ao cientista que o intervalo numérico apresentado, acom-

panhado da porcentagem 95%, não quer dizer que a probabilidade de esse intervalo conter o verdadeiro valor do parâmetro é 95%. Deve na verdade explicar que, se o procedimento pudesse ser repetido uma série de vezes, os intervalos construídos conteriam o verdadeiro valor do parâmetro em 95% das vezes. Para entender-se melhor a dificuldade que esse procedimento representa, o exemplo a seguir mostra a diferença entre probabilidade e confiança.

1.4.1. Exemplo: Uniforme com Média Desconhecida

Considerem-se quatro observações, $\{u, v, x, y\}$, independentes (estatisticamente) e identicamente distribuídas segundo uma uniforme no intervalo $[z - \frac{1}{2}z + \frac{1}{2}]$. O objetivo é construir um intervalo de confiança para z a partir das quatro observações. Represente-se por m e M , respectivamente, o menor e o maior valores amostrais. Não é difícil ver que o intervalo $[m; M]$ é um intervalo com confiança 87,5%, pois $1 - 2(\frac{1}{2})^4 = 0,875$. Lembre-se de que $(\frac{1}{2})^4$ é a probabilidade de que as quatro observações estejam acima (ou abaixo) da média z ; i.e., $p(m > z) = p(M < z) = (\frac{1}{2})^4 = 0,0625$. Os quatro valores efetivamente observados foram: 1,11; 1,27; 1,43; 1,59. Um estatístico clássico concluiria que $[1,11; 1,59]$ é um intervalo com 87,5% de confiança. Note-se que, se no lugar de 1,59 tivéssemos 1,50 ou mesmo 1,62, a confiança seria a mesma, embora os comprimentos dos intervalos fossem respectivamente 0,48, 0,39 e 0,51. Mesmo no caso em que M fosse 1,91 a confiança do intervalo $[1,11; 1,91]$ continuaria sendo 87,5%. O que um estatístico menos comprometido com métodos faria? Certamente consideraria que $M < z + 0,5$ e $m > z - 0,5$. Isto é, o mínimo e o máximo de qualquer amostra devem situar-se no intervalo de variação da variável que está sendo estudada. Com isso seria possível dizer que com probabilidade 1, certeza total, $M - 0,5 < z < m + 0,5$. Isto é, com certeza, o intervalo $[M - 0,5; m + 0,5]$ contém z ! Resumindo: se $m = 1,11$, os casos em que $M = 1,59$, $M = 1,50$, $M = 1,62$ e $M = 1,91$ produziriam, respectivamente, os intervalos $[1,09; 1,61]$, $[1,00; 1,61]$, $[1,12; 1,61]$ e $[1,41; 1,61]$, cujos comprimentos seriam 0,52, 0,61, 0,49 e 0,21. Todos esses intervalos, com certeza – com probabilidade 1 –, conteriam o verdadeiro valor do parâmetro, z . Assim, o resultado amostral efetivamente observado pode ser mais ou menos informativo. Isto é, quanto maior a diferença $M - m$, mais informativo deve ser o resultado da inferência sobre o valor de z . Note-se que, com o procedimento do clássico intervalo de confiança de 87,5%, os intervalos seriam $[1,11; 1,59]$, $[1,11; 1,50]$, $[1,11; 1,62]$ e $[1,11; 1,91]$. Os comprimentos dos intervalos seriam 0,48, 0,39, 0,51 e 0,81. A conclusão é que o procedimento apresentado oferece menos informação ao cliente – intervalo de grande comprimento – justamente quando a amostra é mais informativa, M e m distantes. **Isto é uma incoerência!**

Para uma solução Bayesiana do problema estudado no Exemplo 5, considere-se *a priori* que o parâmetro z se distribui uniformemente no intervalo $[-1000; 1000]$. Com respeito às amostras descritas acima, as distribuições *a posteriori* seriam uniformes nos intervalos $[1,09; 1,61]$, $[1,00; 1,61]$, $[1,12; 1,61]$ e $[1,41; 1,61]$. De posse dessas distribuições, poder-se-iam construir intervalos centrais com 0,875 de probabilidade. Esses intervalos são conhecidos como intervalos de credibilidade com 87,5% de credibilidade. Conseqüentemente, $[1,12; 1,58]$, $[1,04; 1,57]$, $[1,15; 1,58]$ e $[1,42; 1,60]$ seriam os intervalos com credibilidade 87,5% e comprimentos 0,46, 0,53, 0,43 e 0,18. Observe-se a coerência desse tipo de inferência em que, amostras mais informativas oferecem respostas superiores; ao escrevermos 87,5% de credibilidade

estamos, como é desejável, falando de o intervalo conter o valor desconhecido do parâmetro, z , com probabilidade $0,875$.

A Seção 5 conclui este artigo com uma discussão sobre o método estatístico “teste de significância”, o mais contemplado nas diversas áreas de aplicação. Novamente, o interesse é mostrar que certos cuidados devem ser tomados e os resultados corretamente interpretados quando um determinado método é utilizado.

1.5. SIGNIFICÂNCIA E EVIDÊNCIA

Neste capítulo discute-se, por meio de dois exemplos, o conceito de significância. Foram usadas as recomendações comuns para o cálculo dos valores-p em problemas de teste de significância. Trata-se de dois exemplos simples e talvez comuns ao dia-a-dia de um estatístico.

É comum aos cursos de estatística o fato de não ser dada ênfase à hipótese alternativa quando testes de hipóteses baseiam-se no conceito de significância. Após a obtenção das observações amostrais, apenas a hipótese nula é usada no momento da definição da distribuição nula e suas caudas. Muitas vezes não se consideram o fato de distribuições sob as hipóteses alternativas possuírem caudas menos pesadas do que aquelas sob a hipótese nula. Em alguns casos, os níveis de significância não devem ser áreas das caudas, mas sim áreas centrais da distribuição nula. Os exemplos discutidos a seguir mostram quão importantes podem ser as hipóteses alternativas na definição da regra aceita/rejeita.

1.5.1. Exemplo: Bolinhas na Urna

Apresenta-se uma urna com três bolinhas de diferentes cores: preta, verde e branca. Após três retiradas de uma bola da urna, anotaram-se as quantidades x = número de pretas (0 ou 1) e y = número de verdes (0 ou 1). Observado o resultado $(x,y) = (1,1)$, o interesse é testar a hipótese **H**: a amostra foi obtida com reposição, contra a alternativa **A**: a amostra foi obtida sem reposição. A Tabela 1.4.2 apresenta a distribuição de (x,y) , multiplicada por 27, quando **H** é verdadeira. Note-se que o valor-p é igual a 100%, se a definição da cauda incluir o observado $(1,1)$, ou 77,8%, no caso de sua exclusão. Seguindo-se as regras ensinadas em cursos de estatística, **H** não seria rejeitada e **A** não seria considerada. Isto é, se $(x,y)=(1,1)$, **H** é aceita! A questão é: **A** pode ser desconsiderada nesse caso? Observe-se que amostras que teriam probabilidade de ocorrência baixa seriam aquelas diferentes de $(1,1)$. Entretanto, qualquer amostra diferente de $(1,1)$ teria probabilidade zero de ocorrer no caso de **A** verdadeira. Isto é, os pontos que poderiam levar à rejeição de **H** seriam pontos impossíveis de ocorrer sob **A**. **Novamente, observa-se uma incoerência!**

Analisando-se a questão com os olhos de um estatístico preocupado com o problema a ser resolvido e não com um método disponível, a razão de verossimilhanças seria algo a ser considerada. A razão de verossimilhanças para um resultado (x,y) é a probabilidade desse resultado, quando **A** é verdadeira, dividido pela probabilidade, quando **H** é verdadeira. Isto é, $r(x,y) = p(x,y|\mathbf{A}) \div p(x,y|\mathbf{H})$. Para todos os resultados possíveis, exceto o ponto $(1,1)$, o valor de r seria zero. No caso do resultado observado, $r(1,1) = 9/2 = 4,5$. Isto prova que privilegiar **A** em detrimento de **H** é a conclusão

mais acertada com o resultado $(x,y)=(1,1)$. Qualquer outro resultado indicaria a aceitação de **H**.

1.5.2 Tabela: Distribuição nula de (x,y) do Exemplo 6.

| y | 0 | 1 | 2 | 3 |
|-----|---|---|---|---|
| x | | | | |
| 0 | 1 | 3 | 3 | 1 |
| 1 | 3 | 6 | 3 | 0 |
| 2 | 3 | 3 | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 |

Sob a perspectiva Bayesiana a análise fica mais simples. Supondo-se equi-probabilidades *a priori*, $p(\mathbf{H}) = p(\mathbf{A}) = 1/2$, a conseqüente distribuição *a posteriori* de **A** seria $p(\mathbf{A}|1,1) = 1 - p(\mathbf{H}|1,1) = 1 - 6/33 = 9/11$. No caso de qualquer outro resultado, **H** certamente seria verdadeira. Assim, com probabilidade *a priori* considerada, a observação de $(x;y) = (1;1)$ tem como conseqüência o fato de a probabilidade de **A** passar a ser quatro vezes maior do que a probabilidade de **H**. Isto é, a amostra observada favorece **A** em detrimento de **H**.

O exemplo final deste manuscrito é mais comum ao professor de estatística, pois se refere à distribuição normal, a mais discutida em salas de aula.

1.5.3. Exemplo: Normal com variância desconhecida

Considerem-se três observações, x , y e z , estatisticamente independentes e identicamente distribuídas segundo uma normal com média zero e variância desconhecida, representada por v , $N(0,v)$. A estatística suficiente neste caso é a soma dos quadrados das observações, $s = x^2 + y^2 + z^2$. As hipóteses a serem confrontadas são a nula **H**: $v = 2$ contra a alternativa **A**: $v \neq 2$. É importante ressaltar-se que a variável aleatória s/v é distribuída segundo uma χ^2 com 3 graus de liberdade (*gl*), independente do valor que v pode assumir. Considere-se que os valores observados sejam $x = 0,6$, $y = 0,3$ e $z = -0,2$, conseqüentemente, $s = 0,49$. Sob a hipótese **H**, $N(0,2)$ é a distribuição nula e suas caudas, delimitadas por essas observações, seriam $C(x) = 89\%$, $C(y) = 83\%$ e $C(z) = 67\%$. Multiplicando-se as áreas das três caudas, o resultado seria $C(x,y,z) \cong 50\%$. Por outro lado, a correspondente área da cauda da distribuição nula (χ^2 com *gl* = 3) de $s/2$ é calculada como $p(s > 0,49 | \mathbf{H}) = p_3(\chi^2 > 0,245) = 0,97$. Isto é, a amostra $(x;y;z) = (0,6;0,3;-0,2)$ produz uma significância (valor-p) de 97%, favorecendo **H** contra **A**. Todos esses cálculos padrões privilegiam a hipótese **H**, em detrimento de **A**. Outra forma de obter-se um valor-p seria concentrar as atenções apenas em s , abandonando a amostra original, visto que s é uma estatística suficiente para

(x,y,z) com respeito ao parâmetro z . Ao considerar-se que valor da significância é a área das caudas da distribuição nula, o primeiro passo seria a definição das caudas da distribuição nula de $s/2$, de acordo com seu valor obtido na amostra, **0,245**. O ponto com mesma densidade na distribuição χ^2 com $gl = 3$ é **2.6112**. Com as caudas definidas pelos limites (de igual densidade) **0,245** e **2,6112**, o novo valor-p seria

$$p_3(\chi^2 < 0,245) + p_3(\chi^2 > 2,6112) = 0,02998 + 0,45553 = 0,4855.$$

Dessa forma, a significância é **48,55%**, novamente privilegiando **H** contra **A**.

Usando-se métodos padrões, como os descritos acima, as observações encontradas favoreceriam **H** contra **A**. Contudo, usando a intuição, não é absurdo concluir-se que a observação $s = 0,49$ deveria privilegiar valores de ν inferiores a 2: As três observações pertencem a um intervalo contido em **(-1;1)**. Relevante seria o leitor entender que valores baixos de s correspondem a pontos amostrais centrais. Tais pontos são mais prováveis de ocorrer – maior densidade – quando ν for baixo, como, por exemplo, $\nu \leq 1$. Neste caso, a hipótese alternativa é que deveria ser privilegiada e não a nula. A conclusão desta discussão indicaria que as regiões críticas do teste seriam aquelas que contivessem valores amostrais próximos de zero (caso da amostra estudada) ou bem distantes de zero, isto é nas caudas da distribuição nula. O problema aqui é a definição de região crítica: apenas a hipótese nula é usada na delimitação das caudas, usadas como região crítica. A seguir, ambas as hipóteses serão usadas na definição de regiões que beneficiam **A**, em detrimento de **H**, mais do que a amostra efetivamente observada.

Para uma visão menos objetiva do Exemplo 7, considere-se que, sob **A**, $u = \nu^{-1}$ tem uma distribuição *a priori* χ^2 com um grau de liberdade. Ao usar-se a função de densidade $h(u)$ dessa distribuição como um sistema de preferências no espaço (paramétrico) dos possíveis valores de u , a densidade preditiva sob **A**, $f(s|\mathbf{A})$, é definida como a média – ponderada por $h(u)$ – da densidade de s , $f(s|u)$. Obtém-se assim,

$$f(s|\mathbf{A}) = \int h(u)f(s|u)du = \left(\frac{2}{\pi}\right)(\sqrt{s})(s+1)^{-3}.$$

Com as densidades nula e preditiva – $f(s|\mathbf{H}) = f(s|\nu=2)$ e $f(s|\mathbf{A})$ – define-se a razão de verossimilhanças ou Fator de Bayes:

$$r(s) = f(s|\mathbf{A}) \div f(s|\mathbf{H}) = \left(\frac{8}{\sqrt{\pi}}\right) [(s+1)^{-2}] \exp\left(\frac{s}{4}\right).$$

O Fator de Bayes pode ser usado como um índice de ordem no espaço amostral. Este índice, $r(s)$, permite o cálculo um valor-p que leva em consideração ambas as hipóteses. Com a observação $s = 0,49$, procura-se o outro ponto s' com mesmo valor de r . No caso do exemplo, o ponto $s' = 22,5845$ satisfaz essa propriedade; isto é, $r(0,49) = r(22,5845) = 4,073$. Com esses pontos delimitando as caudas da distribuição nula, obtém-se o valor-p modificado, $p\nu = 4\%$. Isto é,

$$1 - \Pr(0,49 \leq s \leq 22,5845 | \mathbf{H}) = 1 - p_3(0,245 < \chi^2 < 11,292) = 0,04.$$

Seguindo a intuição do estatístico competente, este valor de pv , se considerado como *índice de significância*, privilegia a hipótese alternativa em detrimento da hipótese nula. Por outro lado, se as probabilidades *a priori* das duas hipóteses fossem consideradas equiprováveis, a probabilidade *a posteriori* seria $p(\mathbf{H}|s) = (1+4,073)^{-1} = 0,197$, novamente favorecendo **A** contra **H**.

Concentrando a atenção apenas na função de verossimilhança, que carrega toda a informação contida na amostra sobre o parâmetro, o estatístico na maioria das vezes tira as conclusões corretas e de forma coerente. Com a parametrização considerada, a hipótese nula é re-escrita como $\mathbf{H}: u = v^{-1} = 1/2$. A função de verossimilhança de u é proporcional a uma densidade gama com parâmetros $A = 5/2$ e $B = 0,245$. Por outro lado, os valores assumidos por esta função nos pontos $u = 1/2$ e $u = 24,2696$ são iguais. Conseqüentemente, todo valor de u no intervalo $(0,5;24,2696)$ é mais verossímil do que $1/2$, o valor de u da hipótese nula. O conjunto \mathbf{T} formado por todos os pontos paramétricos mais verossímeis do que o da hipótese nula é denominado *conjunto tangente*: $\mathbf{T} = (0,5;24,2696)$ no presente caso. Como a função $g(u)$, a função de verossimilhança normalizada por sua integral, é a densidade Gama com parâmetros $A = 5/2$ e $B = 0,245$, a área sob a função de verossimilhança em \mathbf{T} é igual a **96,23%** da área total. Isto é um indicativo de que há uma predominância de pontos mais verossímeis do que a hipótese nula. Como uma alternativa ao valor-p, define-se o valor da *evidência* da amostra em favor de **H**, *valor-e ou ev*, como o complemento da área do conjunto tangente; i.e., $ev(t) = 3,77\%$. Novamente, como esperado, apresenta-se um indicativo contra a hipótese nula em favor da alternativa.

Esta ultima avaliação da amostra, com relação à hipótese nula, foi baseada exclusivamente na função de verossimilhança, em completa concordância com o princípio da verossimilhança. De acordo com esse princípio, duas funções proporcionais de verossimilhança devem produzir a mesma inferência estatística. Evidentemente, ev pode ser calculado sempre que uma densidade de probabilidades esteja sendo definida no espaço paramétrico, como é o caso de uma densidade *a posteriori*, ou mesmo de uma função de verossimilhança cuja integral seja finita. No caso discutido acima, os valores-e são baseados nas caudas da verossimilhança, caudas essas definidas a partir da hipótese nula. Uma das vantagens deste método em relação a outros é que não existe restrição ao número de parâmetros envolvidos. Além disso, não há a necessidade de eliminação de parâmetros *indesejáveis*, uma prática comum ao estatístico que atua sob o paradigma freqüentista. Não há, teoricamente, restrições de dimensão nem ao espaço paramétrico nem ao amostral. Evidentemente, quanto maior a dimensão, maior o desafio computacional.

Embora pv e ev sejam probabilidades calculadas, respectivamente, no espaço amostral e no espaço paramétrico, seus valores apontam para a mesma conclusão: evidência contra **H** favorecendo **A**. As conclusões obtidas com os três métodos são:

1. ao usar-se a receita freqüentista de olhar apenas a densidade sob a hipótese nula, desconsiderando a hipótese alternativa, os dados observados favorecem **H** em detrimento de **A**;

2. ao considerar-se a hipótese alternativa no cálculo do valor-p modificado, os mesmos dados desfavorecem **H** em favor de **A** e
3. sob uma visão genuinamente Bayesiana, onde a única contribuição da amostra é a função de verossimilhança, a conclusão baseada no valor da evidência também favorece **A** em detrimento de **H**.

Resumindo: as significâncias de **H**, $pv \cong ev \cong 4\%$, obtidas com os dois métodos, clássico modificado e Bayesiano, indicam a rejeição de **H**, favorecendo **A**, enquanto o método padrão, com o valor-p de **49%**, aponta para a direção inversa.

1.6. ILUSTRAÇÃO GRÁFICA

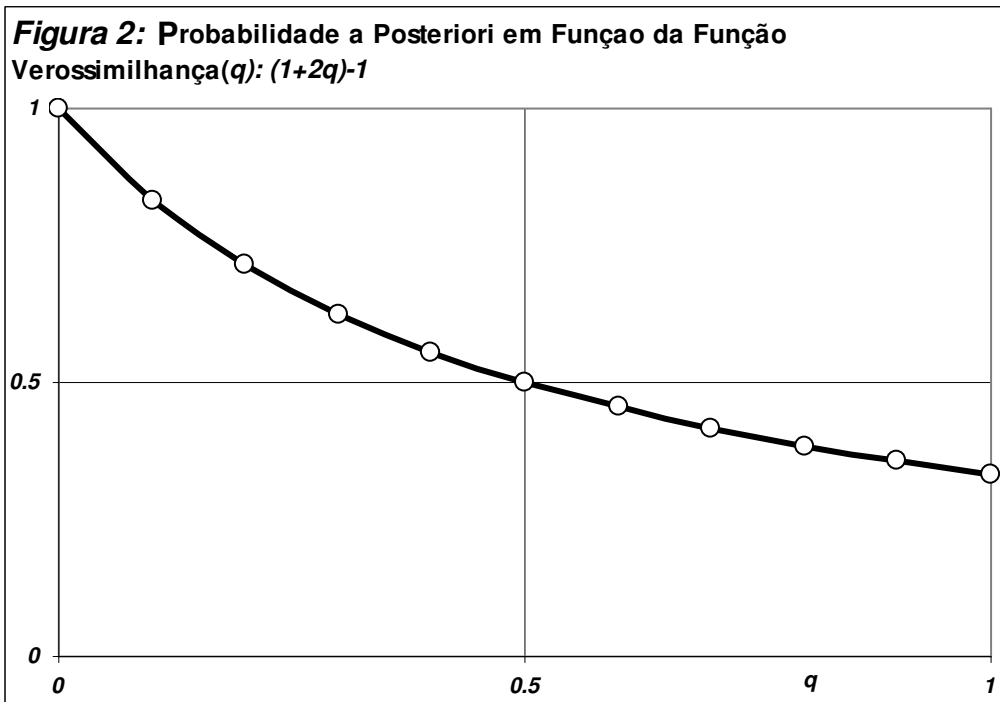
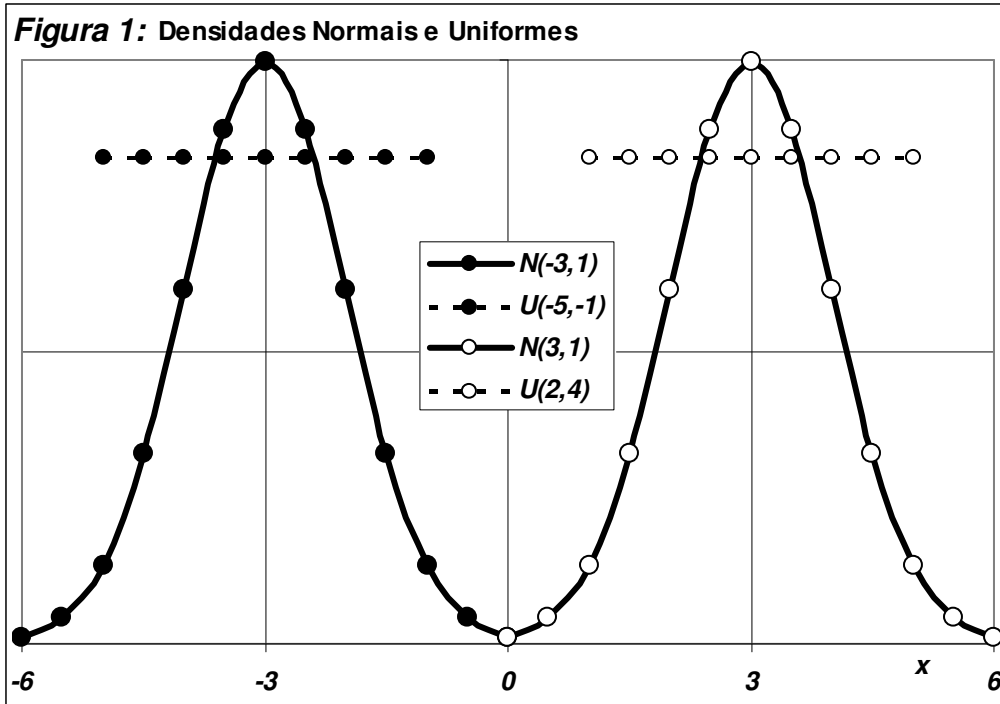
Nesta seção o leitor encontra uma série de ilustrações que visam auxiliá-lo no entendimento dos exemplos apresentados. Densidades normais e uniformes, discutidas nos Exemplos 1.2(2 e 3), são apresentadas na Figura 1. A Figura 2 apresenta os valores da probabilidade *a posteriori* do Exemplo 1.3.1 em função das possíveis verossimilhanças. O objetivo principal é ressaltar que a subjetividade pode estar, muitas vezes, na licitação da função de verossimilhança e não apenas da distribuição *a priori*.

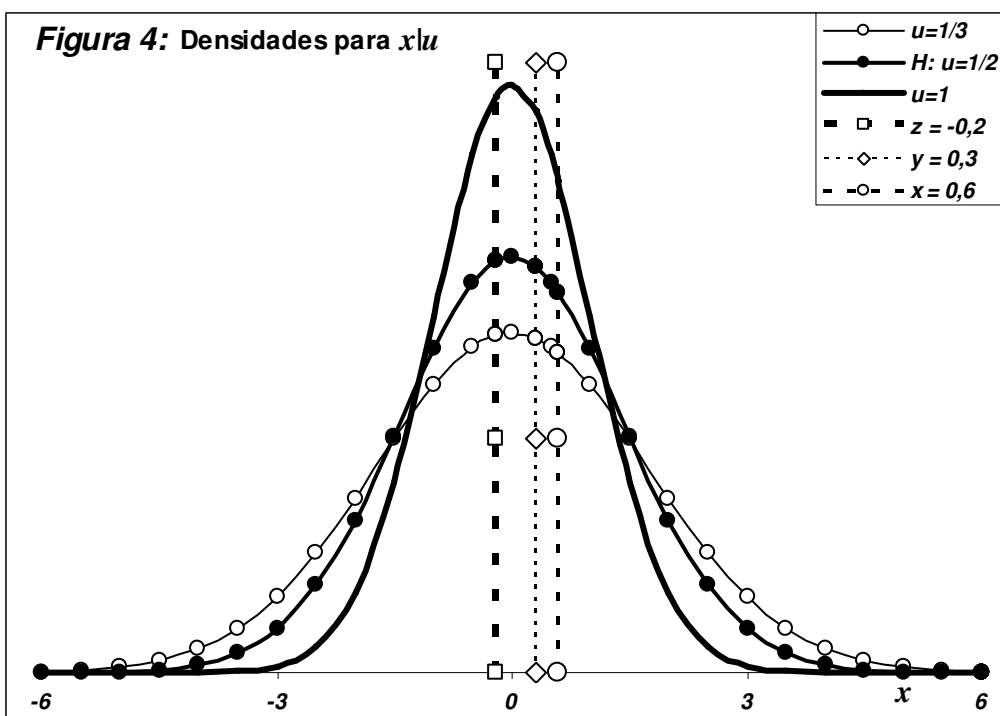
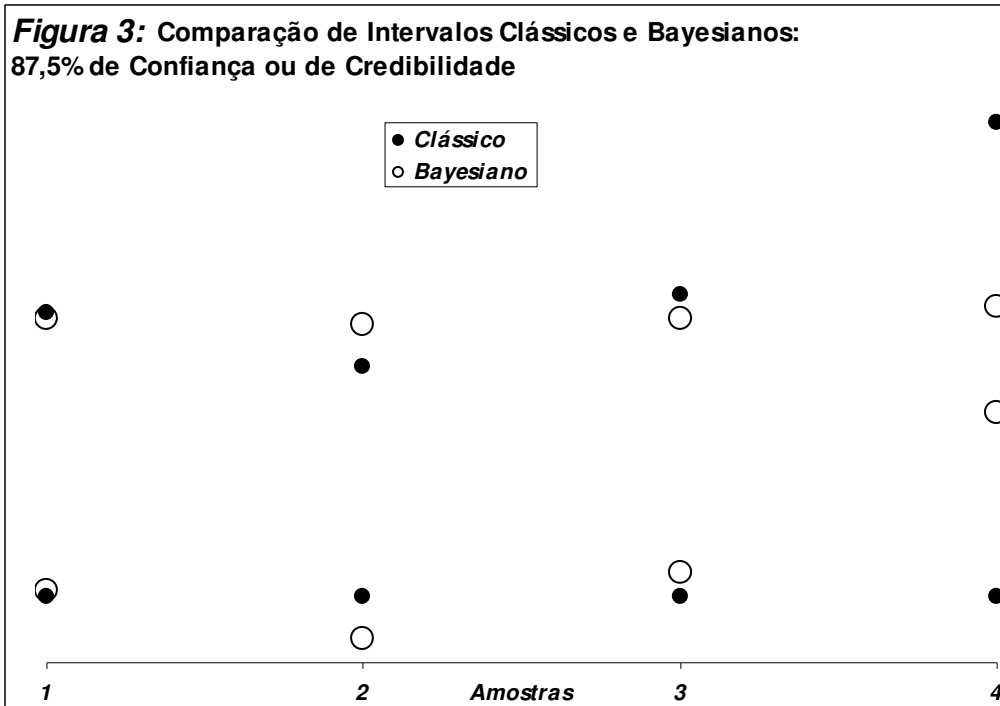
O Exemplo 1.4.1 apresentou o caso onde intervalos de confiança efetivamente observados produzem uma incoerência de ordem prática. Quanto melhor a amostra for, pior é a precisão produzida pelo método. Em outras palavras, o intervalo de confiança aumenta seu comprimento, diminuindo a precisão, se uma nova amostra for mais informativa - o máximo **M** mais distante do mínimo **m**. O método de intervalos de credibilidade discutido, caminha na direção oposta, isto é quanto mais informativa a amostra, maior a precisão obtida. A Figura 3 apresenta os dois tipos de intervalos para as quatro amostras discutidas. Com a amostra 4, a mais informativa, a incoerência do método padrão fica evidente. Com uma amostra ainda mais informativa; por exemplo, se $m \cong 1,1$ e $M \cong 2,1$, qualquer pessoa deveria concluir que certamente $z \cong 1,55$. Mas, o intervalo com confiança de **87,5%** seria $[1,1;2,1]$, produzindo uma inferência quase totalmente não informativa.

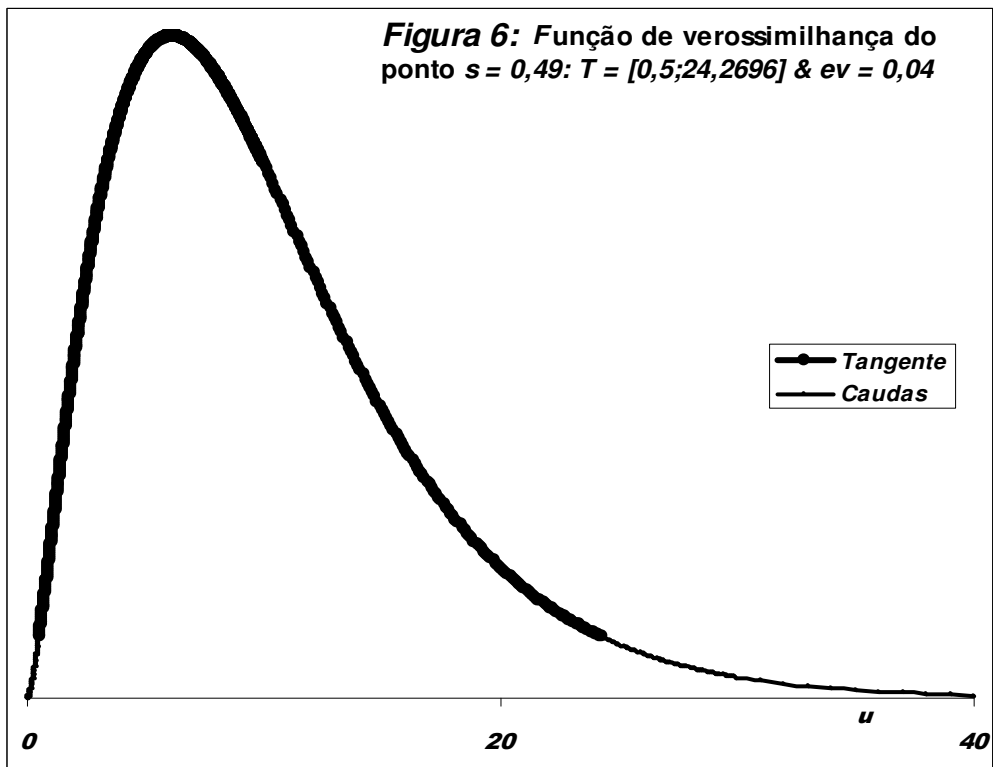
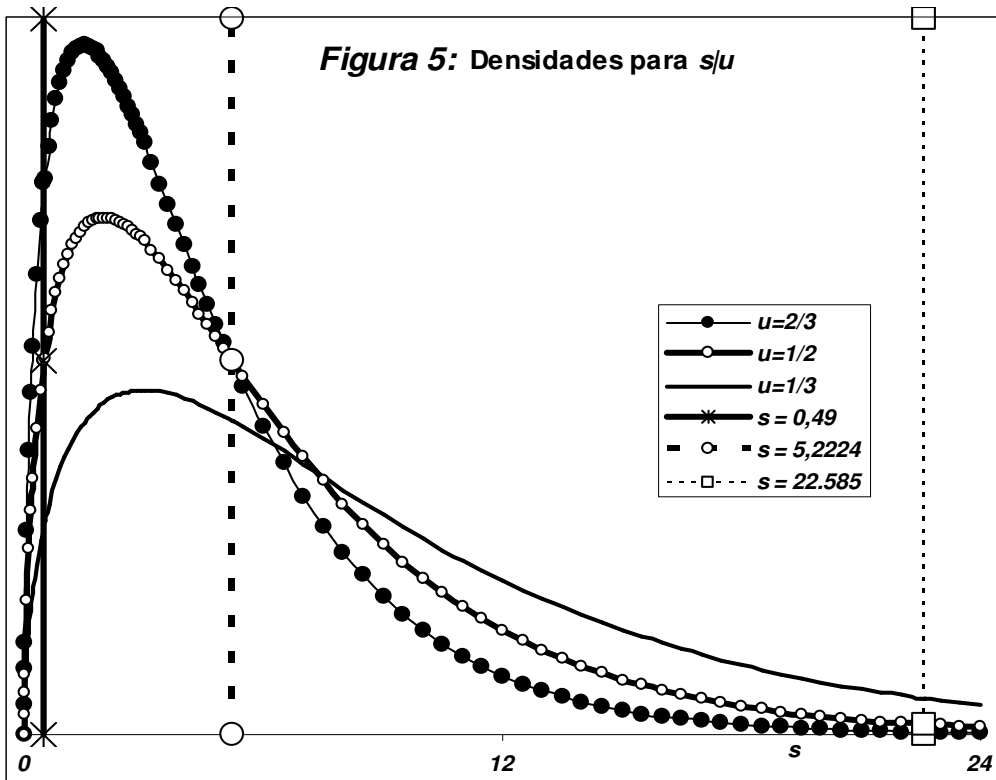
Como na Figura 1, a Figura 4 apresenta possíveis densidades normais para x , pertencentes ao conjunto de densidades que formam o modelo amostral de $x|u$. Note-se que todas as três observações possuem maior densidade quando $u = 1$, um ponto da hipótese alternativa. Como os procedimentos de inferência são obtidos a partir da estatística suficiente s , a Figura 5 apresenta alternativas das densidades amostrais de $s|u$.

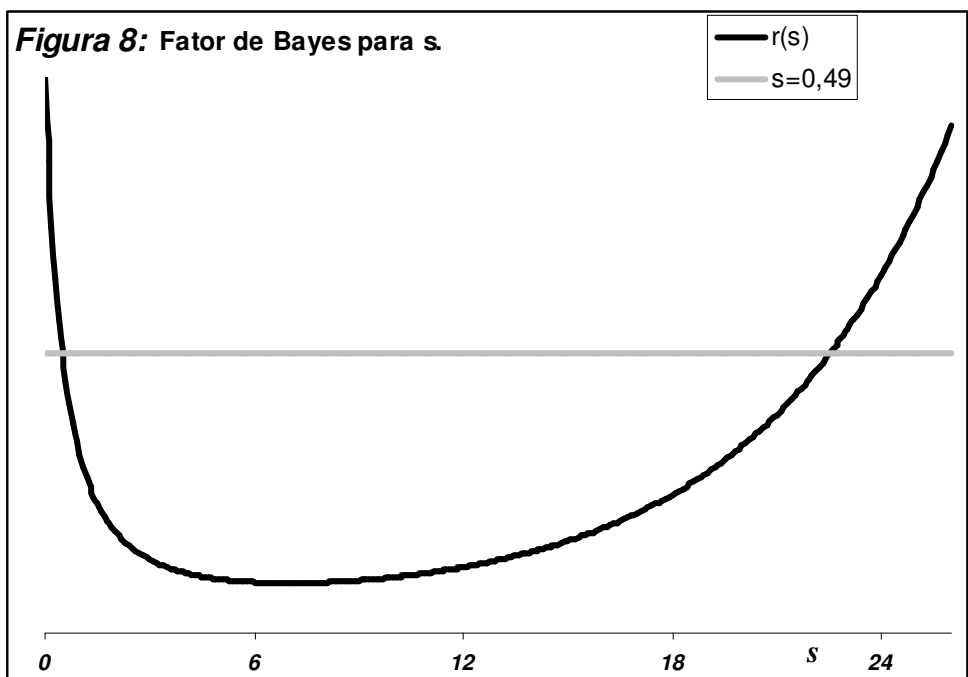
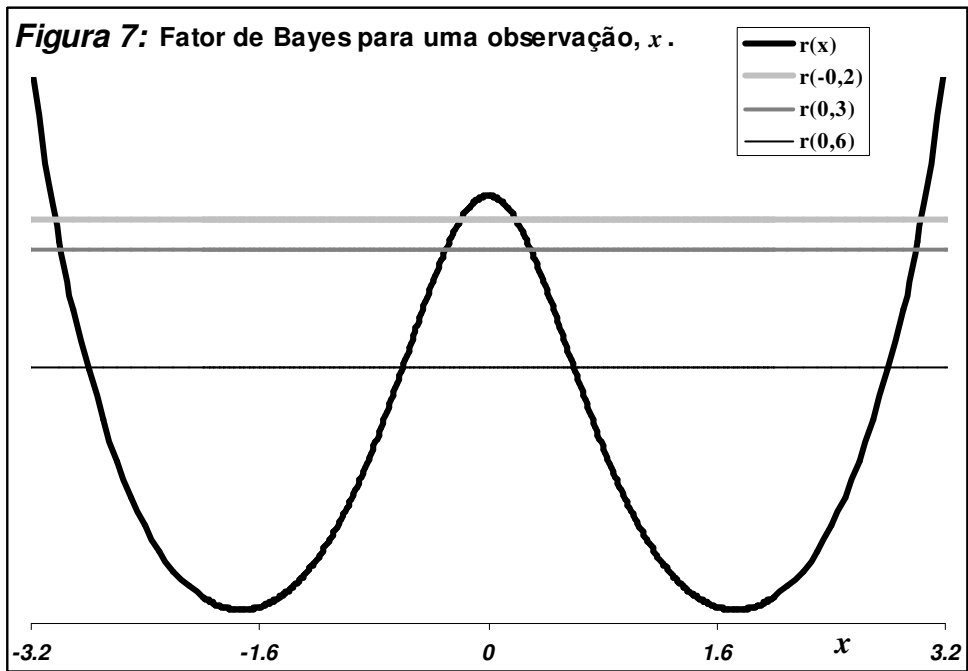
A Figura 6 apresenta a função de verossimilhança normalizada por sua área. As caudas e o conjunto tangente são ressaltados. O conjunto tangente é definido em relação à hipótese **H**: $u = 1/2$.

A figura 7 apresenta a razão de verossimilhança $r(x)$ para amostras unitárias. Finalmente, para amostras de três observações, a Figura 8 é a razão de verossimilhança $r(s)$ para a estatística suficiente s .









1.7. CONSIDERAÇÕES FINAIS

Este capítulo nasceu de desafios e incentivos que o autor recebeu de seus colegas. Encontros informais, durante a participação em bancas ou seminários, permitiram ao autor expor idéias e discussões que aparentemente surpreendiam os presentes.

O professor Dev Basu, supervisor do autor durante o programa de doutorado, foi o mestre dos contra-exemplos. Muitos dos seus artigos tinham a perspectiva do desafio intelectual sobre o entendimento dos fundamentos da estatística. Dev e Kemp - Oscar Kempthorne outro ilustre professor - criavam um ambiente de conflito que muito auxiliou o desenvolvimento da estatística. A impressão é a de que não concordavam entre si, mas certamente se desafiavam mutuamente para o desenvolvimento das idéias que tão brilhantemente defenderam. Interessante é ressaltar que a admiração do autor pelos professores Dev e Kemp, bem como por I J Good e Bruno De Finetti, talvez se deva ao fato de todos, sem exceção, terem feito história por suas controvérsias e pela apresentação de alternativas metodológicas coerentes.

O presente capítulo é a tentativa do autor de criar um ambiente de conflito de idéias. O objetivo é impedir que métodos sejam usados apenas porque foram criados. O desafio é fazer com que todos saibam defender, com conhecimento dos fundamentos, os métodos que usam.

2. A likelihood Approach to Diagnostic Tests

Basilio & Carlos Alberto de Bragança Pereira

2.1. INTRODUÇÃO

The likelihood function plays a central role in parametric statistical inference since it contains all the information in the observed data. It is used in both frequentist antagonistic approaches, Fisherian and Neyman-Pearson-Wald (NPW), but in neither methodology it is the main tool. On the other hand, the only experimental source of information to the ones fond of Bayesian methodology is exactly the likelihood function. Hence, for Bayesians using uniform (proper or improper) densities, the only relevant tool for analysis is the likelihood function.

Most Bayesians and frequentists may disagree with the views presented here (see Basu, 1988 and Kempthorne & Folks, 1971) since they are close to the ideas described by Ronald Fisher in his last and controversial book, Fisher (1956). We believe that it is closer to the Bayesian perspective than to the standard frequentist approaches. A recent revival of interest in the likelihood approach is in action; see King (1998), Lindsey (1996), Pawitan (2001), Royall (2001), Severini (2000), and Sprott (2001) for instance. The site <http://www.stat.unipd.it/LIKASY/biblio.html> presents a comprehensive list of references.

A brief history is presented in Section 2. The likelihood perspective is presented and discussed in Sections 3 and 4. In Section 5, diagnostic tests results are compared to the contingencies of statistical results of the different views. In Section 6 we present an index for the diagnostic ability of a clinical test. Section 7 contains the likelihood view of a diagnostic test with a graphical illustration. Finally, in Section 8 we present a real example to illustrate the ideas discussed in previous sections.

2.2. STATISTICAL TESTS: A BRIEF HISTORY

Some of the material of this section can be found in Stern (2002). The idea of significance tests was proposed by Fisher, who introduced the *p-value* as an index of agreement between the data and the null hypothesis: the greater the *p-value*, the greater the evidence in favor of the null hypothesis. A *p-value* of 5% is commonly used as a standard threshold for deciding against H_0 ($p < 0.05$) or in favor of H_1 ($p \geq 0.05$). However, we strongly support the idea that the choice of the threshold should depend on the problem currently faced by the scientist, the sample size, and the amount and type of information being collected. This is in fact the idea of significance tests as prescribed by Cox (1977) and Kempthorne (1976).

The subjective judgment of an observed *p-value* to decide against or in favor of H_0 led Neyman and Pearson (1936) and Wald (1939 & 1950) to proposing the theory of Test of Hypotheses. This theory, contrarily to Fisher's significance tests, was designed to replace the subjective judgment of the strength of evidence in favor of the null hypothesis, provided by a *p-value* judgment, with an objective decision-theoretical approach. By fixing, in advance, the Type I error rate, α , and minimizing the Type II error rate, β , the number of wrong decisions, made over many different repetitions of the experiment, would be limited. This may generate some controversy since only in very few medical applications repetitions are possible.

Originally, the NPW theory required the specification of single point null, H , and alternative, A , hypotheses. By fixing Type I and Type II error rates, the sample size could be determined. Sample size determination is an area in which NPW theory has been appropriately used in medicine (and also in industrial quality control), although a confuse mixture of the Fisher and NPW approaches to hypothesis testing may be found in the medical literature. Statements such as “ p -values smaller than 5% were considered statistically significant”, without specifying the alternative hypothesis and the type II error rate, are common. It is usual to have a table with p -values and intervals obtained by summing and subtracting twice the sample standard error from the sample mean.

Jeffreys (1939) attacked the problem under a Bayesian perspective. Let \mathbf{x} denote the observations, π and $\pi(\mathbf{x})$ the prior and posterior probabilities for H . Alternatively the corresponding probabilities for A are $1-\pi$ and $1-\pi(\mathbf{x})$. Defining the prior and posterior odds by $\rho = \pi(1-\pi)^{-1}$ and $\rho(\mathbf{x}) = \pi(\mathbf{x})[1-\pi(\mathbf{x})]^{-1}$, Jeffreys proposed to look at the posterior odds, also called Bayes Factor, as the index of evidence in favor of H .

In the case of single point hypotheses, let $f_H(\mathbf{x})$ and $f_A(\mathbf{x})$ be the two alternative densities being compared. The likelihood ratio is $R(\mathbf{x}) = f_H(\mathbf{x})/f_A(\mathbf{x})$. Hence, one can easily prove that $\rho(\mathbf{x}) = \rho R(\mathbf{x})$. Also, for $\pi=1/2$ we would have $\rho(\mathbf{x}) = R(\mathbf{x})$. Hence, for the case of single point hypotheses, judging H based on the likelihood ratio corresponds to a Bayesian judgment with very particular prior choices. On the other hand, recall that the likelihood ratio is the function used by the Neyman-Pearson theorem of optimal decision. Also, note that one can use $R(\mathbf{x})$ to order the sample space, Dempster (1997), Montoya-Delgado et al. (2001) and Pereira & Wechsler (1993). If the computation of the p -value were performed under this ordering, the alternative hypothesis would be taking into consideration. As one may see, the three methods have their conclusions based on the likelihood ratio, $R(\mathbf{x})$.

Real controversial problems emerge with the consideration of composite hypotheses. Many of the practical problems in medicine involve sharp null hypotheses. That is, the dimension of the subspace where H is defined is smaller than the dimension of the subspace where A is defined. Let us consider the well-known standard problem of the test for independence in a 2 by 2 contingency table. Let C_1 and C_2 be two populational characteristics and $\mathbf{x}=(x_{11},x_{12},x_{21},x_{22})$ be the vector of the sample frequencies for the respective combination of the levels of categories C_1 and C_2 . The parameter space associate with this experiment is the simplex

$$\Theta = \left\{ (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) \mid \theta_{ij} > 0, \sum_{i,j=1}^2 \theta_{ij} = 1 \right\}$$

and the null hypothesis is defined by the subset

$$\Theta_H = \{ (pq, p(1-q), (1-p)q, (1-p)(1-q)) \mid 0 \leq p, q \leq 1 \} \subset \Theta.$$

Note that the two hypotheses are composite. The sets that define the null and the alternative hypotheses, Θ_H and $\Theta_A = \Theta - \Theta_H$, have different dimensions, i. e.,

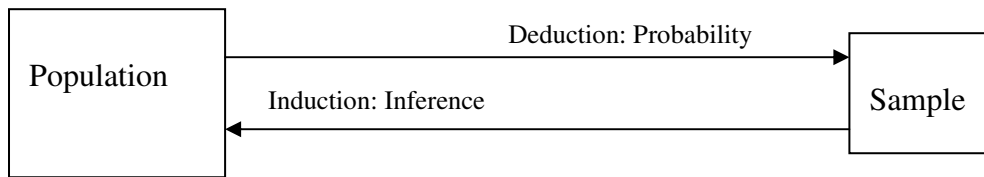
$$\dim(\Theta) = 3 > \dim(\Theta_H) = 2.$$

Letting $f(x|\theta)$ denote the likelihood function, frequentists will define $S_H(x)$ and $S_A(x)$ as the suprema of $f(x|\theta)$ under Θ_H and Θ_A , respectively. The profile likelihood ratio is defined as $PR(x) = S_H(x)/S_A(x)$. Bayesians, on the other hand, in addition to the prior probabilities for H and A , namely $\pi = P\{H\}$ and $1-\pi = P\{A\}$, define densities over Θ_H and Θ_A . Considering these densities as weighing systems and taking the weighted likelihood averages, $M_H(x)$ and $M_A(x)$, under Θ_H and Θ_A respectively, they define the Bayes Factor $BF(x) = \rho MR(x)$ where $\rho = \pi(1-\pi)^{-1}$ is the prior odds and $MR(x) = M_H(x)/M_A(x)$ is the weighted likelihood ratio. Irony et al. (2000) uses this approach for a Bayesian version of the McNemar test for also comparing two composite hypotheses of different dimensions in a 2×2 contingency table. NPW (Jeffreys's Bayesian) approach for hypothesis testing consists in the evaluation of $PR(x)$ ($BF(x)$). The Fisher approach for testing independence is a modification based on a conditional distribution of the data in the basic cells of the table given the marginal cells. It does not seem appropriate to consider that the marginal cells are known before the data were observed.

The fourth approach to hypothesis tests is that of (pure) likelihood, which is described in the next section.

2.3. LIKELIHOOD APPROACH

The deductive nature of probability versus the inductive nature of statistical inference is clearly reflected in the dual concepts of probability distributions and likelihood (Lindsey, 1995 and Fenders, 1999).



Given a probability model and the corresponding parameter values, we may calculate the probabilities associated to all possible observations, x . Suppose that before observing the performance of the next 10 patients that will be submitted to a drug known to have efficacy of 60%, a doctor writes his probability model function for “the number of recovered patients, X ” as:

$$f(x | 0.6) = Pr\{X = x | \theta = 0.6\} = \binom{10}{x} (0.6)^x (0.4)^{10-x}$$

The probability of having 7 recovered patients is $f(7|0.6) = Pr\{X=7|\theta=0.6\} = 0.215$. Note that $f(x|\theta)$ is a function of two variables: x , the observation, and θ , the parameter. For fixed θ , f is a probability function of x and for fixed x , f is a function of θ called likelihood function associated to the observed value, x . Suppose that we observe 7 success and 3 failures for this sample of 10 patients. The likelihood function is

$$L(\theta | X = 7) = Pr\{X = 7 | \theta\} = \binom{10}{7} \theta^7 (1-\theta)^3 = 120\theta^7 (1-\theta)^3$$

In order to illustrate the differences between probability and likelihood functions, in Figure 1 we present the corresponding probability functions for $\theta=0.6$ and for $\theta=0.3$, while in Figure 2 we present the likelihood functions for $x = 7$ and for $x=2$.

Note that the two probability functions in Figure 1 are discrete. Since the parameter space Θ is the interval $[0;1]$, the likelihood functions depicted in Figure 2 are continuous. A statistical model has two arguments, the possible observations and the possible values of the parameter. The likelihood function is not a probability density function. However, dividing it by its integral over the parameter space (whenever this integral exists), the resulting normalized likelihood is a probability density over Θ , and corresponds to the Bayesian posterior density under a uniform prior. Areas under this curve define probabilities of subsets of the parameter space.

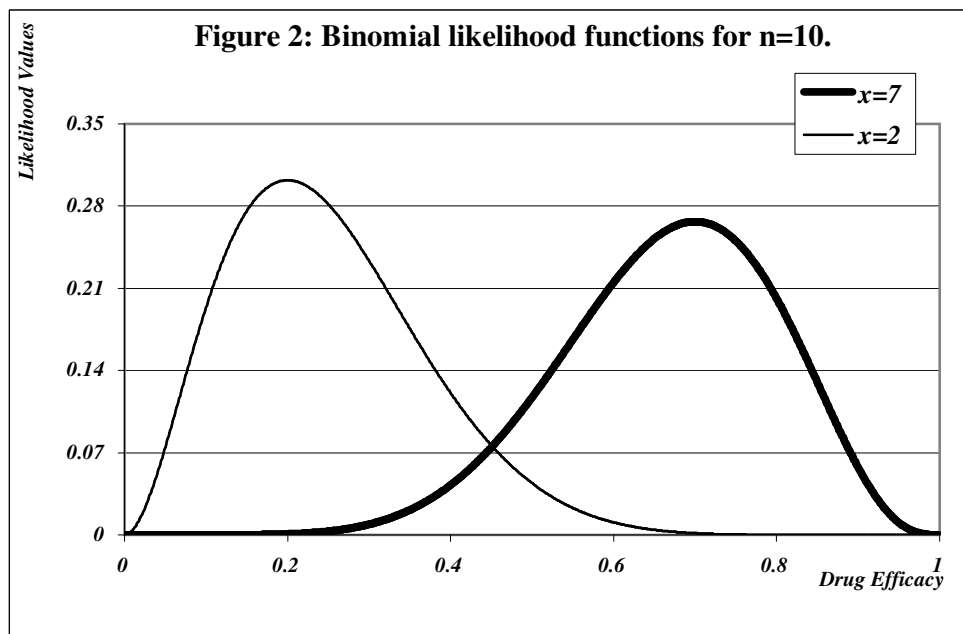
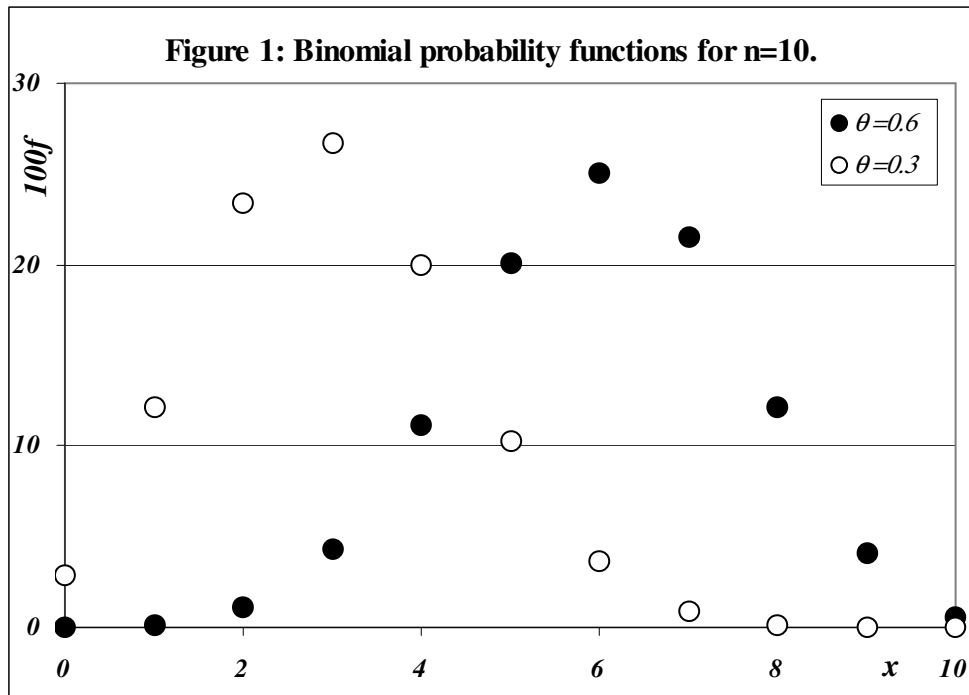
The likelihood function, L , induces an ordering of preferences about the possible parameter points. Note that this order is not changed if a proportional function is defined. This means that we can divide L by any constant without modifying the conclusions about parameter point preferences. We can divide L by its integral obtaining the normalized likelihood, the Bayesian way, or divide it by the maximum value of L whenever it exists, obtaining what we call relative likelihood. Comparing two parameter values, we would say that the one with higher (normalized or relative) likelihood is more plausible than the other.

An important feature of the Likelihood approach is that it is independent of stopping rules. That is, it does not violate the likelihood principle, Barnard (1949), Berger & Wolpert (1984) and Birnbaum (1962). For instance, suppose that another doctor in another clinic decided to start his analysis only when he obtain 3 failures, i. e., 3 patients that do not recover. As soon he obtained his 3rd failure, corresponding to the 10th patient, he realizes that he had 10 patients with 7 successes and 3 failures. Although he has the same results as his colleague, the underlying statistical model is completely different but his (normalized) relative likelihood is equal to the one obtained from the previous models. Here the probability model is a negative binomial distribution. That is, the random variable is the number Y of failures to be observed since the number of failures k was fixed in advance. The model here is given by

$$Pr\{Y = y \mid \theta\} = \binom{y+k-1}{k} \theta^y (1-\theta)^k$$

For the sample with $k=3$ and $y=7$, the likelihood is proportional to the one illustrated in Figure 2. Figure 3 shows the negative binomial probability distributions for $k=3$, $\theta = 0.6$ and $\theta = 0.3$.

Note that for both Figures 1 and 3, the probabilistic models, Binomial and Negative Binomial, have their sample space well defined since the stopping rules were defined previously. However there are many cases in medical statistics where the sample space is not well defined. For instance, suppose that a doctor wants to write a paper and decides to look at the data he has collected up to that moment. In this case, neither the sample size nor the number of success (or failures) was fixed a priori. However, if he had observed 7 recoveries in 10 patients, his likelihood would be $\theta^3(1-\theta)^7$, which is proportional to both observed Binomial and Negative Binomial likelihoods. Hence, in all 3 cases, the relative (normalized) likelihoods are exactly the same and then the inference would be the same as prescribed by the likelihood principle. We emphasize that the normalized likelihood for the example of 3 failures and 7 successes is a beta density with parameters $a=4$ and $b=8$. The relative likelihood is the beta density divided by the density evaluated at its mode.



In Figure 4 we illustrate the relative likelihood for 3 failures and 7 successes, with a solid line intercepting it at points with plausibility equal to $1/3$ (relative to the maximum) and a dotted line at points with plausibility equal to 0.8057 . Recall that the maximum of the likelihood function is attained at $\theta=0.7$. Also, at $\theta=0.6$, the suggested drug efficacy, the plausibility is 0.8057 . Note that both $\theta_1=0.4681$ and $\theta_2=0.8777$ have plausibility equal to $1/3$. Any parameter point inside (outside) the interval

$$I(1/3)=(.4681;.8771)$$

has plausibility larger (smaller) than $1/3$. If one uses the normalized likelihood as the posterior density, the (posterior) probability that the unknown parameter θ lies in $I(1/3)$ is equal to 0.8859 . That is, $I(1/3)$ is a credible interval for θ with credibility 88.59% . This probability (or credibility) is calculated by computing the area under the curve limited by the vertical segments at 0.4681 and 0.8771 divided by the total area under the curve.

Consider the other point, $\underline{\theta}_0 = 0.7886$, with the same plausibility as the suggested drug efficacy, $\theta_0 = 0.6$. These two points have plausibility equal to 0.8057 and the interval $I(0.8057) = (0.6000; 0.7886)$ has credibility 51.65% . Considering now $\theta_0 = 0.4$, the corresponding parameter point with the same plausibility is $\underline{\theta}_0 = 0.9124$. These points have plausibility equal to 0.1592 and the interval $I(0.1592) = (0.4000; 0.9124)$ has credibility 95.90% .

Observing the low (high) probability of having a parameter value with more plausibility than 0.6 (0.4), we would say that the hypothesis $H: \theta = 0.6$ ($H: \theta = 0.4$) should be not rejected (rejected). We suggest that the credibility of the interval $(\theta_0, \underline{\theta}_0)$ may be interpreted as an index of evidence against the null sharp hypotheses $H: \theta = \theta_0$ or $H: \theta = \underline{\theta}_0$. The probability of the complement of this credibility interval is an index (like a p -value) of evidence in favor of H , see Pereira & Stern (1999) and Madruga et al. (2003) for more on this measure of evidence. For the two cases presented here, the evidence in favor of H is 48.35% for $H: \theta = 0.6$ and 4.10% for $H: \theta = 0.4$.

We end this section by stating a rule to be used by Pure Likelihood followers.

Pure Likelihood Law: If the relative likelihood function of two points, θ_0 and θ_1 , satisfy $RL(\theta_0) < (>) RL(\theta_1)$, we say that θ_0 is more (less) plausible than θ_1 . We say they have the same plausibility if equality of the likelihood functions holds. For single point hypotheses $H: \theta = \theta_0$ versus $A: \theta = \theta_1$, if $RL(\theta_0) < (>) RL(\theta_1)$, we reject (accept) H . The strength of evidence of the data x in favor of H against A is measured by the likelihood ratio, $LR(\theta_0; \theta_1) = RL(\theta_0)/RL(\theta_1)$.

For the example above, we have

$$LR(0.6; 0.7) = 0.8057 \text{ and } LR(0.6; 0.4) = 5.0625.$$

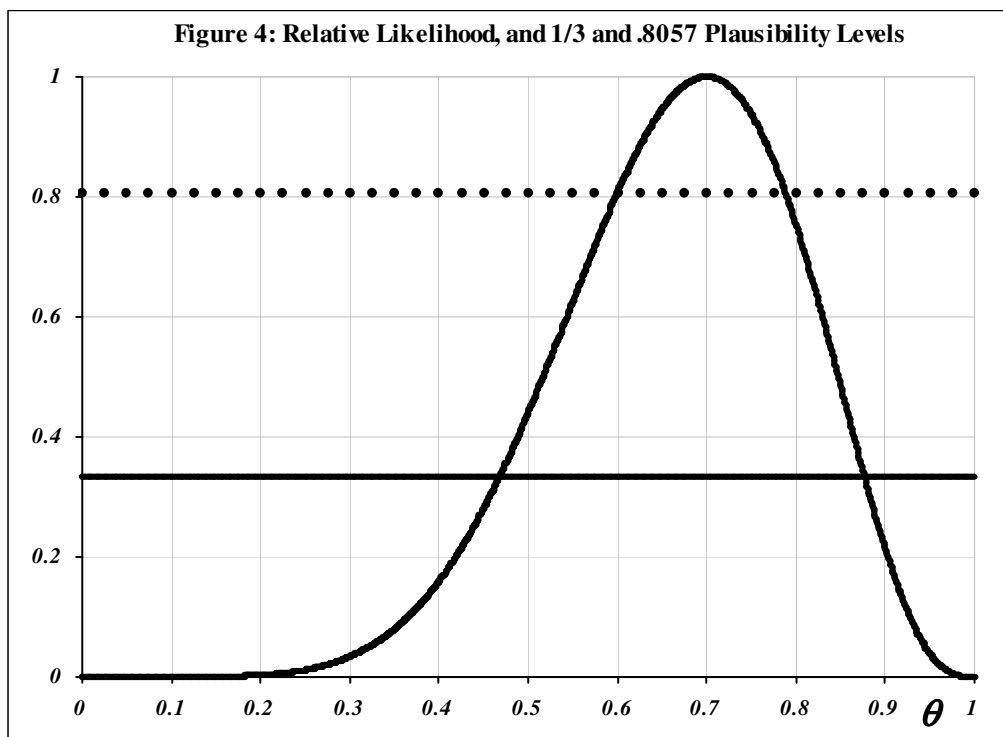
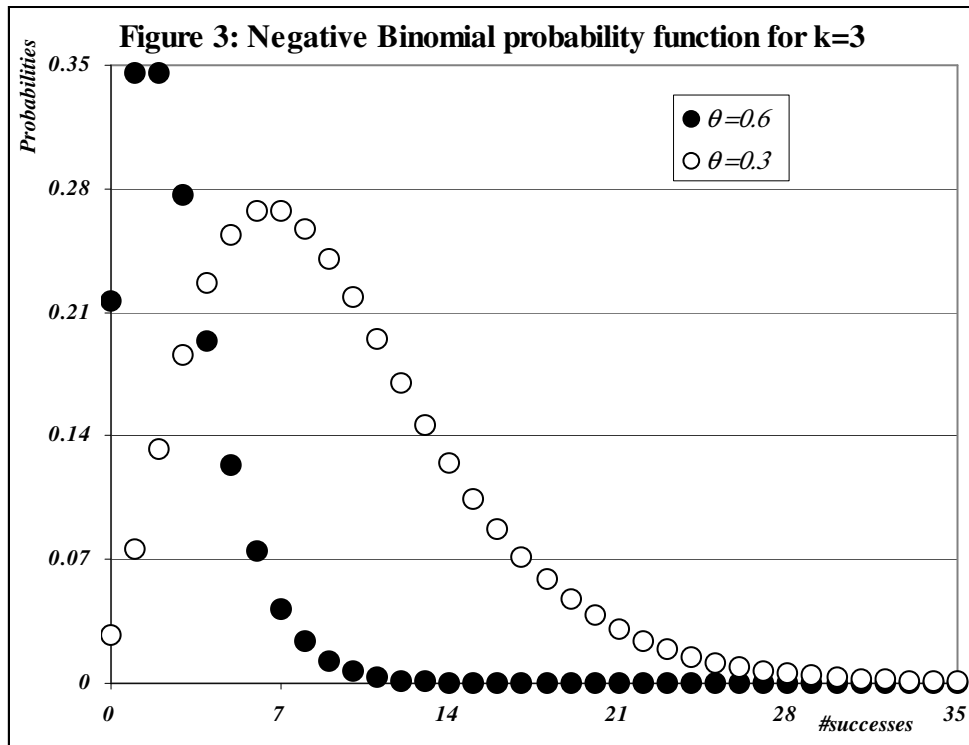
2.4. LADDER OF UNCERTAINTY AND CONTROVERSIES

Tests of hypotheses are decision procedures based on judgments and one can only judge something in relation to the alternatives. The concept of statistical evidence of some data, x , in favor or against some hypothesis must be relative in nature. We should not talk about evidence for or against H without mentioning the alternative A . Pereira & Wechsler (1993) shows how to build a p -value that takes the two antagonistic hypotheses into consideration.

An implication of the pure law of likelihood is that: “uncertainty about x given θ ” and “statistical evidence in x about θ ” have different mathematical forms. The statistical model is based on a trinity of mathematical elements: the sample space X , the parameter space Θ , and a function $f(\cdot, \cdot)$ of two arguments $(x, \theta) \in X \times \Theta$. For every fixed

$\theta \in \Theta$, $f(\cdot|\theta)$ is a probability (density) function on X and for every fixed $x \in X$, $f(x|\cdot) = L(\cdot|x)$ is the likelihood function. The following sets characterize the statistical model:

- i) $\mathcal{F} = \{f(x|\theta) | x \in X \text{ and } \theta \in \Theta\}$ is the overall statistical model,
- ii) for $\forall \theta \in \Theta$, $\mathcal{F}_\theta = \{f(x|\theta) | x \in X\}$ are the probability models, and
- iii) for $\forall x \in X$, $\mathcal{F}_x = \{f(x|\theta) | \theta \in \Theta\}$, are the likelihood functions.



Uncertainty is measured by probabilities, \mathcal{I}_θ , and evidence is measured by the likelihood, \mathcal{I}_x . This is a critical insight: the measure of the strength of evidence and the frequency with which such evidence occurs are distinct mathematical quantities (Blume 2002). Royall (1997) clearly explains alternative areas of Statistics where these concepts appear. Suppose a patient has a positive result in a diagnostic test, the physician might draw one of the following conclusions:

1. The person probably has the disease,
2. The person should be treated for the disease,
3. The test results are evidence that the person has the disease.

These possible attitudes front the tests results may represent, respectively, answers to different questions:

- 1'. What should I believe?
- 2'. What should I do?
- 3'. How should I interpret this body of observation as evidence about having the disease against not having the disease?

These questions involve distinct aspects of statistical methods, namely: frequentist or Bayesian inference, decision theory and, lastly, interpretation of statistical data as containing evidence, the significance test of hypothesis.

The correctness of the answer for the first question requires, the additional information of the behavior of the test in other (exchangeable) patients or the personal opinion about the probability of the disease before the test (prior probability). For the second question, in addition to the requirements of the first, one also needs knowledge about the costs or utilities of the decisions to be made. Only the third one does not require additional information other than data. Biggerstaff (2000) consider these arguments to suggest that the role of the likelihood in Statistics is equivalent to the role of diagnostic tests used in Medicine.

Royall (1977) also discusses a possible paradox in the use of the pure likelihood approach through the following example:

“ We pick a card at random out of a deck of 52 cards and observe an ace of clubs. Then consider two alternative hypotheses H : it is a deck with 52 aces of clubs or A : it is a standard deck of cards. The likelihood ratio of H against A is 52. Some find this disturbing. What this results shows is that this strong evidence is not strong enough to overcome the prior improbability of H . A Martian faced with this problem would find H most appealing.”

Clearly, the Martian's ignorance about card decks does not permit him to use the tools used by both Bayesian and frequentist statisticians. These people may achieve stronger results than pure likelihood statisticians do, but at the price of more assumptions in their applications. Pawitan (2000) tentatively tried some reconciliation among the different approaches using the Fisherian idea of ladder of uncertainty. It remains to be proved that his ideas will succeed in Statistics by means of practical applications.

2.5. DIAGNOSTIC TESTS AND STATISTICAL VERDICTS

The inadequacy in relying only and strongly on p -values in medicine has been widely emphasized in recent years. Worst yet, is the lack of understanding of what p -

values are. In this section we present quantities that may be of more interest to medicine than the p -values are. For more discussion on the subject we refer to Diamond & Forrest (1983) and Pereira & Louzada-Neto (2002).

We use the following notation: D^+ = Disease, D^- = No Disease, T^+ = Positive test result and T^- = Negative test result. For the populational parameters let $N(++)$ be the frequency of units in category ($D^+ T^+$), $N(+ -)$ the units in category ($D^+ T^-$), $N(- +)$ the units in category ($D^- T^+$), and $N(--)$ the units in category ($D^- T^-$). $N(+ \bullet)$ denote the number of units with the disease, $N(- \bullet)$ the number of units without the disease, $N(\bullet +)$ the number of units with positive test result, and $N(\bullet -)$ the number of units with negative test result. Table 1 summarizes this notation.

Table 1: Populational frequencies related to a Diagnostic Test

| Test Results | State of Health | | Total |
|-----------------|-----------------|-----------------|------------------------|
| | Absense, D^- | Presence, D^+ | |
| Negative, T^- | $N(- -)$ | $N(+ -)$ | $N(\bullet -)$ |
| Positive, T^+ | $N(- +)$ | $N(+ +)$ | $N(\bullet +)$ |
| Total | $N(- \bullet)$ | $N(+ \bullet)$ | $N=N(\bullet \bullet)$ |

The following quantities are of great interest for physicians evaluating patients. For a randomly selected unit from the population we define

- Sensitivity is the conditional probability of responding positively to the test given that the patient has the disease, i. e., $S = Pr\{T^+ | D^+\} = N(++)/N(+ \bullet)$.
- Specificity is the conditional probability of responding negatively to the test given the absence of the disease, i.e., $E = Pr\{T^- | D^-\} = N(--)/N(- \bullet)$.
- Prevalence is the probability that the patient has the disease, i. e., $\pi = Pr\{D^+\} = N(+ \bullet)/N$ Alternatively, $1 - \pi = Pr\{D^-\}$ is the probability that the patient does not have the disease.
- Test Positivity and Test Negativity are the probabilities of positive and negative test results, i. e., $\tau = Pr\{T^+\} = N(\bullet +)/N$ and $1 - \tau = N(\bullet -)/N$.
- Diagnostic Parameters are the posterior probabilities of the states of a patient given the response to the clinical test:
 PPV: Positive Predictive Value is the conditional probability of presence of disease given positive test response: $\pi(T^+) = Pr\{D^+ | T^+\} = N(++)/N(\bullet +)$ and
 NPV: Negative Predictive Value ($\bullet -$) is the conditional probability of absence of disease given negative test response: $1 - \pi(T^-) = Pr\{D^- | T^-\} = N(--)/N$.

The quantities of higher interest in clinical practice are the predictive values, PPV and NPV. Using Bayes formula, we obtain important relations between the predictive values and the other terms of the model, namely

$$PPV = \pi(T^+) = \frac{\pi S}{\pi S + (1-\pi)(1-E)} = \left\{ 1 + \left[\left(\frac{\pi}{1-\pi} \right) \left(\frac{S}{1-E} \right) \right]^{-1} \right\}^{-1} \text{ and}$$

$$NPV = 1 - \pi(T^-) = \frac{(1-\pi)E}{(1-\pi)E + \pi(1-S)} = \left\{ 1 + \left[\left(\frac{1-\pi}{\pi} \right) \left(\frac{E}{1-S} \right) \right]^{-1} \right\}^{-1}$$

Denoting the likelihood ratio for positive results by $LR(+)=S/(1-E)$, the likelihood ratio for negative results by $LR(-) = (1-S)/E$ and the prevalence odds by $\rho = \pi/(1-\pi)$ we have:

$$PPV = \left\{ 1 + [\rho LR(+)]^{-1} \right\}^{-1} \text{ and } NPV = \left\{ 1 + \rho LR(-) \right\}^{-1}$$

Let ρ denote the prior odds in favor of the disease and $1/\rho$ the prior odds against it; similarly, let the posterior odds in favor and against the disease be $\rho(+)=PPV/(1-PPV)$ and $\rho(-)=NPV/(1-NPV)$. We have the following interesting formulas:

$$\begin{aligned} \rho(+) &= \frac{\pi}{1-\pi} \frac{S}{1-E} = \frac{\text{prevalence}}{1-\text{prevalence}} \frac{\text{sensitivity}}{1-\text{specificity}} = \\ &= \text{prior odds} \times \text{likelihood ratio for +} = \rho LR(+) \text{ and} \end{aligned}$$

$$\begin{aligned} \rho(-) &= \frac{1-\pi}{\pi} \frac{E}{1-S} = \frac{1-\text{prevalence}}{\text{prevalence}} \frac{\text{specificity}}{1-\text{sensitivity}} = \\ &= (\text{prior odds} \times \text{likelihood ratio for -})^{-1} = [\rho LR(-)]^{-1}. \end{aligned}$$

Other important relations are as follows:

$$PPV = \frac{\rho(+)}{1+\rho(+)}; \quad NPV = \frac{\rho(-)}{1+\rho(-)}; \quad \rho(+)=\frac{PPV}{1-PPV}; \text{ and } \rho(-)=\frac{NPV}{1-NPV}.$$

The important question for a physician working with diagnostic tests is to decide what to do when the result is positive (or negative). In fact, measures of sensitivity and specificity, when available, would be of great help to him since they may yield other valuable quantities, see Diamond & Forrest (1983) and Pereira & Louzada-Neto (2002).. Note that if there is a big change from prior to posterior odds the test will be considered of great value. In the next section we discuss a way of defining diagnostic power of clinical evaluations. This index is of great value to state an order of preference in a set of clinical procedures

2.6. DIAGNOSABILITY

In this section we discuss the diagnostic power of a medical test, T . To evaluate the diagnostic ability of a test T , we should focus on the change from ρ to $\rho(+)$ and from $1-\rho$ to $\rho(-)$. This is related with the weight of evidence provided by T^+ (T^-) in favor of D^+ (D^-) and denoted by $w^+ = w(D^+; T^+)$ [$w^- = w(D^-; T^-)$]. Good (1968) showed that the function w , to follow reasonable requirements, ought to be an increasing function of the

ratio of posterior to prior odds, or, equivalently, an increasing function of the likelihood ratio. That is, w^+ and w^- must be increasing functions of $\rho(+)/\rho = LR(+)=S/(1-E)$ and $\rho\rho(-)=[LR(-)]^{-1}=E/(1-S)$, respectively.

The usual odds ratio (cross-product in the context of contingency tables), useful in measure association, is simply

$$R = LR(+)/LR(-) = \frac{SE}{(1-S)(1-E)} = \frac{(PPV)(NPV)}{(1-PPV)(1-NPV)}$$

As we will see in the sequel, the larger R is, the better the test for detecting disease D , i. e., the better its diagnosability.

As a consequence of the requirement of additivity of information, Good (1968) proves that the weights of evidence, w^+ and w^- , are the natural logarithms of $LR(+)$ and $LR(-)$. Good (1950) also points out that the expected value of the weight of evidence is more meaningful than the likelihood ratio. Hence, the measure of the ability of a medical test, T , to discriminate in favor of D^+ (D^-), given that the true state of nature is D^+ (D^-) is the conditional expectation of w^+ (w^-) given S , E and the state of the patient, D^+ or D^- . We denote these conditional expectations by ε^+ and ε^- . Finally, the diagnosability of T is given by $\Delta = \varepsilon_+ + \varepsilon_-$. Let us write these formulas explicitly:

Weight of Evidence

- a) in favor of D^+ , $w(D^+; T^+) = w^+ = \ln(LR(+))$ and $w(D^+; T^-) = -w^- = \ln(LR(-))$.
- b) in favor of D^- , $w(D^-; T^+) = -w^+ = -\ln(LR(+))$ and $w(D^-; T^-) = w^- = -\ln(LR(-))$.

Average Weight of Evidence

- c) in favor of $D^+ = \varepsilon_+ = Sw^+ - (1-S)w^-$
- d) in favor of $D^- = \varepsilon_- = Ew^- - (1-E)w^+$

Diagnosability Index

- e) $\Delta = \varepsilon_+ + \varepsilon_- = (S+E-1)\ln R$.

We would like to call the attention to the fact that all these indices depend strongly on the values of many parameters that are in fact not completely known. Usually the prevalence, the sensitivity and the specificity have to be estimated with sample data. Pereira & Pericchi (1990) introduced Bayesian techniques for such purposes. They also consider the case where a set of clinical tests are observed in the same subject and show how a combination of them improves the diagnosability of the medical procedure. In a predictivist context, Pereira (1990) and Pereira & Barlow (1990), show that if we look at a particular patient, the computation of her/his posterior probability of having the disease simplifies significantly the diagnostic calculus.

In order to decide if a new (possibly expensive) test must be considered in lieu of some other test, one must collect, observe, and analyze a new sample. Usually the size of a sample of patients, known to have the disease, is the number of patients under treatment at the clinic and the test is applied to all possible patients. A control group of units without the disease is also selected and tested after all ethical procedures have been fulfilled. Based on the two samples, S and E are estimated. Estimates of $LR(+)$, $LR(-)$, and R are then obtained.

The association measure R plays the most important role in the determination of the diagnostic power of a test T . In the next section, we present plots that will help to use only the likelihood ratios to define situations where a test is of interest for the clinician. We end this section with an analogy linking different schools of statistics and the clinician's interest in the properties of diagnostic tests:

- A Fisherian clinician would be mainly concerned with the false positive rate, cases where the treatment is harmful for the patients (e.g. prescribing a surgery when it is not necessary).
- A Neyman-Pearson-Wald clinician would be concerned with the false positive and false negative rates.
- A Bayesian clinician would be concerned with the positive and negative predictive values.
- A likelihood clinician view would be concerned with positive and negative likelihood ratios, which will be discussed further in the next section.

2.7. LIKELIHOOD ANALYSIS OF A DIAGNOSTIC TEST AND LIKELIHOOD RATIO PLOTS

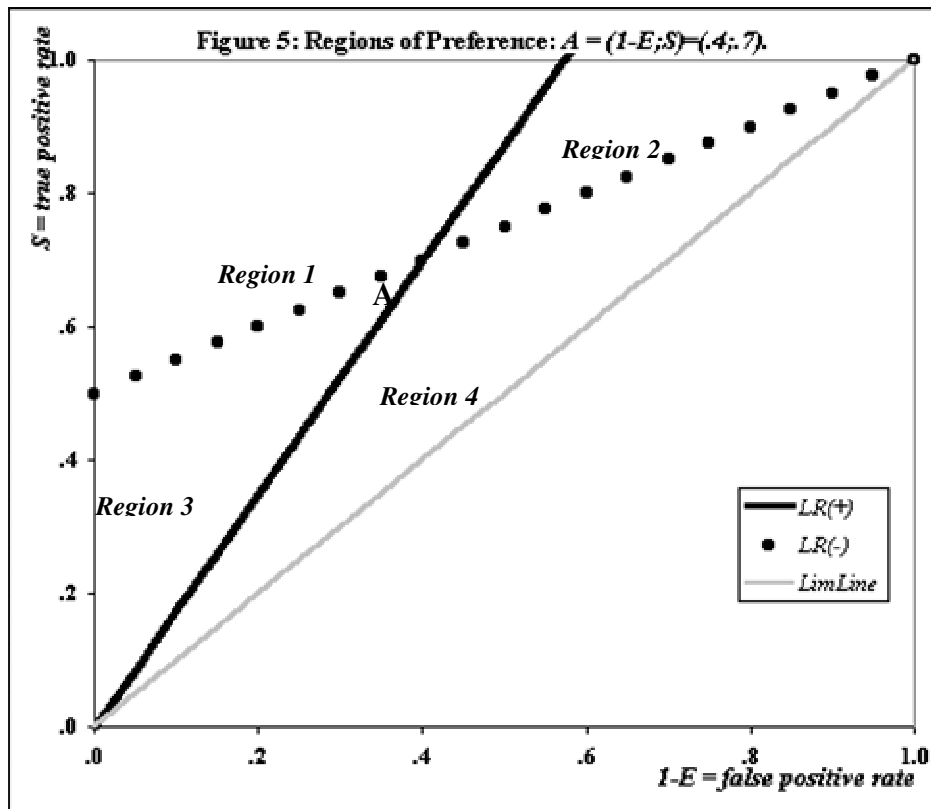
For a given diagnostic test we have defined, respectively, the likelihood ratios of positive and negative test results as $LR(+)$ and $LR(-)$. We also show how to measure the diagnosability of a test, which is based on the change of the pre-test to the post-test odds ratios. According to Jaeschke et al. (1994), the directions and magnitudes of the pre to post changes using likelihood ratio values as a rough guide are as follows:

- i. LR 's larger than 10 or smaller than 0.1 generate conclusive changes.
- ii. LR 's in the interval $(5;10]$ or $[0.1;0.2)$ generate moderate shifts.
- iii. LR 's in the interval $(2;5]$ or $[0.2;0.5)$ generate small (important sometimes) shifts.
- iv. LR 's in the interval $(1;2]$ or $[0.5;1)$ generate small (rarely important) shifts.

Jaeschke et al (1994) also presented a modification of a monogram suggested by Fagan (1975). The monogram is as an old calculus rule where in the left side we have values for the prevalence, in the middle the likelihood ratio and in the right side the PPV values. By drawing a straight line from the prevalence value throughout the likelihood ratio value and ending the line at the right side, the value obtained at this end is just the PPV observed.

Biggerstaff (2000) presented another interesting graphical method for comparing diagnostic tests. A large value of $LR(+)$ indicates that the test has good sensitivity and a small value of $LR(-)$ means that the test has good specificity. If both situations hold we have that R is large and the test has a high diagnostic ability or equivalently high diagnosability. In many situations, due to costs or the health conditions of a patient, one must choose among a set of diagnostic tests a subset that will be performed. In this way ordering the tests by their diagnosability becomes important. To order a set of diagnostic tests according to their diagnostic ability one should have in mind the risks, the costs

and the likelihood ratio values. Note that ordering the tests according to $LR(+)$, high to low values, is equivalent to ordering them based on the values of their PPV's. On the other hand, ordering the tests according to $LR(-)$, low to high values, is equivalent to ordering them based on the values of their NPV's.



Similarly to the ROC (Receiver Operator Characteristic Curve), in Figure 5 we plot, for a diagnostic test T_1 , the point $A = (1-E_1; S_1)$. That is, the false positive rate, $X = 1-E_1$, against the true-positive rate, $Y = S_1$. Additionally we draw two lines through this point; (i) a solid line-segment through $(0;0)$ and A , ending in the horizontal line $(X;1)$ and (ii) a dotted line-segment through $(1;1)$ and A , ending in the vertical line $(0;Y)$. It is not difficult to prove that the slopes of the solid and the dotted lines are, respectively, $LR_1(+)$ and $LR_1(-)$, the likelihood ratios for the test T_1 . The diagonal line delimitates the area where a test is useful. Also, it is easy to show that, for a test, if the point A is below the diagonal line the test is useless. We end this section with the following example:

Example: Consider a diagnostic test T_1 where $S_1 = 0.7$ and $E_1 = 0.6$. For this case we have $A = (.4; .7)$, the solid line is $Y = 1.75X$ and the dashed line is $Y = (1+X)/2$. We have then $LR_1(+)=1.75$ and $LR_1(-)=0.5$. If a new test T_2 is considered we have four possible locations for the point $A_2 = (1-E_2; S_2)$:

- i. $A_2 \in \text{Region 1}$ which implies that T_2 is better than T_1 overall, since $LR_2(+)>LR_1(+)$ and $LR_2(-)<LR_1(-)$.
- ii. $A_2 \in \text{Region 2}$, which implies that T_2 is better (worse) than T_1 for confirming absence (presence) of the disease, since $LR_2(-)<LR_1(-)$ [and $LR_2(+)<LR_1(+)$].

- iii. $A_2 \in \text{Region 3}$ which implies that T_2 is better (worse) than T_1 for confirming presence (absence) of the disease, since $LR_2(+)>LR_1(+)$ [and $LR_2(-)>LR_1(-)$].
- iv. $A_2 \in \text{Region 4}$ which implies that T_2 is worse than T_1 overall, since $LR_2(+)<LR_1(+)$ and $LR_2(-)>LR_1(-)$.

2.8. FINAL REMARKS

We would like to end this report with an optimistic view for the future of pure likelihood approach of Statistics. Let us recall that the work of a statistician lies in a trinity of problems; design of experiments, estimation, and hypotheses testing. We want to show how the likelihood approach works well for the three problems.

In the domain of design of experiments, consider the problem of determination of the number of patients to be tested in order to estimate S , the sensitivity of a clinical test. The maximum of the likelihood is the prescribed estimate. However, we would also need to fix an interval around this estimate in order to guarantee the control of our sampling error. For this purpose we use the normalized likelihood and would like to have the smallest interval with relative plausibility (or credibility) around 95%. Since the binomial distribution is an adequate model, the normalized likelihood follows a beta distribution with parameter $(X+1;Y+1)$ where X (Y) is the number of true positive (false negative) results in a sample of size n , to be determined. Recall that the mean and the variance of this beta distribution are, respectively, $m = (X+1)/(n+2)$ and $v = m(1-m)/(n+3)$. Note that $v \leq [4(n+3)]^{-1}$ since $0 \leq m \leq 1$. Hence, the worst case ($m = 1/2$) is a symmetric beta distribution; i. e. $X=Y$. In this case the mean and the mode (the maximum likelihood estimate) are equal to $1/2$. Adding and subtracting twice the standard deviation to m , we obtain a fair plausible interval (as usually we do when considering normal distributions). Let us represent this interval by $[I_1;I_2]$, where

$$I_1 = 1/2 - (n+3)^{-1/2} \text{ and } I_2 = 1/2 + (n+3)^{-1/2}.$$

Let us now fix the length of the interval of highest plausibility as $I_2-I_1= 2(n+3)^{-1/2} = 0.1$. For this value we obtain $n = 397$. In order to satisfy the restriction $X=Y$, we would take $n=398$ as the sample size. Note that, for $n = 398$, the normalized likelihood would be a beta density with parameter $(200;200)$; that is, $X=Y=199$. Considering this case, the interval $[.45;.55]$ would have credibility 95.49% and length 0.1. Now suppose that we perform the experiment and observe that $X=53=398-Y$. The parameter of the corresponding beta density is $(54;346)$. This is not a symmetric density around its maximum, $53/398$, and the smallest interval with a fixed credibility has equal plausibility in its limits, I_1 and I_2 . For this non-symmetric case we would have the interval $[.1033;.1703]$ with credibility 95.01% and length .067. To obtain this interval we recall that a beta distribution with parameters larger than 1 is uni-modal. Hence, to every parameter point there is a corresponding one with the same plausibility. Considering a pair, say I_1 and I_2 , with the same plausibility in such a way that the interval $[I_1;I_2]$ has posterior probability equal to the fixed credibility, say 95%, we obtain our interval. For bi-dimensional parameter spaces, obtaining a set of 95% of credibility, corresponds to obtaining a level curve where its interior has posterior probability of 0.95.

In the above discussion we have shown how a likelihood approach will solve the sample size determination and both point and interval estimation problems. We now

discuss the testing problem. We use here real data presented in Pereira & Pericchi (1990). Two samples of size 150 were taken respectively from the patients having a disease D and from a healthy control group. A new clinical test was applied to these samples. For the patients, we observed $x = 20 = 150 - y$ true positive cases and for the control sample we obtained $x' = 3 = 150 - y'$ false positive cases. We have here two likelihood functions, one for the sample of patients and another for the control sample. We want to compare this new test, T_1 , with a standard one, T_0 , known to have sensitivity $S_0 = 0.15$ and specificity $E_0 = 0.91$. To replace T_1 for T_0 , we would like to have $S_1 > S_0$ and $E_1 > E_0$. To make a decision about the use of the new test we first identify the set of parameter points with plausibility higher than $S_1 = 0.15$ in the sample of patients and then compute its credibility. For the control sample we identify the set of parameter points with plausibility higher than $E_1 = 0.91$ and then compute its credibility. Note that the normalized likelihood for S_1 (E_1) obtained in the patient (control) sample is a beta density with parameters 21 and 131 (148 and 4). Before we describe the computations let us recall that $LR_0(+)=5/3=1.67$ and $LR_0(-)=85/91=0.93$. On the other hand, the maximum likelihood estimates for the likelihood ratios of the new test are $LR_1(+)=20/3=6.67$ and $LR_1(-)=130/147=0.88$. The odds ratio for the standard test is $R_0=1.78$ and the maximum likelihood estimate for the odds ratio of the new test is $R_1=7.54$. The Good weights of evidence are $\Delta_0=0.0347$ and $\Delta_1=0.2289$. These values already provide evidence that the new test is superior. However, to quantify this superiority we proceed as follows:

1. For the sample of patients, the set of possible values of S_1 with plausibility higher than $S_0=.15$ is the open interval $(.1178;.1500)$; this set has credibility 43.92%. Hence, the evidence in favor of $H: S_1=.15$ is 56.09%. With these figures we cannot reject the hypothesis that the two tests have equivalent sensitivities.
2. For the control sample, the set of possible values of E_1 with plausibility higher than $E_0=.91$ is the interval $(.910;.999)$; this interval has credibility 99.95%. Hence, the evidence in favor of $H: E_1=.91$ is 0.05%. The conclusion here is that the new test is far more specific than the old one.
3. Finally, constructing a plot like in Figure 1 with $A = (1-E_0;S_0) = (.09;.15)$, one would show that the estimated value of $A_1 = (1-E_1;S_1)$, $(1/50;2/15)$, belongs to *Region 1*, supporting the superiority of the new test, T_1 .

We believe to have covered the three problems without using other elements than the likelihood function. We did not have to bring into consideration sample points that could be observed but were not, as in the usual frequentist techniques of unbiased estimation, confidence interval construction or standard significance and hypothesis testing. The most important feature of the methods described in this chapter is that the likelihood principle is never violated.

We finalize this chapter by presenting p -values for the hypothesis

$$H: S_1=.15 \text{ and } H: E_1=.91.$$

In the first case we have 64.78% and in the second case 0.02% as exact p -values. Had we used the chi-square test, we would have 56.76% and 0.42%, respectively. Recall that

our evidence values, based only on the likelihood function (defined on the parameter space, not on possible sample points), for these two hypotheses are **56.09%** and **0.05%**.

3. Informação Estatística e Análise Pré-Posteriori

3.1. INTRODUÇÃO

Neste capítulo discutimos conceitos fundamentais para um planejamento adequado sobre coleta de informações sobre o estado da natureza ou parâmetro θ . O material aqui apresentado é mais conceitual do que operacional embora apresente soluções para alguns problemas específicos. Ressaltamos que o objetivo é entender o que seja diminuição da incerteza sobre θ . Isto é, a busca por boas inferências sobre θ é parte relevante do trabalho do estatístico.

Este capítulo é dividido em cinco seções incluindo esta introdução. Seção 3.2 apresenta o conceito de informação como visto por Basu (1975). Seção 3.3 introduz um conceito de informação mais operacional apresentado por DeGroot (1962). Seção 3.4 é dedicada ao conceito de experimento mais informativo, introduzido por David Blackwell (1951). O conceito de informação apresentado por Blackwell tem caráter mais teórico, mas ainda assim pode ser utilizado na escolha de experimentos. Com caráter mais aplicado, Seção 3.5 mostra o cálculo do tamanho da amostra na perspectiva Bayesiana. Duas situações são discutidas: a amostragem por atributos, tanto para o modelo binomial quanto para o hiper-geométrico.

3.2. INFORMAÇÃO SEGUNDO BASU

A procura por uma definição de informação nos levou inicialmente a considerar que o conceito apresentado por Basu (1975), embora não operacional, parece ser o que melhor descreve o que se entende por informação.

Informação é o que ela faz por você, muda a sua opinião.

O caráter subjetivo desse conceito está intrínseco com a inclusão da pessoa que está tentando obter informação. Muitas vezes um conjunto de observações pode não alterar em nada o conhecimento de um indivíduo, mas pode ser bastante relevante para um outro com paradigma diferente do primeiro.

Para operacionalizar esse conceito abstrato, devemos tentar responder as seguintes questões fundamentais para um bom trabalho estatístico:

- I. Informação sobre o que?
- II. Onde está a informação?
- III. Quanto de informação é usada?
- IV. Como a informação é extraída?

Vejamos como poderíamos responder essas perguntas.

A informação é sobre o valor de θ e é descrita pela distribuição atual de θ . Essa descrição é baseada em avaliações probabilísticas. Informações adicionais podem ser culturais ou experimentais. Isto é, o maior envolvimento na área onde a pesquisa está sendo realizada produz um ganho cultural que pode modificar a distribuição de θ . Já a informação experimental é relativa a realização de experimentos (variáveis aleatórias), X , Y , Z etc., que são passíveis de serem realizados (observadas). Na verdade, o processo de incorporação da informação sobre θ , contida nos resultados experimen-

tais, é um processo de adestramento, diferente do cultural. Aqui temos condições de discutir apenas esse segundo tipo de informação, a experimental.

Vamos supor que o experimento X está sendo realizado e o resultado observado é x . Como consequência natural, passamos de $P(\theta)$ para $P(\theta|x)$. Recordando apenas que essa calibração da informação é obtida pela operação de Bayes:

$$P(\theta|x) \propto L(\theta|x) \times P(\theta).$$

Quanto ao valor da quantidade de informação extraída, devemos definir, de acordo com os interesses do problema que está sendo resolvido, uma distância entre a distribuição a priori e a posteriori. Essa distância muitas vezes está relacionada com custos ou utilidades definidas em determinadas situações como por exemplo em estudos de qualidade de vida. Na próxima seção utilizaremos a função utilidade como medida dessa distância.

Estamos aptos agora a responder as perguntas apresentadas acima:

- i. Informação sobre o que?
Ri: Sobre θ !
- ii. Onde está a informação?
Rii. Está em $L(\theta|x)$!
- iii. Quanto de informação é usada?
Riii. O valor da distância $D\{P(\theta|x), P(\theta)\}$!
- iv. Como a informação é extraída?
Riv. Pela operação de Bayes, $P(\theta|x) \propto L(\theta|x) \times P(\theta)$!

A seguir voltamos ao exemplo das bolas de gude descrito no Capítulo 1:

3.2.1. Exemplo: Bolas na Caixa.

Consideremos quatro bolinhas de gude, duas brancas e duas pretas. Escolho três e coloco em uma caixa. Você deve adivinhar qual o número, θ , de bolinhas brancas dentro da caixa, um ou dois. Você pode, inicialmente, retirar ao acaso uma bolinha da caixa para ganhar informação adicional. Para tentar ganhar mais informação, pode retirar uma segunda bolinha da caixa e essa retirada pode ser com ou sem a reposição, na caixa, da primeira bolinha retirada. Para cada um dos resultados experimentais, calculamos as probabilidades a posteriori que foram apresentadas no Capítulo 1. Ao escolher o experimento, esperamos produzir maior informação com a nossa escolha. Vamos considerar a distância Euclidiana, DE , entre os vetores de probabilidades. Se calcularmos a distância entre priori e posteriori para cada resultado possível e depois avaliarmos o valor esperado em cada um dos experimentos, encontraremos o mesmo valor (**0.24**), independente de qual experimento, X , (X,Y) ou (X,Z) .

Observando que a maior distância possível (maior ganho?) é $DE[(.5,.5);(0,1)] = 0.71$, podemos concluir o seguinte:

- i. Com apenas uma retirada, ao observarmos X , o ganho é **33%** do possível, tanto para $X=0$ quanto para $X=1$. Isto é, ganho pequeno mas garantido!
- ii. Com duas retiradas com reposição, ao observarmos (X,Y) , o ganho pode ser **60%**, resultados $(0,0)$ ou $(1,1)$, com probabilidade **0.56** ou pode não

haver ganho, 0% com probabilidade 0.44, caso dos resultados $(0,1)$ e $(1,0)$. Isto é, ganho moderado com risco moderado!

- iii. Com duas retiradas sem reposição, ao observarmos (X,Z) , o ganho pode ser 100% com probabilidade 0.33, caso dos resultados $(0,0)$ e, novamente, pode não haver ganho, 0% com probabilidade 0,67 se o resultado for $(0,1)$ ou $(1,0)$. Isto é, ganho máximo com risco alto!

O exemplo acima descreve bem a racionalidade do cotidiano na sociedade. Contudo, a distância entre distribuições, como apresentada, não nos ajuda na análise pré-posteriori, para escolher qual o experimento que desejamos realizar. A intuição nos diz que talvez o fato de não utilizarmos a topologia do espaço paramétrico nos impede de uma avaliação mais apropriada. A próxima seção apresenta uma forma de considerarmos outra nuance do problema.

3.3. INFORMAÇÃO SEGUNDO DEGROOT

Em um artigo fundamental da teoria da decisão, DeGroot (1962) considerou uma função real que tentativamente ordena o conjunto das distribuições possíveis para o parâmetro θ . Essa função foi denominada de Função Incerteza. A idéia é a seguinte: conforme quantidades, X, Y, \dots , associadas a θ são observadas, a incerteza sobre este parâmetro deve diminuir. A seguir apresentamos mais formalmente a metodologia de DeGroot.

Consideremos um experimento genérico X cujo valor observado é x . Se P é a função de probabilidade a priori, P_x a função de probabilidade a posteriori, U uma função incerteza e E o operador média então $U(P) \geq E\{U(P_x)\}$. Isto é, espera-se que a incerteza a posteriori seja menor do que a incerteza a priori. Usando a conhecida desigualdade de Jensen, DeGroot mostra que uma função real é uma função de incerteza se e somente se é côncava. Podemos escrever o resultado da seguinte forma: Para quaisquer priori para θ e experimento X , temos o seguinte:

$$I\{X, P, U\} = U(P) - E[U(P_x)] \geq 0 \Leftrightarrow U \text{ é côncava.}$$

O operador I é a quantidade de informação sobre θ contida no experimento X quando a priori é P e a incerteza é U . Note que anteriormente nos referíamos a informação contida no resultado x efetivamente observado de um experimento X . Aqui estamos falando da informação que se espera de um experimento em relação ao parâmetro.

No atual contexto o objetivo é a escolha de experimentos. Consideramos os casos onde existam uma série de experimentos e nem todos podem ser realizados. Os experimentos escolhidos para serem observados devem ser aqueles com maiores valores de I .

É importante ressaltar que as escolhas de P e de U , além de subjetivas, podem estar interligadas e muitas vezes podemos ter casos onde a escolha de uma implica na caracterização da outra. Voltaremos a essa questão após discutir alguns exemplos.

3.3.1. Exemplo: Variância como Incerteza.

Consideremos um parâmetro cujo espaço paramétrico está contido na reta. Vamos considerar uma distribuição a priori com média e variância finitas. Se a variância da distribuição de θ é considerada como função incerteza, podemos provar que a variância é uma função côncava. Isto é, $I\{X, P, V\} = V(\theta) - E\{V(\theta | X)\} = V(E\{\theta | X\}) \geq 0$.

Um exercício necessário para o leitor é provar que $V(\theta) = E\{V(\theta|X)\} + V(E\{\theta|X\})$. Chamamos atenção para o fato de $E\{\theta|X\}$ ser um estimador de Bayes e $V(E\{\theta|X\})$ sua variância. Assim estamos considerando um estimador, $E\{\theta|X\}$, que certamente é viciado no sentido da estatística clássica. Ao procurarmos qual experimento deve ser realizado, escolhemos o que produz um estimador com maior variância. Completamos esse exemplo retornando ao Exemplo 3.2.1 das bolas na caixa. Lembremos que θ assume apenas os valores 1 ou 2. Após avaliarmos $E\{\theta|X\}$, $E\{\theta|X,Y\}$ e $E\{\theta|X,Z\}$ em cada um dos resultados possíveis, calculamos as respectivas variâncias e obtivemos .03, .05 e .08. Assim, o experimento mais informativo no sentido de DeGroot é (X,Z) que produziu a maior variância. Por maior variância devemos entender um maior risco de não produzir informação e possibilidade de obter toda a informação possível!

Antes de finalizar esta seção gostaríamos de ressaltar o caráter subjetivo da escolha tanto da distribuição a priori quanto da função incerteza. Por exemplo, vamos supor que na estimação da proporção de uma característica, o modelo binomial é o modelo estatístico do problema. Consideremos que a família de distribuições beta é de onde a priori será escolhida. Em conjunto com a média, que define a estimativa inicial, a fixação da variância da priori estabelece a priori do problema. Neste caso, a escolha da variância como função incerteza está diretamente relacionada com a escolha da priori.

Na próxima seção discutiremos o conceito de suficiência de Blackwell. Novamente o objetivo é escolher o experimento mais informativo para inferências sobre um parâmetro de interesse. Mostraremos que, mesmo para estatísticos que não aceitam o uso de probabilidade para parâmetros, a escolha de um experimento para inferências, em presença de alternativas, pode ser feita racionalmente. Infelizmente, ao contrário do que apresentamos na presente seção, por evitar-se o uso de medidas de probabilidade sobre os espaços paramétricos, não se pode dizer quanto de informação um experimento X tem a mais, quando comparado com um experimento menos informativo, Y . Isto é, o conceito de suficiência de Blackwell introduz uma forma de identificar-se, quando existe, o experimento mais informativo dentre um conjunto de concorrentes.

3.4. INFORMAÇÃO SEGUNDO BLACKWELL: SUFICIÊNCIA DE BLACKWELL

O conceito de estatística suficiente está relacionado com o objetivo de reduzir a dimensão amostral sem perda de informação. Por exemplo, a média e a variância da amostra podem conter toda a informação contida em uma amostra de uma distribuição normal com média e variância desconhecidas. Note que neste caso o espaço de observações sofre apenas uma redução, pois no lugar de precisarmos de toda a amostra, podemos trabalhar apenas com um par de variáveis, sem perder informações relevantes sobre os parâmetros populacionais. Blackwell (1951) estendeu o conceito de suficiência ao considerar diferentes espaços estatísticos cujo único elemento em comum seria exatamente o parâmetro θ de interesse e desconhecido. A melhor forma de discutir a suficiência de Blackwell é por meio do seguinte exemplo. O material apresentado nesta seção pode ser encontrado em Basu & Pereira (1990).

3.4.1. Exemplo:

Vamos supor que uma nova empresa afirma estar produzindo um produto com o dobro da qualidade de seu concorrente. Isto é, a taxa de falha de seus produtos é metade da taxa de falha, π , de seu concorrente. Para estimar π devemos coletar amostras da nova empresa ou da antiga? Para responder a essa questão, consideremos dois experimentos de Bernoulli: $X|\pi \sim \text{Ber}(\pi)$ e $Y|\pi \sim \text{Ber}(\pi/2)$. Consideremos agora um experimento adicional que é o resultado de um lançamento de uma moeda: $Z \sim \text{Ber}(1/2)$. Note que Z não é relacionado a π e pode ser realizado em qualquer lugar e em qualquer tempo, em resumo, o exercício mais simples de aleatorização. Note que se considerarmos o experimento $Y'=XZ$ temos que Y' e Y são igualmente distribuídos: $Y' \sim Y$. Assim se observarmos X e depois realizarmos um exercício simples de aleatorização, chegamos a um experimento equivalente a Y .

O exemplo acima ilustra o conceito de suficiência de Blackwell. Isto é, suponha que dois experimentos X e Y são tais que suas distribuições dependem dos parâmetros $g(\theta)$ e $h(\theta)$. Dizemos que X é suficiente para Y em relação a θ , se existir um experimento Y' , obtido por uma composição de X e algum experimento Z definido por um exercício de aleatorização (distribuição conhecida após a realização de X), tal que Y e Y' são identicamente distribuídos. Neste caso escrevemos $X \gg Y$. No exemplo, a primeira questão que se apresenta é se $Y \gg X$. A resposta é negativa como veremos a seguir.

Novamente, voltamos ao conceito de informação. Na verdade continuamos a procura por um experimento mais informativo sobre um parâmetro θ . Como estatísticos Bayesianos, poderíamos comparar experimentos olhando para as possíveis distribuições a posteriori. Nossa procura por definições intuitivas nos levou a considerar o conceito de experimento mais informativo na perspectiva Bayesiana. No contexto acima e considerando todas as distribuições a priori dentro da classe de alternativas, vamos definir suficiência comparando as distribuições a posteriori. Na perspectiva Bayesiana, escrevemos $X \gg Y$ se para qualquer ponto amostral y de Y , existir um ponto amostral x de X , tal que $\Pr\{\theta|Y=y\} = \Pr\{\theta|X=x\}$. Fica para o leitor mostrar que suficiência de Blackwell implica na suficiência Bayesiana. Se estivermos comparando experimentos bem comportados e pudermos falar em todas as priores possíveis, as duas definições seriam equivalentes. Para isso teríamos de ser mais cuidadosos e lembrarmos dos conceitos de dominância e modelos discretos da teoria da medida. O argumento aqui apresentado faz-nos lembrar e entender a razão de Basu considerar que a suficiência de Blackwell é um conceito Bayesiano e a denominou de suficiência Bayesiana de Blackwell. Além disso, podemos dizer com essa equivalência que a suficiência de Blackwell não viola o princípio da verossimilhança, pois uma inferência sobre θ com a observação $Y=y$ produz o mesmo resultado se for realizada com $X=x$, pois as verossimilhanças decorrentes são proporcionais. A seguir apresentamos um pouco da teoria de Blackwell.

2.4.2. Definição: Função de Transição

Consideremos dois experimentos X e Y cujos espaços amostrais são representados por X e Y . Uma função de transição, F , de X para Y é uma família $F = \{f_x(\cdot); x \in X\}$ de funções (densidade) de probabilidade $f_x(y)$ definidas em Y e indexadas por $x \in X$.

Por exemplo, a família de funções de probabilidade Hipergeométrica, $H(y;x,n,N)$, é uma função de transição de $\{0,1,\dots,N\}$ para $\{0,1,\dots,n\}$, onde $n < N$ são inteiros positivos. Poderíamos aqui pensar em uma máquina produzindo peças com taxa de falha π . Consideremos um lote de N dessas peças e uma amostra de n peças desse lote. A função de transição Hipergeométrica é na verdade a distribuição condicional de $Y|X$, onde Y é o número de defeituosas da amostra e X é o número de defeituosas do lote. Importante notar que H é a mesma função para todos os valores de π .

3.4.3. Definição: Suficiência de Blackwell

Sejam X e Y dois experimentos, como pensados acima, e com funções (densidade) de probabilidade $g(x|\theta)$ e $h(y|\theta)$. Dizemos que X é suficiente para Y , com respeito a θ , no sentido de Blackwell, se existir uma função de transição tal que $h(y|\theta) = \sum_x f_x(y)g(x|\theta)$.

(No caso de modelos dominados, devemos trocar somatórios por integrais.)

O teorema a seguir resolve o problema da comparação de experimentos de Bernoulli e mostra como o problema da amostragem em populações finitas privilegia a amostragem sem reposição.

3.4.4. Teorema: Comparações de Experimentos de Bernoulli

Sejam $X|\theta \sim \text{Ber}[p(\theta)]$ e $Y|\theta \sim \text{Ber}[q(\theta)]$. Os experimentos X e Y são comparáveis no sentido de Blackwell se o conjunto $\{(p(\theta), q(\theta)) : \theta \in \Theta\}$ estiver contido em uma reta que corta dois lados opostos do quadrado unitário $[0,1]^2$: $X \gg_{\theta} Y$ ($Y \gg_{\theta} X$) se a reta corta os dois lados verticais (horizontais) do quadrado.

Conseqüentemente, se a reta é diagonal então X e Y são equivalentes e escrevemos $X \approx_{\theta} Y$.

Demonstração:

Quando tratamos de distribuições de Bernoulli, X e Y , a existência de uma função de transição corresponde à existência de uma matriz de transição. Isto é, vamos supor que X é suficiente para Y . Isto quer dizer que $P(Y=1|\theta) = P(X=0|\theta)f_0(1) + P(X=1|\theta)f_1(1)$. Ou seja, $q(\theta) = (1-p(\theta))f_0(1) + p(\theta)f_1(1) = f_0(1) + p(\theta)[f_1(1) - f_0(1)]$. Por outro lado $1-q(\theta) = [1-p(\theta)][1-f_0(1)] + p(\theta)[1-f_1(1)]$.

Isto é: $[1-q(\theta) ; q(\theta)] = [1-p(\theta) ; p(\theta)]F$, onde F é a matriz de transição com primeira linha $[1-f_0(1); f_0(1)]$ e segunda linha $[1-f_1(1); f_1(1)]$.

Para entender que a prova está completa basta escrever $q(\theta) = a + bp(\theta)$ onde $a = f_0(1)$ e $b = f_1(1) - f_0(1)$.

Finalizamos essa seção mostrando que a amostragem sem reposição é mais informativa que a com reposição quando retiramos uma amostra de peças de um lote para estimar o número de defeituosas do lote.

3.4.5. Exemplo: Amostragem com e sem reposição

Consideremos um lote de N peças e θ o parâmetro de interesse, número de defeituosas do lote. Se vamos retirar duas peças do lote, devemos decidir se a segunda retirada deve ser com reposição ou sem reposição. Representemos por (X, Y) o experimento com reposição e por (X, Z) o sem reposição. Isto é:

$$X|\theta \sim \text{Ber}(\theta/N), Y|X, \theta \sim Y|\theta \sim \text{Ber}(\theta/N) \text{ e } Z|X, \theta \sim \text{Ber}([\theta - X]/[N - 1]).$$

De fato temos de estudar se o conjunto formado pelos pontos

$$(P_X(\theta); Q_X(\theta)) = ([\theta - X]/[N - 1]; \theta/N)$$

pertence a uma reta que corta os eixos opostos de do quadrado unitário, tanto para $X=0$ quanto para $X=1$. Mas é evidente que

$$Q_0(\theta) = P_0(\theta)(N-1)/N \text{ e } Q_1(\theta) = P_1(\theta)(N-1)/N + 1/N$$

Isto mostra que, para estimar q , (X, Z) é mais informativo do que (X, Y) no sentido de Blackwell. Falta ainda mostrarmos qual seria o exercício de aleatorização que nos permitiria construir um experimento equivalente a (X, Y) a partir de (X, Z) . O leitor fica encarregado de mostrar que: definindo o experimento $Y^*|X, Z$ distribuído como Bernoulli com probabilidade $[X + (N-1)Z]/N$, então $(X, Y^*)|\theta \sim (X, Y)|\theta$. Por outro lado, as matrizes de transição correspondentes a $X=0$ e $X=1$ são as seguintes:

$$T_0 = \begin{pmatrix} 1 & 0 \\ \frac{1}{N} & \frac{N-1}{N} \end{pmatrix} \quad T_1 = \begin{pmatrix} \frac{N-1}{N} & \frac{1}{N} \\ 0 & 1 \end{pmatrix}$$

Vamos finalizar essa seção com um problema proposto por Blacwell (1951). Suponha que desejamos estimar o número, θ , de mulheres fumantes na população de professores da USP. Sabemos o tamanho das populações de Professores (N), de Mulheres (M) e de Fumantes (F). De qual população devemos retirar a primeira pessoa a ser estudada? Da população de professores, da população de mulheres, da população de homens, da população de fumantes ou da população de não fumantes? Por simplicidade e sem perda de generalidade vamos supor que $\theta < M < F < N - F < N - M < 1$.

3.5. TAMANHO DE AMOSTRA

Não é raro o estatístico ter de tentar responder a pergunta “Qual o tamanho de amostra que devo selecionar?” No entanto esta é uma pergunta que pode ser difícil de ser respondidas visto que depende de muitas considerações, suposições e restrições sobre o problema tratado. Recomendar a seleção da maior amostra possível é sempre uma resposta que satisfaz as qualidades estatísticas. Restrições de custo, no entanto, podem inviabilizar essa resposta. Por outro lado, restrições na precisão e na incerteza irão colaborar para um compromisso entre todos os limites estabelecidos. Neste capítulo mostraremos como podemos estabelecer o tamanho de amostra respeitando as diversas restrições do problema da estimação de proporções populacionais. Consideraremos os dois casos conhecidos, Populações Infinitas (tamanho desconhecido) e Populações Finitas (tamanho conhecido). Finalizamos o capítulo com o cálculo do tamanho da amostra para uma situação muito particular de modelos dominados, estimação da correlação entre duas variáveis normais.

3.5.1. Amostras de Populações Infinitas.

Aqui o nosso interesse é o parâmetro π de um modelo Binomial. Isto é, estamos interessados na taxa de falha, π , a proporção de peças defeituosas produzidas por uma máquina. Consideramos o caso em que o processo de produção está sob controle, Isto é, quando as unidades podem ser consideradas como sendo permutáveis. Equivalentemente, consideramos que as unidades são produzidas de acordo com um processo de Bernoulli com probabilidade π . Representando o processo por $\{U_i\}_{i>1}$, as variáveis U_i são, condicionalmente a π , independentes e igualmente distribuídas segundo uma $Ber(\pi)$. Dessa forma temos um processo de Bernoulli, U_1, U_2, \dots , definido da seguinte forma: $Pr(U_i=1|\pi)=1-Pr(U_i=0|\pi) = \pi$ onde a sequência (U_1, U_2, \dots) é composta de variáveis mutuamente independentes dado π .

O nosso objetivo é estabelecer o número, n , de elementos da seqüência que serão observadas. Por populações infinitas entendemos aquelas que não conhecemos seu tamanho.

Os números observados de defeituosas e de boas da amostra serão denotados por X e Y , respectivamente e o tamanho da amostra por $X+Y = n$. Os parâmetros da priori beta são denotados por a e b (≥ 1) e a soma por $n_0 = a+b$. Após observada a amostra, x e y são os valores assumidos por X e Y . Os parâmetros da posteriori Beta serão denotados por $A = x+a$ e $B = y+b$ e assim $A+B = n+n_0$.

O modelo estatístico na verdade é formado por uma trinca de modelos que são descritos a seguir:

| Priori: | Verossimilhança: | Posteriori: |
|-------------------------------|-------------------------------------|---|
| Modelos: | | |
| $\pi \sim Beta(a,b)$ | $X (\pi,n) \sim Binomial(n, \pi)$ | $\pi (x,n) \sim Beta(A,B)$ |
| médias: | | |
| $e = E(\pi) = a/n_0$ | $m = E(X \pi,n) = n\pi$ | $e_n = E(\pi n,x) = A/(n+n_0)$ |
| variâncias: | | |
| $v = V(\pi) = e(1-e)/(n_0+1)$ | $Var(X \pi,n) = n\pi(1-\pi)$ | $v_n = V(\pi n,x) = e_n(1-e_n)/(n+n_0+1)$ |

Note que $v_n < \{4(n+n_0+1)\}^{-1}$. Se considerarmos apenas valores de a e b maiores que a unidade temos que: $v_n < \{4(n+3)\}^{-1}$. Assim a pior situação será obtida quando a priori é não informativa $a = b = 1$ e a posteriori for simétrica, $A=B$.

Vamos supor que para ser informativo, o intervalo que será fornecido para a inferência sobre π não deva ser superior a $0,1$. Assim o erro não pode ser superior a $0,05$. Vamos imaginar que (I_1, I_2) seja o intervalo final e como a pior situação é o caso simétrico, com priori não informativa teríamos que $I_1 = e_n - td$ e $I_2 = e_n + td$. Aqui, t é um multiplicador do desvio padrão, d , da posteriori. A escolha de t depende da credibilidade que se deseja para o intervalo (I_1, I_2) . Vamos imaginar que $Pr\{I_1 \leq \pi \leq I_2 | x\} = 0,95$. Imitando a normal vamos considerar que $t = 1,96$. Assim, $1,96[4(n+3)]^{-0,5} \leq 0,05$ é a desigualdade que usaremos para encontrar o tamanho de nossa amostra. Teríamos assim, $n > 381,16$. Tomando-se então $n=382$, uma posteriori simétrica resultaria em uma $Beta(192,192)$. O intervalo simétrico da pior amostra possível, $[0,45;0,55]$, teria uma credibilidade de

$$1-2Pr\{\pi \leq 0,45\} = 1-2(0.024809) = 0.950381705.$$

Tabela 1: Tamanhos de amostra em função do comprimento e da credibilidade.

| Lenght | Credibility | | | | | | | |
|--------|-------------|--------|--------|--------|--------|--------|--------|--------|
| | 90,00% | 95,00% | 99,00% | 99,50% | 99,80% | 99,90% | 99,95% | 99,99% |
| 0,05 | 1080 | 1534 | 2650 | 3149 | 3817 | 4327 | 4844 | 6050 |
| 0,06 | 749 | 1065 | 1839 | 2186 | 2650 | 3004 | 3363 | 4201 |
| 0,07 | 550 | 781 | 1351 | 1605 | 1946 | 2206 | 2470 | 3086 |
| 0,08 | 420 | 598 | 1033 | 1229 | 1489 | 1689 | 1891 | 2362 |
| 0,09 | 331 | 472 | 816 | 970 | 1176 | 1334 | 1493 | 1866 |
| 0,1 | 268 | 382 | 660 | 785 | 952 | 1080 | 1209 | 1511 |
| 0,11 | 221 | 315 | 545 | 649 | 787 | 892 | 999 | 1248 |
| 0,12 | 185 | 264 | 458 | 545 | 660 | 749 | 839 | 1048 |
| 0,13 | 158 | 225 | 390 | 464 | 562 | 638 | 714 | 893 |
| 0,14 | 135 | 193 | 336 | 399 | 485 | 550 | 616 | 769 |
| 0,15 | 118 | 168 | 292 | 348 | 422 | 478 | 536 | 670 |
| 0,16 | 103 | 147 | 256 | 305 | 370 | 420 | 471 | 589 |
| 0,17 | 91 | 130 | 227 | 270 | 328 | 372 | 417 | 521 |
| 0,18 | 81 | 116 | 202 | 241 | 292 | 331 | 371 | 464 |
| 0,19 | 72 | 104 | 181 | 216 | 262 | 297 | 333 | 417 |
| 0,2 | 65 | 93 | 163 | 194 | 236 | 268 | 300 | 376 |

Valores da distribuição normal.

| | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 1,645 | 1,960 | 2,575 | 2,807 | 3,090 | 3,290 | 3,481 | 3,890 |
|-------|-------|-------|-------|-------|-------|-------|-------|

Vamos supor agora que com a amostra de **382** observações, foram obtidos **92** sucessos e **290** fracassos. Com uma uniforme como priori a posteriori seria *Beta(93,291)*. Com essa distribuição a posteriori, o intervalo de credibilidade teria um comprimento menor do que **0,1**. De fato, o intervalo com aproximadamente **95%** de credibilidade seria: **[0,1998;0,2854]** com comprimento **0.0856** e com credibilidade de **0.950432424**. Lembremos que na posteriori, a média é **0.2421875** e o desvio padrão **0.021833652**. Usando-se a aproximação normal para esta distribuição teríamos uma credibilidade de **0.949995818**. Para o intervalo simétrico **[0.1994;0,2850]** a aproximação normal daria a credibilidade de **0,95**. Note que os valores calculados tanto para a média como para o desvio padrão são exatos e obtidos com a distribuição Beta.

A Tabela 1 apresenta tamanhos amostrais para diferentes restrições:

3.5.2. Amostras de Populações Finitas.

Por populações finitas entendemos aquelas que conhecemos seu tamanho. Vamos imaginar que temos uma população com N unidades, onde N é um inteiro positivo e conhecido. O nosso interesse agora é a estimação do número θ de unidades defeituosas em um lote de N peças. Evidentemente, θ é um inteiro positivo com valor desconhecido com limite superior N . Sem perda de generalidade podemos ordenar as unidades e identifica-las por sua ordem. Isto é o conjunto de unidades será $\{1,2,\dots,N\}$. Assim para a unidade i , estaremos associando uma variável U_i que assume o valor 1 se i for defeituosa e 0 se for uma peça boa. Dessa forma nosso vetor de valores popula-

cionais pode ser representado por $\Psi = (U_1, U_2, \dots, U_N)$ e $\theta = U_1 + U_2 + \dots + U_N$. É natural supormos que trocamos a ordem das unidades não altere o que se espera do vetor da distribuição de Ψ . Dessa forma, consideremos que a distribuição de Ψ é permutável; isto é,

$$\Pr\{U_1 = u_1, U_2 = u_2, \dots, U_N = u_N\} = \Pr\{U_{p1} = u_1, U_{p2} = u_2, \dots, U_{pN} = u_N\},$$

$(p1, p2, \dots, pN)$ é uma permutação de $(1, 2, \dots, N)$ e (u_1, u_2, \dots, u_N) é um vetor cujos elementos são iguais a zero ou a unidade.

Sob a condição de permutabilidade podemos sempre dizer que o vetor Ψ é uma observação finita de um processo permutável de zeros e uns. Com essa restrição básica e pelo teorema de De Finneti, vai existir uma constante $0 < \pi < 1$ tal que, condicionado a π , a seqüência de variáveis aleatórias U_1, U_2, \dots é um processo de Bernoulli com probabilidade comum de sucesso igual a π . Dessa forma é natural considerarmos que $\theta|\pi$ tem distribuição binomial com parâmetros (N, π) .

Com o objetivo de estimar o parâmetro de interesse θ , consideremos uma amostra de unidades populacionais, digamos $\{1, 2, \dots, n\}$, sem perda de generalidades, pois estamos nos restringindo ao processo permutável. Assim, temos a independência condicional, dado π , entre a amostra (U_1, U_2, \dots, U_n) e o complemento populacional da amostra $(U_{n+1}, U_{n+2}, \dots, U_N)$. Se denotarmos por $X = U_1 + U_2 + \dots + U_n$ o total amostral, temos a principal consequência do teorema de De Finneti:

$$\Pr\{X = x, \theta - X = k | \pi\} = \Pr\{X = x | \pi\} \Pr\{\theta - X = k | \pi\}.$$

Isto é, X e $\theta - X$ são condicionalmente independentes dado π .

Lembremos que π é apenas um parâmetro que entrou no desenvolvimento apenas por conveniência teórica e assim não existe algum interesse em seu valor. O nosso primeiro interesse é a obtenção do modelo estatístico, $\Pr\{X = x | \theta\}$.

Notemos que:

$$\Pr\{X = x | \theta = t, \pi\} = \Pr\{X = x, \theta = t | \pi\} / \Pr\{\theta = t | \pi\} =$$

$$\Pr\{X = x, \theta - X = t - x | \pi\} / \Pr\{\theta = t | \pi\} = \Pr\{X = x | \pi\} \Pr\{\theta - X = t - x | \pi\} / \Pr\{\theta = t | \pi\}.$$

Por outro lado, pela permutabilidade do processo definido, sabemos que:

- i. X e $\theta - X$ são condicionalmente independentes dado π
- ii. $X | (\pi, n) \sim \text{Bin}(n, \pi)$,
- iii. $\theta - X | (\pi, n, N) \sim \text{Bin}(N - n, \pi)$ e
- iv. $\theta | (\pi, N) \sim \text{Bin}(N, \pi)$.

Com essas propriedades podemos escrever a seguinte igualdade:

$$P\{X=x | \theta=t, \pi\} = \Pr\{X=x, \theta=t | \pi\} / \Pr\{\theta=t | \pi\} = \Pr\{X=x, \theta - X = t-x | \pi\} / \Pr\{\theta=t | \pi\}.$$

Então,

$$P\{X=x | \theta=t, \pi\} = \Pr\{X=x, \theta=t | \pi\} / \Pr\{\theta=t | \pi\} = \Pr\{X=x | \pi\} \Pr\{\theta - X = t-x | \pi\} / \Pr\{\theta=t | \pi\}.$$

Isto é, $P\{X=x | \theta=t, \pi\} = \text{Bin}(n, \pi) \text{Bin}(N-n, \pi) / \text{Bin}(N, \pi) = H(x, t; n, N)$, onde H é a função de probabilidade hipergeométrica avaliada no ponto x , com parâmetro t e tamanhos de amostra e população dados por n e N . Como esta função não depende de π , concluímos que $P\{X=x | \theta=t, \pi\} = P\{X=x | \theta=t\}$.

Notamos também que ao observarmos na amostra o valor de X , o valor de θ - X passa a ser o parâmetro de interesse. Utilizando a mesma notação do caso infinito, ao considerarmos uma distribuição a priori $Beta(a,b)$ para π , a distribuição a posteriori seria $Beta(A,B)$. Como nosso interesse é a distribuição de $(\theta-X)|X$, utilizamos a distribuição de $(\theta-X)|(\pi,X) \sim (\theta-X)|\pi \sim Bin(N-n,\pi)$ e calculamos a preditiva de $\theta-X$ ao eliminarmos π por sua posteriori $Beta(A,B)$. Finalmente temos $(\theta-X)|X \sim BetaBin(N-n;A,B)$, como apresentada no Capítulo 2. Isto é,

$$Pr\{(\theta - X) = k | X = x\} = \frac{\binom{N-n}{k} \Gamma(n+n_0) \Gamma(k+A) \Gamma(N-n-k+B)}{\Gamma(A) \Gamma(B) \Gamma(N+n_0)}$$

A média e a variância dessa distribuição são, respectivamente,

$$\begin{aligned} E\{\theta-X | X = x\} &= E\{E\{\theta-X | X = x, \pi\} | X = x\} = E\{E\{\theta-X | \pi\} | X = x\} = \\ &= (N-n)E\{\pi | X = x\} = (N-n)A/(n+n_0) = (N-n)e_n \\ V\{\theta-X | X = x\} &= \{(N-n)(N+n_0)/(n+n_0+1)\}e_n(1-e_n). \end{aligned}$$

Novamente aqui podemos notar que o pior caso é aquele onde a média a posteriori de π , e_n , é igual a 0,5. Consideremos agora um novo parâmetro definido por $\delta = \theta/N$. Vamos nos concentrar nessa proporção para estabelecer que o comprimento de um intervalo de credibilidade para δ não deve exceder a um valor prefixado, digamos $2\varepsilon = 0.1$. Note que e_n é a média a posteriori de δ cuja variância a posteriori seria:

$$v_n = Var\{\delta | X = x\} = \frac{(N-n)(N+n_0)e_n(1-e_n)}{(n+n_0+1)N^2} \leq \frac{(N-n)(N+n_0)}{4(n+n_0+1)N^2}$$

Como antes iremos procurar um intervalo (I_1, I_2) tal que

$$Pr\{I_1 \leq \delta \leq I_2 | x\} = 0,95.$$

Isto é, iremos encontrar o valor de t tal que $I_1 = e_n - td$ e $I_2 = e_n + td$. Aqui, t é um multiplicador do desvio padrão, d , da posteriori de δ . Neste caso, além da variável observada, também o parâmetro de interesse é discreto. Assim, os padrões da distribuição normal podem não funcionar. No entanto se conseguirmos fixar o valor do desvio padrão do parâmetro δ encontraremos o valor de n necessário para atingir o limite fixado. Por exemplo, suponha que fixamos $td = 2d = 0.05$. Neste caso, considerando $d = 0,025$ e $a = b = 1$ ($n_0=2$), para $N=5000$, teríamos $n=394$. Com esse tamanho de amostra, o caso simétrico seria aquele onde $A = B = 198$ que corresponderia a $x = y = 197$. A tabela abaixo apresenta os tamanhos de amostra para diferentes restrições.

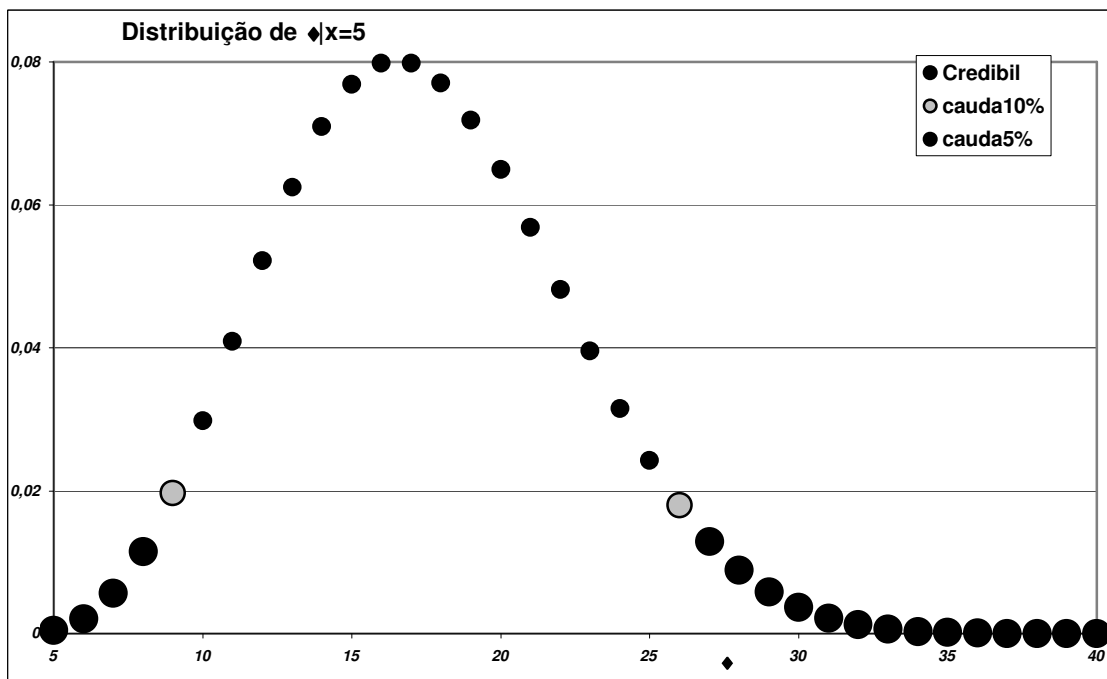
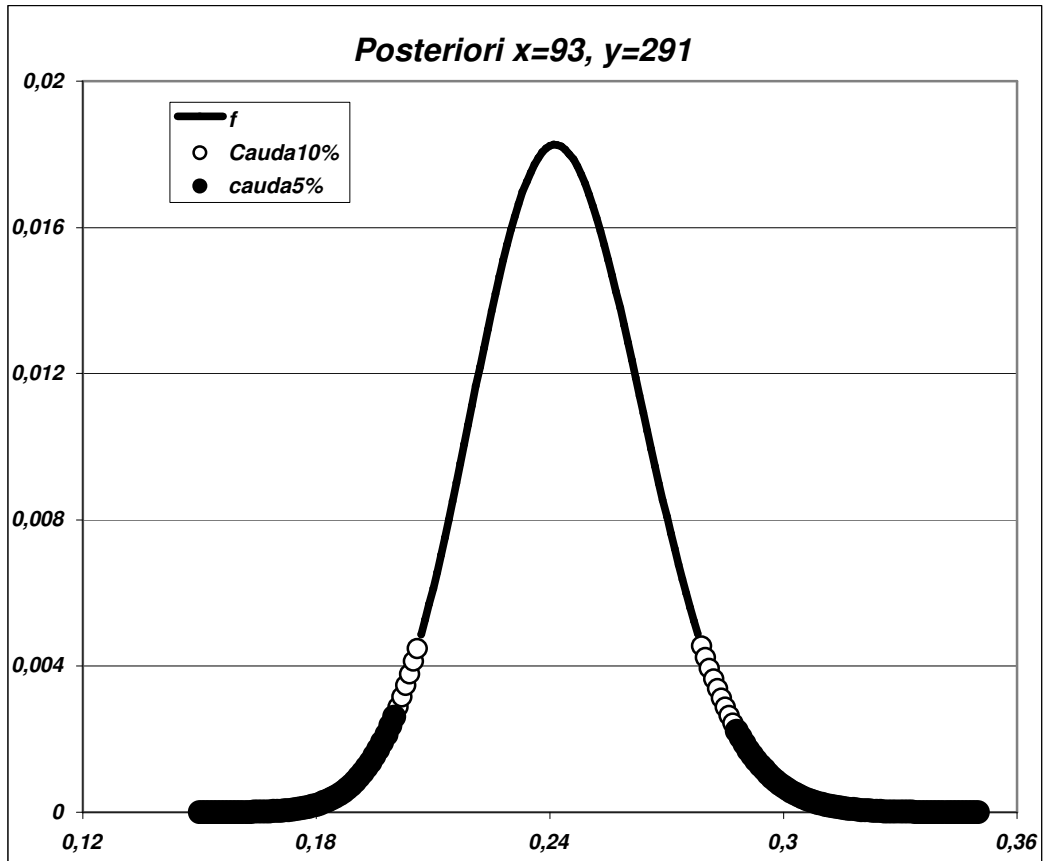
Como exemplo, vamos supor que desejamos considerar uma amostra de um lote de $N = 50$ peças. Consideremos uma amostra de $n = 15$ peças o número X de peças defeituosas nessa amostra. Suponha que encontramos $x = 5$. Concluimos aqui que o conjunto $\Theta_0 = \{10,11,\dots,25\}$ é um conjunto de 90,69% de credibilidade para θ . Isto é,

$$Pr\{\theta \in \Theta_0\} = Pr\{0,2 \leq \delta \leq 0,5\} = 0,906891$$

Os gráficos das distribuições de probabilidades a posteriori de π no exemplo de populações infinitas e de θ , no caso acima são apresentados na seqüência.

Tabela 2: Tamanho da amostra em populações finitas.

| Parâmetro | População | Desvio padrão | | | | |
|-----------|-----------|---------------|-------|------|------|-----|
| | | 0,01 | 0,025 | 0,05 | 0,07 | 0,1 |
| π | ∞ | 2473 | 394 | 95 | 48 | 22 |
| δ | 5000 | 1655 | 365 | 93 | 48 | 22 |
| δ | 4000 | 1530 | 358 | 93 | 47 | 22 |
| δ | 3000 | 1358 | 348 | 92 | 47 | 22 |
| δ | 2000 | 1108 | 329 | 90 | 47 | 22 |
| δ | 1500 | 934 | 313 | 89 | 47 | 22 |
| δ | 1000 | 712 | 284 | 86 | 46 | 22 |
| δ | 750 | 578 | 261 | 84 | 45 | 21 |
| δ | 500 | 416 | 221 | 80 | 44 | 21 |
| δ | 400 | 345 | 199 | 78 | 43 | 20 |
| δ | 300 | 268 | 171 | 73 | 41 | 20 |
| δ | 200 | 185 | 133 | 65 | 39 | 20 |
| δ | 100 | 96 | 83 | 49 | 33 | 18 |
| δ | 50 | 49 | 44 | 33 | 24 | 15 |



4. FBST por Gráficos

4.1. INTRODUÇÃO

Este capítulo não tem mil palavras, mas alguns gráficos. Como o FBST já surgiu em discussões nos Capítulos 1 e 2, o presente capítulo tem o objetivo de ilustrar o que é chamado de FBST – Full Bayesian Significance Test. De fato a nomenclatura correta deveria ser Teste de Significância Genuinamente Bayesiano. Por genuíno entenda-se aquele que usa apenas a distribuição a posteriori na avaliação da evidência em favor da hipótese. Assim que a distribuição a posteriori é definida, todo o passado pode ser esquecido incluindo-se o espaço amostral, que deu origem a definição do modelo estatístico. O espaço paramétrico continua sendo o original, a mudança ocorrida, após as observações experimentais, recai apenas na nova distribuição obtida após a operação Bayesiana, *prior to posterior*. Ao leitor resta a tarefa de interpretar e entender o que o FBST faz por ele, apresenta o valor de um índice de evidência. A decisão em função do valor assumido pelo índice fica na mão do cientista. Regras de decisão serão apresentadas, mas as perdas possíveis devem ser sempre consideradas.

4.2. TEORIA DO FBST

Suponha que $f(\theta)$ é a densidade a posteriori sobre o espaço paramétrico Θ . Considere a hipótese precisa $H: \theta \in \Theta_H \subset \Theta$, onde a dimensão de Θ_H é menor do que a dimensão de Θ .

4.2.1. Definição: Conjunto Tangente e Evidência de H

O conjunto tangente de H é definido por

$$T = \{t \in \Theta \mid f(t) > \text{Sup}_H f(\theta)\}.$$

Por Sup_H entenda-se o supremum de f restrito a Θ_H . A evidência contra H é a probabilidade a posteriori do evento $\theta \in T$. Consequentemente, a evidência a favor de H é o complemento dessa probabilidade. Se $ev(H)$ é o valor da evidência em favor de H então, $ev(H) = 1 - E\{I_T\}$ onde E é o operador esperança na distribuição a posteriori, e I é a função indicadora do conjunto T .

4.2.2. Definição: Intuitiva do FBST

A hipótese H não deve ser rejeitada sempre que o valor de $ev(H)$ for alto.

A dificuldade com a definição acima é decidir-se quando o valor observado de ev é alto. Com esse objetivo Madruga, Esteves e Wechsler (2001) mostraram a solução baseada na teoria da decisão. Assim ao se escolher um limite v de tal forma que se aceita (rejeita) H se $ev \geq v$ ($ev < v$). Os autores mostraram que qualquer que seja o valor de v escolhido, existe uma função de perda p tal que a regra acima, com limite v , é ótima. Isto é, minimiza a perda esperada!

4.2.3. Teorema: O Bayesianismo do FBST

O procedimento FBST é a minimização da esperança a posteriori da seguinte função de perda:

$$\mathbf{p}(\text{rejeição de } H; \theta) = \mathbf{a}(1-I_T) \text{ \& } \mathbf{p}(\text{aceitação de } H; \theta) = \mathbf{b} + \mathbf{c}I_T$$

a, **b** e **c** são números reais positivos.

Importante ressaltar que com a definição apresentada do cálculo de evidência, o FBST não seria um procedimento totalmente invariante sob transformações do parâmetro. Madruga, Pereira & Stern (2006) apresentaram uma adaptação do teste que passou então a ser invariante. Foi introduzida uma função de referência, r , para a obtenção do conjunto \mathbf{T} . Como, nessa adaptação, o conjunto \mathbf{T} não tem mais o papel tangencial, passou a ser denominado de conjunto da surpresa relativa, um conceito que foi introduzido por Good (1983). De fato consideramos no espaço paramétrico, onde a densidade *a priori* foi definida, a distribuição *não informativa* – a de máxima entropia em certos casos –. Defina agora o conjunto \mathbf{T} . Denote por $r(\theta)$ a densidade de referência. Substitua então na definição de \mathbf{T} , a função de densidade f por $s = f/r$. Com isto, na maioria dos casos, o conjunto de relativa surpresa já não é mais o conjunto tangente. Todos esses aspectos estão discutidos em detalhes em Madruga et al (2006). A função de perda \mathbf{p} só se modifica por conta do novo conjunto surpresa. Um aspecto que não é comum em teoria da decisão é o fato da função de perda \mathbf{p} depender, além da densidade f e da função de referencia r , também da amostra \mathbf{x} . Contudo essa característica já vem sendo usada parte da literatura: Bernardo e Smith (1994) por exemplo.

Mesmo com todos os aspectos positivos e práticos apresentados, alguém pode olhar \mathbf{ev} como sendo uma estatística. Isto é, a distribuição amostral de \mathbf{ev} pode ser de interesse para aqueles acostumados uso da estatística frequentista. Para um estatístico Genuinamente Bayesiano, essa é uma questão irrelevante! Contudo, existem aspectos teóricos interessantes para serem discutidos. É claro que, como a distribuição a posteriori depende do ponto amostral observado e sendo \mathbf{ev} obtido com essa posteriori, \mathbf{ev} é uma estatística. Uma visão clássica de \mathbf{ev} exige a consideração de sua distribuição amostral. A questão então é a obtenção da probabilidade $\mathbf{Pr}\{\mathbf{ev} \leq \mathbf{v} | \theta\}$ qualquer que seja $0 < \mathbf{v} < 1$.

As afirmações que aparecem na seqüência são conjecturas que estão sendo provadas aos poucos pelos colegas e alunos do autor. Além disso, as simulações realizadas até o momento comprovam essas afirmações.

Vamos imaginar uma situação bem comportada onde tanto a verossimilhança como as densidades a posteriori satisfazem as muitas condições de contorno listadas em Schervish (1995). Com respeito à convergência, para grandes amostras de tamanho n , podemos afirmar que a densidade a posteriori se aproxima da verossimilhança normalizada, como no Capítulo 2. Represente por \mathbf{L} a verossimilhança normalizada, por \mathbf{M} a moda ou o estimador de máxima verossimilhança e por \mathbf{m} o máximo restrito a Θ_H . Com essa representação, o conjunto \mathbf{T} pode ser escrito como

$$\mathbf{T} = \{\mathbf{t} \in \Theta \mid \mathbf{L}(\mathbf{m}) < \mathbf{L}(\mathbf{t}) < \mathbf{L}(\mathbf{M})\}$$

Lembrando o bom comportamento de \mathbf{L} , pode-se fazer uso de aproximações normais para o cálculo das probabilidades a posteriori de subconjuntos do espaço paramétrico; em particular o conjunto \mathbf{T} . Assim, se $\|\mathbf{t}-\mathbf{M}\| = (\mathbf{t}-\mathbf{M})\Sigma^{-1}(\mathbf{t}-\mathbf{M})'$ é a norma a posteriori onde Σ^{-1} é a inversa (generalizada) da matriz de covariância a posteriori, Σ , no caso de ser não singular (singular), podemos escrever o conjunto \mathbf{T} da seguinte forma:

$$\mathbf{T} = \{\mathbf{t} \in \Theta \mid \|\mathbf{t}-\mathbf{M}\| < \|\mathbf{m}-\mathbf{M}\|\}$$

Se a dimensão de θ for $k \geq 1$ então, com o uso da aproximação normal, pode-se concluir que a densidade a posteriori de $\|\theta - \mathbf{M}\|$ pode ser aproximado por uma densidade Chi-quadrado com k graus de liberdade. Assim, $ev = F_k(\|\mathbf{m}-\mathbf{M}\|)$, onde F_k é a função de distribuição Chi-quadrado. De fato, o que se está afirmando é que essa aproximação permite escrever que

$$ev = 1 - \Pr\{\mathbf{T}|x\} \approx F_k(\|\mathbf{m}-\mathbf{M}\|).$$

Lembrando que Θ e \mathcal{X} são os espaços paramétrico e amostral, essa ultima probabilidade deve ser olhada como uma probabilidade condicional definida no espaço produto $\Theta \times \mathcal{X}$. Isto é, o evento definido na sigma álgebra produto é \mathbf{T} e a sub-sigma álgebra do condicionante é aquela equivalente a sigma álgebra de \mathcal{X} . A seguir uma outra representação de \mathbf{T} é apresentada.

Se $\ell(\mathbf{t}) = \mathbf{L}(\mathbf{t})/\mathbf{L}(\mathbf{M})$ é a verossimilhança relativa e seu logaritmo natural for representado por $\lambda(\mathbf{t}) = \ln \ell(\mathbf{t})$, o conjunto tangente tem a seguinte representação:

$$\mathbf{T} = \{\mathbf{t} \in \Theta \mid \lambda(\mathbf{m}) < \lambda(\mathbf{t}) < 0\} = \{\mathbf{t} \in \Theta \mid -2\lambda(\mathbf{m}) > -2\lambda(\mathbf{t}) > 0\}.$$

Se k e h são as dimensões Θ e Θ_H e lembrando que a distribuição amostral assintótica de $-2\lambda(\mathbf{m})$ é Chi-quadrado com $k-h$ graus de liberdade, a distribuição amostral de \mathbf{T} também pode ser obtida. Note que o evento $\{ev \leq ev_0\}$ é equivalente ao evento $\{-2\lambda(\mathbf{m}) \geq -2\lambda(\mathbf{m}_0)\}$, onde o índice 0 indica o valor observado da estatística em questão, ev e \mathbf{m} no presente caso. Assim, na distribuição amostral de \mathbf{T} , o valor p da estatística ev , quando ev_0 for sua observação efetiva, é a cauda superior da densidade Chi-quadrado com $k-h$ graus de liberdade, a partir do ponto $-2\lambda(\mathbf{m}_0)$, onde \mathbf{m}_0 é o valor observado do máximo da verossimilhança restrito ao conjunto Θ_H . Agora, com o condicionante sendo a sub-sigma álgebra equivalente a sigma álgebra do espaço paramétrico, pode-se escrever:

$$pv_0 = \Pr\{ev < ev_0 \mid \theta\} = 1 - F_{k-h}(-2\lambda(\mathbf{m}_0)).$$

Concluindo, se o índice 0 indica valores efetivamente observados e

$$d_0 = \|m_0 - M_0\| \ \& \ -2\lambda(m_0)$$

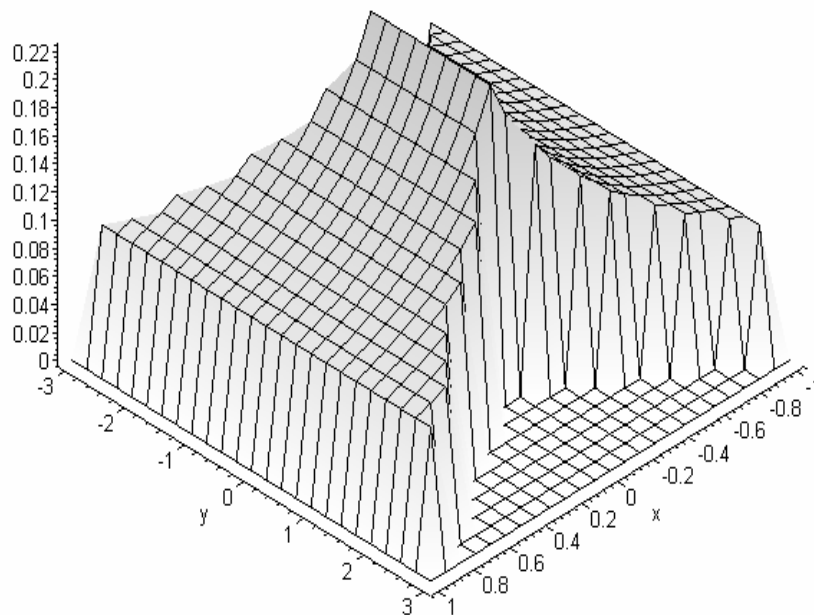
são os valores amostrais das estatísticas Chi-quadrado, as seguintes probabilidades produzem tanto o teste de significância Bayesiano quanto o frequentista:

$$ev_0 = F_k(d_0) \ \& \ pv_0 = 1 - F_{k-h}(-2\lambda(m_0)).$$

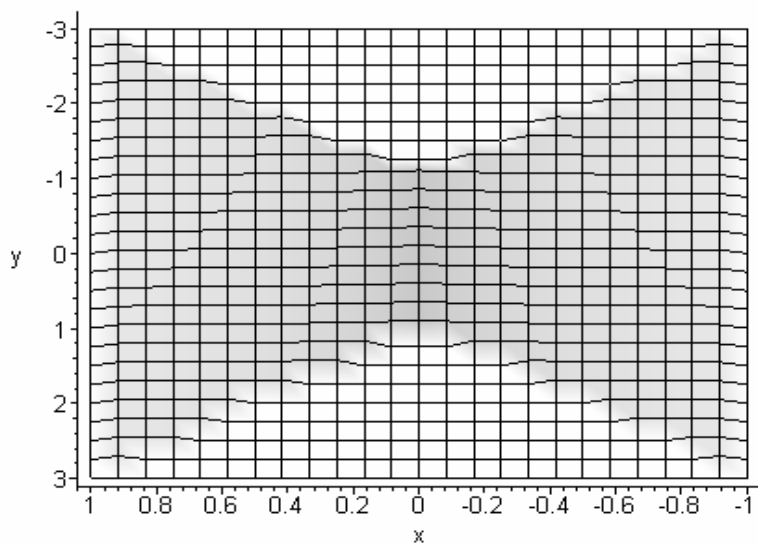
Para um Bayesiano, a decisão é tomada em relação ao valor observado de ev_0 e baseada na função de perda p . Para um frequentista, a decisão é tomada em relação ao valor observado pv_0 . Interessante seria encontrar-se a relação existente entre ev e pv . Lembremos que tanto ev quanto pv são obtidos a partir do mesmo conjunto tangente T , que varia com os valores amostrais x .

Note-se que em nenhum momento foi necessário falar-se em redução de dimensão ou eliminação de parâmetros indesejáveis. Uma das razões para esta atitude é apresentada a seguir por meio de gráficos. O exemplo que deu origem a estes gráficos é por nós denominado de Gravata Borboleta. A Figura 4.2.4 apresenta a densidade conjunta de (x,y) . Figura 4.2.5 apresenta a mesma densidade em uma visão por cima e de onde originou a denominação. Por último, a Figura 4.2.6 apresenta a densidade marginal de x . Notem que um conjunto de credibilidade na densidade conjunta é a parte central do gráfico. Por outro lado um conjunto de credibilidade para x na marginal irá produzir um conjunto formado pelas caudas da densidade. Acreditamos que o correto é definir-se credibilidade na densidade conjunta e não na marginal.

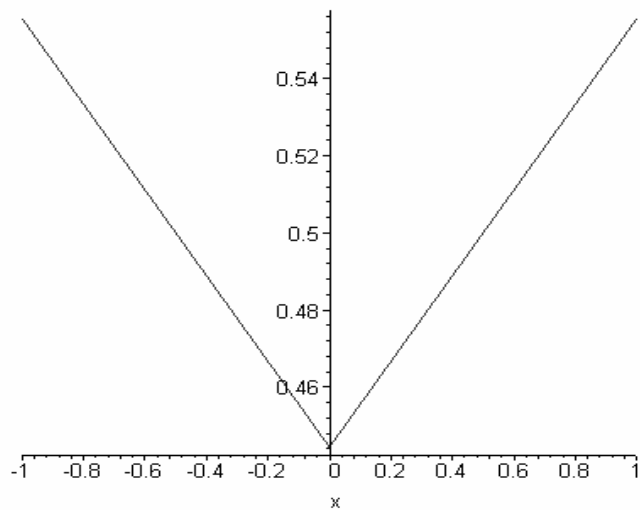
4.2.4 *Figura:* Densidade conjunta de um vetor aleatório (x,y)



4.2.5 *Figura:* Visão superior da densidade conjunta de um vetor aleatório (x,y)



4.2.6 *Figura:* Densidade marginal de x



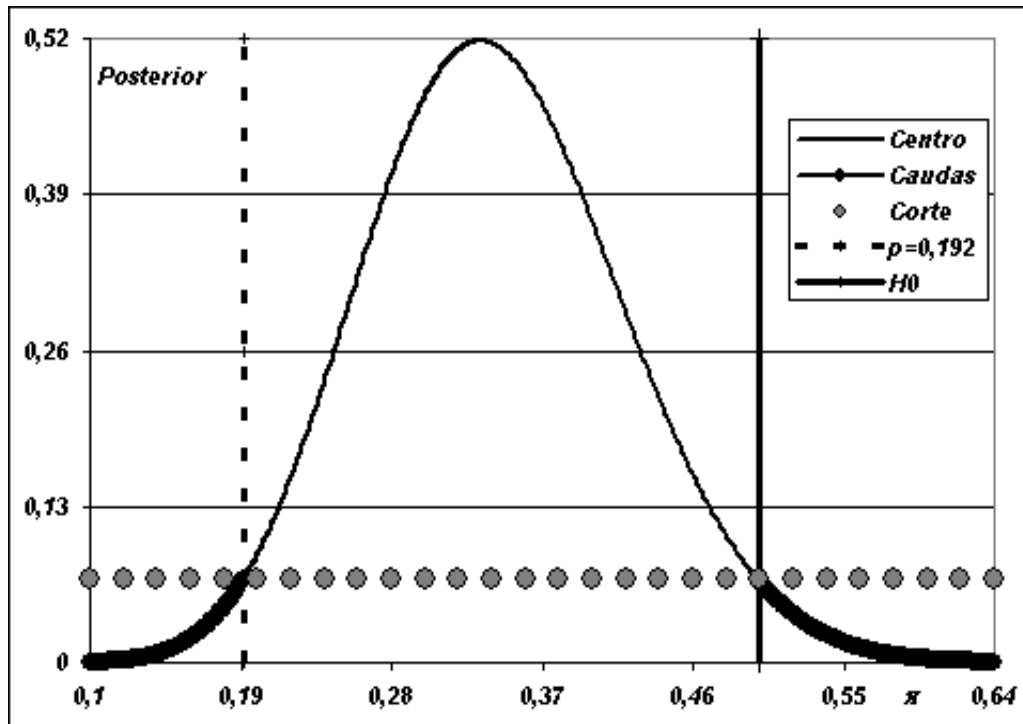
4.3. FBST: GRÁFICOS ILUSTRATIVOS

Nesta seção serão apresentados problemas clássicos da estatística. No primeiro exemplo o FBST é aplicado para testar se uma proporção π populacional pode ser considerada igual a um valor específico p . Não se faz referência à regra de parada.

4.3.1. Exemplo: Teste de Proporção

Neste exemplo uma moeda é lançada 36 vezes, o número de caras, $x=12$ e o número de coroas, $y=24$, são registrados. Não é relevante registrar-se, por exemplo, qual das seguintes regras foi tomada no início: i) o total de lançamentos, $n = x+y = 36$, foi fixado anteriormente ou ii) o número de fracassos $y = 12$ foi o fixado anteriormente. Na verdade, a única estatística relevante para a análise é o kernel da verossimilhança (o fator que depende dos resultados amostrais): $L(\pi|x,y)=\pi^x(1-\pi)^y$. A hipótese nula que se deseja testar é **H**: $\pi = 1/2$. A hipótese alternativa neste caso é **A**: $\pi \neq 1/2$. Figura 4.3.1.1 é auto explicativa e apresenta o conjunto tangente e o respectivo valor da evidência **ev**.

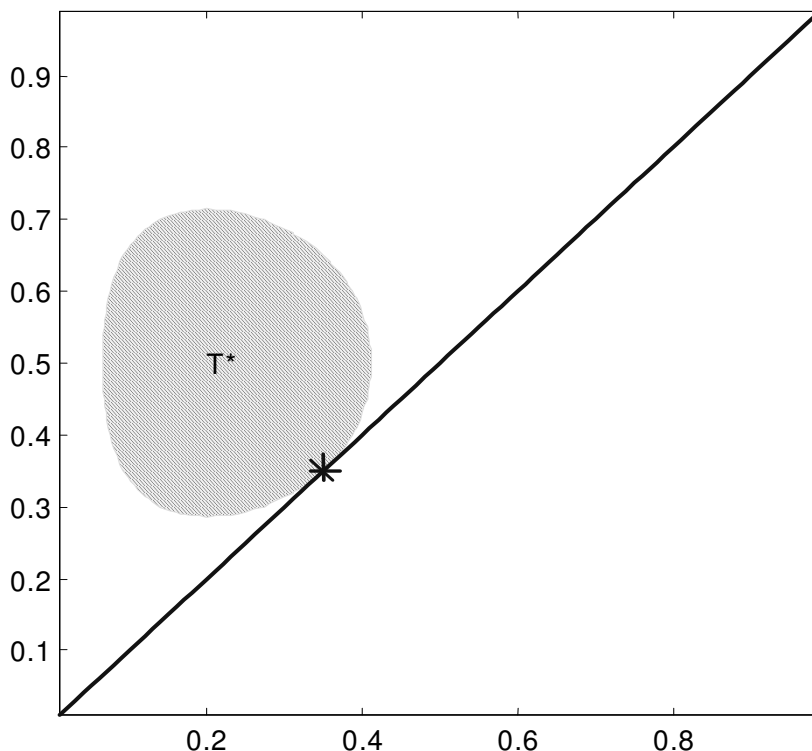
4.3.1.1 Figure: FBST para proporção – $(x;y) = (12;24)$ e **ev** = 4,1%



4.3.2. Exemplo: Teste de Homogeneidade

O presente exemplo considera duas amostras independentes onde observou-se em cada uma o número de sucessos e fracassos, $(x;y)$. O objetivo é testar se as proporções populacionais de sucesso são iguais. A Figura 4.3.2.1 ilustra o conjunto tangente que dá origem ao cálculo da evidência.

4.3.1.1 Figure: FBST para homogeneidade
 $(x;y)_1 = (4;16)$, $(x;y)_2 = (10;10)$, e $\mathbf{ev} = 1,2\%$



Na seqüência o FBST é usado para testar os parâmetros de uma distribuição normal. Para uma melhor composição gráfica decidiu-se pelo uso do programa power point da microsoft.

4.3.3. Exemplo: FBST: Modelo Normal

O presente exemplo é o mais celebrado dentro da comunidade estatística; Normal com média e variância desconhecida. A seguir estão listados as diversas etapas do trabalho estatístico.

- **Parâmetros de Interesse:**

(μ, τ) é o vetor de parâmetros de distribuição normal. μ é a media e τ , conhecido como *precisão*, é o inverso da variância, $\tau = 1/\sigma^2$.

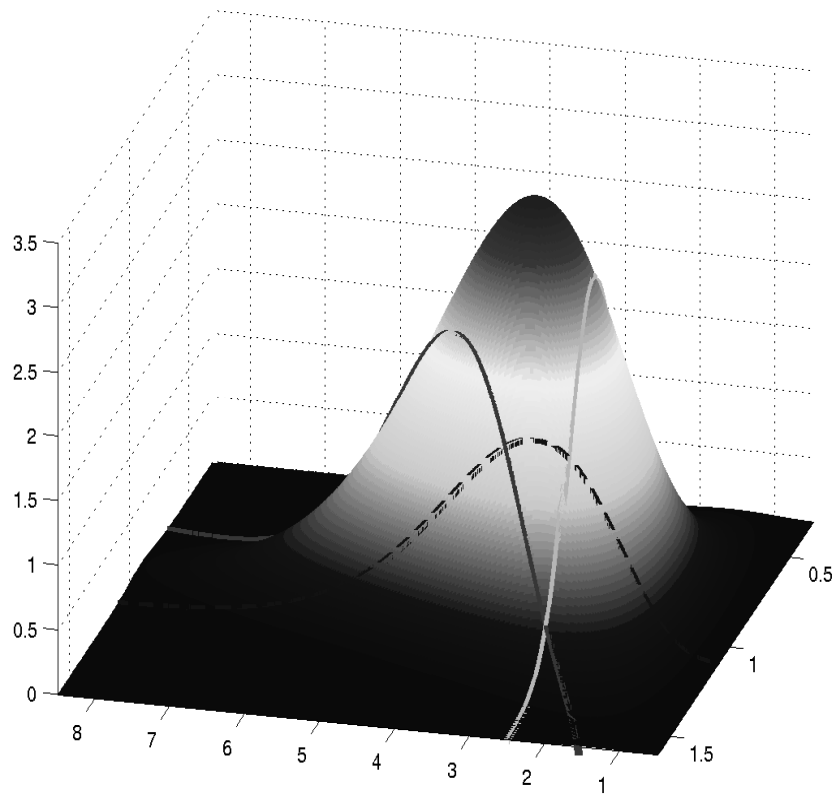
- **Densidade Posteriori:**

$$f(\mu, \tau) \propto \tau^{6.5} \exp \left\{ -\frac{\tau [4 + 11(\mu - 0.9)^2]}{2} \right\}$$

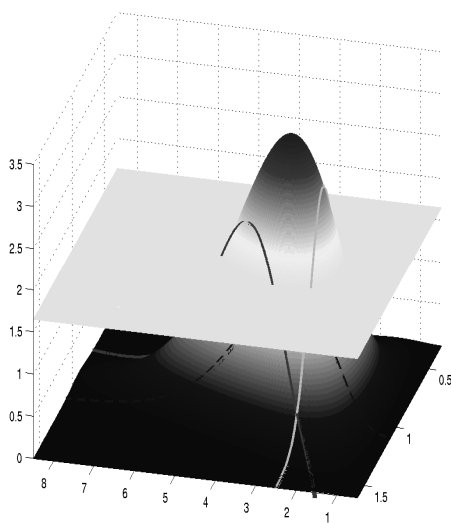
- Hipóteses e Evidências:
- $H_0: \mu = 1.1$ $ev_0 = 51\%$
- $H_1: \tau = 2.5$ $ev_1 = 70\%$
- $H_2: cv = (\mu^2 \tau)^{-0.5} = 1/2$ $ev_2 = 81\%$
- $H_3: \mu = 1.1 \text{ \& } \tau = 2.5$ $ev_3 = 47\%$
- $H_4: \mu = 1.1 \text{ \& } cv = 1/2$ $ev_4 = 50\%$
- $H_5: \tau = 2.5 \text{ \& } cv = 1/2$ $ev_5 = 14\%$

Os gráficos que apresentamos a seguir são auto-explicativos. Todas as hipóteses e evidências podem ser contempladas analisando os gráficos.

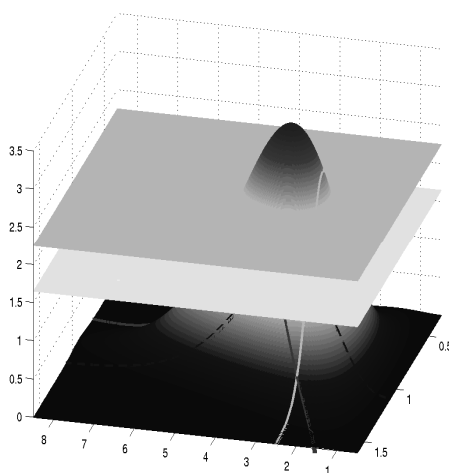
4.3.3.1 *Figura: Densidade posteriori de $f(\mu, \tau)$:
Curvas das hipóteses 0,1 e 2.*



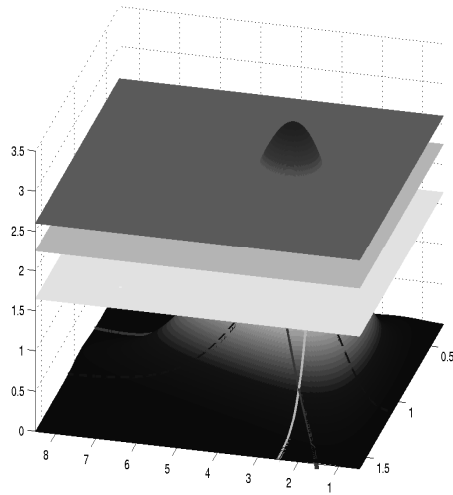
4.3.3.2 Figura: Cortando a densidade posteriori para definir o conjunto T_0



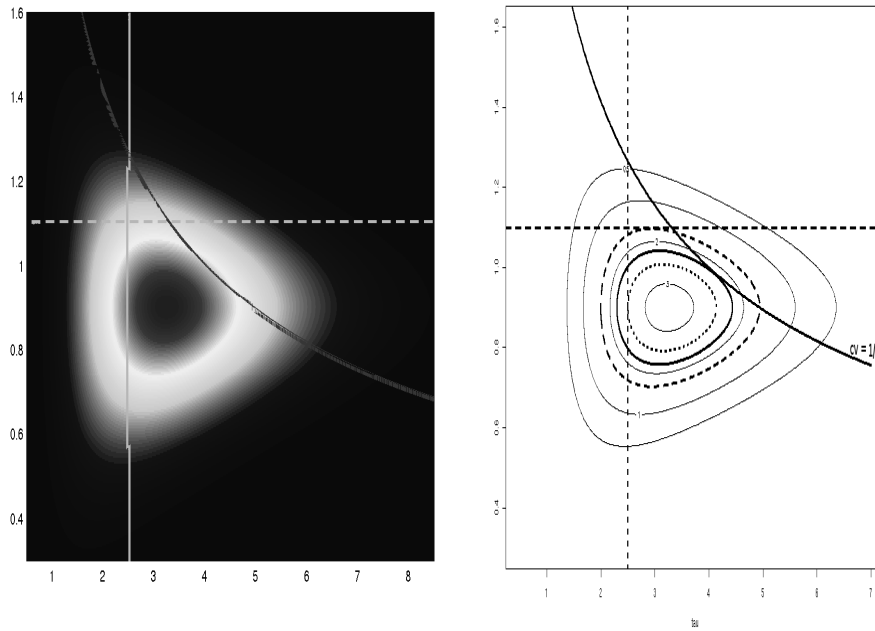
4.3.3.3 Figura: Cortando a densidade posteriori para definir os conjuntos T_0 e T_1



4.3.3.4 *Figura:* Cortando a densidade posteriori para definir os conjuntos T_0 , T_1 e T_2



4.3.3.5 *Figura:* Curvas de Nível: Tangentes T_0 , T_1 , T_2 , T_3 , T_4 e T_5



5. Soluções de Problemas Reais: O Bayesianismo na Consultoria

5.1. INTRODUÇÃO

Este capítulo apresenta exemplos de aplicação, objetivando ilustrar o poder do método Bayesiano ou indutivo. O primeiro exemplo faz parte da tese de mestrado do autor, apresentada e defendida em 1971. A solução aqui apresentada é diferente da original, Pereira (1971). Esta é uma solução genuinamente Bayesiana e a original, mais complexa, mistura idéias Frequentistas e Bayesianas. O leitor poderá comprovar a simplicidade da solução aqui apresentada. O exemplo trata da preferência sexual na população de uma espécie de inseto (mosca da Banana) denominada *Drosophila melanogaster*. Na seqüência apresentamos os detalhes do experimento e do tipo de dado analisado.

O segundo e o terceiro exemplos são casos discutidos e resolvidos no Centro de Estatística Aplicada (CEA) do Instituto de Matemática e Estatística (IME) da Universidade de São Paulo (USP), durante o segundo semestre de 2005, na disciplina Estatística Aplicada. Os dados dizem respeito à incidência ou prevalência de determinada enfermidade ou presença de mecanismos agressivos a saúde. Descrever os problemas e as soluções sugeridas é a melhor forma de divulgação de um método.

5.2. PREFERÊNCIA SEXUAL

No estudo de populações, sob o ponto de vista da genética, torna-se importante, por diversos motivos, a verificação da ocorrência da falta de equilíbrio populacional, com respeito aos alelos de um determinado loco genético. Diz-se que um loco está em equilíbrio populacional quando as frequências gênicas de seus alelos (ou genes) não se alteram ao longo das gerações. Isto é, se a é um gene de um loco estudado, a proporção $p(a)$ desse gene, no total de genes (duas vezes o tamanho da população se o loco é dialélico), é a mesma ao longo das gerações. As formas usuais de avaliação dessa característica são os testes de equilíbrio populacional de Hardy-Weinberg. Para um estudo detalhado desse conceito veja Pereira (1985).

A perda de equilíbrio populacional pode ser causada por vários fatores e um dos mais comuns é a existência de preferência sexual, o que quer dizer que um indivíduo com um determinado genótipo (ou fenótipo) procure parceiro do mesmo genótipo (ou fenótipo). Acredita-se que em populações não humanas essa característica seja comum. Evidentemente, que se esta for a causa do desequilíbrio, a preocupação com deformações genéticas diminui. Esta é a razão da importância do estudo. A seguir descrevemos o problema.

Na época que o autor iniciou seu trabalho no Laboratório do Professor Luiz Edmundo Magalhães, no Instituto de Biologia da USP, experimentos com as moscas eram realizados em pequenas caixas de população. Inicialmente produziam-se populações com genótipos conhecidos. Note que ao cruzar uma fêmea Ébony, (e,e), com um macho Ébony, (e,e), todos os descendentes serão necessariamente (e,e) a menos que exista mutação, o que neste caso é raro. Já o indivíduo Selvagem pode ser de dois

genótipos possíveis, $(e,+)$ e $(+,+)$, pois o gene $+$ é dominante. Assim, se na descendência de um Ébony com um selvagem resultar algum Ébony, é sinal de que o progenitor selvagem seja heterozigoto e , além disso, todos os descendentes selvagens são heterozigotos, $(e,+)$. Com este entendimento, fica claro que há como construir tanto uma população de heterozigóticos, $(e,+)$, quanto uma população Ébony, (e,e) .

Para o estudo da preferência, colocavam-se, em uma caixa de população, 20 fêmeas selvagens heterozigóticas, 20 machos do mesmo tipo e 20 machos Ébony. Quando ocorria um acasalamento, o casal era sugado por meio de um canudo e registrado o tipo de cruzamento, se do Tipo 1, $(e,+) \times (e,+)$, ou do Tipo 2, $(e,+) \times (e,e)$. Contavam-se assim as frequências observadas de cada tipo de cruzamento nos 20 possíveis. Quanto mais as proporções se distanciassem de $\frac{1}{2}$ maior seria a taxa de preferência sexual. Notou-se que quase não ocorriam valores distantes de $\frac{1}{2}$, talvez pelo fato de haver forte dependência na ocorrência do tipo de cruzamento. Isto é, ao sugar-se um casal, a população do tipo de macho sugado fica diminuída quando comparada com a população rival. Assim, ao retirar-se um macho do tipo $(e,+)$ a população do tipo (e,e) fica em vantagem e assim há maior oferta desse tipo de macho. Uma melhor representação da população natural se tornou necessária. Isso só seria possível com o aumento do número de indivíduos nos experimentos.

Com os argumentos apresentados acima, observou-se que os descendentes de um cruzamento seguem um processo de Bernoulli com probabilidade definida pelo tipo de cruzamento. Isto é, a ocorrência de um descendente Ébony tem probabilidade $\frac{1}{4}$ sob o cruzamento do Tipo 1 e $\frac{1}{2}$ sob o cruzamento do Tipo 2. Assim, poderiam ser usadas as frequências fenotípicas dos descendentes de cada cruzamento para se avaliar o tipo de cruzamento ocorrido. De forma indireta, com essa interpretação, ter-se-ia uma idéia das frequências de cruzamentos após olhar os descendentes obtidos em um experimento. Optou-se então por passar de $N=20$ para $N=250$ e assim definiram-se caixas de populações com 750 indivíduos, sendo 250 fêmeas e 250 machos heterozigotos, $(e,+)$, e 250 machos homozigotos do tipo (e,e) .

No experimento aqui apresentado, das 250 fêmeas, apenas 232 tiveram descendentes e puderam assim ser estudadas. Considerem-se as seguintes variáveis aleatórias do modelo: i) I é a indicadora do tipo de cruzamento ocorrido, $I=1$ se do Tipo 1 e $I=0$ se do Tipo 2, ii) x é a frequência de descendentes selvagens, $(e,+)$ ou $(+,+)$, do cruzamento e iii) y é o número de descendentes Ébony, (e,e) , do cruzamento. Isto é, as variáveis de interesse podem ser representadas pelo vetor (I,x,y) , onde x e y são observadas em cada cruzamento e I é uma variável de interesse e que deve ser predita. A esperança condicional de I dado (x,y) é a probabilidade a posteriori de $E\{I|x,y\} = Pr\{I=1|x,y\} = \pi(x,y)$. Para esse cálculo precisamos estabelecer a probabilidade a priori $\pi = E\{I\}$, que é a esperança do indicador I a priori. Note que, ao considerar-se que todas as fêmeas da amostra possuem a mesma probabilidade π , pode-se admitir que π seja o parâmetro de interesse, a taxa de preferência sexual, e que será modificado a posteriori. Note que, como cada fêmea receberá a posteriori um valor diferente de π , pode-se pensar que a esperança de π é que define nosso parâmetro de interesse. Na seqüência discutiremos esse problema com o cuidado necessário.

Ao observarmos uma fêmea de nossa amostra, com o vetor observado (x,y) , a função de verossimilhança associada, L , pode ser escrita como

$$L(\phi|x,y) \propto \phi^x (1-\phi)^y,$$

onde $\phi = 3/4$ se o cruzamento é do Tipo 1 e $\phi = 1/2$ se do Tipo 2. Assim se π é a esperança de I , então é igual à probabilidade a priori de $\phi = 3/4$. Um simples exercício (recomendado para o leitor) é mostrar que a razão de chances (odds) a posteriori é igual a razão de chances a priori multiplicada pela razão das verossimilhanças. Assim, se $o = \pi(1-\pi)^{-1}$ e $O(x,y) = \pi(x,y)[1-\pi(x,y)]^{-1}$ representam os odds a priori e a posteriori pode-se escrever que $O(x,y) = o(3^x 2^{-n})$, onde $n=x+y$. Por outro lado, como probabilidade de um evento, π , e a respectiva razão de chances, o , estão ligadas como $1-\pi=(1+o)^{-1}$, a probabilidade a posteriori da fêmea selecionada ter tido um cruzamento do Tipo 2 será

$$1-\pi(x,y)=[1+o(3^x 2^{-n})]^{-1}$$

A Tabela 5.2.1 apresenta resultados amostrais e probabilidades a posteriori para algumas fêmeas da amostra, considerando-se como probabilidade a priori do cruzamento do Tipo 1 para cada fêmea, $\pi = 1/2$. A tabela com os resultados das 232 fêmeas pode ser obtida na página <http://www.ime.usp.br/~cpereira/livro/Mack.xls>.

5.2.1. Tabela: Observações amostrais, probabilidade a posteriori do cruzamento do Tipo 1 e proporção de descendentes selvagens.

| fêmea | n | x | y | odds | $\pi(x,y)$ | $PA(x)$ |
|-------|-----|-----|-----|---------|------------|---------|
| 217 | 108 | 86 | 22 | 3,3E+08 | 1,00 | 0,80 |
| 218 | 77 | 52 | 25 | 4,3E+01 | 0,98 | 0,68 |
| 219 | 75 | 61 | 14 | 3,4E+06 | 1,00 | 0,81 |
| 220 | 129 | 80 | 49 | 2,2E-01 | 0,18 | 0,62 |
| 221 | 83 | 59 | 24 | 1,5E+03 | 1,00 | 0,71 |
| 222 | 86 | 72 | 14 | 2,9E+08 | 1,00 | 0,84 |
| 223 | 100 | 48 | 52 | 6,3E-08 | 0,00 | 0,48 |
| 224 | 70 | 48 | 22 | 6,8E+01 | 0,99 | 0,69 |
| 225 | 118 | 61 | 57 | 3,8E-07 | 0,00 | 0,52 |
| 226 | 142 | 75 | 67 | 1,1E-07 | 0,00 | 0,53 |
| 227 | 121 | 90 | 31 | 3,3E+06 | 1,00 | 0,74 |
| 228 | 74 | 50 | 24 | 3,8E+01 | 0,97 | 0,68 |
| 229 | 81 | 60 | 21 | 1,8E+04 | 1,00 | 0,74 |
| 230 | 118 | 66 | 52 | 9,3E-05 | 0,00 | 0,56 |
| 231 | 90 | 41 | 49 | 2,9E-08 | 0,00 | 0,46 |
| 232 | 119 | 100 | 19 | 7,8E+11 | 1,00 | 0,84 |

Preferência Sexual = 61,4%

Para completar o exemplo, denote-se por I_1, I_2, \dots, I_{232} os valores do indicador de cruzamento do Tipo 1 nas 232 fêmeas e a soma por s . Caso os valores dos indicadores pudessem ser observados, com θ sendo o valor da taxa de preferência sexual, s seria distribuída como uma binomial de parâmetros $n = 232$ e θ . Dessa forma, qualquer boa estimativa de θ seria função da proporção amostral, $p = s/n$. Na visão fre-

quentista, p é uma estimativa de máxima verossimilhança e não viciada de θ , desde que não seja absurdo considerar-se que 232 seja um tamanho de amostra previamente fixado. Como não é possível observar os verdadeiros valores assumidos pelas variáveis indicadoras, I_i , mas apenas suas esperanças, não é absurdo considerar-se a esperança, S , de s para a estimativa, P , de p e conseqüentemente de θ . Lembremos que a esperança de s é simplesmente a soma S das esperanças de cada uma das indicadoras; isto é, a soma das probabilidades a posteriori de cada fêmea é a melhor predição do valor de s . É claro que se pode discutir o uso de $1/2$ como probabilidade a priori para cada fêmea ter cruzado com um macho de seu tipo. No entanto, observou-se que os números de descendentes dos cruzamentos foram todos significativos e, conseqüentemente, informativos. Por informativo entendemos um tamanho de prole suficiente para eliminar informações a priori distorcidas. A menor prole observada foi de 40 *Drosophilas*. Antes de discutir-se uma nova proposta de trabalho, note que a soma das probabilidades a posteriori foi de $S=142,49$ e assim o valor estimado da taxa de preferência sexual seria $P = S/n = 0,6114$. Com este número, pode-se dizer que a preferência sexual é, sim, uma das possíveis causas de desequilíbrio encontrada na população estudada.

Para finalizar esta seção, imaginemos que a preferência sexual, π_i , (não observável) associada a i -ésima fêmea, é um número no intervalo (0,1) gerado por uma densidade beta com parâmetros $\alpha > 0$ e $\beta > 0$. Uma definição razoável de taxa de preferência associada a uma determinada população seria a média da distribuição Beta geradora dos números π_i . Isto é, $\mu = \alpha/(\alpha+\beta)$ seria a taxa de preferência sexual. Para um acesso de uma priori, considere-se uma parametrização mais bem comportada visto que o espaço paramétrico passa a ser limitado. Além da média μ , um real no intervalo (0,1), considere o desvio padrão $\sigma = [\mu(1-\mu)/(\alpha+\beta+1)]^{1/2}$ e assim o parâmetro de interesse e desconhecido seria o vetor (μ, σ) limitado ao quadrado unitário. O espaço paramétrico é então de variação dependente e definido por um semicírculo de raio unitário.

$$\Theta = \{(\mu, \sigma); 0 < \mu < 1 \text{ \& } 0 < \sigma < [\mu(1-\mu)]^{1/2}\}.$$

Definindo-se uma priori neste espaço, obteríamos a densidade de (μ, σ) condicionada a matriz de observações (X, Y) de descendentes. Cada linha dessa matriz corresponde a uma fêmea, onde o correspondente elemento da coluna X (Y) é o número de descendentes selvagens (Ébony). É evidente que o trabalho computacional é intenso e, além disso, o resultado não será significativamente diferente do apresentado acima. Para o tratamento desse modelo recomendamos o artigo Bueno et all (2002).

Outros modelos poderiam ser discutidos, mas talvez não fossem tão realistas como os apresentados aqui. Por exemplo, uma pessoa poderia pensar em um modelo de mistura das duas binomiais com probabilidades de sucessos $3/4$ e $1/4$ e mantendo o mesmo elemento misturador, π_i , para todos os cruzamentos. Mas esse modelo não iria satisfazer o fato de as proporções de machos disponíveis poderem ser diferentes para as fêmeas da amostra. Embora este modelo seja mais simples, talvez não seja realista.

5.3. PADRÕES DE ECTOPARASITISMO EM AVES DE CERRADO E MATA DE GALERIA DO DISTRITO FEDERAL

O problema apresentado nesta seção é fruto de um trabalho de consultoria realizado pelo CEA-IME-USP. A responsável pelo trabalho de pesquisa e pela coleta dos dados é a pesquisadora Mieko Ferreira Kanegae do Instituto de Biologia, IB-USP. Os estudantes, atualmente estatísticos, que me acompanharam nessa consulta foram Aurora Kyoko Nakati e Marco Aurélio Hirata, hoje. Aurora e Marco realizaram os cálculos aqui apresentados.

5.3.1. O problema e o experimento: Apresentação

As aves são hospedeiras de inúmeros tipos de parasitas que apresentam diversos graus de patogenicidade. Os ectoparasitas são parasitas, externos ao corpo do hospedeiro, que afetam as aves de forma direta através da ingestão de tecidos, causando anemia, danos à pele e inoculando toxinas.

Apesar das parasitoses serem responsáveis por extinções de espécies silvestres, as suas presenças são apenas detectadas ou observadas após causarem danos ao ambiente. Apenas recentemente estão sendo reconhecidos os danos provocados às populações silvestres por parasitas e patógenos. Entre as principais causas dessa presença incômoda pode-se destacar o desmatamento e a fragmentação. A consequência da ocorrência desses fatos é um aumento de contato com espécies de algum habitat adjacente, como animais domésticos. Além disso, a introdução de espécies, novas ou comuns no passado, pode proporcionar a entrada de novos patógenos e parasitos, produzindo consequências ainda mais desastrosas.

Os ectoparasitas considerados neste estudo são os ácaros de plumícolas, carrapatos e malófagos. Os ácaros de plumícolas são ectoparasitas obrigatórios que vivem exclusivamente nas aves, podendo ocorrer em diversas partes da plumagem, principalmente nas penas de vôo e nas coberteiras das asas. Os carrapatos são quase onipresentes nos vertebrados terrestres, sendo de grande importância como reserva e vetores de organismos patogênicos. O seu ciclo de vida engloba três instares: larva, ninfa e adulto, sendo que cada muda é precedida de uma refeição com sangue.

O problema aqui discutido é o da estimação e avaliação da incidência (ou prevalência) de ectoparasitismo por carrapatos e ácaros plumícolas em aves de cerrado e mata de galeria. Os dados foram coletados de janeiro a dezembro de 2002, na Fazenda Água Limpa, localizada a 20 km a sudeste de Brasília.

O estudo foi desenvolvido em dois ambientes: mata de galeria e cerrado. Cada um dos ambientes foi amostrado pelo menos quatro dias por mês. A captura das aves foi realizada com, cerca de, 15 redes ornitológicas de 12×2 metros, dispostas em transectos lineares, totalizando 180m de redes em cada local. As redes foram abertas nas primeiras horas do dia, após o nascer do sol e fechadas às 13:00. As aves capturadas foram marcadas com anilhas.

Em todas as aves capturadas anotaram-se a presença ou ausência de carrapatos, de ácaros plumícolas e de malófagos. Os carrapatos foram contados na região da

cabeça, pescoço e ventre das aves, soprando-se as penas até que a pele pudesse ser visualizada.

Para verificar a presença e quantificar os ácaros plumícolas, foi realizada uma inspeção das penas das asas e cauda contra a luz do sol. As amostras de ácaros obtidas foram retiradas de diferentes locais das penas. Os ectoparasitas foram coletados com uma pinça e conservados em álcool 70%, sendo posteriormente enviados para identificação laboratorial.

Para cada ave capturada, anotaram-se os valores das componentes do vetor indicador de presença ou ausência dos 3 tipos de ectoparasitas, $I = (c, a, m)$. As componentes do vetor assumem os valores 0 ou 1, de acordo com a ausência (0) ou presença (1) de carrapatos (c), ácaros (a) e malófagos (m).

Considerações iniciais são necessárias para que o modelo utilizado possa ser aceito. A primeira suposição é a de que toda ave do ambiente tenha a mesma chance de ser capturada. A segunda é a de que não exista preferência de espécies no que diz respeito ao ataque dos diversos tipos de parasita. A terceira e última, talvez a mais rigorosa, é a de que as capturas foram realizadas de forma independente. Isto é, o fato de um tipo de ave ser selecionado não aumenta ou diminui a chance de uma outra ave ser selecionada, principalmente as da mesma espécie. Em resumo, a coleta é feita segundo um modelo de seleção independente e identicamente distribuída.

A seguir, a notação utilizada é apresentada:

1. As frequências amostrais são denotadas por y_{cam} . Isto é, o número de aves com vetor de incidência (c, a, m) . Por exemplo, y_{011} é o número de unidades amostrais com ausência de carrapatos e presença de ácaro e malófagos. A Tabela 1 descreve todas as frequências observadas. O vetor das observações será então

$$Y = (y_{000}, y_{001}, y_{010}, y_{100}, y_{011}, y_{101}, y_{110}, y_{111}).$$

2. As proporções populacionais de aves em cada uma das 8 possíveis categorias, (c, a, m) , são denotadas por θ_{cam} . Assim, o vetor populacional, das proporções nas respectivas categorias, é representado por

$$\Theta = (\theta_{000}, \theta_{001}, \theta_{010}, \theta_{100}, \theta_{011}, \theta_{101}, \theta_{110}, \theta_{111}).$$

O vetor Θ é o alvo principal do estudo. É o parâmetro desconhecido e de interesse.

5.3.2. Estimação e testes de significância

Com as restrições apresentadas, a função de verossimilhança é definida como:

$$L(\Theta | Y) = \Theta^Y = \prod (\theta_{cam})^{y_{cam}}.$$

O produto é sob o conjunto $\{(c, a, m); c, a, m = 0, 1\}$.

A verossimilhança, correspondente à amostra observada, será então

$$L(\Theta | Y) = \theta_{000}^{208} \times \theta_{001}^{79} \times \theta_{010}^{276} \times \theta_{100}^{91} \times \theta_{011}^{140} \times \theta_{101}^{15} \times \theta_{110}^{108} \times \theta_{111}^{36}$$

O máximo dessa função – a estimativa de máxima verossimilhança – ocorre no ponto

$$\hat{\theta} = (208, 79, 276, 91, 140, 15, 108, 36) \frac{1}{953}$$

$$= (0,2183; 0,083; 0,2896; 0,0955; 0,1469; 0,0157; 0,1133; 0,0378)$$

5.3.2.1 Tabela. Frequências observadas

| <i>c</i> | <i>a</i> | <i>m</i> | Θ | <i>Y</i> | Observado |
|----------|----------|----------|----------------|-----------|-----------|
| 0 | 0 | 0 | θ_{000} | y_{000} | 208 |
| 0 | 0 | 1 | θ_{100} | y_{100} | 79 |
| 0 | 1 | 0 | θ_{010} | y_{010} | 276 |
| 1 | 0 | 0 | θ_{001} | y_{001} | 91 |
| 0 | 1 | 1 | θ_{110} | y_{110} | 140 |
| 1 | 0 | 1 | θ_{011} | y_{011} | 15 |
| 1 | 1 | 0 | θ_{101} | y_{101} | 108 |
| 1 | 1 | 1 | θ_{111} | y_{111} | 36 |

Como o tamanho da amostra não pode ser decidido a priori, o espaço amostral – o espaço de todas as frequências possíveis – não é bem determinado. Isto dificulta o uso das técnicas usuais da estatística clássica ou frequentista. Consequentemente, a opção pelo uso das técnicas Bayesianas de análise é naturalmente adequada. De fato, com a especificação da verossimilhança seu estudo e análise são simplificados.

Utilizando a função de verossimilhança correspondente a amostra e a partir de uma densidade *a priori* Dirichlet de ordem 8 e com vetor de parâmetros formado por componentes iguais a 0,25 nas oito posições, encontra-se a densidade *a posteriori* que, utilizando-se o símbolo de proporcionalidade, pode ser escrita como:

$$f(\Theta | Y) \propto \theta_{000}^{207,25} \times \theta_{001}^{78,25} \times \theta_{010}^{275,25} \times \theta_{100}^{90,25} \times \theta_{011}^{139,25} \times \theta_{101}^{14,25} \times \theta_{110}^{107,25} \times \theta_{111}^{35,25}$$

Com esta densidade, a média e a moda a posteriori de Θ são os vetores:

$$\begin{aligned}\tilde{\Theta} &= \left(\frac{1}{955} \right) \times (208,25; 79,25; 276,25; 91,25; 140,25; 15,25; 108,25; 36,25) = \\ &= (0,2181; 0,0830; 0,2893; 0,0955; 0,1469; 0,0160; 0,1134; 0,0380) \quad \&\end{aligned}$$

$$\begin{aligned}\hat{\Theta} &= \left(\frac{1}{947} \right) \times (207,25; 78,25; 275,25; 90,25; 139,25; 14,25; 107,25; 35,25) = \\ &= (0,2188; 0,0826; 0,2907; 0,0953; 0,1470; 0,0150; 0,1133; 0,0372)\end{aligned}$$

Com a fórmula abaixo, calcula-se a matriz de covariância do vetor Θ :

$$Cov(\Theta | Y) = \left(\frac{1}{956} \right) \left(\text{diag}(\tilde{\Theta}) - \tilde{\Theta} \otimes \tilde{\Theta}^t \right)$$

Aqui, sendo A um vetor de 8 componentes, a notação $\text{diag}(A)$ representa a matriz quadrada de ordem 8, cujos elementos da diagonal são os componentes de A e os outros elementos da matriz são iguais a zero. O vetor coluna, formado pelos componentes de A , é representado por A^t . O símbolo \otimes representa o produto matricial de dois vetores ou matrizes (o número de colunas do primeiro deve ser igual ao número de linhas do segundo). Este produto é também uma matriz quadrada de ordem 8. .

Com o exposto acima, a conclusão é a de que podem ser calculadas todas as características desejáveis para descrever as informações sobre Θ que os dados oferecem. Conforme visto, a distribuição *a posteriori* de probabilidades de Θ é totalmente especificada. Para facilitar o entendimento das inferências, os cálculos foram realizados por etapas. Consideremos, inicialmente, a proporção populacional das aves sem algum dos 3 parasitas, θ_{000} . Esse parâmetro tem distribuição marginal Beta com parâmetros

$$a = 208,25 \text{ e } b = 955 - 208,25 = 746,75.$$

A média e o desvio padrão serão, respectivamente,

$$M_{000} = 0,2181 \text{ e } D_{000} = 0,0134.$$

Usando-se a distribuição beta com esses parâmetros ou a aproximação para uma normal com mesma média e mesmo desvio padrão, o intervalo com 95% de credibilidade para θ_{000} é $(0,1919; 0,2442)$. Isto é, a proporção populacional de aves saudáveis, com alta probabilidade, é um número pertencente a este intervalo.

Considerando agora a população das aves com pelo menos um dos 3 tipos de parasitas. O parâmetro de interesse agora é um vetor de 7 componentes. Os componentes serão definidos como

$$\lambda_{cam} = \frac{\theta_{cam}}{1 - \theta_{000}}$$

Dessa forma a soma dos componentes do novo vetor será igual à unidade. Foquemos agora a atenção na proporção das aves que possuem todos os 3 parasitas dentro da população das aves com pelo menos um parasita, isto é, o parâmetro $\lambda_{111} = \theta_{111}/(1 - \theta_{000})$. Novamente, este parâmetro é distribuído segundo uma Beta, mas agora com parâmetros $a = 36,25$ e $b = 746,75 - 36,25 = 710,5$. A média e o Desvio padrão desse parâmetro são, respectivamente, os seguintes:

$$M_{111}^* = 0,04854 \quad \& \quad D_{111}^* = 0,00786$$

Com esta distribuição, obtemos o seguinte intervalo de 95% de credibilidade para λ_{111} :

$$(0,03337; 0,06395).$$

Para o restante da população, a proporção de unidades com apenas um ou com dois tipos de parasitas é representada por $\delta = 1 - \theta_{000} - \theta_{111}$. As seguintes quantidades são agora o foco das inferências:

$$\Pi = \frac{1}{\delta}(\theta_{001}, \theta_{010}, \theta_{100}, \theta_{011}, \theta_{101}, \theta_{110}) \quad \& \quad \pi = \frac{\theta_{001} + \theta_{010} + \theta_{100}}{\delta}$$

O parâmetro Π é o vetor das proporções da população das aves com apenas um ou dois parasitas e π é a proporção total das aves com apenas um parasita nesta sub-população. O vetor Π tem distribuição de Dirichlet com vetor de parâmetros igual à $(79,25; 275, 25, 91,25; 140,25; 15,25; 108,25)$. Já o parâmetro π tem distribuição Beta com parâmetros $a = 446,75$ e $b = 263,75$. O intervalo com 95% de credibilidade para π é facilmente obtido: $(0,5933; 0,6643)$.

Completando as estimações de interesse, o vetor Π é agora particionado no vetor de proporções de aves com apenas um parasita e no vetor de proporções de aves com exatamente dois tipos de parasitas. Denotemos esses dois vetores por:

$$\Sigma = (\sigma_{001}; \sigma_{010}; \sigma_{100}) = \left(\frac{1}{\theta_{001} + \theta_{010} + \theta_{100}} \right) (\theta_{001}; \theta_{010}; \theta_{100}) \quad \&$$

$$\Omega = (\omega_{011}; \omega_{101}; \omega_{110}) = \left(\frac{1}{\theta_{011} + \theta_{101} + \theta_{110}} \right) (\theta_{011}; \theta_{101}; \theta_{110}).$$

Ambos os vetores são distribuídos como Dirichlet, respectivamente, com vetores de parâmetros $(79,25; 276,25; 91,25)$ e $(140,25; 15,25; 108,25)$.

Lembremos que a soma dos elementos, tanto do vetor Σ quanto do vetor Ω , é igual à unidade. Assim, a região de credibilidade a ser construída é de dimensão 2 embora os vetores sejam de ordem 3. Por razões apenas estéticas foram escolhidos os pares $(\sigma_{010}; \sigma_{100})$ e $(\omega_{011}; \omega_{110})$ para a definição dos conjuntos de credibilidade das Figuras 5.3.2.2 e 5.3.2.3.

É natural que um leitor menos cuidadoso pergunte se não houve exagero teórico, ao considerar todo esse processo hierárquico para estimação dos parâmetros.

Por que não considerar os intervalos de credibilidade para os parâmetros originais do vetor Θ ? A razão para tal atitude é o fato de os elementos de Θ estarem associados negativamente. Isto limita a realização de um trabalho exclusivamente nas distribuições marginais de Θ . No entanto, como as regiões de credibilidade ilustram esse problema, foram apresentados na Tabela 5.3.2.4 os intervalos marginais para todos os componentes de Θ . Pode-se notar que a soma dos limites superiores dos intervalos irão produzir um número maior do que a unidade o que é um absurdo, pois a soma das proporções das diversas classificações somam 1.

Esse exemplo termina com um teste de significância. O interesse é procurar saber qual dos parasitas apresenta a maior incidência (prevalência) na população de aves. Os parâmetros de interesse neste caso são as proporções de aves que apresentam cada uma das parasitas. Isto é, os parâmetros a serem comparados são os seguintes:

$$\begin{aligned}\theta_{1\bullet\bullet} &= \theta_{111} + \theta_{110} + \theta_{101} + \theta_{100} \approx \text{Beta}(251;704), \\ \theta_{\bullet 1\bullet} &= \theta_{111} + \theta_{110} + \theta_{011} + \theta_{010} \approx \text{Beta}(561;394) \\ \theta_{\bullet\bullet 1} &= \theta_{111} + \theta_{011} + \theta_{101} + \theta_{001} \approx \text{Beta}(271;684).\end{aligned}$$

As hipóteses envolvendo estes parâmetros podem ser definidas como

$$\begin{aligned}H_{CA} : \theta_{1\bullet\bullet} < \theta_{\bullet 1\bullet} &\Leftrightarrow H_{CA} : \theta_{101} + \theta_{100} < \theta_{011} + \theta_{010}, \\ H_{AM} : \theta_{\bullet\bullet 1} < \theta_{\bullet 1\bullet} &\Leftrightarrow H_{AM} : \theta_{101} + \theta_{001} < \theta_{110} + \theta_{010} \quad (5.2.2.5) \\ H_{CM} : \theta_{\bullet\bullet 1} < \theta_{\bullet 1\bullet} &\Leftrightarrow H_{CM} : \theta_{011} + \theta_{001} < \theta_{110} + \theta_{100}.\end{aligned}$$

Para tornar simples o cálculo das probabilidades de cada uma destas hipóteses, os seguintes parâmetros foram definidos:

$$\begin{aligned}\Pi_{CA} &= \frac{\theta_{101} + \theta_{100}}{\theta_{101} + \theta_{100} + \theta_{011} + \theta_{010}} \approx \text{Beta}(106,5;416,5); \\ \Pi_{AM} &= \frac{\theta_{101} + \theta_{001}}{\theta_{101} + \theta_{001} + \theta_{110} + \theta_{010}} \approx \text{Beta}(94,5;384,5) \quad \& \\ \Pi_{CM} &= \frac{\theta_{011} + \theta_{001}}{\theta_{011} + \theta_{001} + \theta_{110} + \theta_{100}} \approx \text{Beta}(219,5;199,5).\end{aligned}$$

O leitor deve entender que o cálculo das probabilidades de cada uma das hipóteses é equivalente ao cálculo das probabilidades destes últimos parâmetros serem menores do que 0,5. Note que para estes cálculos foram considerados vetores divididos pela soma de seus elementos. Como antes, foram produzidas outras distribuições de Dirichlet ou distribuições Betas. Os cálculos necessários para tomadas de decisão sobre as hipóteses estão apresentados a seguir.

$$\begin{aligned}\Pr\{H_{CA}\} &= \Pr\{\Pi_{CA} < 1/2\} = 1 \\ \Pr\{H_{AM}\} &= \Pr\{\Pi_{AM} < 1/2\} = 1 \quad \& \\ \Pr\{H_{CM}\} &= \Pr\{\Pi_{CM} < 1/2\} = 0,164.\end{aligned}$$

A conclusão é a de que os Ácaros são as parasitas com maior prevalência entre os estudados. Os carrapatos são aqueles com menor prevalência.

Para aqueles acostumados com testes de significância, é aconselhável usar o teste baseado no valor da evidência, $ev = e\text{-value}$, apresentado no Capítulo 4. Testando se as prevalências são iguais, calculam-se as evidências, EV , apresentadas a seguir. Para definir as hipóteses sendo testadas basta substituir nas fórmulas 5.2.2.5 os símbolos de desigualdade pelos de igualdade.

1 – Hipótese 1: Carrapatos e Ácaros possuem a mesma prevalência:

$$EV(C=A) = 0,004\%$$

2 – Hipótese 2: Ácaros e Malófagos possuem a mesma prevalência:

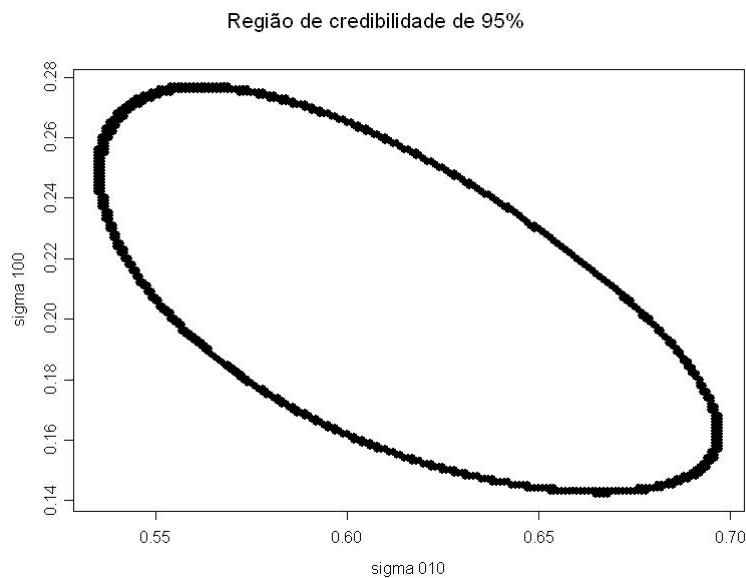
$$EV(A=M) = 0,004\%$$

3 – Hipótese 3: Carrapatos e Malófagos possuem a mesma prevalência:

$$EV(C=M) = 0,005\%$$

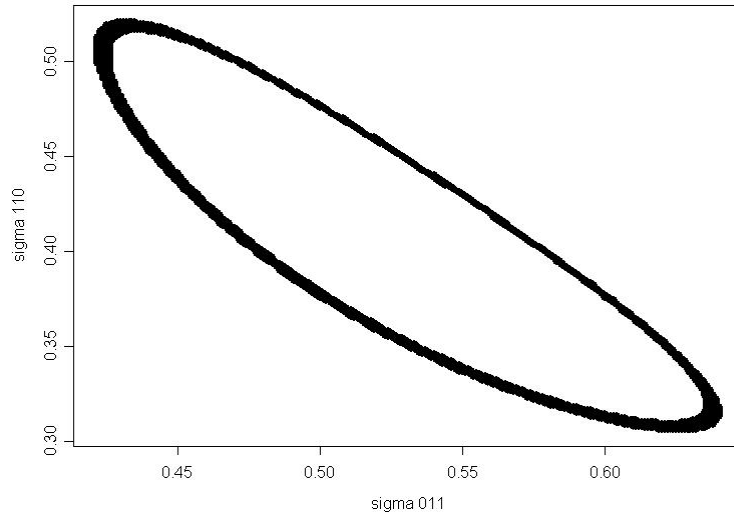
Com estes resultados pode-se afirmar que as prevalências são diferentes. A ordem da maior para a menor é a seguinte: Ácaros, Malófagos e Carrapatos. Completando o conjunto de técnicas estatísticas Bayesianas utilizadas, a Tabela 3 apresenta os intervalos de credibilidade para as proporções de incidência de cada um dos tipos parasitas.

5.3.2.2. Figura: Região de credibilidade de 95% para $(\sigma_{010}; \sigma_{100})$.



5.3.2.3. Figura: Região de credibilidade de 95% para $(\omega_{011}; \omega_{110})$.

Região de credibilidade de 95%



5.3.2.4. Tabela: Intervalos de credibilidade, com marginais Beta de parâmetros α e β .
LimInf e LimSup são os limites inferiores e superiores dos intervalos.

| Θ | α | β | Média | D. Padrão | LimInf | LimSup | Credibilidade |
|----------------|----------|---------|-------|-----------|--------|--------|---------------|
| θ_{000} | 208,25 | 746,75 | 0,218 | 0,013 | 0,192 | 0,244 | 95,02% |
| θ_{001} | 79,25 | 875,75 | 0,083 | 0,009 | 0,065 | 0,100 | 95,06% |
| θ_{010} | 276,25 | 678,75 | 0,289 | 0,015 | 0,261 | 0,318 | 95,01% |
| θ_{100} | 91,25 | 863,75 | 0,096 | 0,010 | 0,077 | 0,114 | 95,05% |
| θ_{011} | 140,25 | 814,75 | 0,147 | 0,011 | 0,124 | 0,169 | 95,03% |
| θ_{101} | 15,25 | 939,75 | 0,016 | 0,004 | 0,008 | 0,024 | 95,35% |
| θ_{110} | 108,25 | 846,75 | 0,113 | 0,010 | 0,093 | 0,133 | 95,04% |
| θ_{111} | 36,25 | 918,75 | 0,038 | 0,006 | 0,026 | 0,050 | 95,15% |

5.3.2.6. Tabela: Intervalos de credibilidade para a incidência dos 3 tipos de parasitas, utilizando novamente Betas com parâmetros α e β . **LimInf** e **LimSup** são os limites Inferior e superior dos intervalos.

| Parâmetro | α | β | Média | D. Padrão | LimInf | LimSup | Credibilidade |
|----------------|----------|---------|-------|-----------|--------|--------|---------------|
| $\theta_{1..}$ | 251 | 704 | 0,26 | 0,014 | 0,235 | 0,291 | 95% |
| $\theta_{.1.}$ | 561 | 394 | 0,59 | 0,016 | 0,556 | 0,619 | 95% |
| $\theta_{..1}$ | 271 | 684 | 0,28 | 0,015 | 0,255 | 0,312 | 95% |

5.4. PREVALÊNCIA E DIAGNÓSTICO DA FIBROMIALGIA NO MUNICÍPIO DO EMBU, SÃO PAULO.

Estas notas terminam com um problema no qual o autor participou desde a elaboração do projeto, passando pelo planejamento amostral, até a análise dos dados. Apenas não pode participar da coleta de dados, pois, além de não ter o tempo necessário, é leigo nos aspectos importantes do tratamento e avaliação da dor causada por problemas muscular esqueléticos. A coleta dos dados aqui discutidos exigiu uma expertise, principalmente em dor, que poucos fisioterapeutas possuem.

O envolvimento do autor com o trabalho científico do grupo de fisioterapeutas do departamento de fisioterapia da USP foi intenso. Este envolvimento foi iniciado quando o autor teve de se submeter a um tratamento longo e contínuo para diminuir o sofrimento com o problema reumático que o irá acompanhar para o resto da vida. Para poder juntar as necessidades do tratamento com o convívio agradável e produtivo das fisioterapeutas, o autor passou a se envolver com os trabalhos do grupo, justamente nas etapas das análises estatísticas. A Professora Amélia Pasqual Marques, responsável pela melhora de qualidade de vida do autor, juntamente com suas alunas, principalmente a mestranda Ana Assunção, dirigiu todo o trabalho, desde o projeto inicial até a observação minuciosa dos pacientes do EMBU. A análise estatística descrita nesta seção será parte de alguns artigos científicos que estão sendo elaborados. Fará parte também da dissertação de mestrado da Ana. O autor é grato a todas essas pessoas maravilhosas que dedicam a carreira ao trabalho da diminuição do sofrimento humano. Somente com a presença no ambiente é que alguém pode apreciar a alegria e dedicação que estas pessoas colocam no trabalho diário do grupo. É uma honra para o autor ser considerado membro do grupo. Não somente como paciente, mas também como observador e evidentemente estatístico!

5.4.1. Fibromialgia e Qualidade de Vida

Fibromialgia é uma síndrome reumática caracterizada por dores muscular-esqueléticas difusas e crônicas – em várias partes do corpo e durante um longo período.

do. O diagnóstico é estabelecido quando se observam pelo menos 11 pontos dolorosos à palpação dos 18 pontos anatômicos específicos, conhecidos como *tender points*. Estes sinais estão freqüentemente associados à fadiga, distúrbios do sono e rigidez matinal. O acompanhamento dos pacientes com Fibromialgia baseia-se, principalmente, em aspectos de natureza clínica. Conseqüentemente, observa-se um alto grau de subjetividade, tanto por parte do paciente como do médico (Martinez et all, 1995).

A Fibromialgia causa um impacto negativo na qualidade de vida de pessoas em idade produtiva de trabalho, pois acarreta um decréscimo de energia, atingindo diretamente a capacidade para o trabalho. Este quadro agrava os aspectos psicológicos, como depressão e ansiedade.

O objetivo principal deste trabalho é a estimativa da prevalência da Fibromialgia na população adulta do município de Embu. Um segundo objetivo é a obtenção de um índice diagnóstico baseado em questionários usuais no ambiente Hospitalar ou ao profissional da saúde. O diagnóstico da Fibromialgia é realizado por profissionais especialistas que estudam os 18 *tender points* e verificam quanto destes pontos suportam pressão acima de um determinado limite estabelecido. Se mais de 11 pontos estiverem com valores inferiores ao limite, o paciente é diagnosticado como Fibromialgico. A denominação síndrome é justamente pelo fato do diagnóstico se basear apenas nos sintomas. Por outro lado, postos de saúde em geral não possuem em seus quadros profissionais com a expertise necessária para esse diagnóstico. Torna-se assim relevante o estudo de outras formas de diagnósticos que independam da presença do profissional. A tentativa é a aplicação de questionários que medindo qualidade de vida podem produzir suspeita da presença de Fibromialgia. A tentativa de definir-se um índice diagnóstico, baseado nos resultados da aplicação de questionários sobre qualidade de vida, pode ser um bom caminho para uma melhor avaliação da prevalência dessa doença na população em geral.

A pesquisa foi realizada com a população adulta na faixa etária de 35 a 60 anos do município Embu. O trabalho de coleta dos dados foi estruturado em duas etapas. Na primeira, foi realizado o levantamento das pessoas cadastradas nas nove Unidades Básicas de Saúde do município nos anos de 2002 e 2003. Em seguida, através de contato telefônico, os indivíduos selecionados (768 pessoas) responderam a um questionário com as seguintes informações:

- Dados pessoais: idade, sexo;
- Características da dor: presença ou não de dor, tempo de dor, local da dor;
- Qualidade do sono: presença de distúrbio no sono (acorda cansado ou não).

A segunda etapa consistiu na convocação dos pacientes, entrevistados na primeira etapa, para o comparecimento nas Unidades Básicas de Saúde (UBS) mais próximas de suas residências. Na visita as UBS, os indivíduos submeteram-se e responderam a um questionário mais completo do que o primeiro. Os indivíduos compareceram (305 pessoas) foram avaliados pelas pesquisadoras nas UBS do Embu: todos estes indivíduos foram submetidos aos mesmos protocolos, que incluíram as seguintes informações ou respostas à aplicação dos instrumentos:

- Dados pessoais: idade, peso, altura, escolaridade, profissão e estado civil;

- Escala Analógica Visual da Dor (VAS): Avalia a intensidade da dor, consistindo-se de uma reta de 10 centímetros de comprimento desprovida de números, na qual há apenas indicação no extremo esquerdo de “ausência de dor” e no extremo direito “dor insuportável”. Quanto maior o escore, maior a intensidade de dor. O paciente é instruído a marcar um ponto que indique a intensidade de dor que sente;

- Post Sleep Inventory (PSI): Avalia a qualidade do sono, sendo dividido em três categorias: pré-sono (ao deitar-se), durante o sono e pós-sono (ao acordar), sendo 8, 13 e 9 itens, respectivamente. Cada item é composto por dois extremos possíveis (bom sono ou mau sono). Em cada resposta, a pessoa é orientada a graduar a sua percepção da noite de sono em uma escala de 1 a 13. Para cada item, as respostas “Não sei” e “Não importante” são consideradas como informação ausente. Ao final da aplicação, a graduação apontada pelos participantes em cada item é somada, obtendo-se a pontuação final (escala numérica). Quanto maior o escore obtido, melhor será a qualidade do sono;

- Fibromyalgia Impact Questionnaire (FIQ): Avalia os principais sintomas da Fibromialgia e o seu impacto na qualidade de vida. Este questionário envolve 10 questões relacionadas à capacidade funcional, situação profissional, distúrbios psicológicos e, principalmente, os sintomas. A primeira questão produz um inteiro de 0 a 30, a segunda e a terceira questões variam de 0 a 7. As sete questões restantes se utilizam da escala analógica visual, semelhante à utilizada pelo questionário VAS, que varia de 0 a 10. As contagens resultantes passarão por um procedimento de normalização, de modo que todas as contagens sejam expressas em unidades similares (Burckhardt, C.S.; Clark, S.R. e Bennett, R.M., 1991). Quanto maior o escore, maior é o impacto da Fibromialgia (escala numérica), exceto para o item “sentiu-se bem”, onde um maior escore indica melhora. Para a versão brasileira veja Marques et al (2006);

- Avaliação funcional pelo Stanford Health Questionnaire (HAQ): Avalia o grau de dificuldade de realizar as atividades de vida diária. Este questionário apresenta questões referentes a 8 atividades: vestir, levantar, comer, caminhar, higiene, alcançar, apreensão e outras. As oito questões possíveis possuem de 2 a 3 itens. De fato, o número de itens das 8 questões são, respectivamente, 2, 2, 3, 2, 3, 2, 3 e 3. Para cada item é atribuído um grau quanto à capacidade de realização, que varia de 0 a 3 (sem dificuldade, pouca dificuldade, muita dificuldade e incapaz, respectivamente). A pontuação varia de 0 a 60 no questionário total (escala numérica). Quanto maior a pontuação, maior será o grau de dificuldade de realizar as atividades.

- Escala de fadiga de Chalder modificada: Avalia a fadiga sob dois aspectos: sintomas físicos, constituído por 8 itens, e sintomas emocionais, constituído por 6 itens. As pontuações variam de 8 a 32 e de 6 a 24, respectivamente. No total, temos a pontuação variando de 14 a 56, sendo que os menores valores estão relacionados à ausência de fadiga. As respostas das questões são obtidas através das alternativas: melhor do que o habitual, não mais que o habitual, pior que o habitual e muito pior que o habitual (escala numérica);

- SF – 36: É um instrumento genérico de avaliação de qualidade de vida (Ware e Sherborne, 1992). Este questionário consiste de 36 itens, estratificados em oito grupos: capacidade funcional (10 itens), aspectos físicos (4 itens), dor (2

itens), estado geral de saúde (5 itens), vitalidade (4 itens), aspectos sociais (2 itens), aspectos emocionais (3 itens), saúde mental (5 itens) e mais uma questão de avaliação comparativa entre as condições de saúde atual e a de um ano atrás. Para cada alternativa de resposta é dado um número. Após a obtenção da contagem de pontos de cada um dos oito componentes específicos, uma transformação é aplicada para que os valores variem no intervalo de 0 a 100. Quanto maior o escore, melhor é a qualidade de vida;

- Depressão: Para este estudo, utilizou-se escala de Depressão de BECK (Beck, et. Al., 1961). Este instrumento consiste em 21 grupos de afirmações. Ao final, considera-se o número total de afirmações que recebem concordância e obtemos o escore final, sendo que a pontuação máxima é 63. Segundo Meyer (2000), um escore acima de 16 indica possível depressão;

- Ansiedade: Considerada como um sintoma secundário, a ansiedade é muito comum e freqüentemente severa no Fibromiálgico. Um dos instrumentos mais utilizados para avaliar a ansiedade é o Inventário de Ansiedade Traço-Estado (IDATE) proposto por Spielberger et all (1979) para medir o traço da ansiedade - propensão à ansiedade - e o estado da ansiedade (tensão, nervosismo, preocupação e apreensão). Este inventário é composto por duas escalas distintas:

- *A escala de traço de ansiedade:* consiste de 20 afirmações, requisitando dos sujeitos a descrição de como, geralmente, eles se sentem;
- *A escala de estado de ansiedade:* consiste em 20 afirmações, requisitando dos sujeitos à indicação de como eles se sentem em determinado momento.

O IDATE foi planejado para ser auto-aplicável, o que pode ser feito de maneira individual ou em grupo. Cada afirmação é composta por escores de 1 a 4. Nas duas escalas, escores altos indicam mais ansiedade. A pontuação em cada questionário varia de 20 a 80.

- Dolorimetria: Avalia o limiar da dor nos 18 pontos sensíveis, tender points, através de uma avaliação física com um aparelho chamado Dolorímetro, que mede a pressão exercida pelo examinador em cada ponto. A unidade de medida é kg/cm². Desta forma, um paciente com limiar de dor baixo será caracterizado por suportar apenas baixos valores de pressão em pontos específicos do corpo. Por convenção, um ponto é considerado doloroso se a pressão suportada é inferior a 2,6 kg/cm².

Note que o último instrumento acima é o que realiza o diagnóstico da Fibromialgia. Se houver uma forte associação entre as respostas à dolorimetria e aos questionários, poder-se-ia saber, mesmo sem a presença do especialista, a taxa de ocorrência da síndrome em uma UBS conveniada e que aplica os outros instrumentos de avaliação: questionários, na maioria, auto-aplicáveis.

5.4.2. Experimento e Coleta de Dados

Por convenção iremos denominar todos os indivíduos do estudo de pacientes, pois em algum momento no passado passaram por atendimento nas UBS's do EMBU. Repetindo a descrição anterior, A pesquisa foi realizada com a população adulta na faixa etária de 35 a 60 anos do município Embu. O trabalho de coleta dos dados foi estruturado em duas etapas.

- i. Etapa 1: Realizou-se inicialmente um levantamento dos pacientes cadastrados nas nove UBS do município do Embu no período de 2002 e 2003. Em seguida, por contato telefônico, 768 pacientes responderam a um questionário com perguntas cujas respostas permitiram classificar os pacientes de duas formas: tempo de dor e tipo da dor. A classificação por tempo foi em três categorias: SD (sem dor), DA (dor aguda, por menos de três meses) e DD (dor crônica, por mais de três meses). A classificação por tipo foi também em três categorias: sd (sem dor), dr (dor regional) e dd (dor difusa). Além disso, relativo ao sono, o entrevistado respondeu se acorda, em geral, cansado ou com disposição normal. O cansaço ao acordar é uma característica presente na vida do Fibromialgico.

- ii. Etapa 2: O entrevistado na primeira etapa foi convidado para submeter-se a uma avaliação de qualidade de vida na UBS mais próxima de sua residência. Como esperado, nem todos responderam ao convite, mas o comparecimento foi expressivo: 305 pacientes. É evidente que um paciente em bom estado de saúde deve preferir participar de outras atividades no final de semana do que comparecer para uma avaliação em posto de saúde. Relativamente, o comparecimento maior foi do grupo dos que sofrem problemas de dor e depressão. Todos esses pacientes participaram, inicialmente, de uma entrevista onde se confirmaram as informações básicas como idade, peso e principalmente as classificações anteriores. Em seguida realizaram-se as pesquisas sobre qualidade de vida. Todos os instrumentos descritos anteriormente foram aplicados e assim foi possível conhecer o número, 19, de Fibromiálgicos nesta amostra de 305 pacientes. Entre os estudados nessa segunda etapa, 106 pacientes confirmaram a classificação DC/dd; isto é, dor crônica e difusa. Essa etapa termina assim com a classificação em 4 categorias, a saber: a) C_0 formada por pacientes sem dor (SD/sd); b) C_1 , formada por pacientes com dor crônica e regional ou dor aguda regional ou difusa (DC/dr + DA/dr + DA/dd); c) C_2 , formada por pacientes com dor crônica e difusa (DC/dd) mas sem Fibromialgia; e finalmente, d) C_3 formada por pacientes DC/dd com Fibromialgia. Cada um dos pacientes dessa amostra foi passado pela aferição dos 10 instrumentos de avaliação de qualidade de vida.

Para a análise apresentada aqui descrita, utilizamos normalizações dos instrumentos, de forma que cada um dos índices de qualidade de vida assumam valores no mesmo intervalo de 0 a 10. Alguns dos instrumentos são constituídos por diversos itens que também variam em diferentes intervalos. Na construção do resumo desses itens, efetuamos uma normalização para que todos os itens possuam o mesmo peso no número final. Obtivemos assim valores correspondentes aos 10 instrumentos usados. Alguns instrumentos produzem números que altos indicam boa qualidade de vida e baixos pouca qualidade de vida. Outros instrumentos estão exatamente postos no sentido oposto. Dessa forma, realizamos primeiro a normalização dos números para que todos fossem números compreendidos entre 0 e 10. Em seguida invertemos a direção de alguns instrumentos de forma que todos os números finais indiquem baixa qualidade de vida quando apresentam valores mais próximos de 10 e boa qualidade de vida quando estão próximos de zero. Os resumos finais dos resultados dos 10 instrumentos, obtidos pelos pacientes que compareceram na Etapa 2 estão dispostos em parte na Tabela 5.3.2.2. Os resultados de todos os pacientes estão apresentados no

site <http://www.ime.usp.br/~cpereira/livro/Embu.xls>.

A Tabela 5.4.2.1. descreve os grupos como observados nas duas etapas da coleta de dados. Estes números formam a base do trabalho de estimação da prevalência da Fibromialgia, objetivo principal desta seção.

Com respeito à composição dos grupos, a primeira etapa é mais informativa, entretanto somente na segunda etapa é que pacientes com fibromialgia são diagnosticados. Esta é uma típica situação de dados censurados. A intuição nos obriga a usar, evidentemente, a segunda etapa para a estimação da proporção de Fibromiálgicos na população de pacientes com dor difusa e crônica, característica principal do fibromiálgico. Para a estimação da proporção de pacientes com dor difusa e crônica deve-se usar a primeira etapa que além de usar uma amostra maior, esta foi obtida sem alguma indução. Já a segunda etapa certamente motivou mais os indivíduos com sofrimento de dor do que os sem dor: uma indução motivada pela ética das pesquisadoras que realizaram um atendimento (gratuito) intenso com as pacientes. É claro que essa indução não prejudicou a composição dos grupos que já sofrem de dor. Como verão na seqüência, as estimativas produzidas são concordantes com os números apresentados na literatura.

5.4.2.1. Tabela: Composição da Amostra nas Duas Etapas

| Categorias | D O R | | E T A P A | |
|----------------|----------|--------------|-----------|------------|
| | Tempo | Tipo | Primeira | Segunda |
| C_0 | Sem | Sem | 185 | 47 |
| C_1 | Crônica | Regional | 349 | 144 |
| C_1 | Aguda | Regional | 39 | 7 |
| C_1 | Aguda | Difusa | 7 | 0 |
| C_2 ou C_3 | Crônica | Difusa | 188 | 106 |
| C_2 | Sem | Fibromialgia | Sem Info | 87 |
| C_3 | Com | Fibromialgia | Sem Info | 19 |
| Tamanho | Amostrai | | 768 | 304 |

5.4.2.2. Tabela: Exemplos de valores observados para cada instrumento.

| Paciente | UBS | Categ | Age | IMC | IVAS | Isono | IHAQ | IFIQ | ISF-36 | IFadiga | IAnsie | IDepre |
|----------|-------------|-------|-----|-------|------|-------|------|------|--------|---------|--------|--------|
| 242 | Pinheirinho | 3 | 47 | 19,15 | 9,20 | 4,63 | 2,29 | 6,45 | 8,89 | 8,68 | 8,83 | 7,30 |
| 461 | S Eduardo | 3 | 55 | 31,96 | 7,70 | 4,74 | 2,71 | 6,17 | 5,57 | 7,01 | 5,33 | 1,43 |
| 759 | Centro | 3 | 47 | 33,23 | 8,90 | 5,53 | 3,68 | 6,52 | 6,59 | 6,60 | 5,17 | 2,22 |
| 95 | Centro | 2 | 55 | 33,78 | 5,50 | 4,61 | 0,35 | 3,79 | 6,69 | 5,90 | 6,75 | 3,49 |
| 209 | Itatuba | 2 | 37 | 27,10 | 4,10 | 5,90 | 0,90 | 5,35 | 5,45 | 7,01 | 5,75 | 3,17 |
| 498 | S Eduardo | 2 | 45 | 30,46 | 6,70 | 6,21 | 2,29 | 7,33 | 6,29 | 6,81 | 5,67 | 3,17 |
| 574 | S Marcos | 1 | 51 | 21,00 | 6,80 | 3,65 | 2,71 | 5,62 | 6,20 | 8,89 | 4,92 | 2,86 |
| 734 | V Alegre | 1 | 39 | 23,94 | 7,30 | 5,18 | 1,74 | 5,47 | 6,69 | 9,10 | 6,83 | 3,17 |
| 742 | Ressaca | 1 | 48 | 24,13 | 9,80 | 4,31 | 2,36 | 5,24 | 5,19 | 5,21 | 4,08 | 3,49 |
| 3 | Centro | 0 | 53 | 28,52 | 0,10 | 2,86 | 0,00 | 3,02 | 1,93 | 6,11 | 2,75 | 1,43 |
| 743 | S Tereza | 0 | 60 | 33,39 | 0,00 | 3,23 | 0,21 | 1,06 | 1,50 | 4,10 | 3,75 | 0,95 |
| 764 | S Emília | 0 | 55 | 31,02 | 1,00 | 2,52 | 1,39 | 1,90 | 2,97 | 3,61 | 1,42 | 1,11 |

5.4.3. Estimativa de Prevalência da Fibromialgia

Como discutido anteriormente, a população foi dividida em quatro subpopulações ou categorias: C_0 , C_1 , C_2 e C_3 . O vetor das proporções de indivíduos em cada uma das classes, $(\pi_0, \pi_1, \pi_2, \pi_3)$, é o objeto principal do estudo. A prevalência da Fibromialgia na população é o parâmetro π_3 , que aqui é estimado.

Usando fatos descritos na literatura, para a definição da densidade *a priori* considerou-se que o que se espera de uma população de idosos, caso do presente estudo, seria a observação de um vetor de proporções $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ correspondentes as subpopulações de saudáveis, C_0 , com dor aguda, C_1 e com dor crônica, $C_{23} = C_2 \cup C_3$, respectivamente. Assim, como densidade *a priori* para a estimação das proporções populacionais, π_0 , π_1 , e $\pi_{23} = \pi_2 + \pi_3$, foi considerada uma distribuição de Dirichlet com parâmetros. $\frac{1}{2}$, 1 e $\frac{1}{2}$.

A distribuição a posteriori, do vetor (π_0, π_1, π_{23}) , consequência da operação Bayesiana da priori com a verossimilhança (amostra da primeira etapa com 768 pacientes), é uma distribuição de Dirichlet com parâmetros 185,5, 396 e 188,5. Recordando mais uma vez as características da distribuição de Dirichlet, a densidade *a posteriori* neste caso tem a seguinte expressão:

$$f(\pi_0, \pi_1, \pi_{23}) = \frac{\Gamma(770) \pi_0^{184,5} \pi_1^{395} \pi_{23}^{188,5}}{\Gamma(185,5) \Gamma(396) \Gamma(188,5)}$$

A moda e a média a posteriori do vetor de parâmetros (π_0, π_1, π_{23}) , consideradas como estimativas Bayesianas alternativas, neste caso praticamente coincidem e são iguais ao vetor $(M_0; M_1; M_{23}) = (0,2409; 0,5143; 0,2448)$.

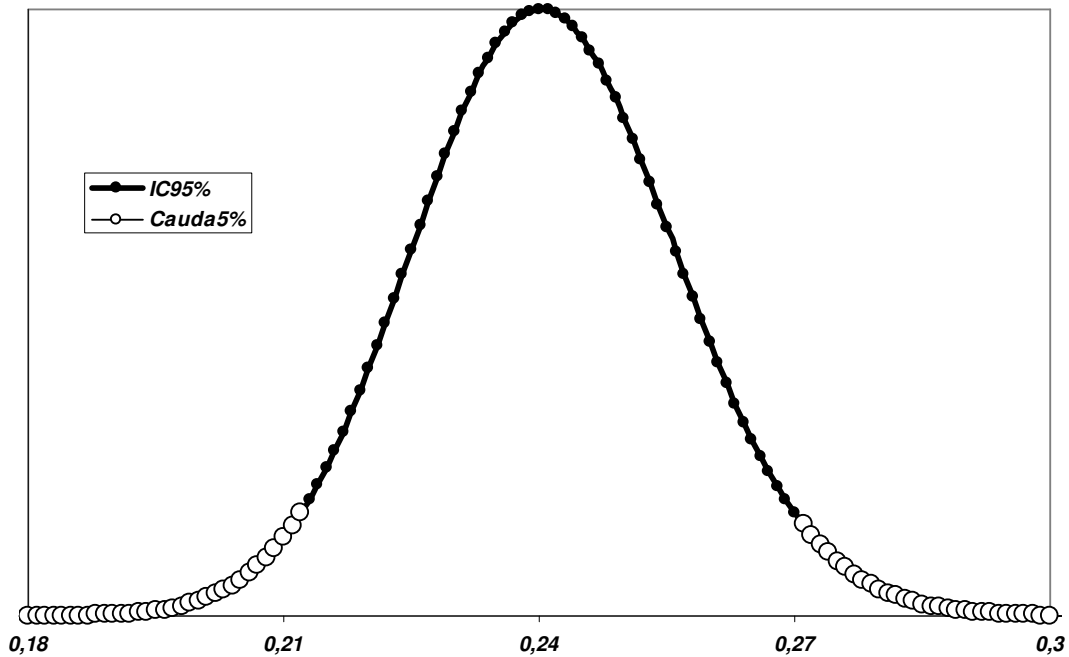
Completando essa primeira parte da análise apresentamos o vetor dos desvios

padrões e das correlações a posteriori do vetor de parâmetros acima.

Desvio Padrão: $(D_0, D_1, D_{23}) = (0,0154; 0,0180; 0,0155)$
 Correlação: $(R_{01}, R_{0(23)}, R_{1(23)}) = (-0,5797; -0,3207; -0,5859)$

Como ilustração, encontre abaixo, Figura 5.4.3.1.A, a densidade marginal da proporção π_0 de saudáveis. Contudo, o gráfico ilustrativo, Figura 5.4.3.1.B, é o bidimensional que apresenta o conjunto de credibilidade para (π_0, π_1) . As diversas curvas de nível da distribuição desse parâmetro estão também ilustradas na figura.

5.4.3.1.A Figura: Densidade da Proporção de saudáveis com intervalo de 95% de Credibilidade: [0,2114;0,2717]



- i. A distribuição a priori de $(\pi_0, \pi_1, \pi_2, \pi_3)$ é $D_4(0,5; 1; 0,4; 0,1)$, Dirichlet com parâmetro $(0,5; 1; 0,4; 0,1)$;
- ii. A distribuição a posteriori de (π_0, π_1, π_{23}) , considerando os dados obtidos na primeira etapa é $D_3(185,5; 396; 188,5)$;
- iii. A distribuição de λ é beta com parâmetro $(0,1; 0,4)$ a priori e, considerando os dados da segunda etapa, é beta com parâmetro $(19,1; 87,4)$, a posteriori.
- iv. Os parâmetros (π_0, π_1, π_{23}) e λ são independentes tanto a priori quanto a posteriori.
- v. O parâmetro de interesse, prevalência de Fibromialgia, é definido como $\pi_3 = \lambda\pi_{23}$. Assim, a média a posteriori é

$$M_3 = E(\pi_3) = E(\lambda\pi_{23}) = E(\lambda)E(\pi_{23})$$

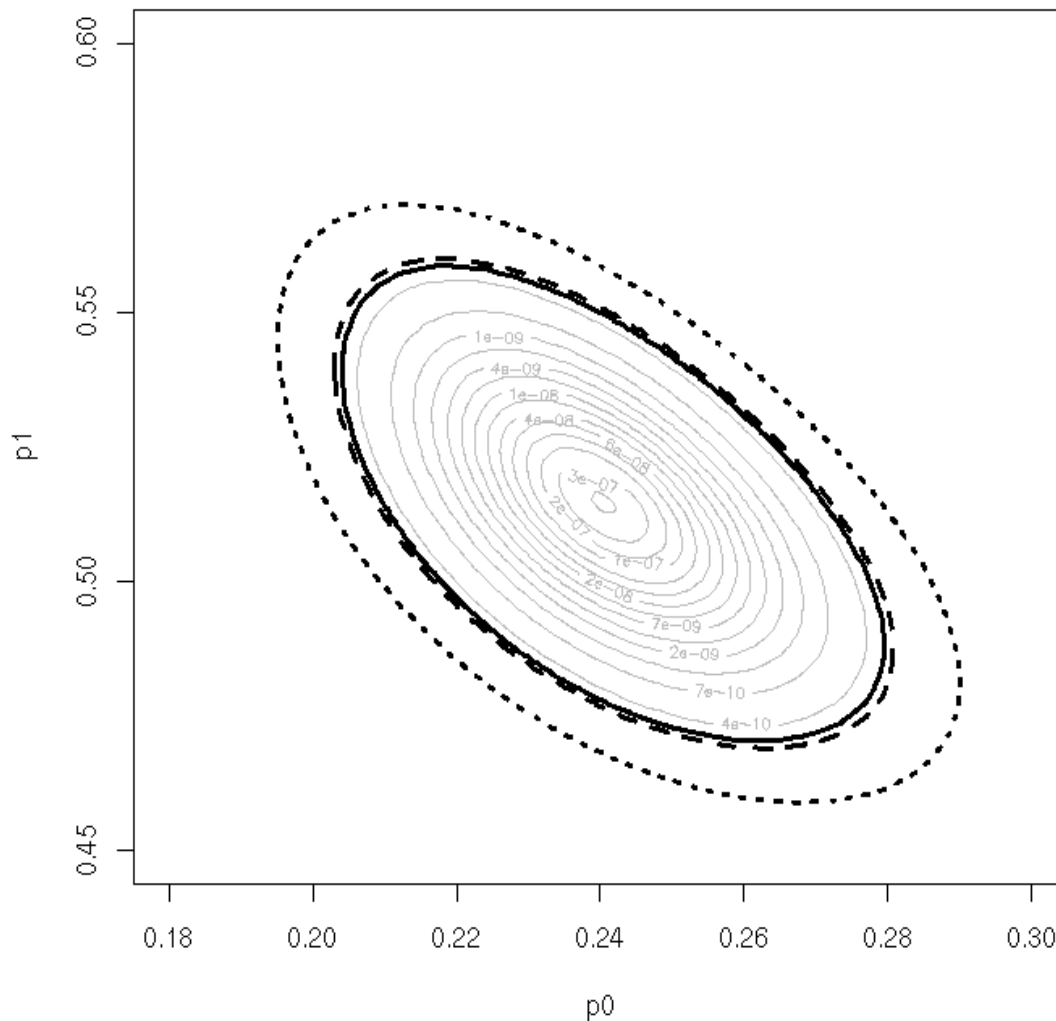
Concluindo, $(19,1 \div 106,5)(188,5 \div 770) = 0,18 \times 0,24 = 0,0439$ é uma estimativa de Bayes da prevalência. Como a literatura sugere que a prevalência esteja entre 0,03 e 0,06, a população idosa do EMBU segue o que é

esperado de uma população de idosos. Para completar o trabalho de estimação e como $\pi_2 = \pi_{23} - \pi_3$, note-se que

$$M_2 = E(\pi_2) = E(\pi_{23}) - E(\pi_3) = M_{23} - M_3 = 0,2448 - 0,0439 = 0,2009.$$

Isto é, $(M_0; M_1; M_2; M_3) = (0,2409; 0,5143; 0,2009; 0,0439)$ é a estimativa de Bayes do parâmetro de interesse, $(\pi_0, \pi_1, \pi_2, \pi_3)$.

Figura 5.4.3.1.B: Conjunto de Credibilidade de 95% para (π_0, π_1) .

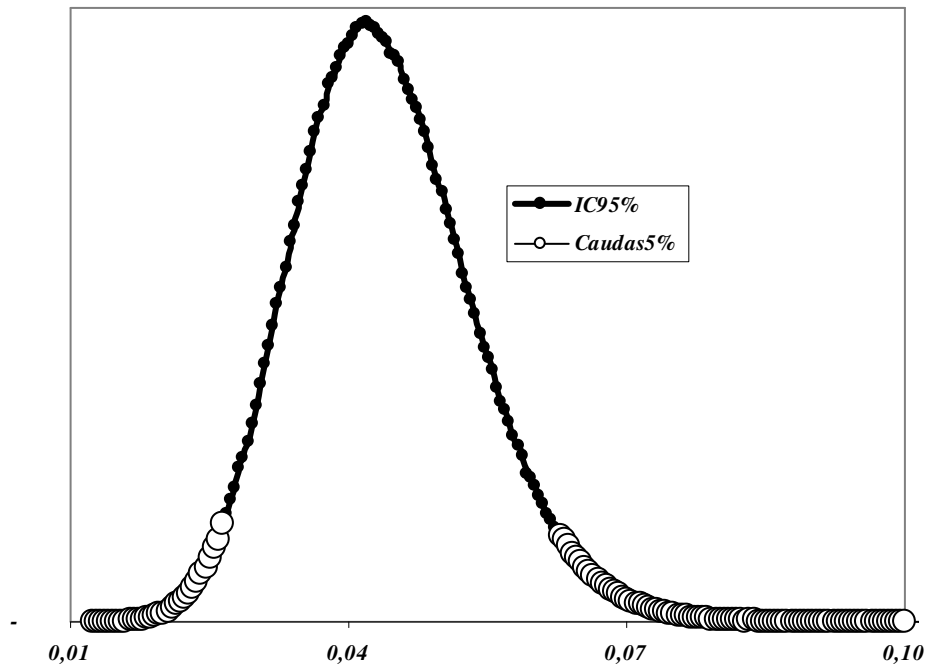


- vi. A variância da distribuição a posteriori de π_3 é obtida usando-se a seguinte propriedade:

$$V(\pi_3) = E(\lambda^2)E[(\pi_{23})^2] - [E(\lambda)E(\pi_{23})]^2 = \\ V(\pi_{23})V(\lambda) - V(\pi_{23})[E(\lambda)]^2 - V(\lambda)[E(\pi_{23})]^2 = 9 \times 10^{-5}.$$

O desvio padrão neste caso é $0,0095$. Usando uma aproximação normal o intervalo de credibilidade de **95%** seria $[0,0253; 0,0625]$. O intervalo calculado, usando-se as distribuições exatas e simulações de 1 milhão de pontos, é $[0,026; 0,063]$. A distribuição simulada da prevalência é apresentada pela Figura 5.4.3.2.

5.4.3.2 Figura: Densidade simulada da prevalência e intervalo de credibilidade de 95% (0,026;0,063)



5.4.4. Índice Diagnóstico

Nesta seção os dados da segunda etapa serão usados para a construção de um índice diagnóstico. Como discutido anteriormente, a Fibromialgia é definida após a observação dos valores dos pontos sensíveis, *the tender points*. Mas para esse tipo de classificação de um paciente necessita-se da presença de um especialista. Contudo, com o uso de um índice, baseado em resultados da aplicação de questionários padrões sobre qualidade de vida, será mostrado que o diagnóstico pode ser direcionado.

O índice foi construído da seguinte forma:

i. Inicialmente normalizou-se os questionários de forma que todos sejam de severidade: isto é, quanto maior mais severo é o caso. Além disso, após a normalização, todos os oito questionários produzem números entre zero e um.

ii. Com os resultados obtidos pela aplicação dos oito questionários considerou-se a média de resultados, denominado de Índice Geral ou IG. Como diferentes clínicas usam diferentes grupos de questionários, consideramos além da média geral, a média entre os resultados do VAS, HAQ, FIQ, Sono e Fadiga. Esse índice foi denominado de Índice Ana, IA, em homenagem à fisioterapeuta que assim o definiu. O terceiro índice é denominado Índice Cris, IC, em homenagem a outra fisioterapeuta. Sua definição é a média entre VAS, Depressão, Ansiedade e SF-36. O SF-36, um questionário sobre qualidade de vida, é consagrado entre os profissionais da saúde.

Para que se entenda que os índices são consistentes com o problema do diagnóstico, a Tabela 5.4.4.1 apresenta as médias por categorias de severidade. Com essas médias pode-se deduzir que os índices servem como medida de qualidade de vida. Com os percentis de ordens 47/304, (47+51)/304 e (304-19)/304 definiu-se os cortes

de classificação para cada um dos índices. Os cortes de classificação para cada um dos índices são:

- i. IG: 2,25; 4,73; e 6,20
- ii. IA: 2,33; 4,90; e 6,47 e
- iii. IC: 2,05; 4,83; e 6,78.

5.4.4.1. Tabela: Média dos índices por categoria

| Categoria | Amostra | IG | IA | IC |
|------------------|----------------|-------------|-------------|-------------|
| C0 | 47 | 2,47 | 2,53 | 2,33 |
| C1 | 151 | 3,93 | 4,14 | 3,97 |
| C2 | 87 | 4,59 | 4,79 | 4,71 |
| C3 | 19 | 5,39 | 5,69 | 5,64 |
| Geral | 304 | 3,98 | 4,17 | 4,03 |

Após escolher o índice para uso, a clínica faz anotações da seguinte forma: se o paciente obtiver um escore não superior ao primeiro percentil será designado (---); se o escore for acima do primeiro e não superior ao segundo percentil será designado (--+); se for acima do segundo e não superior ao terceiro percentil será (-++) e finalmente, se for acima do terceiro terá uma anotação de (+++).

A seguir o trabalho é feito utilizando-se o índice geral, IG, e ao leitor recomendamos repetir os cálculos com os outros índices para entender a simplicidade do procedimento. Lembre-se que os dados estão disponíveis em

<http://www.ime.usp.br/~cpereira/livro/Embu.xls>.

A seguir descrevemos as etapas dos cálculos. A Tabela 5.4.4.2 apresenta as frequências amostrais em cada classificação para todas as categorias: C₀, C₁, C₂ e C₃. Para cada uma das categorias, a priori, utilizaremos como parâmetro da distribuição de Dirichlet de ordem 4 o vetor (0,5;1;0,4;0,1). Note que essa escolha tem o objetivo de manter o mesmo nível de imprecisão de uma uniforme no caso de apenas uma dimensão. Como na uniforme, a soma dos parâmetros é 2. A Tabela 5.4.4.3 apresenta as médias das distribuições a posteriori, Dirichlet, para cada uma das categorias. Essas médias são simplesmente as frequências relativas, calculadas com as frequências amostrais acrescidas dos parâmetros da priori.

Com os dados da literatura, as proporções descritas de indivíduos em cada classe podem ser usadas como probabilidades a priori de um futuro indivíduo pertencer a cada uma dessas classes: 25% para C₀, 50% para C₁, 20% para C₂ e 5% para C₃. Com essas probabilidades e as médias a posteriori dentro de cada categoria, obtém-se a distribuição conjunta de classificação e categoria. Tabela 5.4.4.4 apresenta a distribuição conjunta com a marginal da classificação. Com esses dados, podem-se obter as probabilidades a posteriori de um indivíduo pertencer a uma categoria tendo recebido uma determinada classificação. A seguinte seqüência de tabelas ilustra bem o procedimento de obtenção das probabilidades de um indivíduo pertencer às diversas categorias alternativas conhecida a sua classificação.

5.4.4.2. Figura: Frequências amostrais: Classificação por Categoria

| Classif. Catego | (---) | (--+) | (-++) | (+++) | Total |
|----------------------|-----------|------------|-----------|-----------|------------|
| C₀ | 25 | 20 | 1 | 1 | 47 |
| C₁ | 20 | 82 | 43 | 6 | 151 |
| C₂ | 2 | 43 | 36 | 6 | 87 |
| C₃ | 0 | 6 | 7 | 6 | 19 |
| Soma | 47 | 151 | 87 | 19 | 304 |

5.4.4.3. Figura: Distribuição conjunta: Categoria e Classificação

| Classif. Catego | (---) | (--+) | (-++) | (+++) | Priori |
|----------------------|---------|---------|---------|---------|--------|
| C₀ | 0,13010 | 0,10459 | 0,00765 | 0,00765 | 0,25 |
| C₁ | 0,06699 | 0,26961 | 0,14216 | 0,02124 | 0,5 |
| C₂ | 0,00562 | 0,09775 | 0,08202 | 0,01461 | 0,2 |
| C₃ | 0,00119 | 0,01548 | 0,01786 | 0,01548 | 0,05 |
| Marginal | 0,20390 | 0,48743 | 0,24969 | 0,05898 | 1 |

5.4.4.4. Figura: Distribuições posteriori

| Classif. Catego | (---) | (--+) | (-++) | (+++) | Priori |
|----------------------|---------|---------|---------|---------|--------|
| C₀ | 0,63806 | 0,21458 | 0,03065 | 0,12976 | 0,25 |
| C₁ | 0,32855 | 0,55312 | 0,56933 | 0,36017 | 0,5 |
| C₂ | 0,02755 | 0,20055 | 0,3285 | 0,24766 | 0,2 |
| C₃ | 0,00584 | 0,03175 | 0,07152 | 0,26241 | 0,05 |
| Soma | 1 | 1 | 1 | 1 | 1 |

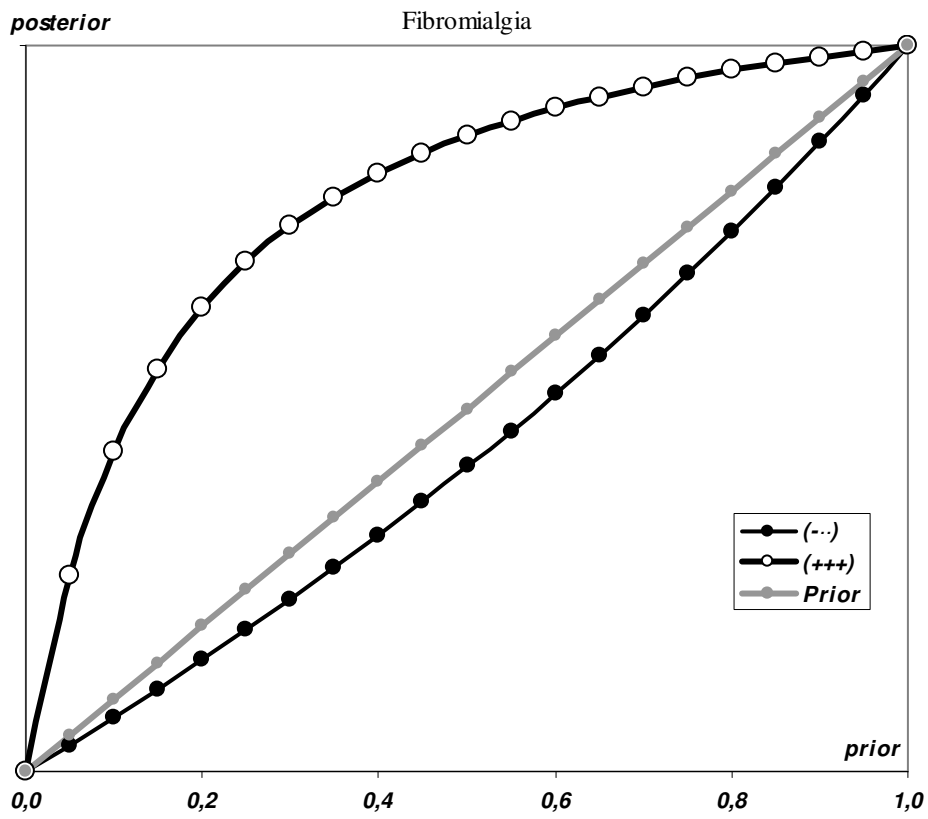
Tabela 5.4.4.2 são as frequências amostrais obtidas após a definição dos cortes de classificação que foram baseados nos tamanhos das categorias obtidos na segunda etapa da coleta de dados. Evidentemente, a próxima etapa de estudos tem como objetivo a obtenção de um procedimento de regressão logística genuinamente bayesiano. Deseja-se na verdade identificar aqueles instrumentos que realmente são efetivos na identificação da verdadeira categoria a qual o paciente pertence.

Note-se que após a correta alocação dos pacientes pelas categorias, definiu-se amostras de subpopulações definidas por ordem de gravidade do problema de saúde dos pacientes. Com os resultados obtidos com as respostas aos questionários ou instrumentos, decidiu-se por uma forma de partição que define o método de classificação do paciente. Dessa forma, modelos multinomiais de quatro classificações são estabelecidos para cada uma das categorias. A primeira etapa do trabalho é assim a estimação dos parâmetros dessas multinomiais definidas para cada uma das categorias. As estimativas, médias a posteriori, multiplicadas pelas probabilidades de cada categoria, fornecem as probabilidades conjuntas de categoria e classe e estão apresentadas na Tabela 5.4.4.3. A partir dessa tabela, ao dividir-se os números do corpo da tabela pela marginal inferior, obtem-se as distribuições a posteriori descritas pela Tabela 5.4.4.4. Para ilustrar o uso dos resultados, considere um paciente que vem para uma consulta na clínica. A priori as probabilidades deste paciente pertencer a cada uma das categorias estão listadas pela marginal direita da Tabela 5.4.4.4. Após responder aos questionários o paciente é classificado, por exemplo em (+++). Neste caso suas probabilidades de pertencer a cada uma das categorias está listada na coluna correspondente ao resultado do classificador, a coluna (+++). Note que neste caso a probabilidade do paciente ser fibromiálgico passou de 5% para 26%. Por outro lado, suponha que o paciente foi classificado como (---). A probabilidade de ser um indivíduo saudável passou de 25% para 64%.

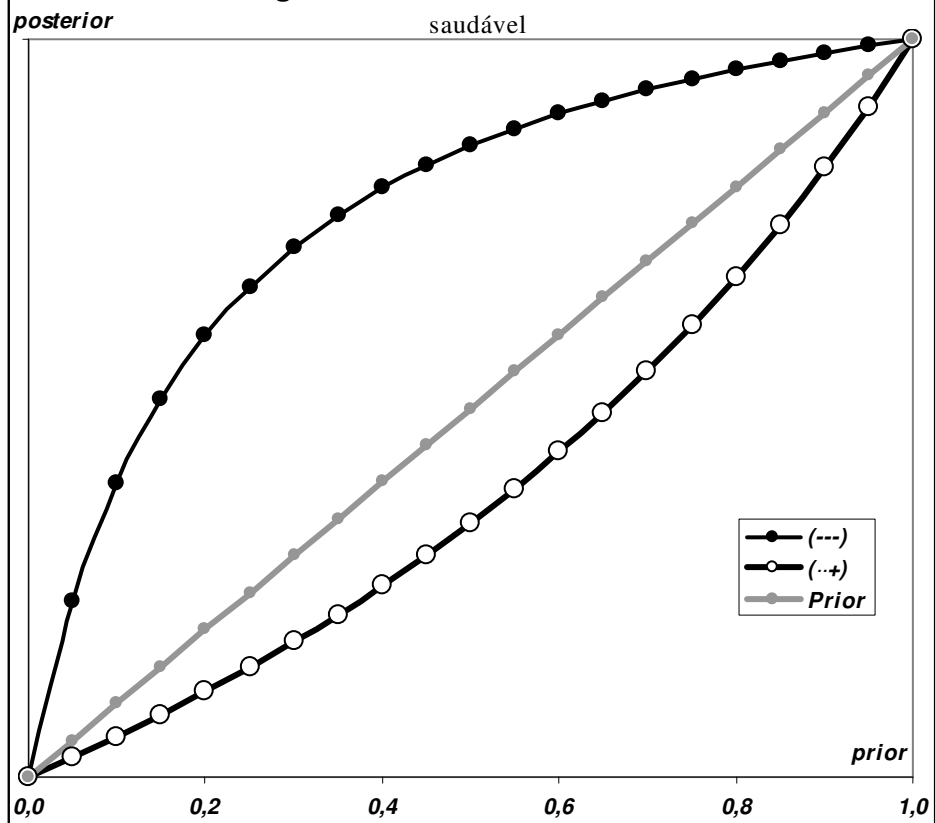
Com o objetivo de melhor ilustrar o poder da metodologia apresentada, Figura 5.4.4.5 e Figura 5.4.4.6 mostram a mudança das probabilidades, prior to posterior, ao considerarmos todas as possíveis probabilidades a priori do paciente pertencer a uma categoria. O primeiro gráfico, Figura 5.4.4.5, diz respeito a probabilidade do paciente ser Fibromiálgico em relação a classificação (+++) contra todas as outras. Note que se o resultado é diferente de (+++) a diminuição da probabilidade de fibromiálgico não é tão forte. Justamente pelo fato da negação de (+++) incluir classificações muito diferentes. O mesmo ocorrerá com a classificação (---) com indivíduos saudáveis. Figura 5.4.4.6. é muito semelhante mostrando que o resultado (---) aumenta consideravelmente a probabilidade do paciente ser saudável, enquanto a negação da classificação (---) não diminui muito a probabilidade do paciente ser saudável. Como antes, a razão é justamente o confundimento de três diferentes categorias.

O autor gostaria que os leitores entendessem que o que foi apresentado aqui é uma forma muito simplista de darmos o primeiro passo para o diagnóstico (ou prognóstico em outras situações) de um paciente. Metodologias mais sofisticadas já estão sendo desenvolvidas. Por exemplo os alunos do autor já iniciaram o desenvolvimento de uma ferramenta de regressão logística bayesiana na qual se faz uso do FBST para a escolha das variáveis. É um bom desafio aos leitores a tentativa de soluções frequentistas para os problemas aqui tratados. Isto irá permitir um julgamento mais adequado da qualidade do pensamento bayesiano.

5.4.4.5.Figura: Prior to Posterior: Classificação com ou sem



5.4.4.6.Figura: Prior to Posterior: Classificação ser ou não



6. Referências Bibliográficas

- Barlow R (1998), *Engineering Reliability*, SIAM.
- Barnard GA (1949), Statistical Inference (w discussion), *JRSS 11*(2):115-49.
- Basu D (1988), *Statistical Information & Likelihood: A Collection of Critical Essays*, JK Gosh editor, Springer-Verlag. Lecture Notes in Statistics #45.
- Basu D & Pereira CAB (1983), Conditional Independence in Statistics, *Sankhya A* 45(3): 324-37.
- Basu D & Pereira CAB (1983) A note on Blackwell sufficiency and a Shibinsky characterization of distribution. *Sankhya A* 45(1): 99-104.
- Basu D & Pereira CAB (1990), Blackwell Sufficiency and Bernoulli Experiments, *Braz J Probability and Statistics* 4:137-45.
- Beck AT; Ward CH; Mendelson M; Mock J; Erbaugh J (1961). An inventory for measuring depression. *Arch Gen Psychiatry* 5:61-71.
- Berger JO & Wolpert R (1988), *The likelihood Principle: A Review and Generalizations*, IMS Lectures Notes-Monograph Series 6 (2nd ed), Hayward, CA.
- Bernardo J; Smith A (1994), *Bayesian Theory*, N York: Wiley
- Berry DA (1996), *Statistics: A Bayesian Perspective*, Belmont, CA: Duxbury
- Biggerrstaff BJ (2000), Comparing diagnostic tests: A simple graphic using likelihood ratios, *Statistics in Medicine*, 19:649-663.
- Birnbaum A (1962), On the foundations of statistical inference (with discussion), *JASA* 32;414-435.
- Blackwell D (1954), *Theory of Games and Statistical Decisions*, N York, Wiley.
- Blackwell D (1959), Comparison of experiments, in: *Proceedings of the 2nd Berkeley Symposium*, 93-102.
- Blackwell D (1969), *Basic Statistics*, McGraw-Hill.
- Blume JD (2002), Likelihood methods for measuring statistical evidence, *Statistics in Medicine*, 21: 2563-2599.
- Bueno AMS; Pereira CAB; Rabello-Gay MN.; Stern JM (2002) Environmental Genotoxicity Evaluation: Bayesian Approach for a Mixture Statistical Model, *Stochastic Environmental Research & Risk Assessment* 16(4): 267-78.
- Burckhardt CS; Clark SR; Bennett RM (1991), The Fibromyalgia Impact Questionnaire: Development and validation. *J Rheumatology* 18: 728 -34.
- Chiann C; Faria L; Pereira CAB; Machado IP (2005), *Prevalência da Fibromialgia na população do município Embu – São Paulo: Dor, Capacidade Funcional, Fadiga, Sono e Qualidade de Vida*. RAE 05P19, São Paulo, IME-USP, 2005.
- Cox DR (1977), The role of significance tests, *Scand J Statist* 4:49-70.
- Cunningham A (1995), Disease risks of wildlife translocations, *Conservation Biology* 10:349-53.
- De Finetti B (1972), *Probability, induction, and statistics*, Wiley.
- DeGroot MH (1962), Uncertainty, Information, and Sequential Experiments, *Annals of Mathematical. Statistics* 33:404-19.

Referências Bibliográficas

- DeGroot MH (1970), *Optimal Statistical Decisions*, McGr;NY
- DeGroot MH (1986), *Probability & Statistics*, Adison-Wesley
- Dempster AP (1997), The direct use of likelihood for significance testing, *Statistics and Computing* 7:247-52.
- Diamond GA & Forrester JS (1983), Clinical trials and statistical verdicts: probable grounds for appeal, *Annals of Internal Medicine*, 98:385-394.
- Fagan TJ (1975), Nomograms for Bayes theorem, *N Engl J Med.* 293:257.
- Fenders AJ (1999), Statistical Concepts, in: M. Berthold and D. Hand (editors), *Intelligent Data Analysis*, Ch 2, N. York: Springer.
- Fisher RA (1956), *Statistical Methods and Scientific Inference*, London: Oliver & Boyd
- Flechtmann CHW (1990), *Ácaros de importância médico veterinária*, Nobel, São Paulo.
- Good IJ (1950), *Probability and the Weighing of Evidence*, London; Griffin.
- Good IJ (1968), Corroboration, explanation, involving probability, simplicity and a sharpened razor, *Br. J. Phil. Sci.* 19:123-43.
- Good IJ (1983), *Good Thinking*, U. Minnesota Press.
- Hill G; Forbes W; Kozak J; MacNeill I (2000), Likelihood and clinical trials, *J Clinical Epidemiology* 53:223-7.
- Irony TZ; Lauretto MS; Pereira CAB; Stern JM (2002), A Weibull wearout test: FBST approach, in: Hayakawa, Irony & Shie eds, *Systems and Bayesian Reliability*, 287-300, World Scientific, Singapura
- Irony TZ & Pereira CAB (1986), Exact tests for equality of two proportions: Fisher vs. Bayes. *J Statistical Computation and Simulation* 25:93-114.
- Irony TZ & Pereira CAB (1995), Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses, *Resenhas* 2:27-46.
- Irony TZ; Pereira CAB; Tiwari RC (2000), Analysis of Opinion Swing: Comparison of Two Correlated Proportions, *Am Statist* 54(1):57-62.
- Jaeschke R; Guyatt G; Sackett D (1994), User's guide to the medical literature: III How to use an article about diagnostic test: B What are the results and will they help me in caring for my patients? *JAMA* 271(9):703-7.
- Jeffreys, H (1939), *Theory of Probability*, Oxford: Clarendon Press.
- Kanegae MF; Hirata MA; Nakati AK; Pereira CAB (2005), *Padrões de ectoparasitismo em aves de cerrado e mata de galeria do Distrito Federal - Brasil'* RC - 05E18 – CEA, São Paulo, IME – USP
- Kempthorne O (1976), Of what use are tests of significance and tests of hypothesis, *Commu. Statist. Theory Methods* 8(A5):763-777
- Kempthorne O & Folks L (1971), *Probability, Statistics & Data Analysis*, Iowa U Press
- King G (1998), *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, Mineapolis; U Michigan Press.

- Lauretto MS; Pereira CAB; Stern JM; Zacks S (2003), Comparing parameters of two bivariate normal distributions using the invariant FBST, *Braz. J Probab & Statist* **17**:147-68
- Lindley DV (1985), *Making Decisions*, London: Wiley
- Lindsey JK (1995), *Introductory Statistics: A Modeling Approach*, N. York: Claredon.
- Lindsey JK (1996), *Parametric Statistical Inference*, Oxford: Oxford U Press.
- Lindsey JK (1999), Relationship among sample size, model selection and likelihood regions and scientifically important differences, *The Statistician* **48**:4001-11.
- Madruga MR; Esteves LG; Wechsler S (2001), On the Bayesianity of Pereira-Stern Tests, *Test* **10**:291-9.
- Madruga MR; Pereira CAB; Stern JM (2003), Bayesian Evidence Test for Precise Hypotheses, *J Statistical Planning and Inference* **117**:185-198.
- Marques AP; Santos AMB; Assumpção A; Matsutani LA; Lais VL; Pereira CAB (2006), Validation of the Brazilian version of the Fibromyalgia Impact Questionnaire (FIQ), *Rev Bras Reumatologia* **46**(1);24-31.
- Marques AP; Ferreira EAG; Matsutani LA; Pereira CAB; Assumpção A (2005), Quantifying Pain Threshold and Quality of Life of Fibromyalgia Patients, *Clinical Rheumatology* **24**(3); 266-71.
- Martinez JE; Ferraz MB; Sato EI; Atra E (1995), Fibromyalgia vs Rheumatoid arthritis: a longitudinal comparison of quality of life. *J Rheumatology* **22**:201-4.
- Montoya-Delgado LE; Irony TZ; Pereira CAB; Whittle MR (2001), An Unconditional Exact Test for the Hardy-Weimberg Equilibrium Law: Sample Space Ordering Using the Bayes Factor, *Genetics* **158**:875-83.
- Neyman J & Pearson ES (1936), Sufficient statistics and uniformly most powerful tests of statistical hypotheses, *Stat. Res. Memoirs*, **1**:133-137.
- Pawitan Y (2000), Likelihood: consensus and controversies, in: *Conference in Applied Statistics in Ireland* (Pawitan's HP)
- Pawitan, Y. (2001), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford: Oxford U Press.
- Paulino CDM; Turkman MAA; Murteira B (2003), *Estatística Bayesiana*, Lisboa: Fundação Calouste Gulbenkian
- Pereira BB & Louzada-Neto F (2002), Statistical Inference (in Portuguese), in: Medronho RA (ed), *Epidemiologia*, Ch 19, Rio de Janeiro: Atheneu.
- Pereira BB & Pereira CAB (2005), A likelihood Approach to Diagnostic Tests in Clinical Medicine, *Statistical J* **3**(1):77-98.
- Pereira CAB (1971), Estimativa da Probabilidade a Priori em Um Problema de Classificação, *Ciência e Cultura* **23**:773-86.
- Pereira CAB (1985), *Testes de hipóteses definidas em espaços de diferentes dimensões: Visão Bayesiana e interpretação clássica*, Tese de Livre Docência, IME – USP.
- Pereira CAB (1990), Influence diagrams and medical diagnosis, in: Oliver & Smith (eds), *Influence Diagrams, Belief Networks, and Decision Analysis*, N York: Wiley.

Referências Bibliográficas

- Pereira CAB & Barlow RE (1990), Medical diagnosis using influence diagrams, *Networks* **20**:565-77.
- Pereira CAB & Lindley DV (1987), Examples questioning the use of partial likelihood, *JRSS D (The Statistician)* **36**:15-20.
- Pereira CAB & Pericchi LR (1990), Analysis of diagnosability, *JRSS C (Applied Statistics)* **39**:189-204.
- Pereira CAB & Stern JM (1999), Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy J* **1**:69-80.
- Pereira CAB; Stern JM (2001), Model Selection: **FBST** approach, *Environmetrics* **12**:559-68
- Pereira CAB & Stern JM (2005), *Inferência indutiva com dados discretos: uma visão genuinamente Bayesiana*. Antofagasta : U de Antofagasta (XV COMCA). 104p.
- Pereira CAB & Wechsler S (1993), On the concept of *P-value*, *Braz J Probability and Statistics*, **7**: 159-77
- Rodrigues J (2006), FBST for zero-inflated distributions, *Communications in Statistics – TM* **35**:1-9.
- Royall, R.M. (1997), *Statistical Evidence: A Likelihood Paradigm*, N York; Chapman Hall.
- Schervish M (1992), *Theory or Statistics*, Springer, NY
- Severini TA (2000), *Likelihood Methods in Statistics*, Oxford: Oxford U Press.
- Spielberger DC; Gorsuch LR; Lushene ER (1979), *Inventário de Ansiedade Traço-Estado*. Rio de Janeiro: Cepsa.
- Sprott DA (2001), *Statistical Inference in Science*, N. York: Springer.
- Sterne JAC (2002), Teaching hypothesis test – time for significance change? (With discussion), *Statistics in Medicine*, **21**:985-94.
- Wald A (1939), Contributions to the theory of statistical estimation and testing hypotheses, *Annals of Probability and Statistics*, **10**:299-326.
- Wald A (1950), *Statistical Decision Functions*, N. York: Wiley.
- Ware JE & Sherborne CD (1992). The MOS 36-item Short-Form Health Survey (SF-36): Conceptual framework and item selection. *Med Care* **30**: 473-3.