PRINCÍPIOS ESTATÍSTICOS E A TEORIA DE KIEFER

Franciné dos Santos Pessoa

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

ÁREA DE CONCENTRAÇÃO: INFERÊNCIA ESTATÍSTICA

ORIENTADOR: PROF. DR. CARLOS ALBERTO BRAGANÇA PEREIRA

SÃO PAULO(SP), JANEIRO DE 1986

À memória de meu pai

Agradeço

Ao *Carlinhos*, pelo competente trabalho de orientação e revisão.

Ao Prof. *Formiga*, pelas facilidades que me proporcionou durante a realização deste trabalho.

Aos colegas de pós-graduação que muito me incentivaram durante toda essa longa jornada.

À *Sônia Regina*, pelo seu trabalho incansável e eficiente de datilografia.

Em especial à *Gracinha*, minha esposa, a quem pelo incentivo, insistência e até broncas nas horas certa, dedico este trabalho.

ÍNDICE

Pág.

Introdução 01

Capítulo I: Princípios Estatísticos 04

1.1 - Introdução..... 04

1.2 - Experimento Estatístico e Informação..... 04

1.3 - Definições Básicas..... 06

1.4 - Princípios Estatísticos..... 13

1.5 - Consideraçoẽs Finais..... 23

Capítulo II: A Metodologia de Kiefer..... 25

2.1 - Introdução..... 25

2.2 - Sub-álgebras e Partições..... 26

2.3 - A Metodologia de Kiefer (um exemplo)..... 35

2.4 - Notação e Definição..... 38

2.5 - Admissibilidade e Escolha de Partições..... 42

Capítulo III: Aplicações..... 45

3.1 - Introdução..... 45

3.2 - O Problema Genético..... 45

3.3 - O Experimento..... 47

3.4 - O Problema Genético e o Princiípio da Verossimilhança..... 48

3.4.1 - Metodologia..... 48

3.4.2 - Aplicação aos Dados Genéticos..... 51

3.5 - O Problema Genético e a Teoria de Kiefer..... 54

3.6 - Análise Comparativa das Estimativas..... 62

Pág.

Bibliografia 70

Anexo "A"

Anexo "B"

Introdução

Desde que R.A. Fisher, D.R. Cox e J.W. Tukey elucidaram a idéia daquilo que mais tarde seria chamado de Princípio da Condicionalidade, uma grande polêmica foi criada entre diversos autores. Essa controvérsia devia-se (deve-se?) não propriamente ao que postula tal princípio, mas principalmente às implicações matemáticas dele decorrentes, como por exemplo o Princípio da Verossimilhança. Esse último princípio passou a ser o pivô de todo o conflito, porque de certa forma ele sugere uma reformulação de alguns conceitos da Inferência Clássica, tais como, níveis de confiança e significância, uma vez que ele estabelece que a informação de um resultado experimental é completamente caracterizada pela função de verossimilhança, independentemente da estrutura do experimento. Até mesmo a palavra "informação" passou a ser interpretada de modo diferente pelos simpatizantes deste princípio, pois a literatura estatística, até então, tratava de informação relacionada a um experimento e não a uma observação experimental.

O Capítulo 1 é dedicado a apresentação de alguns dos Princípios da Inferência Estatística e suas implicações. Para uma melhor compreensão de tais princípios, algumas definições foram apresentadas usando-se a linguagem de "informação contida em um resultado experimental". Na demonstração do Teorema de Birnbaum, que estabelece a equivalência entre o Princípio da Verossimilhança e os Princípios

da Suficiência e Condicionalidade, seguimos em linhas gerais a orientação de Basu (1975).

No Capítulo 2 apresentamos a Teoria de Kiefer, que faz uso do Princípio da Condicionalidade (decisão condicionada ao resultado de uma estatística ancilar) na tentativa de medir a importância do valor observado diante da decisão adotada. Essa idéia de que uma estatística ancilar, quando usada em conjunto com uma outra estatística , pode ser de grande valia para se medir a "qualidade" do resultado amostral, já havia sido apontada por R.A. Fisher (1935) que escreveu. "It is shown that some, or sometimes all of the lost information may be recovered by calculating what I call ancillary statistics, which themselves tell us nothing about the value of the parameter, but instead, tell us how good an estimate we have made of it. Their function is, in fact, analogous to the part which the size of the sample is always expected to play, in telling us what reliance to place on the result".

Usando dados experimentais observados pelo Prof. Dr. Luiz Edmondo Magalhães (Inst. de Biociências -USP), sobre o comportamento sexual do inseto Drosóphila Melanogaster, apresentamos, no Capítulo 3, duas soluções para o problema de estimar a preferência sexual da fêmea diante de dois tipos de macho. A primeira solução é baseada no Princípio da Verossimilhança, na qual fazemos uso da variabilidade da função de verossimilhança para chegarmos a um intervalo de credibilidade e testar hipótese acerca da preferência sexual de tal inseto por um dos dois tipos de macho. A segunda solução

faz uso da Metodologia de Kiefer e tem como propósito mostrar que em algumas ciscunstâncias essa metodologia também é aplicável a testes não simétricos para distribuições discretas.

Capítulo I
Princípios Estatísticos

1.1 - Introdução

O objetivo deste capítulo é levar ao leitor um resumo dos Princípios Estatísticos, bem como algumas implicações existentes entre eles. Para tanto, definições familiares àqueles que lidam com estatística, serão apresentadas usando-se, no entanto, uma linguagem mais compatível com o conceito de informação.

1.2 - Experimento Estatístico e Informação

Basicamente, um problema estatístico surge quando procuram-se informações acerca de um estado da natureza (parâmetro) w, que é desconhecido (não totalmente conhecido). Em busca de tal conhecimneto, um experimento E é planejado e realizado, obtendo-se assim um resultado x (normalmente um vetor). Neste trabalho, chamaremos de "observação estatística", ou simplesmente "observação", ao par (E, x) para indicar não somente o resultado x, mas também seu processo de obtenção.

A seguir, as estruturas formais de um experimento estatístico são apresentadas e discutidas com o intuito de tornar matematicamente natural o conceito de informação.

Um experimento E fica bem caracterizado quando representado por um terno $(\mathcal{X}, \Omega, f)$, onde:

a) $\mathcal{X}$ representa o conjunto de todos os possíveis resultados x, quando o procedimento que define o experimento E é realizado. Esse conjunto é comumente chamado de "Espaço Amostral".

b) Ω, normalmente chamado de "Espaço Paramétrico", é o conjunto de todos os possíveis (conforme o conhecimento de quem o define) valores do estado da natureza, w.

c) $f: \mathcal{X} \times \Omega \to [0, \infty)$, que a cada ponto (x, w) e $\mathcal{X} \times \Omega$ associa o valor $f(x|w) \geq 0$, é uma distribuição de probabilidade ou uma função densidade de probabilidade em $\mathcal{X}$ para cada $w \in \Omega$.

De posse da observação (E, x), o objetivo do estatístico é dela retirar "toda a informação relevante" (expressão de R.A. Fisher) acerca do parâmetro w. É verdade que a expressão de Fisher, ainda nos dias de hoje, não possui uma definição universalmente aceita, porém tem havido uma certa aceitação da notação Inf (E, x) como significado de "toda a informação relevante sobre o parâmetro w, contida na observação (E, x)".

Desta forma diz-se que as observações (E_1, x_1) e (E_2, x_2) são "equivalentes" ou "igualmente informativas" sobre w (e denota-se por $Inf(E_1, x_1) = Inf(E_2, x_2)$), se E_1 e E_2 são experimentos relacionados ao mesmo estado da natureza w, e os resultados x_1 e x_2 de E_1 e E_2 respectivamente, conduzem à mesma inferência.

Nas próximas seções apresentaremos definições, exemplos e algumas condições necessárias para se estabelecer a equivalência de observações. Em alguns dos exemplos que se seguem, utilizamos a linguagem de caras e coroas de uma moeda para representar situações envolvendo ensaios de Bernoulli. No entanto, consideramos a probabilidade de obtenção de um sucesso (cara) como sendo $p \in (0, 1)$ no lugar de 1/2 da moeda balanceada.

1.3 - Definições Básicas

Com o intuito de tornar esse trabalho um pouco auto-suficiente, no sentido de conter boa parte dos pré-requisitos necessários, recordaremos algumas definições utilizadas na linguagem estatística formal. Nossa referência básica para este capítulo é: Basu (1975), Berger (1980) e Birnbaum (1962).

Definição 1: Seja $E = (\mathcal{X}, \Omega, f)$ um experimento. Diz-se que $T: \mathcal{X} \to \tau$ é uma "estatística relacionada com o experimento E", se T é uma função (mensurável) de $\mathcal{X}$ em $\tau = \{t: T(x) = t,$ para

algum x e $\mathcal{X}$}.

Note que cada ponto t e τ define um sub-conjunto $\mathcal{X}_t = \{ x \in \mathcal{X} : T(x) = t\}$ de $\mathcal{X}$ e que a família $\{\mathcal{X}_t : t \in \tau\}$ constitui uma partição de $\mathcal{X}$. Reciprocamente toda partição de $\mathcal{X}$ é gerada por uma estatística T convenientemente definida. Assim, algumas vezes, usaremos uma partição de $\mathcal{X}$ em lugar de uma estatística e vice-versa.

Definição 2: seja $E = (\mathcal{X}, \Omega, f)$ um experimento e T uma estatística relacionada com E. Define-se o "experimento marginal" E_T como sendo o terno (τ, Ω, f_T) onde:

$$f_T(t|w) = \int_{\mathcal{X}_t} f(x|w)\, dx, \quad t \in \tau$$

Observação: O símbolo de integração está sendo usado de maneira bastante geral, de forma que ele pode representar um somatório quando necessário.

Exemplo 1: Com a finalidade de estimar a probabilidade p de se obter cara no lançamento de uma moeda, um estatístico decidiu lançar a moeda três vezes e observar a sequência (x_1, x_2, x_3) de caras (1) e coroas (0). Um segundo estatístico, achou por bem anotar somente o número de caras nos três lançamentos, isto é, usar a estatística $T(x_1, x_2, x_3) = x_1 + x_2 + x_3$.

No primeiro caso, temos o experimento $E = (\mathcal{X}, \Omega, f)$, onde:

$\mathcal{X} = \{(x_1, x_2, x_3): x_i = 0 \text{ ou } x_i = 1: i = 1, 2, 3\}$

$\Omega = \{p : 0 \leq p \leq 1\}$

e

$$f(x|p) = \begin{cases} (1-p)^3 & \text{se } x \in \{(0, 0, 0)\} \\ p(1-p)^2 & \text{se } x \in \{(0, 0, 1); (0, 1, 0); (1, 0, 0)\} \\ p^2(1-p) & \text{se } x \in \{(0, 1, 1); (1, 0, 1); (1, 1, 0)\} \\ p^3 & \text{se } x \in \{(1, 1, 1)\} \end{cases}$$

No outro caso, temos o experimento $E_T = (\tau, \Omega, f_T)$, onde:

$\tau = \{0, 1, 2, 3\}$

$\Omega = \{p: 0 \leq p \leq 1\}$ e

$f_T(t|p) = \binom{3}{t} p^t (1-p)^{3-t}, \ t \in \tau.$

<u>Definição 3</u>: Seja T uma estatística relacionada com um experimento E. Para cada $t \in \tau$, define-se o "experimento condicional", $E^{(T=t)}$ como sendo o terno:

$E^{(T=t)} = (\mathcal{X}_t, \Omega, f^{(T=t)})$, onde $f^{(T=t)}$ é dada por:

$$f^{(T=t)}(x|w) = \frac{f(x|w)}{f_T(t|w)}, \ x \in \mathcal{X}_t$$

<u>Exemplo 2</u>: Considere o experimento E e a estatística T do exemplo anterior. Para cada $t \in \tau$, podemos definir o experimento condicional $E^{(T=t)} = (\mathcal{X}_t, \Omega, f^{(T=t)})$, com componentes:

$\mathcal{X}_t = \{x \in \mathcal{X} : T(x) = t\}$

$\Omega = \{p: 0 \leq p \leq 1\}$ e

$$f^{(T=t)}(x|p) = \frac{f(x|p)}{f_T(t|p)} = \frac{p^t(1-p)^{3-t}}{\binom{3}{t}p^t(1-p)^{3-t}} = \frac{1}{\binom{3}{t}}, \quad x \in \mathcal{X}_t.$$

<u>Definição 4</u>: Um experimento E é dito "não informativo" (sobre w), se para cada $x \in \mathcal{X}$, $f(x|w)$ é constante em $w \in \Omega$.

<u>Exemplo 3</u>: Com relação ao exemplo 2, para cada $t \in \tau$, o experimento condicional $E^{(T=t)}$ é não informativo, uma vez que $f^{(T=t)}(x|p) = \frac{1}{\binom{3}{t}}$ é constante em p, para todo $x \in \mathcal{X}_t$.

<u>Definição 5</u>: Diz-se que uma estatística $T: \mathcal{X} \to \tau$ é "ancilar" (com respeito a w), se o experimento marginal E_T é não informativo.

<u>Exemplo 4</u>: Seja X uma v.a. com distribuição normal de média μ e variância 1, onde $\mu \in \{-1, 1\}$. Considere a estatística T, baseada em um único resultado x, definida por:

$T(x) = I_{\{|x| \geq c\}}(x)$, sendo c uma constante positiva e $I_A(\cdot)$ a função indicadora do conjunto A. Nosso experimento inicial é dado por: $E = (\mathcal{X}, \Omega, f)$, onde: $\mathcal{X} = R$, $\Omega = \{-1, 1\}$, $f(c|\mu) = \Psi(x-\mu)$ e Ψ a densidade associada à distribuição normal padrão N (0, 1). O experimento marginal E_T é constituído por:

$\tau = \{0, 1\}$; $\Omega = \{-1, 1\}$ e f_T dada por:

$$f_T(0|\mu) = P_\mu(|x| < c) = \Phi(c-\mu) - \Phi(-c-\mu) =$$

$= \Phi\ (c-\mu) + \Phi(c+\mu) - 1$, onde Φ é a função de distribuição da $N\ (0,\ 1)$. Assim: $f_T\ (1|\mu) = 1 - f_T\ (o|\mu) = \Phi\ (-c - \mu) + \Phi\ (-c + \mu)$. Logo:

$$f_T\ (0|-1) = \Phi\ (c+1) + \Phi\ (c-1) - 1 = f_T\ (0|1) \text{ e}$$

$$f_T\ (1|-1) = \Phi\ (-c + 1) + \Phi\ (-c - 1) = f_T\ (1|1).$$

Como $f_T\ (t|\mu)$ é constante em μ para $t \in \{0,1\}$, conclui-se que o experimento E_T é não informativo, e consequentemente T é uma estatística ancilar.

De maneira mais geral, se $X \sim N\ (\mu, \sigma_0^2)$, $\mu \in \{-\mu_0, \mu_0\}$ (μ_0 e σ_0, reais positivos e conhecidos), então a estatística $T(x) = |X|$ é ancilar.

<u>Definição 6</u>: Uma estatística T é dita "suficiente" (para (E, x) com respeito a w) se, para todo $t \in \tau$ o experimento condicional $E^{(T=t)}$ é não informativo.

Podemos verificar, então, que a estatística T do exemplo 1 é suficiente, pois de acordo com o exemplo 3, para cada $t \in \tau$, $f^{(T=t)}\ (x|p) = \dfrac{1}{\binom{3}{t}}$ é constante em p.

<u>Definição 7</u>: Suponha que a realização de um experimento $E = (\mathcal{X}, \Omega, f)$, resultou na observação (E, x). A função $L: \Omega \to R^+$ dada por $L(w|x) = f(x|w)$ é chamada "função de verossimilhança" gerada pela observação (E,x).

Exemplo 5: Com o objetivo de estimar a probabilidade p de obtenção de cara de uma moeda, um estatístico decidiu lançar a moeda 20 vezes e anotar o número de caras obtidas. Assim, seu experimento é dado por:

$E_1 = (\mathcal{X}_1, \Omega, f_1)$, onde: $\mathcal{X}_1 = \{0, 1, 2, \ldots, 20\}$

$\Omega = \{p: 0 \leq p \leq 1\}$ e

$f_1(x|p) = \binom{20}{x} p^x (1-p)^{20-x}, \; x \in \mathcal{X}_1$

Ao realizar seu experimento, ele obteve 15 caras (e 5 coroas). Desse modo a função de verossimilhança gerada pela observação $(E_1, 15)$ é:

$L(p|15) = \binom{20}{15} p^{15} (1-p)^5, \; 0 \leq p \leq 1$

Definição 8: Dizemos que duas funçoẽs de verossimilhança L_1 e L_2, definidas em um mesmo espaço paramétrico Ω (porém podendo corresponder a distintas observaçoẽs (E_1, x_1) e (E_2, x_2) respectivamente), são "equivalentes" se, $L_1(w|x_1) = C. L_2(w|x_2)$, para alguma constante positiva C (que pode depender de E_1, x_1, E_2 e x_2) e todo $w \in \Omega$.

Exemplo 6: Suponha que diante de situação do exemplo anterior, um outro estatístico decida lançar a moeda tantas vezes quantas sejam necessárias para obter 15 caras. Dessa forma, seu experimento $E_2 = (\mathcal{X}_2, \Omega, f_2)$ é constituído por:

$\mathcal{X}_2 = \{15, 16, 17, \ldots\}$

$\Omega = \{ p: 0 \leq p \leq 1 \}$ e

$f_2(x|p) = \binom{x-1}{14} p^{15} (1-p)^{x-15}, \quad x \in \mathcal{X}_2.$

Suponha ainda que ao realizar esse experimento, verificou que a 15ª cara ocorreu no 20º lançamento, ou seja, sua observação foi $(E_2, 20)$, gerando assim a função de verossimilhança.

$$L_2(p|20) = \binom{19}{14} p^{15} (1-p)^5, \quad 0 \leq p \leq 1.$$

Comparando L_2 com L_1 do exemplo 5, verificamos que:

$$L_1(p|15) = \frac{\binom{20}{5}}{\binom{19}{14}} L_2(p|20), \quad 0 \leq p \leq 1.$$

Assim L_1 e L_2 são funções de verossimilhança equivalentes.

<u>Definição 9</u>: Suponha que dispomos de uma coleção enumerável de experimentos $E_i = (\mathcal{X}_i, \Omega, f_i)$, $i = 1, 2, \ldots$. E seja $\Pi_1, \Pi_2, \ldots$ uma distribuição de probabilidade sobre o conjunto $\{1, 2, \ldots\}$, que independe do conhecimento do estado de natureza w. A mistura E dos experimentos $E_1\ E_2 \ldots$, de acordo com as probabilidades de seleção $\Pi_1, \Pi_2, \ldots$, é definida como um experimento nos dois seguintes estágios:

(1) Seleciona-se um dos experimentos $E_1, E_2, \ldots$, com probabilidade de seleção $\Pi_1, \Pi_2, \ldots$

(2) Realiza-se o experimento selecionado em (1).

Desta forma, o experimento mistura $E = (\mathcal{X}, \Omega, f)$ possui as seguintes componentes:

$\mathcal{X} = \{x = (i, x_i);\ x_i \in \mathcal{X}_i,\ i = 1, 2, \ldots\}$, que é a união disjunta dos espaços amostrais $\mathcal{X}_1, \mathcal{X}_2, \ldots$ de $E_1, E_2, \ldots$, respectivamente, Ω o espaço paramétrico comum a $E_1, E_2, \ldots$ e $f(x|w) = \Pi_i f_i(x|w)$, onde $x = (i, x_i)$ e $x_i \in \mathcal{X}_i$.

Exemplo 7: Se um terceiro estatístico na dúvida de adotar o experimento E_1 do exemplo 5 ou o experimento E_2 do exemplo 6, resolve lançar um dado honesto e realizar E_1 se o resultado for par e E_2 se o resultado for impar, na realidade ele estará realizando um experimento E que é uma mistura de E_1 e E_2 com probabilidades de seleção $\Pi_1 = \Pi_2 = 1/2$.

1.4 - Princípios Estatísticos

Com as definições exibidas acima, estamos agora de posse de todas as ferramentas necessárias para apresentarmos alguns dos mais importantes princípios da inferência estatística, princípios esses que estabelecem condições para a comparação (equivalência) de observações e verificação da consistência de métodos estatísticos.

P1. Princípio da Suficiência: Seja $E = (\mathcal{X}, \Omega, f)$ um experimento estatístico e T uma estatística suficiente. Então para cada $x \in \mathcal{X}$.

$\text{Inf}\ (E, x) = \text{Inf}\ (E_T, t),\ t = T(x)$

Assim o princípio P1 estabelece que se T é suficiente, então para cada $t \in \tau$, todos os elementos de $\mathcal{X}_t$ teem a mesma informação.

Exemplo 8: Com referência aos exemplos 1 a 3, uma vez que $T\ (x_1, x_2, x_3) = x_1 + x_2 + x_3$ é suficiente, temos:

Para t = 0

$\text{Inf}\ (E, (0,0,0)) = \text{Inf}\ (E_T, 0)$

Para t = 1

$\text{Inf}\ (E, (0,0,1)) = \text{Inf}\ (E, (0,1,0)) = \text{Inf}\ (E, (1,0,0)) = \text{Inf}\ (E_T, 1)$

Para t = 2

$\text{Inf}\ (E, (0,1,1)) = \text{Inf}\ (E, (1,0,1)) = \text{Inf}\ (E, (1,1,0))$

E para t = 3

$\text{Inf}\ (E, (1,1,1)) = \text{Inf}\ (E_T, 3)$.

P2. Princípio Fraco da Condicionalidade: Se o experimento E é uma mistura dos experimentos $E_1, E_2, \ldots$, então para cada resultado $x = (i, x_i)$ de E temos:

$$\text{Inf}\ (E, x) = \text{Inf}\ (E_i, x_i),$$

o que significa que a informação de qualquer resultado $x = (i, x_i)$ do experimento E, é a mesma que a do resultado x_i do experimento E_i, ignorando-se totalmente a estrutura (probabilidade de seleção) do experimento E.

Exemplo 9: Suponha que o estatístico do exemplo 7 ao lançar o dado obteve o número 4 e ao realizar o experimento E_1 (devido a sua regra de seleção) obteve 15 caras. Assim a observação do seu experimento, $(E, x) = (E, (1, 15))$ é tão informativa quanto a observação $(E_1, 15)$ feita pelo estatístico do exemplo 5, ou seja:

$$\text{Inf}\ (E, (1, 15)) = \text{Inf}\ (E_1, 15)$$

P3. Princípio da Condicionalidade: Se $T: \mathcal{X} \to \tau$ é uma estatística ancilar associada ao experimento $E = (\mathcal{X}, \Omega, f)$, então, para todo $x \in \mathcal{X}$ e $t = T(x)$.

$$\text{Inf}\ (E, x) = \text{Inf}\ (E^{(T=t)}, x)$$

Exemplo 10: Com relação ao exemplo 4, tomemos $c = 2$. Assim a partição induzida pela estatística $T(x) = I_{\{|x| \geq 2\}}(x)$ é

$\mathcal{X}_0 = \{x \in \mathcal{X} : |x| < 2\}$ e

$\mathcal{X}_1 = \{x \in \mathcal{X} : |x| \geq 2\}$

Suponha que ao realizar-se o experimento E (do exemplo 4) foi observado $(E, 3)$. Se por outro lado uma realização de $E^{(T=1)}$ resulta em $x = 3$, então

$$\text{Inf}\ (E, 3) = \text{Inf}\ (E^{(T=1)}, 3),$$

uma vez que T é uma estatística ancilar associada ao experimento E.

P4. Princípio da Verossimilhança: Se E_1 e E_2 são dois experimentos quaisquer com o mesmo espaço paramétrico Ω e se (E_1, x_1) e (E_2, x_2) são observações respectivas, determinando funções de verossimilhança equivalente, então:

$$\mathrm{Inf}(E_1, x_1) = \mathrm{Inf}\ (E_2, x_2)$$

Em palavras o princípio P4 estabelece que de posse da observação (E, x), toda a informação relevante sobre w está contida na função de verossimilhança por ela gerada, ou seja, valores que não foram efetivamente observados são irrelevantes.

Exemplo 11: Considerando-se as observações $(E_1, 15)$ do exemplo 5 e $(E_2, 20)$ do exemplo 6, podemos concluir que elas possuem a mesma informação ou seja:

$$\mathrm{Inf}\ (E_1, 15) = \mathrm{Inf}\ (E_2, 20)$$

de vez que $L_1(p|15)$ e $L_2(p|20)$ são equivalentes.

É interessante citar, nesse ponto, um fato curioso em relação aos princípios acima. A grande maioria dos estatísticos tendem a aceitar, para propósitos de inferência sobre w, os princípios P1 e P2. Contudo, poucos desses mesmos estatísticos aceitam o princípio P4. Para surpresa daqueles que não aceitam este último princípio, Birnbaum (1962, pp 284-5) mostra que P4 é matematicamente equivalente a P1 e P2, ou seja, (P1 e P2) $\leftrightarrow$ P4.

Com relação a este resultado, Jerome Cornfield (in Birnbaum (1962, discussion) se referiu da se

guinte maneira "... ainda não estou completamente recuperado do choque de ver que dois princípios que eu achava razoáveis e um que eu achava duvidoso, são equivalentes. É claro que eu preciso ou acreditar nos três, ou desacreditar em pelo menos um dos dois razoáveis. O que ainda não está claro é em que base essa escolha deva ser feita".

Apenas com o propósito de ilustrar e "surpreender" nosso leitor, decidimos apresentar uma discussão da demonstração do teorema de Birnbaum. Para tanto seguimos, em linhas gerais, as sugestões de Basu (1975).

Antes de apresentarmos a demonstração deste teorema, demonstraremos dois lemas que serão de grande utilidade para o nosso propósito.

Lema 1: Uma estatística T é suficiente se, é somente se, para x_1 e x_2 pertencentes a $\mathcal{X}$, $T(x_1) = T(x_2)$ implica em $L(w|x_1)$ equivalente a $L(w|x_2)$.

Demonstração:

a) Seja T uma estatística suficiente e $\{x_1, x_2\} \subset \mathcal{X}$ tal que $T(x_1) = T(x_2) = t$.

Como T é suficiente temos:

$$f^{(T=t)}(x_1|w) = C(x_1), \quad (1)$$

$$e \; f^{(T=t)}(x_2|w) = C(x_2), \quad (2)$$

Considerando-se que $f^{(T=t)}(x|w) = \frac{f(x|w)}{f_T(t|w)}, x \in \mathcal{X}_t,$

segue-se de (1) e (2) que:

$f(x_1|w) = C(x_1) . f_T(t|w)$ e

$f(x_2|w) = C(x_2) . f_T(t|w)$

Assim,

$$f(x_1|w) = \frac{C(x_1)}{C(x_2)} . f(x_2|w), \quad w \in \Omega$$

o que implica que $L(w|x_1)$ é equivalente a $L(w|x_2)$.

b) Suponha que $T(x_1) = T(x_2)$ implica em $L(w|x_1) = C(x_1, x_2)$ $L(w|x_2)$, para todo $\{x_1, x_2\} \subset \mathcal{X}$ e alguma função positiva C. Seja $x_o \in \mathcal{X}$ tal que $T(x_o) = t$ com função de verossimilhança $L(w|x_o)$.

Então para todo $x \in \mathcal{X}_t$ temos:

$$L(w|x) = C(x,x_o) \; L(w|x_o), \; \forall \; w \in \Omega$$

Assim,

$$f^{(T=t)}(x|w) . f_T(t|w) = C(x, x_o) \; f^{(T=t)}(x_o|w) \; f_T(t|w)$$

Consequentemente,

$$1 = \int_{\mathcal{X}_t} f^{(T=t)}(x|w) \, dx = f^{(T=t)}(x_o|w) . \int_{\mathcal{X}_t} C(x, x_o) \, dx$$

Daí

$$f^{(T=t)}(x_o|w) = \frac{1}{\int_{\mathcal{X}_t} C(x, x_o)dx}, \quad \forall\ x_o \in \mathcal{X}_t$$

Logo $E^{(T=t)}$ é não informativo e portanto T suficiente.

<u>Definição 10</u>: Seja $E = (\mathcal{X}, \Omega, f)$ um experimento estatístico e x um resultado de E determinando função de verossimilhança $L(w|x)$. A função $L^*(.|x)$ definida por:

$$L^*(w|x) = \frac{L(w|x)}{\int_\Omega L(w'|x)dw}, \quad w \in \Omega$$

é denominada "função de verossimilhança padronizada" gerada pela observaçaõ (E, x).

<u>Observações</u>: a) Nessa definição estamos admitindo a convergência da integral $\int_\Omega L(w'|x)dw'$, uma vez que em situação práticas, geralmente é possível visualizar Ω como um conjunto limitado. Por outro lado esse denominador pode ser substituído por outro operador. Por exemplo, no caso da existência de máximo, esse poderia ser considerado um bom candidato.

b) Como consequência desta definição podemos afirmar que $L(w|x_1)$ é equivalente a $L(w|x_2)$ se, somente se, $L^*(w|x_1) = L^*(w|x_2)$.

<u>Lema 2:</u> Se x_1 e x_2 são dois resultados de um experimento E determinando funções de verossimilhança equivalentes, então

existe uma estatística suficiente minimal T, tal que $T(x_1) = T(x_2)$.

Demonstração:

Seja $\{x_1, x_2\} \subset \mathcal{X}$ tal que

$L(w|x_1) = C(x_1, x_2)\, L(w|x_2)$ para todo $w \in \Omega$ e alguma função positiva C (3)

Definamos $T: \mathcal{X} \to \mathcal{L}_0^*$ por

$T(X) = L^*(.|X)$, ou seja, T é uma transformação que a cada resultado $x \in \mathcal{X}$ associa a função de verossimilhança padronizada $L(.|x) \in \mathcal{L}_0^*$

a) $T(x_1) = T(x_2)$

Como $L(w|x_1) = C(x_1, x_2)\, L(w|x_2)$, $\forall\, w \in \Omega$

segue-se que $L^*(w|x_1) = L^*(w|x_2)$, $\forall\, w \in \Omega$

Assim,

$T(x_1) = L^*(.|x_1) = L^*(.|x_2) = T(x_2)$

b) T é suficiente

De acordo com o Lema 1, basta mostrar que para qualquer $\{x', x''\} \subset \mathcal{X}$ tal que $T(x') = T(x'')$ teremos $L(w|x')$ equivalente a $L(w|x'')$.

Com efeito, se $T(x') = T(x'')$ segue-se que $L^*(.|x') = L^*(.|x'')$ e assim $L^*(w|x') = L^*(w|x''), \forall w \in \Omega$ e portanto $L(w|x')$ equivalente a $L(w|x'')$.

c) T é minimal

Seja S uma estatística suficiente qualquer, então pelo critério da fatoração de Halmos-Savage temos:

$$T(X) = L^*(.|X) = \frac{L(.|X)}{\int_\Omega L(w'|X)dw'} = \frac{h(X).\ g(s|.)}{h(X) \int_\Omega g(s|w')dw'} =$$

$$= \frac{g(s|.)}{\int_\Omega g(s|w')dw'},$$ onde g é a função densidade de probabilidade de S e h uma função não negativa que independe de w.

Como T é uma função de S que por sua vez é uma estatística suficiente qualquer, concluímos que T é minimal.

<u>Teorema de Birnbaum</u>: P4 implica e é implicado por P1 e P2.

a) P4 → P1

Consequência imediata do Lema 2.

b) P4 → P2

Basta observar que as funções de verossimilhança geradas por $(E,(i, x_i))$ e (E_i, x_i) são equivalentes.

c) P1 e P2 $\rightarrow$ P4

Sejam $E_1 = (\mathcal{X}_1, \Omega, f_1)$ e $E_2 = (\mathcal{X}_2, \Omega, f_2)$ dois experimentos quaisquer com o mesmo espaço paramétrico Ω. Considere o experimento E construido a partir de mistura de E_1 e E_2 com probabilidades de seleção $\Pi_1 = \Pi_2 = 1/2$.

Assim um resultado genérico $x \in \mathcal{X}$ será representado por:

$$x = \begin{cases} (1, x') \text{ se } E_1 \text{ for selecionado} \\ (2, x'') \text{ se } E_2 \text{ for selecionado} \end{cases}$$

e a função densidade de probabilidade (ou função de probabilidade)f será dada por:

$$f(x|w) = \begin{cases} \frac{1}{2} f_1(x'|w) \text{ se } x = (1, x'),\ x' \in \mathcal{X}_1 \\ \\ \frac{1}{2} f_2(x'|w) \text{ se } x = (2, x'),\ x' \in \mathcal{X}_2 \end{cases}$$

segue-se então de P2 que:

$$\begin{aligned} \text{Inf}(E, x) &= \text{Inf}(E_1, x_1),\ x = (1, x_1),\ x_1 \in \mathcal{X}_1 \\ \text{Inf}(E, x) &= \text{Inf}(E_2, x_2),\ x = (2, x_2),\ x_2 \in \mathcal{X}_2 \end{aligned} \qquad (4)$$

Sejam x_1 e x_2 dois resultados quaisquer de E_1 e E_2 respectivamente, os quais determinam funções de verossimilhança equivalentes segundo o experimento E, ou seja:

$L(w|(1,x_1))$ equivalente a $L(w|(2,x_2))$, $\forall\ w \in \Omega$

Então segue-se de P1 e do Lema 2 que

$$\mathrm{Inf}(E,(1, x_1)) = \mathrm{Inf}\ (E,(2, x_2)) \qquad (5)$$

De (4) e (5) concluiu-se que

$$\mathrm{Inf}(E_1, x_1) = \mathrm{Inf}\ (E_2, x_2) \qquad (6)$$

Uma vez que (6) estabelece que dois resultados quaisquer de dois experimentos quaisquer E_1 e E_2 (com o mesmo espaço paramétrico), teem a mesma informação se eles determinam funções de verossimilhança equivalentes, fica provada a implicação de P4 por P1 e P2.

1.5 - Considerações Finais

É importante salientar que em nenhum momento, explicitamos a definição de informação. Contudo, na definição 4 apresentamos o conceito de "não informação". Assim, um experimento é considerado informativo sobre w, se não satisfaz a definição 4. Por outro lado, sendo o nosso objetivo escolher um "melhor" (mais simples) experimento para realizá-lo, apenas nos preocupamos em caracterizar alguns experimentos equivalentes (no sentido de conter a mesma informação sobre w). Dessa forma não nos preocuparemos com a difícil tarefa de encontrar uma medida de informação.

Muitas tentativas de definir uma medida razoável de informação, teem aparecido na literatura estatística. Dentro do espírito de nosso trabalho, poderíamos citar Lindley (1956), De Groot (1962) e Bernardo (1979). Entretanto, todos esse autores apresentam medidas para a informação do experimento e não para a informação contida na observação. Assim, poderíamos encarar essas medidas como sendo o valor esperado da informação contida em uma observação a ser realizada.

Como sugestão para um futuro trabalho, acreditamos que uma "boa" medida de informação, poderia ser obtida a partir do estudo da variabilidade da função de verossimilhança. A motivação para um estudo dessa natureza surge naturalmente quando aceitamos a definição 4 e o fato de que toda a informação relevante, sobre w, contida no resultado x, está também contida no gráfico da função de verossimilhança (P4 e Lema 2).

No Capítulo 3 apresentaremos uma estimativa pontual e um intervalo de credibilidade para a preferência sexual do inseto Drosóphila Melanogaster diante de dois tipos de machos, utilizando apenas a variabilidade da função de verossimilhança. No entanto é necessário frisar que nenhuma tentativa foi feita no sentido de medir o grau de informação contida na observação geradora desta função, embora tenhamos a pretenção de fazê-lo em um trabalho posterior como sugerido anteriormente.

Capítulo II
A Metodologia de Kiefer

2.1 - Introdução

Nas últimas décadas, severas críticas teem sido feitas à teoria de Neyman - Pearson - Wald (NPW) especialmente no que diz respeito aos testes de hipóteses. Um dos pontos comuns entre tais críticos é a insatisfação quanto à forma de avaliação do grau de conclusão de tal metodologia. Por exemplo, na situação mais elementar dos testes de hipóteses simples, H_0 e H_1, a especificação tão somente das probabilidades dos erros tipo I e tipo II e da decisão tomada (α, β, d), tem sido vista como uma forma pouco precisa de se estabelecer a conclusão do teste. Um dos principais motivos dessas críticas é, sem dúvida, o fato do procedimento de NPW atribuir a mesma medida numérica do grau de conclusão a quaisquer dois valores amostrais "que conduzem à uma mesma decisão, embora um desses valores seja, intuitivamente, muito mais conclusivo que o outro. Essa idéia fica clara após observamos o seguinte.

Exemplo 1: Suponha que observa-se uma v.a. X normalmente distribuida com média θ e variância 1, e deseja-se decidir entre as hipóteses simples H_0: $\theta = -1$ e H_1: $\theta = 1$. O teste

simétrico[1] de Neyman-Pearsan (NP) rejeita H_0 se $X \geq 0$ e neste caso $\alpha = \beta = 0,16$. Note que a mesma decisão será tomada se $X = 0,5$ ou $X = 5$, porém a conclusão $(\alpha, \beta, d) = (0,16; 0,16; d_1)$ estabelecida para qualquer dos dois valores amostrais, não exibe qualquer dependência do resultado $X = x$, que traduza a noção intuitiva de que H_1 é mais "provável" quando $X = 5$ do que quando $X = 0,5$.

Com o objetivo de contornar essa "ineficiência" desse tipo de conclusão e consequentemente tornar a decisão mais dependente da observação amostral. J. Kiefer (1975-1977) em uma série de artigos, propõe uma metodologia na qual coloca a teoria de NPW como um caso particular.

Antes de apresentarmos a Metodologia de Kiefer, é interessante revermos alguns resultados relativos à teoria de sub-álgebras e participações, ferramentas fundamentais na construção de tal metodologia.

2.2 - Sub-álgebras e Partições

Seja $(\mathcal{X}, \mathcal{A})$ um espaço mensurável, onde $\mathcal{X}$ é o conjunto dos possíveis resultados de um experimento estatístico) e $\mathcal{A}$ uma σ-álgebra de sub-conjuntos de $\mathcal{X}$.

---

(1) Um teste é dito simétrico se a distribuição de $f_0(x) / f_1(X)$ sob H_0 é a mesma que a de $f_1(X)/f_0(X)$ sob H_1.

<u>Definição 1</u>: Toda σ-álgebra $\mathcal{D}$ de sub-conjuntos de $\mathcal{X}$ tal que $\mathcal{D} \subset \mathcal{A}$ é denominada "sub-álgebra de $\mathcal{A}$".

<u>Definição 2:</u> Uma classe $\Pi = \{\pi_\delta \subset \mathcal{X};\ \delta \in \Delta\}$, onde Δ é um conjunto de índices, não necessariamente enumerável, é dita uma partição de $\mathcal{X}$ se:

a) $\pi_\delta \cap \pi_\delta' = \phi$ se $\delta \neq \delta'$ e

b) $\bigcup_\Delta \pi_\delta = \mathcal{X}$

Note que se T é uma estatística de $(\mathcal{X}, \mathcal{A})$ em algum espaço (τ, β), então $\{\mathcal{X}_t = T^{-1}(\{t\});\ t \in \tau\}$ é uma partição de $\mathcal{X}$. Tal partição será denominada "partição induzida por T" e será denotada por $\{\mathcal{X}_t : t \in \tau\}$.

Para cada partição Π de $\mathcal{X}$, defina: $\mathcal{A}(\Pi) = \{A \in \mathcal{A} : A = \bigcup_{\Delta'} \pi_\delta$, para algum $\Delta' \subset \Delta\}$, isto é, $\mathcal{A}(\Pi)$ é a classe de todos os conjuntos $\mathcal{A}$-mensuráveis que podem ser escritos como uniões de elementos de Π. Em seguida mostraremos que $\mathcal{A}(\Pi)$ é uma sub-álgebra de $\mathcal{A}$.

Pela maneira como $\mathcal{A}(\Pi)$ foi definida, fica claro que $\mathcal{A}(\Pi) \subset \mathcal{A}$. Assim, basta mostrar que $\mathcal{A}(\Pi)$ é uma σ-algebra de sub-conjuntos de $\mathcal{X}$.

i) $\mathcal{X} \in \mathcal{A}(\Pi)$.

Como $\mathcal{X} = \bigcup_{\Delta} \pi_\delta$ e $\mathcal{X} \in \mathcal{A}$, fica evidente que $\mathcal{X} \in \mathcal{A}(\Pi)$

ii) $A \in \mathcal{A}(\Pi) \rightarrow A^c \in \mathcal{A}(\Pi)$

$A \in \mathcal{A}(\Pi)$ então existe $\Delta' \subset \Delta$ tal que

$A = \bigcup_{\Delta'} \pi_\delta$, assim $A^c = \bigcup_{\Delta'^c} \pi_\delta$

Como $A \in \mathcal{A}(\Pi) \subset \mathcal{A}$ então $A^c \in \mathcal{A}$ e $\Delta'^c = \Delta - \Delta' \subset \Delta$ segue-se que $A^c \in \mathcal{A}(\Pi)$

iii) $A_1, A_2, \ldots$ é uma coleção enumerável de elementos de $\mathcal{A}(\Pi)$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}(\Pi)$.

Se $A_1, A_2, \ldots$ são elementos de $\mathcal{A}(\Pi)$, então existem sub-conjuntos $\Delta_1, \Delta_2, \ldots$ de Δ, tais que:

$A_n = \bigcup_{\Delta_n} \pi_\delta$, $n = 1, 2, \ldots$

Seja $\Delta' = \bigcup_{n=1}^{\infty} \Delta_n$. Assim

$\bigcup_{n=1}^{\infty} A_n = \bigcup_{\Delta'} \pi_\delta$. Como $\bigcup_{n=1}^{\infty} A_n$ é $\mathcal{A}$-mensurável e $\Delta' \subset \Delta$, concluimos que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}(\Pi)$,

completando assim a demonstração.

<u>Definição 3</u>: Sejam $(\mathcal{X}, \mathcal{A})$ um espaço mensurável e $\Pi = \{\pi_\delta : \delta \in \Delta\}$ uma partição de $\mathcal{X}$. A classe $\mathcal{A}(\Pi)$ dada por

$$\mathcal{A}(\Pi) = \{A \in \mathcal{A} : A = \bigcup_{\Delta'} \pi_\delta;\ \text{para algum } \Delta' \subset \Delta\}$$

é chamada de "Sub-álgebra de $\mathcal{A}$ induzida pela partição Π".

A definição acima, de certa forma, estabelece que "toda partição Π de $\mathcal{X}$ induz uma sub-álgebra $\mathcal{D} = \mathcal{A}(\Pi)$ de $\mathcal{A}$". No entanto, o exemplo abaixo ilustra bem o fato de que "nem toda sub-álgebra $\mathcal{D}$ de $\mathcal{A}$ é gerada por alguma partição de $\mathcal{X}$".

<u>Exemplo 2</u>: Sejam $\mathcal{X} = R$ e $\mathcal{A} = \beta$, onde R e β representam a reta e a σ-álgebra de Borel da reta, respectivamente. Considere também a classe.

$$\mathcal{D} = \{D \in \beta : D \text{ ou } D^c \text{ é enumerável}\}$$

Assim $\mathcal{D}$ contém todos os conjuntos unitários $\{x\}$, $x \in R$, e portanto a unica partição que poderia induzir $\mathcal{D}$ seria $\Pi = \{\{x\};\ x \in R\}$. No entanto, $\mathcal{A}(\Pi) = \beta = \mathcal{A} \neq \mathcal{D}$. Logo $\mathcal{D}$ não pode ser induzida por nenhuma partição de $\mathcal{X}$.

Seja $\mathcal{D}$ uma sub-álgebra de $\mathcal{A}$, para cada $x \in \mathcal{X}$ defina:

$$C_x = \{D \in \mathcal{D} ;\ x \in D\} \text{ e } \pi_x = \bigcap_{D \in C_x} D$$

Em palavras, π_x é a interseção de todos os elementos da sub-álgebra $\mathcal{D}$ que contêm o ponto x. Mostraremos em seguida que $\Pi = \{\pi_x;\ x \in \mathcal{X}\}$ constitui uma partição de $\mathcal{X}$.

Para mostrar que Π é uma partição de $\mathcal{X}$, será suficiente verificar que para todo $\{x', x''\} \subset \mathcal{X}$, temos $\pi_{x'} = \pi_{x''}$ ou $\pi_{x'} \cap \pi_{x''} = \phi$, uma vez que $\bigcup_{x \in \mathcal{X}} \pi_x = \mathcal{X}$.

Seja $\{x', x''\} \subset \mathcal{X}$ então

(a) $C_{x'} = C_{x''}$ ou (b) $C_{x'} \neq C_{x''}$

Se $C_{x'} = C_{x''}$, então para todo $D \in \mathcal{D}$, $x' \in D$ se, e somente se $x'' \in D$ e assim $\pi_{x'} = \pi_{x''}$.

Por outro lado se $C_{x'} \neq C_{x''}$, então existe $D_o \in C_{x'} - C_{x''}$ $(C_{x''} - C_{x'})$, tal que $x' \in D_o$ e $x'' \notin D_o$ $(x' \notin D_o$ e $x'' \in D_o)$. Se $D_o \in C_{x'} - C_{x''}$, então

$x' \in D_o$ (1) e

$x'' \notin D_o$ (2)

De (1) segue-se que $\pi_{x'} \subset D_o$

De (2) decorre que $x'' \in D_o^c$ e consequentemente $\pi_{x''} \subset D_o^c$.

Daí:

$\pi_{x'} \cap \pi_{x''} \subset D_o \cap D_o^c = \phi$.

Se $D_0 \in C_{x''} - C_{x'}$, demonstra-se analogamente que $\pi_{x'} \cap \pi_{x''} = \phi$.

Definição 4: Sejam $(\mathcal{X}, \mathcal{A})$ um espaço mensurável e $\mathcal{D}$ uma sub-álgebra qualquer de $\mathcal{A}$. A classe $\Pi(\mathcal{D}) = \{\pi_x;\ x \in \mathcal{X}\}$, acima descrita, é denominada "partição de $\mathcal{X}$ induzida pela sub-álgebra $\mathcal{D}$ de $\mathcal{A}$".

Da definição acima concluimos que "toda sub-álgebra $\mathcal{D}$ de $\mathcal{A}$ induz uma partição $\Pi(\mathcal{D})$ de $\mathcal{X}$". Em seguida ilustraremos através de um exemplo, que "nem toda partição Π de $\mathcal{X}$ é induzida por alguma sub-álgebra $\mathcal{D}$ de $\mathcal{A}$".

Exemplo 3: Sejam:

$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$,

$\mathcal{A} = \{\phi, \{1, 3, 5\}, \{2, 4, 6\}, \mathcal{X}\}$ e

$\Pi = \{\{1, 2, 3\}, \{4, 5, 6\}\}$

Note que $\Pi(\mathcal{A}) = \{\{1, 3, 5\}, \{2, 4, 6\}\}$, a partição induzida por $\mathcal{A}$, é difrente de Π. Por outro lado, $\mathcal{D} = \{\phi, \mathcal{X}\}$ é a única sub-álgebra própria de $\mathcal{A}$. Como $\Pi(\mathcal{D}) = \{\mathcal{X}\} \neq \Pi$, conclui-se que Π não é induzida por nenhuma sub-álgebra de $\mathcal{A}$.

Definição 5: Seja $(\mathcal{X}, \mathcal{A})$ um espaço mensurável. Diz-se que $A \in \mathcal{A}$ é um átomo de $\mathcal{A}$ se A é não vazio e para todo $A_0 \in \mathcal{A}$ tal que $A_0 \subset A$ tivermos $A_0 = \phi$ ou $A_0 = A$.

<u>Proposição 1</u>: Sejam $\Pi = \{\pi_\delta : \delta \in \Delta\}$ uma partição de $\mathcal{X}$ e $\mathcal{D}$ uma sub-álgebra de $\mathcal{A}$, então:

(a) Se Π é $\mathcal{A}$-mensurável (i,e, todo elemento de Π é também elemento de $\mathcal{A}$), então $\Pi = \Pi\ (\mathcal{A}(\Pi))$

(b) $\mathcal{D} \subset \mathcal{A}\ (\Pi\ (\mathcal{D}))$

<u>Demonstração:</u>

(a) Como Π e $\Pi\ (\mathcal{A}(\Pi)$ são partições de $\mathcal{X}$, basta mostrar que para cada $\pi \in \Pi\ (\mathcal{A}(\Pi))$, existe $\pi_o \in \Pi$ tal que $\pi_o = \pi$.

Seja $C_{x_o} = \{ A \in \mathcal{A}\ (\Pi): x_o \in A\}$, assim

$$\pi_{x_o} = \bigcap_{A \in C_{x_o}} A$$

Seja agora π_o o elemento de Π que contém o ponto x_o. Como Π é $\mathcal{A}$-mensurável, π_o é um átomo de $\mathcal{A}\ (\Pi)$ e consequentemente $\pi_o \subset A$ para todo $A \in C_{x_o}$. Como $\pi_o \in C_{x_o}$ temos:

$$\pi_{x_o} = \bigcap_{A \in C_{x_o}} A = \pi_o \text{ para todo } x_o \in \mathcal{X} \text{ tal que } \pi_{xo} = \pi$$

Logo $\pi_o = \pi$.

(b) Sejam D um elemento de $\mathcal{D}$ e π' um elemento qualquer de $\Pi\ (\mathcal{D})$. Para estabelecer o resultado desejado, basta mos que se $\pi' \cap D \neq \phi$, então $\pi' \subset D$.

Suponha que existe $x_0 \in \mathcal{X}$ tal que $x_0 \in \pi'$ e $x_0 \in D$. E seja x' um elemento genérico de π'. Assim, para qualquer $D' \in \mathcal{D}$, $x' \in D' \leftrightarrow x_0 \in D'$. Como $x_0 \in D$, segue-se que $x' \in D$ e consequentemente $\pi' \subset D$.

De acordo com o que foi mostrado acima, para qualquer $D \in \mathcal{D}$ e $\pi' \in \Pi(\mathcal{D})$ temos que $\pi' \cap D = \phi$ ou $\pi' \cap D = \pi'$. Logo D é necessariamente uma união de elementos de $\Pi(\mathcal{D})$ e por conseguinte $D \in A(\Pi(\mathcal{D}))$.

Assim, $\mathcal{D} \subset A(\Pi(\mathcal{D}))$

<u>Definição 6</u>: Seja $(\mathcal{X}, A)$ um espaço mensurável e sejam Π e Π' duas partições quaisquer de $\mathcal{X}$. Diz-se que Π' é mais fina que Π, e denota-se por $\Pi < \Pi'$, se todo elemento de Π é união de elementos de Π', ou equivalentemente, se para cada $\pi' \in \Pi'$ existe $\pi \in \Pi$ tal que $\pi' \subset \pi$.

<u>Proposição 2</u>: Sejam Π_1 e Π_2 duas partições de $\mathcal{X}$ e $\mathcal{D}_1$ e $\mathcal{D}_2$ duas sub-álgebras de A, então:

(a) $\mathcal{D}_1 \subset \mathcal{D}_2 \rightarrow \Pi(\mathcal{D}_1) < \Pi(\mathcal{D}_2)$

(b) $\Pi_1 < \Pi_2 \rightarrow A(\Pi_1) \subset A(\Pi_2)$

<u>Demonstração</u>:

(a) Sejam $x_0 \in \mathcal{X}$ e π_1 o elemento de $\Pi(\mathcal{D}_1)$ que contém x_0. Precisamos mostrar que existe $\pi_2 \in \Pi(\mathcal{D}_2)$ tal que $\pi_2 \subset \pi_1$.

Considere os conjuntos

$C_1 = \{D \in \mathcal{D}_1 : x_o \in D\}$ e

$C_2 = \{D' \in \mathcal{D}_2 : x_o \in D'\}$

Como $\mathcal{D}_1 \subset \mathcal{D}_2$ temos que $C_1 \subset C_2$

Daí,

$$\pi_1 = \bigcap_{D \in C_1} D \supset \bigcap_{D' \in C_2} D' = \pi_2,$$ onde π_2 é o

elemento de $\Pi(\mathcal{D}_2)$ que contém x_o. Assim, $\Pi(\mathcal{D}_1) < \Pi(\mathcal{D}_2)$

(b) $\Pi_1 < \Pi_2$, então cada $\pi_1 \in \Pi_1$ é união de elementos de Π_2. Se $A \in \mathcal{A}(\Pi_1)$, A é união de elementos de Π_1, que por suas vezes são uniões de elementos de Π_2. Logo A é união de elementos de Π_2, e portanto $A \subset \mathcal{A}(\Pi_2)$. Daí $\mathcal{A}(\Pi_1) \subset \mathcal{A}(\Pi_2)$

São esses os resultados que consideramos de maior importância, dentro da teoria de sub-álgebras e partições, para o nosso trabalho. As seções que se seguem serão dedicadas à apresentação e discussão da Metodologia de Kiefer, no que diz respeito ao problema de testar hipóteses simples contra alternativa simples.

2.3 - A Metodologia de Kiefer (um exemplo)

A forma mais simples da teoria de Kiefer, refere-se ao problema de testar hipóteses do tipo,

H_0: X tem densidade (ou função de probabilidade) f_0

H_1: X tem densidade (ou função de probabilidade(f_1

associadas ao modelo $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, onde $\mathcal{X} = R$, $\mathcal{A} = \beta$ (a σ-álgebra de Borel da reta) e $\mathcal{P} = \{P_0, P_1\}$ uma família de medidas com P_0 e P_1 sendo as medidas de probabilidades correspondentes às densidades (ou funções de probabilidades) f_0 e f_1 respectivamente.

A formulação básica dessa teoria é a seguinte:

(1) Particiona-se o espaço amostral $\mathcal{X}$ em uma família $\Pi = \{\mathcal{X}_i^t, t \in \tau, i = 0,1\}$ de sub-conjuntos de $\mathcal{X}$, onde τ é algum conjunto de índices (veremos que, em geral, τ é a imagem inversa de alguma partição ancilar $\{\mathcal{X}^t, t \in \tau\}$).

(2) Quando o resultado amostral x, cai em $\mathcal{X}_i^t$, adota-se a decisão (Λ_i^t, di) que "H_i é verdadeira com coeficiente condicional de confiança (CCC) Λ_i^t".

(3) Os valores Λ_i^t teem uma interpretação frequentista (embora condicional).

Usando-se P_i para denotar a probabilidade quando H_i é verdadeira e $\mathcal{X}^t = \mathcal{X}_0^t \cup \mathcal{X}_1^t$,

P_i (tomar decisão correta $d_i | \mathcal{X}^t) = \Lambda_i^t$

Assim, supondo-se τ enumerável e $P_i(\mathcal{X}^t) > 0$ para todo $t \in \tau$ e $i = 0,1$, podemos interpretar os valores Λ_i^t da seguinte maneira: "Se esse procedimento for realisado em uma longa sequência de experimentos, e se existem muitos experimentos para os quais os valores amostrais caiam em $\mathcal{X}_t$, a decisão correta d_i será, muito provavelmente, tomada em uma proporção aproximadamente Λ_i^t de tais experimentos".

Exemplo 4: (continuação do exemplo 1). Uma vez que no exemplo 1 $\alpha = \beta$, definimos o conjunto τ de maneira que $\Lambda_0^t = \Lambda_1^t$ para todo $t \in \tau$. Note que, no procedimento incondicional do referido exemplo, o conjunto τ tem apenas um elemento t e o espaço amostral $\mathcal{X} = R$ foi particionado em $\mathcal{X}_0^t = (-\infty, 0]$ e $\mathcal{X}_1^t = (0, +\infty)$. Assim os valores $\Lambda_0^t = \Lambda_1^t$ são incondicionais, e além do mais $1 - \Lambda_0^t = \alpha = \beta = 1 - \Lambda_1^t$.

Exemplo 5: Ainda com referência ao exemplo 1, consideremos agora $\tau = \{u, v, w\}$ e a seguinte partição do espaço amostral $\mathcal{X} = R$.

t	$\mathcal{X}_0^t$	$\mathcal{X}_1^t$	$\mathcal{X}^t$
u	$(-\infty, -2]$	$[2, +\infty)$	$(-\infty, -2] \cup [2, +\infty)$
v	$(-2, -1]$	$[1,2)$	$(-2, -1] \cup [1, 2)$
w	$(-1, 0]$	$(0, 1)$	$(-1, 1)$

Com probabilidades

Sob H_0	u	v	w	Sob H_1
$\mathcal{X}_0^t$	0,1587	0,3413	0,3413	$\mathcal{X}_1^t$
$\mathcal{X}_1^t$	0,0013	0,0215	0,1359	$\mathcal{X}_0^t$
$\mathcal{X}^t$	0,1600	0,3628	0,4772	$\mathcal{X}^t$

Assim

$$\Lambda^u = \Lambda_0^u = \Lambda_1^u = \frac{0,1587}{0,1600} = 0,9919$$

$$\Lambda^v = \Lambda_0^v = \Lambda_1^v = \frac{0,3413}{0,3628} = 0,9407$$

$$\Lambda^w = \Lambda_0^w = \Lambda_1^w = \frac{0,3413}{0,4772} = 0,7152$$

Dessa maneira a conclusão passa a depender mais fortemente do resultado $X = x$ obtido, de tal sorte que a conclusão pode der feita da seguinte maneira:

Se $0 < x < 1$ aceita-se H_1 com CCC 0,72

Se $1 \leq x < 2$ aceita-se H_2 com CCC 0,94

Se $x \geq 2$ aceita-se H_1 com CCC 0,99

e de maneira análoga aceita-se H_0 para os valores negativos de x.

Note que adotando-se essa formulação, a sentença (Λ_i^t, di) faz distinção entre os resultados $x = 0,15$, $x = 1,5$ e $x = 15$, o que não acontecia com a teoria de NPW.

Nesse sentido a metodologia de Kiefer alcança seu objetivo de tornar a conclusão do teste mais dependente da observação amostral, do que no caso particular onde $\tau = \{t\}$ da teoria de NPW.

2.4 - Notação e Definição

Seja $(\mathcal{X}, A, P)$ um modelo estatístico, onde $\mathcal{X}$ é um espaço euclidiano, A uma σ-álgebra de sub-conjuntos de $\mathcal{X}$ e P uma família de medidas como na seção anterior. Seja X a v.a. que representa o experimento $E = (\mathcal{X}, \Omega, f)$ como resultados em $\mathcal{X}$.

O espaço de decisão será representado pelo conjunto $D = \{d_0, d_1\}$ onde d_i $(i = 0,1)$ representa a decisão "H_i é verdadeira".

Uma regra de decisão é uma estatística $Z : \mathcal{X} \to D$, ou seja, Z é uma função $\mathcal{A}$-mensurável, que a cada ponto $x \in \mathcal{X}$ associa uma (e somente uma) decisão $d \in D$.

Com a finalidade de evitar dificuldades de ordem teórica, os modelos estatísticos aqui considerados admitem a existência da medida de probabilidade condicional regular. Assim sendo, toda sub-álgebra $\mathcal{F}$ de $\mathcal{A}$ pode ser considerada uma sub-álgebra condicionante, no entanto é conveniente reservar atenção apenas a sub-álgebras induzidas por estatísticas definidas em $(\mathcal{X}, \mathcal{A})$, uma vez que estamos interessados em uma partição mensurável de $\mathcal{X}$.

Se T é uma tal estatística, então o conjunto de imagens de T será o conjunto de índices τ e a sub-álgebra correspondente: $\mathcal{F} = \{F \in \mathcal{A} : F = T^{-1}(\eta)$, para algum $\eta \subset \tau\}$ é a maior sub-álgebra que induz a mesma partição induzida por T.

Para verificar este fato mostraremos que $\mathcal{F} = \mathcal{A}(\Pi(T))$ onde, $\Pi(T) = \{\mathcal{X}^t : \mathcal{X}^t = T^{-1}(\{t\}), t \in \tau\}$ é a partição induzida pela estatítica T.

(a) Se $F \varepsilon \mathcal{F}$, então existe $\eta \subset \tau$ tal que $\mathcal{F} = T^{-1}(\eta)$, daí $F = T^{-1}(\bigcup_{t \varepsilon \eta} \{t\}) = \bigcup_{t \varepsilon \eta} T^{-1}(\{t\})$. Como $T^{-1}(\{t\}) \varepsilon \Pi(T)$ para todo $t \varepsilon \tau$, F é união de elementos de $\Pi(T)$. Logo $F \varepsilon \mathcal{A}(\Pi(T))$.

(b) $F \varepsilon \mathcal{A}(\Pi(T))$, então existe $\eta \subset \tau$ tal que $F = \bigcup_{t \varepsilon \eta} T^{-1}(\{t\})$, assim $F = T^{-1}(\eta)$ e consequententemente $F \varepsilon \mathcal{F}$.

(c) Uma vez que T é uma estatística em $(\mathcal{X}, \mathcal{A})$, $\Pi(T)$ é uma partição mensurável de $\mathcal{X}$. Portanto, pela proposição 1 (a), $\Pi(T) = \Pi(\mathcal{A}(\Pi(T))) = \Pi(\mathcal{F})$, ou seja, $\mathcal{F}$ induz a mesma partição que aquela induzida por T.

(d) Seja $\mathcal{D}$ uma sub-álgebra qualquer de $\mathcal{A}$ tal que $\Pi(\mathcal{D}) = \Pi(T)$. Pela proposição 1 (b), concluimos que $\mathcal{D} \subset \mathcal{A}(\Pi(\mathcal{D})) = \mathcal{A}(\Pi(T)) = \mathcal{F}$. Assim, $\mathcal{F}$ é a maior sub-álgebra que induz $\Pi(T)$.

A partição $\Pi(T)$ definida acima será denominada "partição condicionante", e neste ponto é interessante lembrar que se T é uma estatística ancilar, então não haverá perda de informação ao se considerar medidas de probabilidades condicionais do tipo $P(.|\mathcal{X}^t)$, $\mathcal{X}^t \varepsilon \Pi(T)$. (Princípio de Condicionalidade - Capítulo 1).

A cada regra de decisão Z está associada uma partição $\Pi(Z) = \{Z^{-1}(d_i);\ d_i \varepsilon D,\ i = 0,1\}$. A partição citada em (1) da seção 2.3 é constituída pelos conjuntos

$\mathcal{X}_i^t = \mathcal{X}^t \cap Z^{-1}(d_i)$, onde $\mathcal{X}^t \in \Pi(T)$, $Z^{-1}(d_i) \in \Pi(Z)$, $t \in \tau$ e $i = 0,1$. A partição $\Pi(Z)$ a pouco definida é denominada "partição básica de $\mathcal{X}$" e é a partição mais grossa, que é mais fina que $\Pi(T)$ e $\Pi(Z)$ simultaneamente.

Um procedimento condicional de confiança é um par (Z, F) ou (Z, T) (ou ainda a partição Π construida a partir de intersecção dos elementos de $\Pi(T)$ e $\Pi(Z)$), e sua função de confiança condicional é o conjunto $\Lambda = \{\Lambda_0^t, \Lambda_1^t\}$ de probabilidades condicionais.

$$\Lambda_i^t = P_i\{X \in Z^{-1}(d_i) \mid T=t\};\ i = 0,1$$

Uma sentença condicional de confiança associado ao procedimento (Z, F) é o par (Z, Λ) e é usada como se segue: se $X = x_0$ e consequentemente $T(X) = T(x_0)$, afirma-se que "para $i = 0,1$ tem-se C.C.C. $\Lambda_i(x_0) = \Lambda_i^{T(x_0)}$ de estar correto se H_i é verdadeira". Assim se $Z(x_0) = d_i$ conclui-se que H_i é verdadeira.

Neste contexro da metodologia de Kiefer se tomarmos $F = \{\phi, \mathcal{X}\}$ estaremos diante da teoria de NPW, onde Λ_i é a probabilidade de uma decisão correta quando H_i é a hipótese verdadeira $(i = 0,1)$.

Até o presente momento foram mostradas técnicas para se tomar decisões e estabelecer coeficientes condicionais de confiança a elas associadas, porém nada foi dito acerca do que pode ser considerada uma boa partição.

Na seção que se segue iremos fazer breves considerações sobre admissibilidade de uma partição e apresentar restrições que possam conduzir à escolha de uma partição conveniente.

2.5 - Admissibilidade e Escolha de Partições

Considere as aleatórias $\Lambda_i(X)$ $i = 0,1$ que assume os valores $\Lambda_i^{T(X)}$. Elementos de grande importância para o estudo de admissibilidade, são as leis de probabilidades das caudas $G_i(t) = P_i(\Lambda_i(X) > t, Z(X) = d_i)$ e os conjuntos F-mensuráveis $Q_i = \{\Lambda_i(X) = 0\}$, que são as uniões dos elementos $\mathcal{X}_i^t$ de $\Pi(T)$ onde nunca se pode estabelecer uma decisão correta para H_i com probabilidade positiva.

Nas condições da anotação introduzidas, diz-se que um procedimento é não degenerado se para cada $i = 0,1$, $P_i(Q_i) = 0$.

Denotaremos por C a classe de todos os procedimentos Π e por C^+ a classe de todos os procedimentos não degenerados. Dado um conjunto de índices τ (conjunto de imagens de uma estatística T), denotaremos por C^τ a classe de todos os procedimentos cujas partições sejam indexadas por τ e por $C^{\tau+} = C^\tau \cap C^+$ a classe de tais procedimentos que sejam não degenerados.

Se o conjunto de índices τ é finito e possui L elementos, usaremos a notação C^L e C^{L+} no lugar de C^τ e $C^{\tau+}$ respectivamente.

Em princípio parece intuitivo considerar-se uma partição Π "melhor" que uma outra Π', quando as leis de probabilidade das caldas, G_i satisfazem:

$$G_i(t) \leqq G_i'(t), \quad i = 0,1, \quad t > 0 \quad (I)$$

com desigualdade estrita em algum caso.

Dizemos que é intuitivo uma vez que o critério acima estabelece que Λ_i em Π é estocasticamente não menor que Λ_i' em Π' para ambos os valores de i e estocasticamente maior para pelo menos um desses valores. Além do mais esse critério em C^1 coincide com o critério usual de admissibilidade na metodologia de NP.

No entanto esse critério sozinho não é sensível em C^L quando $L > 1$, uma vez que ele só considera a lei de probabilidade das caldas sob P_i quando $Z(X) = di$. Assim parece razoável que algum critério adicional a (I) tenha relação com o estudo dessa lei de probabilidades quando $Z(X) = d_1-i$, ou seja

$$P_i(\Lambda_i(X) > t;\ Z(X) = d_1-i), \quad i = 0,1$$

Isso sugere que uma partição Π seja melhor que Π' se as condições em (I) forem satisfeitas e além disso tivermos:

$$P_i\,(Q_i) \leq P_i\,(Q_i'),\ i = 0,1 \qquad \text{(II)}$$

Definição 7: Uma partição Π é chamada uma partição de Neyman-Pearson se sua partição básica $\{\mathcal{X}_0, \mathcal{X}_1\}$ constitui um procedimento admissível em C^1.

Os teoremas que se seguem, estabelecem condições necessárias e suficientes para a obtenção de partições admissíveis.

Teorema 1: Uma partição Π é admissível em C, no sentido de (I) e (II) se, e somente se ela é uma partição de Neyman-Pearson.

Teorema 2: Uma partição não degenerada Π em C^{L^+} (resp. C^+) é admissível entre partições de C^{L^+} (resp. C^+) no sentido de (I) somente, se, e somente se ela é uma partição de Neyman-Pearson.

Devemos observar que diante das condições de admissibilidade impostas em (I) e (II), não é fácil a obtenção de partições admissíveis, a menos que estejamos diante de testes simétricos para distribuições contínuas. Mesmo assim, lançando mão do problema mencionado no final do capítulo 1, tentaremos uma aplicação para a metodologia de Kiefer, comentando sua limitações e vantagens quanto ao grau de influência do resultado amostral sobre a conclusão do teste.

Capítulo III
Aplicações

3.1 - Introdução

Neste capítulo apresentamos uma aplicação de cada um dos métodos sugeridos nos capítulos anteriores. O problema a ser analisado, trata de estimar a "preferência sexual" do inseto Drosóphila Melanogaster e surgiu de pesquisa realizada pelo Prof. Dr. Luiz Edmondo Magalhães no Laboratório de Genética de Populações do Instituto de Biociências da USP. Os dados foram relatados por Pereira (1971).

3.2 - O Problema Genético

Suponha que uma fêmea selvagem heterozigota (e+), seja colocada na presença de dois machos, um selvagem heterozigoto (e+) e o outro ênoby (ee). Nessas condições, tem-se dois possíveis tipos de cruzamentos:

Tipo I: macho (e+) x fêmea (e+)

Tipo II: macho (ee) x fêmea (e+)

Definamos a "preferência sexual" da fêmea (e+) pelo macho (e+) como sendo a probabilidade de haver um cruzamento do tipo I e de modo semelhante a "preferência sexual" pelo macho (ee) com a probabilidade de um cruzamento do tipo II. Se supusermos que sempre haverá um cruzamento, podemos escrever:

$$P\ (\text{cruzamento tipo I}) = \theta$$

$$P\ (\text{cruzamento tipo II}) = 1 - \theta$$

A hipótese básica levantada pelo Prof. Edmondo Magalhães é que a fêmea (e+) cruza mais frequentemente com machos (e+) que com machos (ee), o que nos leva à hipótese estatística $\theta > 1/2$.

Analisando-se os cruzamentos do ponto de vista genético, podemos chegar às seguintes conclusões:

a) Nos cruzamentos do tipo I, a proporção de descendentes selvagens é de 3/4, dos quais 1/4 é homozigoto, e consequentemente 1/4 dos descendentes são ébony. (Fig. 1).

b) Nos cruzamentos tipo II, as proporções de descendentes selvagens e ébony são iguais a 1/2. (Fig. 2).

c) Em qualquer dos cruzamentos os desecendentes selvagens e ébony são produzidos independentemente.

CRUZAMENTO TIPO I

M \ F	e	+
e	ee	e+
+	e+	++

Fig. 1

CRUZAMENTO TIPO II

M \ F	e	+
e	ee	e+
e	ee	e+

Fig. 2

3.3 - O Experimento

Em uma caixa de população foram colocadas 250 fêmeas (e+), 250 machos (e+) e 250 machos (ee), lá permanecendo por um período de três horas. Esse tempo foi estipulado pelo pesquisador, baseando-se no fato de que uma mesma fêmea não cruza mais de uma vez em tal período e que esse tempo é suficiente para que todas elas consigam cruzar.

É importante salientar que um macho pode cruzar mais de uma vez nesse intervalo de tempo, e como o número de machos é o dobro do número de fêmeas, é sempre possível à fêmea "escolher" qualquer dos tipos de macho. Esse fato juntamente com o grande número de fêmeas, torna razoável a suposição de que os cruzamentos sejam efetuados independentemente uns dos outros.

Uma observação a ser feita é que durante a experiência, dezoito fêmeas foram perdidas, ficando nossa amostra reduzida a 232 cruzamentos.

Sendo impossível identificar cada um dos cruzamentos e efetuar a contagem de cada um dos tipos, a estimativa da "preferência sexual", será baseada no tamanho da descendência (N) e o número de descendentes selvagens (X) em cada geração, dado que estes são fortes indicadores do provável tipo de cruzamento.

3.4 - O Problema Genético e o Princípio de Verossimilhança

3.4.1 - METODOLOGIA

Sejam N[1] uma variável aleatória com distribuição qualquer e T uma variável aleatória com distribuição $P(T=1) = \theta$ e $P(T=2) = (1-\theta)$, $0 \leq \theta \leq 1$. Seja X uma variável aleatória tal que:

$$P(X=x \mid N=n) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1\ldots, n$$

$$\text{onde, } p = \begin{cases} p_1 & \text{se } T = 1, \ 0 \leq p_1 \leq 1 \\ p_2 & \text{se } T = 2, \ 0 \leq p_2 \leq 1 \end{cases}$$

(1) N assumindo valores inteiros não negativos

Assim

$$P(X = x | N = n) = P(X = x | N = n, T = 1) . P(T = 1) +$$

$$+ P(X = x | N = n, T = 2) . P(T = 2)$$

Logo

$$P(X = x | N = n) = \binom{n}{x} p_1^x (1 - p_1)^{n-x} \theta + \binom{n}{x} p_2^x (1-p_2)^{n-x} (1-\theta)$$

$x = 0, 1, 2, \ldots n;$ $n = 0, 1, 2, \ldots;$ $0 \leq \theta \leq 1;$ $0 \leq p_1 \leq 1;$

$0 \leq p_2 \leq 1.$

Dessa forma, dispondo-se de uma amostra aleatória $\{(n_i, x_i)\}$ $i = 1, 2, \ldots, K$, temos a seguinte função de verossimilhança.

$$L(\theta | (\underset{\sim}{n}, \underset{\sim}{x})) = \prod_{i=1}^{K} \binom{n_i}{x_i} [p_1^{x_i} (1-p_1)^{n_i - x_i} \theta + p_2^{x_i} (1-p_2)^{n_i - x_i} (1-\theta)],$$

$0 \leq \theta \leq 1$

Consideremos a função de verossimilhança padronizada dada por:

$$L^*(\theta | \underset{\sim}{n}, \underset{\sim}{x})) = \frac{L(\theta | \underset{\sim}{n}, \underset{\sim}{x}))}{\max\limits_{0 \leq u \leq 1} L(u | \underset{\sim}{n}, \underset{\sim}{x}))} \text{ (1)}, \qquad 0 < \theta < 1$$

---

(1) Como essa metodologia está sendo dirigida especificamente à nossa aplicação, veremos mais adiante que esse máximo existe e é único.

De acordo com o Lema 2 do Capítulo I, a estatística $T(\underset{\sim}{N}, \underset{\sim}{X}) = L^*(\theta | (\underset{\sim}{N}, \underset{\sim}{X}))$, $0 \leq \theta \leq 1$ é uma estatística suficiente minimal e pelo princípio de verossimilhança toda informação relevante sobre θ está contida no gráfico desta função.

Assim o estimador natural de θ é o estimador de máxima verossimilhança

$\hat{\Theta}(\underset{\sim}{N}, \underset{\sim}{X}) = \theta_0$ tal que

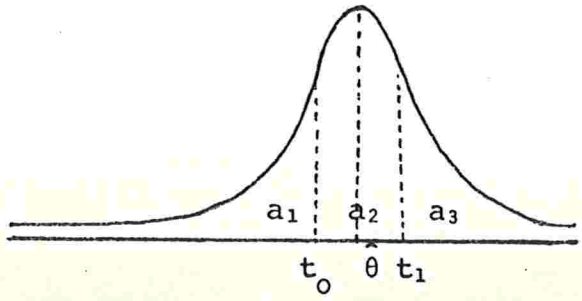
$$L^*(\theta_0 | (\underset{\sim}{N}, \underset{\sim}{X})) \geq L^*(\theta | (\underset{\sim}{N}, \underset{\sim}{X})), \forall\ 0 \leq \theta \leq 1.$$

Para se construir um intervalo de credibilidade, achamos razoável supor que o verdadeiro valor do estado da natureza. θ . esteja contido no menor intervalo que concentre uma certa proporção $(1 - \alpha)$ da área total sob a função de verossimilhança (Fig. 3). Assim, definimos como limites de credibilidade os valores extremos do intervalo tal que:

$$\int_{t_0}^{t_1} L^*(\theta | (\underset{\sim}{N}, \underset{\sim}{X}))\, d\theta = 1 - \alpha \quad \text{e}$$

$t_1 - t_0 \leq t_1' - t_0'$ para todo (t_0', t_1') satisfazendo

$$\int_{t_0'}^{t_1'} L^*(\theta | (\underset{\sim}{N}, \underset{\sim}{X}))\, d\theta = 1 - \alpha$$



t_0 e t_1 são tais que

$$\frac{a_2}{a_1 + a_2 + a} = 1 - \alpha$$

Um teste de hipóteses para $H_o: \theta = \theta_o$ contra a alternativa composta $H_1: \theta > \theta_o$, pode ser construido baseando-se na seguinte regra de decisão: "Rejeitar H_o se, e somente se $t_o > \theta_o$".

3.4.2 - APLICAÇÃO AOS DADOS GENÉTICOS

Consideremos as seguintes variáveis aleatórias: N_i = nº de descendentes do i-ésimo cruzamento, X_i = nº de descendentes selvagens do i-ésimo cruzamento e

$$T_i = \begin{cases} 1 & \text{se o i-ésimo cruzamento é do tipo I} \\ 2 & \text{se o i-ésimo cruzamento é do tipo II} \end{cases}$$

Levando-se em consideração as leis da genética temos:

$$P(x_i = x_i | N_i = n_i) = \begin{cases} \binom{n_i}{x_i} \left(\frac{3}{4}\right)^{x_i} \left(\frac{1}{4}\right)^{n_i - x_i} & \text{se } T_i = 1 \\ \binom{n_i}{x_i} \left(\frac{1}{2}\right)^{n_i} & \text{se } T_i = 2 \end{cases}$$

$$x_i = 0, 1, 2, \ldots, n_i$$

Assim,

$$P\ (X_i = x_i \,|\, N_i = n_i) = \binom{n_i}{x_i} \left(\frac{3}{4}\right) x \left(\frac{1}{4}\right)^{n_i - x_i} \theta + \binom{n_i}{x_i} \left(\frac{1}{2}\right)^{n_i} (1-\theta),$$

onde $\theta = P\ (T_i = 1)$

Após algumas operações elementares esta probabilidade pode ser escrita como:

$$P\ (X_i = x_i \,|\, N_i = n_i) = \binom{n_i}{x_i} \left(\frac{1}{4}\right)^{n_i} [3^{x_i}\ \theta + 2^{n_i}\ (1-\theta)]$$

$$n_i = 0, 1, \ldots;\ x_i = 0, 1, \ldots,\ n_i\ ;\ 0 \leq \theta \leq 1$$

Uma vez que funções de verossimilhança proporcionais contêm a mesma informaçaõ sobre θ (Princípio de Verossimilhança, Capítulo I), podemos definir

$$L\ (\theta \,|\, (\underset{\sim}{n},\ \underset{\sim}{x})) = \prod_{i=1}^{232} [3^{x_i}\ \theta + 2^{n_i}\ (1-\theta)],\ 0 \leq \theta \leq 1$$

como sendo a função de verossimilhança gerada pelo resultado experimental $(\underset{\sim}{n}, \underset{\sim}{x})$.

Baseado nas observações do Prof. Edmondo Magalhães (Tabela 1, Anexo A), o Programa 1 do Anexo B calcula os valores desta função para θ variando de zero até um com incrementos iguais a 0,01. Os valores desta função, bem como o seu gráfico podem ser vistos no Anexo A (Tabela 3 e Fig. 1, respectivamente).

Verificando-se os valores da função de verossimilhança constatamos que $L(\theta \,|\, (\underset{\sim}{n}, \underset{\sim}{x}))$ cresce para valores de θ no intervalo [0; 0,62] e decresce para valores no intervalo (0,62; 1,0).

Assim, o máximo de $L(\theta|(\underset{\sim}{n}, \underset{\sim}{x}))$ ocorre quando $\theta = 0{,}62$, sendo então $\hat{\theta} = 0{,}62$ a estimativa de verossimilhança para a "preferência sexual" da Drosophila Melanogaster.

No entanto o interesse é testar a viabilidade da hipótese levantada pelo Prof. Edmondo Magalhães de que a fêmea (e+) cruza mais frequentemente com machos (e+). Dessa forma, estamos interessado em testar a hipótese $H_o : \theta = 1/2$, contra a alternativa composta $H_1 : \theta > 1/2$.

Uma vez que a estatística

$$T(\underset{\sim}{n}, \underset{\sim}{x}) = \frac{L(\theta|\underset{\sim}{n}, \underset{\sim}{x})}{\max\limits_{\theta \leq \mu \leq 1} L(\mu|(\underset{\sim}{n}, \underset{\sim}{x}))}, \quad 0 \leq \theta \leq 1$$

é suficiente minimal para estimar θ, resolvemos estudar sua variabilidade e considerar como limites de credibilidade, os limites do menor intervalo sob $T(.|(\underset{\sim}{n}, \underset{\sim}{x}))$ que contenha pelo menos 95% da área total sob esta curva.

O Programa 2 do Anexo B calcula a área total sob T e vai acumulando áreas parciais a partir de $\hat{\theta} = 0{,}62$ movendo os limites na direção conveniente para que o intervalo final, no instante em que a razão entre as áreas parcial e total ultrapasse 0,95, seja seguramente o menor intervalo que contenha esssa proporção de área.

Para tais limites de credibilidade foram encontrados os valores $t_o = 0{,}56$ e $t_1 = 0{,}69$ (Tabela 4, Anexo A), assim nosso intervalo de credibilidade é $(t_o, t_1) = (056; 069)$.

Baseando-se na regra de decisão estabelecida na seção 4.3.1, rejeitamos a hipótese H_o uma vez que $t_o = 0,56 > 1/2 = \theta_o$, o que sugere que a hípotese levantada pelo Prof. Edmondo Magalhães seja verdadeira.

3.5 - O Problema Genético e a Teoria de Kiefer

Nessa aplicação nos deteremos em partições pertencentes a C^2, ou seja, partições cujo conjunto de índices τ possuem apenas 2 elementos. Tal conjunto será constituído pelo elementos w e s, para sugerir uma devisão fraca (weak) ou uma decisão forte (Strong), respectivamente.

Tentativas foram feitas no sentido de se obter uma partição conveniente para cada tamanho de descendência (N_i), no entanto não foi possível conseguir partições para as quais Λ_j^t fossem iguais para todo N_i, quando j e t fixados.

Esta restrição se faz importante, uma vez que sua inobservância nos levaria a tomar decisões idênticas com coeficientes condicionais de confiança diferentes. Em outras palavras, poderíamos, por exemplo, para a i-ésima observação tomarmos a decisão d_1 com C.C.C Λ_1^s (n_i, x_i) e para a k-ésima obervação tomar a mesma decisão d_1, porém com CC Λ_1^s $(n_k, x_k) \neq \Lambda_1^s$ (n_i, x_i), se $n_i \neq n_k$.

Essa impossibilidade, deve-se ao fato dos testes não serem simétricos e além do mais as distribuições envolvidas serem discretas.

Diante disto, resolvemos, a exemplo de Pereira (1971), considerar apenas os 40 (menor tamanho de descendência) primeiros descendentes de cada cruzamento. Esses dados se encontram na Tabela 2, Anexo A.

Com o auxílio de tabelas binomiais, particionamos o espaço amostral $\mathcal{X} = \{0, 1, \ldots, 40\}$ da seguinte maneira:

$\mathcal{X}_2^S = [0, 19]$; $\mathcal{X}_2^W = [20, 25]$

$\mathcal{X}_1^W = [26,30]$; $\mathcal{X}_1^S = [31,40]$

Assim nossa partição básica é $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2\}$

onde,

$\mathcal{X}_1 = \mathcal{X}_1^W \cup \mathcal{X}_1^S = [26,40]$, que conduz à decisão d_1 que o cruzamento é do tipo I, e $\mathcal{X}_2 = \mathcal{X}_2^W \cup \mathcal{X}_2^S = [0,25]$, que conduz à decisão d_2 que o cruzamento é do tipo II.

Pode-se verificar que essa partição é de Neymam-Pearson, pois

$$\lambda(x) = \frac{f_1(x)}{f_2(x)} = \frac{\binom{40}{x}\left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{40-x}}{\binom{40}{x}\left(\frac{1}{2}\right)^{40}} = \frac{\frac{3^x}{4^{40}}}{\frac{1}{2^{40}}} =$$

$$= \frac{\frac{3^x}{2^{80}}}{\frac{1}{2^{40}}} = \frac{3^x}{2^{80}} \cdot 2^{40} = \frac{3^x}{2^{40}}$$

Assim $\lambda(x) > 1 \leftrightarrow 3^x > 2^{40} \leftrightarrow x \ln 3 > 40 \ln 2 \leftrightarrow x > 40 . \frac{\ln 2}{\ln 3} \cong$
$\cong 25,24$

ou seja, $x \geqq 26$, que é a partição básica de $\mathcal{X}$.

Considere as probabilidades dos conjuntos $\mathcal{X}_j^t$, $j = 1,2$ e $t = w,s$, sob cada uma das distribuições:

a) Para f_1 [$X \sim B(40, 3/4)$]

t	$\mathcal{X}_1^t$	$\mathcal{X}_2^t$	$\mathcal{X}^t$
w	0,50602	0,05426	0,56028
s	0,43954	0,00018	0,43972
$\mathcal{X}_j$	0,94556	0,05444	1,0

b) Para f_2 [$X \sim B(40, 1/2)$]

t	$\mathcal{X}_1^t$	$\mathcal{X}_2^t$	$\mathcal{X}^t$
w	0,04001	0,52234	0,56235
s	0,00034	0,43731	0,43765
$\mathcal{X}_j$	0,04035	0,95965	1,0

Observando-se que $P_1(\mathcal{X}^W) = 0,56028 \cong 0,56235 = P_2(\mathcal{X}^W)$ e que $P_1(\mathcal{X}^S) = 0,43765 \cong 0,43972 = P_2(\mathcal{X}^S)$, podemos, para efeitos práticos, dizer que a partição $\{\mathcal{X}^W, \mathcal{X}^S\}$ é ancilar.

Os coeficientes condicionais de confiança são:

$$\Lambda_1^W = P_1 \ (X \in \mathcal{X}_1^W \mid X \in \mathcal{X}^W) \cong 0,903$$

$$\Lambda_1^S = P_1 \ (X \in \mathcal{X}_1^S \mid X \in X^S) \cong 0,999$$

$$\Lambda_2^W = P_2 \ (X \in \mathcal{X}_2^W \mid X \in \mathcal{X}^W) \cong 0,929$$

$$\Lambda_2^S = P_2 \ (X \in \mathcal{X}_2^S \mid X \in \mathcal{X}^S) \cong 0,999$$

Para efeito de admissibilidade da partição, seria interessante que $\Lambda_1^W = \Lambda_2^W$ e $\Lambda_1^S = \Lambda_2^S$. Porém, a maior dessas diferenças, $|\Lambda_1^W - \Lambda_2^W| \cong 0,026$, é menor que a menor diferença, $|\Lambda_2^W - \Lambda_2^S| = 0,07$, calculada pela mesma medida sobre w e s, o que nos dá um sentido de "quase admissibilidade" pelo critério (I).

Uma vez que a partição usada é de Neymam-Pearson, podemos considerar que estamos diante de uma partição admissível em C^{2+} (Teorema 2, Capítulo II).

Considere as variáveis

$$v_i = \begin{cases} 1 \text{ se o i-ésimo cruzamento conduz à decisão } d_1 \\ 2 \text{ se o i-ésimo cruzamento conduz à decisão } d_2 \end{cases}$$

Os valores de v_i estão na Tabela 2 do Anexo A, juntamente com os respectivos valores de estatística T. (níveis de classificação forte (s) ou fraca (w)).

O quadro abaixo mostra um resumo das classificações dos 232 cruzamentos.

Níveis de Classificação / Tipo de Cruzamento ↓	w	s	Total
I	74	65	139
II	52	41	93
Total	126	106	232

Antes de partirmos para a estimativa da "preferência sexual" da Drosóphile Melanogaster, iremos testar se existe associação entre os níveis de classificação (forte ou fraca) e o tipo de cruzamento no qual a observação foi classificada (I ou II). Para tanto, utilizaremos o teste de qui-quadrado para análise de tabelas de contigência.

Considere a tabela de frequências esperadas sob a hipótese de independência.

Níveis de Classificação	w	s	Total
Tipo de Cruzamento ↓			
I	75,49	63,51	139
II	50,51	42,49	93
Total	126	106	232

Daí,

$\chi^2_c \cong 0,1606.$

Como $P(\chi^2_{(1)} \geq 0,1606) \cong 0,7$, concluimos que não existe associação entre os níveis de classificação e o tipo de cruzamento (classificado).

Vale a pena salientar que essa "independência" não significa que esta não seja uma boa partição até pelo contrário, pois isso sugere que nenhum dos cruzamentos tenham uma tendência a uma classificação fraca, o que nos faria acreditar numa maior possibilidade de erros de classificação.

Visando encontrar uma estimativa para a "preferência sexual", suponhamos inicialmente que adotamos uma amostragem seletiva, na qual, sempre que o número de descendentes selvagens estiver no intervalo [20, 30] a observação será rejeitada, sob a alegação de conduzir a uma "classificação duvidosa". Assim nossa amostra fica reduzida a

106 observações, das quais, 65 são classificadas como cruzamento do tipo I. Daí:

$$\hat{\theta}_s = \frac{\text{nº de elementos em } \mathcal{X}_2^s}{\text{nº de elementos em } \mathcal{X}^s} \overset{\text{not}}{=} \frac{\#(\mathcal{X}_2^s)}{\#(\mathcal{X}^s)} = \frac{65}{106} \cong 0{,}6132$$

Se por outro lado, selecionamos apenas as observações cujo número de descendentes selvagens se encontrem em [20, 30] (decisão pouco intuitiva), teremos agora uma amostra de 126 cruzamentos, dos quais, 74 foram classificados como do tipo I. Assim

$$\hat{\theta}_w = \frac{\#(\mathcal{X}_2^w)}{\#(\mathcal{X}^w)} = \frac{74}{126} \cong 0{,}5873$$

A proximidade de $\hat{\theta}_w$ e $\hat{\theta}_s$ deve-se à "independência entre níveis e tipos de classificação".

Consideramos agora uma combinação entre as estimativas $\hat{\theta}_w$ e $\hat{\theta}_s$ dada por:

$$\hat{\theta}_c = \hat{\theta}_w . P(X \in \mathcal{X}^w) + \hat{\theta}_s . P(\mathcal{X} \in \mathcal{X}^s)$$

Porém

$$P(X \in \mathcal{X}^w) = P(X \in \mathcal{X}^w \mid \text{cruz.tipo I}) . P(\text{cruz.tipo I}) +$$

$$+ P(X \in \mathcal{X}^w \mid \text{cruz.tipo II}) . P(\text{cruz.tipo II}) =$$

$$= P_1(\mathcal{X} \in \mathcal{X}^w) . \theta + P_2(X \in \mathcal{X}^w) . (1-\theta)$$

Como a partição $\{\mathcal{X}^W, \mathcal{X}^S\}$ é ancilar, segue-se que $P_1 (X \varepsilon \mathcal{X}^W) =$

$= P_2 (X \varepsilon \mathcal{X}^W) = 0,56.$

Assim,

$$P (X \varepsilon \mathcal{X}^W) = 0,56\ \theta + 0,56\ (1-\theta) = 0,56.$$

Consequentemente,

$$P (X \varepsilon \mathcal{X}^S) = 1 - P (X \varepsilon \mathcal{X}^W) = 0,44$$

Logo,

$$\hat{\theta}_C = \hat{\theta}_W\ P (X \varepsilon \mathcal{X}^W) + \hat{\theta}_S .\ P(X \varepsilon \mathcal{X}^S) =$$

$$= 0,5873 \times 0,56 + 0,6132 \times 0,44 \cong 0,5987$$

Do ponto de vista de informação, esta última estimativa "parece" mais confiável, uma vez que ela faz uso de todas as observações amostrais, enquanto que as duas primeiras desprezam boa parte destas mesmas observações.

Um outro estimador, que também leva em conta todas as observações, é o estimador de θ em C^1 quando consideramos apenas a partição básica $\{\mathcal{X}_1, \mathcal{X}_2\}$, ou seja:

$$\hat{\theta} = \frac{\#(\mathcal{X}_2)}{\#(\mathcal{X}_1 \cup \mathcal{X}_2)} = \frac{139}{232} = 0,5991$$

Devemos observar que $\hat{\theta}_s$ é a estimativa que mais se aproxima daquela encontrada na seção anterior, quando usamos a estatística suficiente representada pelo gráfico da função de verossimilhança, e para qual todos os tamanhos de descendência foram considerados. Desta forma nos parece, que mesmo em situação adversa (testes não simétricos para distribuições discretas), conseguimos algum ganho com o condicionamento em uma partição ancilar.

Considerando-se que para a obtenção de $\hat{\theta}_s$, usamos somente as observações para as quais podíamos tomar uma decisão forte, e isso nos levou à mais confiável (ver seção 3.6) das estimativas desta seção, podemos pensar no coeficiente condicional de confiança associado a uma decisão, como sendo um primeiro passo no sentido de se atribuir um valor numérico à informação contida em um resultado experimental.

3.6 - Análise Comparativa das Estimativas

Antes de comparar a estimativa obtida na seção 3.4 com as obtidas na seção 3.5, é importante lembrar que elas foram baseadas em observação diferentes. Enquanto a primeira usa as observações amostrais, tais quais elas foram observadas, as últimas usam um truncamento das observações originais. Dessa forma, para que possamos comparar essas estimativas se faz necessário verificar, antes, se alguma perda significativa ocorreu diante dessa transformação dos dados.

Consideramos, para qualquer observação (n_i, x_i), a seguinte regra de classificação.

$$w_i = \begin{cases} 1 \text{ se } P_1 (\chi_i = x_i | N_i = n_i) > P_2 (\chi_i = x_i | N_i = n_i) \\ 2 \text{ caso contrário} \end{cases}$$

Como P_1 e P_2 são medidas de probabilidades associadas às distribuições $B(n_i, 3/4)$ e $B(n_i, 1/2)$ respectivamente, temos que:

$P_1 (\chi_i = x_i | N_i = n_i) > P_2 (\chi_i = x_i | N_i = n_i)$ se, e somente se

$$x_i > \frac{\ell n2}{\ell n3} \cdot n_i$$

Assim,

$$w_i = \begin{cases} 1 \text{ se } x_i > (\ell n2 | \ell n3) \cdot n_i \\ 2 \text{ se } x_i \leqq (\ell n2 | \ell n3) \cdot n_i \end{cases}$$

onde, $w_i = 1$ significa classificar o i-ésimo cruzamento como do tipo I e $w_i = 2$ como sendo do tipo II.

Diremos que um cruzamento gera uma grande descendência se n_i for não inferior ao valor mediano amostral, ñ (no nosso caso igual a 98) e, consequentemente, gerador de uma pequena descendência se n_i menor que ñ.

Dessa maneira definamos:

$$M_i = \begin{cases} 0 & \text{se} \quad n_i \leq 97 \\ 1 & \text{se} \quad n_i \geq 98 \end{cases}$$

A Tabela 1 do Anexo A mostra, ao lado de cada observação, os valores das classificações W_i e M_i.

Através de uma contagem simples, podemos resumir tais dados no quadro abaixo.

W \ M	0	1	Total
1	83	63	146
2	32	54	86
Total	115	117	232

Usaremos uma vez mais o teste de qui-quadrado para analisar o comportamento dessas frequências.

Considere as frequência esperadas supondo que haja independência entre o tamanho da descendência e o tipo de classificação do cruzamento,

W \ M	0	1	Total
1	72,37	73,63	146
2	42,63	43,37	86
Total	115	117	232

Assim,

$\chi_c^2 = 8,352$

Como $P(\chi^2_{(1)} \geq 8,352) < 0,005$, somos levados a rejeitar a hipótese de independência entre os tipos de classificação e os tamanhos das descendências.

Observando-se os quadros de frequências acima, pode-se notar que $(f_o - f_e)^2$ é constante em todas as celas. Assim as maiores contribuições para o χ_c^2 se encontram nas celas correspondentes às menores frequências esperadas (menores denominadores), sugerindo que a classificação no tipo II ocorre mais frequentemente para grandes descendências.

Levando-se em conta o teste acima, somos levados a acreditar em uma associação entre o tamanho da descendência e a "preferência sexual" da Drosophile Melanogaster. Desta forma nos parece que considerar apenas os 40 primeiros descendentes de cada cruzamento, acarreta uma perda de informação, que poderia estar contida na observação do verdadeiro tamanho da descendência.

Assim, uma vez que o estimador da seção 3.4 utiliza os valores reais dos tamanhos das descendências e que o mesmo é baseado fundamentalmente no Princípio da Verossimilhança, achamos que a estimativa por ele gerada é a que mais merece crédito.

Com relação às estimativas obtidas na seção 3.5, verificamos que $\hat{\theta}_S$ é a que mais se aproxima daquela que acreditamos ser a mais confiável. Assim, como já foi dito, a obtenção de estimativas baseadas apenas em observações fortemente conclusivas, parece nos fornecer algum ganho.

Como $\hat{\textcircled{H}}_S = \dfrac{\#(\mathcal{X}_2^S)}{\#(\mathcal{X}^S)}$ está intimamente ligada à partição usada, pretendemos no futuro, estudar a sensibilidade de tal estimador diante da escolha da partição. No momento porém, nos deteremos a analisar os estimadores $\hat{\textcircled{H}}_S$ e $\hat{\textcircled{H}}_W$ sob a luz desta particular partição.

Sejam:

$Y_1 = \#(\mathcal{X}_2^S)$; $Y_2 = \#(\mathcal{X}_2^W)$; $Y_3 = \#(\mathcal{X}_1^W)$ e

$Y_4 = \#(\mathcal{X}_1^S)$ e mais

$p_1 = P_1 \ (X \ \varepsilon \ \mathcal{X}_2^S) = P_1 \ (0 \leq \chi_i \leq 19) = 0{,}00018$

$p_2 = P_1 \ (X \ \varepsilon \ \mathcal{X}_2^W) = P_1 \ (20 \leq \chi_i \leq 25) = 0{,}05426$

$p_3 = P_1 \ (X \ \varepsilon \ \mathcal{X}_1^W) = P_1 \ (26 \leq \chi_i \leq 30) = 0{,}50602$

$p_4 = P_1 \ (X \ \varepsilon \ \mathcal{X}_1^S) = P_1 \ (31 \leq \chi_i \leq 40) = 0{,}43954$

$q_1 = P_2 \ (X \ \varepsilon \ \mathcal{X}_2^S) = P_2 \ (0 \leq \chi_i \leq 19) = 0{,}43731$

$q_2 = P_2 \ (X \ \varepsilon \ \mathcal{X}_2^W) = P_2 \ (20 \leq \chi_i \leq 25) = 0{,}52234$

$q_3 = P_2 \ (X \ \varepsilon \ \mathcal{X}_1^W) = P_2 \ (26 \leq \chi_i \leq 30) = 0{,}04001$

$q_4 = P_2 \ (X \ \varepsilon \ \mathcal{X}_1^S) = P_2 \ (31 \leq \chi_i \leq 40) = 0{,}00034$

Nessas condições, (Y_1, Y_2, Y_3, Y_4) tem distribuição multinomial.

$M \ (232; \ p_1 \ \theta + q_1 \ (1-\theta); \ p_2\theta + q_2 \ (1-\theta); \ p_3\theta + q_3(1-\theta);$

$p_4\theta + q_4(1-\theta))$

Assim,

$$Y_4 \mid Y_1 + Y_4 \sim B \left(Y_1 + Y_4 ; \frac{p_4\theta + q_4(1-\theta)}{(p_1 + p_4)\theta + (q_1 + q_4) \ (1-\theta)}\right)$$

Como $p_1 + p_4 = q_1 + q_4$ (partição ancilar)

$$Y_4 \mid Y_1 + Y_4 \sim B \left(Y_1 + Y_4 ; \frac{p_4\theta + q_4 \ (1 - \theta)}{p_1 + p_4}\right)$$

Se assumirmos que $P_1 = q_4 = 0$ (1), podemos dizer que

$$Y_4 \,|\, Y_1 + Y_4 \sim B\,(Y_1 + Y_4;\ \theta)$$

Logo

$$E\,[\hat{\Theta}_S] = E\,[\frac{Y_4}{Y_1+Y_4}] = E\,\{E[\frac{Y_4}{Y_1+Y_4} \,|\, Y_1 + Y_4]\} =$$

$$= E\,\{\frac{1}{Y_1+Y_4} \cdot E\,[Y_4 \,|\, Y_1+Y_4]\} = E\,\{\frac{1}{Y_1+Y_4} \cdot (Y_1+Y_4) \cdot \theta\} =$$

$= \theta$, e portanto $\hat{\Theta}_S$ é um estimador não viciado para θ.

De maneira análoga,

$$Y_3 \,|\, Y_2+Y_3 \sim B\,(Y_2+Y_3;\ \frac{p_3\,\theta + q_3\,(1-\theta)}{(p_2+p_3)\,\theta + (q_2+q_3)\,(1-\theta)})$$

como $p_2+p_3 = q_2+q_3$

$$Y_3 \,|\, Y_2+Y_3 \sim B\,(Y_2 + Y_3;\ \frac{p_3\theta + q_3\,(1-\theta)}{p_2+p_3})$$

Neste caso, porém, não é razoável a suposição que $p_2 = q_3 = 0$, assim.

$$E\,[\hat{\Theta}_W] = E\,[\frac{Y_3}{Y_2+Y_3}] = E\,\{E[\frac{Y_3}{Y_2+Y_3} \,|\, Y_2 + Y_3]\} =$$

. . . . . .

(1) Esta suposição está sendo feita para essa partição, particularmente, podendo não ser razoável para uma outra partição.

$$= E\left\{\frac{1}{Y_2+Y_3} \cdot E\left[Y_3 \middle| Y_2 + Y_3\right]\right\} =$$

$$= E\left\{\frac{1}{Y_2+Y_3} \cdot (Y_2+Y_3) \cdot \frac{p_3\theta + q_3(1-\theta)}{p_2+p_3}\right\} =$$

$$= \frac{(p_3 - q_3)\theta + q_3}{p_2 + p_3},$$

Logo,

$$\hat{\Theta}_W^* = \frac{\hat{\Theta}_W \cdot (p_2 + p_3) - q_3}{p_3 - q_3}$$ é um estimador não viciado para θ.

Consideremos, então, a estimativa através de $\hat{\Theta}_W^*$

$$\hat{\theta}_W^* = \frac{0{,}5873 \times (0{,}05426 + 0{,}50602) - 0{,}04001}{0{,}50602 - 0{,}04001} =$$

$$= \frac{0{,}5873 \times 0{,}56028 - 0{,}04001}{0{,}46601} \cong 0{,}6202$$

Observe que $\hat{\theta}_W^*$ é ainda mais próxima da estimativa obtida na seção 3.4 que $\hat{\theta}_S$, porém vale lembrar que para essa partição, $P(\mathcal{X}^W) > P(\mathcal{X}^S)$ e em consequência disto um maior número de cruzamentos foram classificados fracamente, o que possibilitou $\hat{\theta}_W^*$ ser baseada em uma amostra maior, podendo este fato contribuir para uma maior precisão.

Bibliografia

[1] BAHADUR, R.R. (1954). "Sufficiency and Statistical Decision Function", Ann. Math. Stat., 25, 423-62

[2] BAHADUR, R.R. and LEHMANN, E.L. (1955): Two Comments on "Sufficiency and Statistical Decision Function" Ann. Math. Stat., 26, 139 - 42.

[3] BASU, D. (1964). "Recovery of Ancillary Information", Sankhya, 21, 247 - 56.

[4] _____ (1975): "Statistical Information and Likelihood", Sankhya, A37, 1 - 71 (With Discussion).

[5] _____ (1981): "Ancillary Statistics, Pivotal Quantities and Confidence Statements", Presented at the Canadian Conference of Applied Statistics, Concordia University, Monteral, april 29/May 1.

[6] BERGER, J.O. (1980): "Statistical Decision Theory Foundations Concepts and Methods", Spring Verlag.

[7] BERNARDO, J.M. (1979): "Expected Information as Expected Utility", Ann. of Stat. V.7(3), 686-90.

[8] BIRNBAUM, A. (1961): "On the Foundations of Statistical Inference: Binary Experiments", Ann. Math. Stat., 32, 414-35.

[9] BIRNBAUM, A. (1962): "On the Foundations of Statistical Inference", J.A.S.A, V.57, 298, 269 - 326 (With Discussion).

[10] _____ (1972): "More on Concepts of Statistical Evidence", J.A.S.A, 67, 858 - 61.

[11] BLACKWELL, D. (1951): "Comparison of Experiments", Proc. Second Berkeley Symp., U. of Calif. Press, 93 - 102.

[12] BUEHLER, R.J. (1979): "Some Examples of Ancillary Statistics and Their Properties", U. of Minnesota (Technical Report, 364).

[13] COX, D.R. (1958): "Some Problems Connected with Statistical Inference", Ann. Math. Stat., 29, 357-72.

[14] De GROOT, M.H. (1962): "Uncertainty, Information and Sequential Experiments", Ann. Math Stat., 33, 404-19.

[15] EFRON, B.A. and HINCKLEY, D.V. (1978): "Assessing the Accuracy of the Maximun Likelihood Estimator: Observed versus Expected Fisher Information" Biometrika, 65(3), 457-87.

[16] FISHER, R.A. (1935): "The Logic of Inductive Inference". Jour. of Royal Stat. Soc.,Ser. B, 98, 39-54 (with discussion).

[17] FISHER, R.A. (1939): " The Comparison of Sample with Possibly Unequal Variances", Ann of Eugenics, 9, 174-80.

[18] _____. (1956): "Statistical Methods and Scientific Inference", Edinburgh, Oliver and Boyd.

[19] HACKING, I. (1965): "Logic of Statistical Inference", Camb. Univ. Press.

[20] KIEFER, J. (1975): "Conditional Confidence Approach in Multidecision Problems", Proc. of the Fourth Dayton Multivariate Conference, ed. P.B. Krishnaiah, Amsterdam: North Holland Publishing Co., 143-158.

[21] _____, (1976): "Admissibility of Conditional Confidence Procedures", Ann. of Stat., 4, 836-65.

[22] _____, (1977): "Conditional Confidence Statements and Confidence Estimators", J.A.S.A., 72(360), 789-827 (with Comments).

[23] LINDLEY, D.V. (1956): "On the Measure of the Information Provided by an Experiment", Ann. Math. Stat., 27, 986-1005.

[24] PEREIRA, C.A.B. (1971): "Estimativa da Probabilidade a Priori em um Problema de Classificação", V. 23(6), Separata da Revista Ciência e Cultura.

[25] SAVAGE, L.J. (1961): "The Foundation of Statistics Reconsidered", Proc. Fourth Berkeley Symp. V. of Calif. Press, 1, 575-86.

[26] TUKEY, J.W. (1960): "Conclusion vs. Decision", Technometrics, 2, 423-33.

[27] _____. (1962): "The Future of Data Analysis", Ann. Math. Stat., 33, 1 - 47.

ANEXO "A"

TABELA 1

I	N(I)	X(I)	W(I)	M(I)
1	141	77	2	1
2	75	57	1	0
3	121	62	2	1
4	93	76	1	0
5	65	42	1	0
6	75	52	1	0
7	98	54	2	1
8	88	49	2	0
9	74	57	1	0
10	67	36	2	0
11	64	53	1	0
12	116	62	2	1
13	94	44	2	0
14	88	63	1	0
15	107	58	2	1
16	74	55	1	0
17	97	49	2	0
18	137	107	1	1
19	76	45	2	0
20	97	55	2	0
21	111	51	2	1
22	81	64	1	0
23	118	81	1	1
24	79	52	1	0
25	43	28	1	0
26	58	50	1	0
27	122	92	1	1
28	58	32	2	0
29	127	92	1	1
30	122	65	2	1
31	146	62	2	1
32	135	61	2	1
33	87	76	1	0
34	105	74	1	1
35	95	60	1	0
36	98	56	2	1
37	75	48	1	0
38	86	69	1	0
39	109	55	2	1
40	133	60	2	1
41	108	57	2	1
42	106	54	2	1
43	126	63	2	1
44	107	62	2	1
45	105	85	1	1
46	102	80	1	1
47	90	70	1	0
48	79	60	1	0
49	127	76	2	1
50	91	48	2	0

CONTINUA...

TABELA 1

I	N(I)	X(I)	W(I)	M(I)
51	118	67	2	1
52	123	63	2	1
53	91	70	1	0
54	144	111	1	1
55	103	76	1	1
56	115	79	1	1
57	122	62	2	1
58	110	55	2	1
59	137	103	1	1
60	109	87	1	1
61	102	70	1	1
62	138	74	2	1
63	115	84	1	1
64	135	69	2	1
65	110	93	1	1
66	114	88	1	1
67	99	74	1	1
68	90	69	1	0
69	93	41	2	0
70	118	63	2	1
71	121	61	2	1
72	91	63	1	0
73	122	89	1	1
74	133	68	2	1
75	84	64	1	0
76	90	64	1	0
77	100	71	1	1
78	70	53	1	0
79	163	84	2	1
80	40	31	1	0
81	100	81	1	1
82	115	64	2	1
83	103	60	2	1
84	73	33	2	0
85	93	66	1	0
86	96	59	2	0
87	135	104	1	1
88	104	79	1	1
89	62	46	1	0
90	98	78	1	1
91	74	60	1	0
92	104	74	1	1
93	99	76	1	1
94	71	52	1	0
95	82	61	1	0
96	106	56	2	1
97	144	68	2	1
98	83	63	1	0
99	60	46	1	0
100	107	79	1	1

CONTINUA...

TABELA 1

I	N(I)	X(I)	W(I)	M(I)
101	74	50	1	0
102	103	81	1	1
103	62	44	1	0
104	123	66	2	1
105	89	56	2	0
106	75	51	1	0
107	54	44	1	0
108	72	38	2	0
109	101	51	2	1
110	105	41	2	1
111	78	60	1	0
112	108	67	2	1
113	84	47	2	0
114	112	63	2	1
115	84	43	2	0
116	116	93	1	1
117	41	33	1	0
118	109	77	1	1
119	92	63	1	0
120	64	39	2	0
121	87	64	1	0
122	110	88	1	1
123	113	84	1	1
124	112	48	2	1
125	84	34	2	0
126	64	34	2	0
127	57	33	2	0
128	91	67	1	0
129	73	57	1	0
130	82	62	1	0
131	60	50	1	0
132	80	60	1	0
133	112	79	1	1
134	101	57	2	1
135	104	54	2	1
136	98	76	1	1
137	111	85	1	1
138	110	57	2	1
139	108	86	1	1
140	110	74	1	1
141	103	70	1	1
142	110	52	2	1
143	119	90	1	1
144	118	88	1	1
145	72	41	2	0
146	107	74	1	1
147	111	84	1	1
148	80	61	1	0
149	126	91	1	1
150	49	33	1	0

CONTINUA...

TABELA 1

I	N(I)	X(I)	W(I)	M(I)
151	95	51	2	0
152	83	66	1	0
153	93	44	2	0
154	99	72	1	1
155	115	85	1	1
156	84	45	2	0
157	80	63	1	0
158	106	81	1	1
159	94	68	1	0
160	97	70	1	0
161	43	30	1	0
162	113	61	2	1
163	102	68	1	1
164	118	92	1	1
165	124	88	1	1
166	108	77	1	1
167	77	58	1	0
168	95	53	2	0
169	129	69	2	1
170	115	82	1	1
171	46	27	2	0
172	92	78	1	0
173	124	100	1	1
174	77	41	2	0
175	79	41	2	0
176	50	34	1	0
177	69	35	2	0
178	103	52	2	1
179	113	76	1	1
180	92	64	1	0
181	81	58	1	0
182	107	63	2	1
183	96	69	1	0
184	106	79	1	1
185	96	73	1	0
186	96	46	2	0
187	117	91	1	1
188	63	52	1	0
189	99	76	1	1
190	69	46	1	0
191	87	67	1	0
192	114	57	2	1
193	97	76	1	0
194	93	73	1	0
195	105	76	1	1
196	86	42	2	0
197	77	61	1	0
198	116	47	2	1
199	96	72	1	0
200	84	64	1	0

CONTINUA...

TABELA 1

I	N(I)	X(I)	W(I)	M(I)
201	135	61	2	1
202	106	88	1	1
203	90	67	1	0
204	65	46	1	0
205	98	40	2	1
206	94	70	1	0
207	90	63	1	0
208	120	61	2	1
209	48	32	1	0
210	69	51	1	0
211	78	65	1	0
212	75	35	2	0
213	94	49	2	0
214	90	68	1	0
215	122	91	1	1
216	113	69	2	1
217	108	86	1	1
218	77	52	1	0
219	75	61	1	0
220	129	80	2	1
221	83	59	1	0
222	96	72	1	0
223	100	48	2	1
224	70	48	1	0
225	118	61	2	1
226	142	75	2	1
227	121	90	1	1
228	74	50	1	0
229	81	60	1	0
230	118	66	2	1
231	90	41	2	0
232	119	100	1	1

TABELA 2

I	X(I)	V(I)	T(I)
1	17	2	(S)
2	30	1	(W)
3	21	2	(W)
4	36	1	(S)
5	29	1	(W)
6	32	1	(S)
7	23	2	(W)
8	25	2	(W)
9	34	1	(S)
10	19	2	(S)
11	38	1	(S)
12	21	2	(W)
13	15	2	(S)
14	31	1	(S)
15	23	2	(W)
16	26	1	(W)
17	20	2	(W)
18	39	1	(S)
19	24	2	(W)
20	21	2	(W)
21	13	2	(S)
22	27	1	(W)
23	34	1	(S)
24	28	1	(W)
25	27	1	(W)
26	35	1	(S)
27	30	1	(W)
28	25	2	(W)
29	27	1	(W)
30	20	2	(W)
31	10	2	(S)
32	17	2	(S)
33	38	1	(S)
34	27	1	(W)
35	24	2	(W)
36	21	2	(W)
37	21	2	(W)
38	22	2	(W)
39	20	2	(W)
40	11	2	(S)
41	18	2	(S)
42	23	2	(W)
43	20	2	(W)
44	25	2	(W)
45	37	1	(S)
46	39	1	(S)
47	28	1	(W)
48	33	1	(S)
49	23	2	(W)
50	15	2	(S)

CONTINUA...

TABELA 2

I	X(I)	V(I)	T(I)
51	23	2	(W)
52	21	2	(W)
53	30	1	(W)
54	34	1	(S)
55	28	1	(W)
56	30	1	(W)
57	21	2	(W)
58	18	2	(S)
59	29	1	(W)
60	34	1	(S)
61	24	2	(W)
62	19	2	(S)
63	32	1	(S)
64	17	2	(S)
65	33	1	(S)
66	32	1	(S)
67	32	1	(S)
68	29	1	(W)
69	16	2	(S)
70	20	2	(W)
71	20	2	(W)
72	26	1	(W)
73	32	1	(S)
74	17	2	(S)
75	26	1	(W)
76	27	1	(W)
77	26	1	(W)
78	26	1	(W)
79	20	2	(W)
80	31	1	(S)
81	30	1	(W)
82	21	2	(W)
83	16	2	(S)
84	19	2	(S)
85	28	1	(W)
86	23	2	(W)
87	31	1	(S)
88	30	1	(W)
89	27	1	(W)
90	36	1	(S)
91	27	1	(W)
92	26	1	(W)
93	35	1	(S)
94	27	1	(W)
95	29	1	(W)
96	18	2	(S)
97	13	2	(S)
98	29	1	(W)
99	31	1	(S)
100	30	1	(W)

CONTINUA...

TABELA 2

I	X(I)	V(I)	T(I)
101	26	1	(W)
102	31	1	(S)
103	34	1	(S)
104	22	2	(W)
105	26	1	(W)
106	27	1	(W)
107	38	1	(S)
108	25	2	(W)
109	17	2	(S)
110	14	2	(S)
111	26	1	(W)
112	23	2	(W)
113	26	1	(W)
114	21	2	(W)
115	20	2	(W)
116	34	1	(S)
117	33	1	(S)
118	29	1	(W)
119	27	1	(W)
120	25	2	(W)
121	22	2	(W)
122	34	1	(S)
123	34	1	(S)
124	15	2	(S)
125	12	2	(S)
126	17	2	(S)
127	13	2	(S)
128	26	1	(W)
129	30	1	(W)
130	34	1	(S)
131	37	1	(S)
132	39	1	(S)
133	30	1	(W)
134	18	2	(S)
135	18	2	(S)
136	27	1	(W)
137	34	1	(S)
138	18	2	(S)
139	34	1	(S)
140	26	1	(W)
141	24	2	(W)
142	18	2	(S)
143	35	1	(S)
144	34	1	(S)
145	26	1	(W)
146	26	1	(W)
147	34	1	(S)
148	31	1	(S)
149	35	1	(S)
150	27	1	(W)

CONTINUA...

TABELA 2

I	X(I)	V(I)	T(I)
151	20	2	(W)
152	26	1	(W)
153	18	2	(S)
154	32	1	(S)
155	34	1	(S)
156	18	2	(S)
157	27	1	(W)
158	31	1	(S)
159	28	1	(W)
160	30	1	(W)
161	26	1	(W)
162	20	2	(W)
163	29	1	(W)
164	34	1	(S)
165	36	1	(S)
166	29	1	(W)
167	29	1	(W)
168	26	1	(W)
169	23	2	(W)
170	32	1	(S)
171	20	2	(W)
172	36	1	(S)
173	40	1	(S)
174	14	2	(S)
175	23	2	(W)
176	30	1	(W)
177	20	2	(W)
178	13	2	(S)
179	25	2	(W)
180	30	1	(W)
181	31	1	(S)
182	27	1	(W)
183	30	1	(W)
184	33	1	(S)
185	29	1	(W)
186	11	2	(S)
187	39	1	(S)
188	31	1	(S)
189	28	1	(W)
190	27	1	(W)
191	29	1	(W)
192	13	2	(S)
193	34	1	(S)
194	37	1	(S)
195	26	1	(W)
196	25	2	(W)
197	38	1	(S)
198	15	2	(S)
199	26	1	(W)
200	27	1	(W)

CONTINUA...

TABELA 2

I	X(I)	V(I)	T(I)
201	19	2	(S)
202	37	1	(S)
203	26	1	(W)
204	32	1	(S)
205	12	2	(S)
206	28	1	(W)
207	25	2	(W)
208	19	2	(S)
209	26	1	(W)
210	26	1	(W)
211	29	1	(W)
212	23	2	(W)
213	21	2	(W)
214	28	1	(W)
215	33	1	(S)
216	20	2	(W)
217	34	1	(S)
218	29	1	(W)
219	39	1	(S)
220	20	2	(W)
221	26	1	(W)
222	39	1	(S)
223	10	2	(S)
224	28	1	(W)
225	20	2	(W)
226	25	2	(W)
227	39	1	(S)
228	36	1	(S)
229	33	1	(S)
230	14	2	(S)
231	19	2	(S)
232	24	2	(W)

TABELA 3

TETA	L(TETA)	TETA	L(TETA)
0.00	.28532D+06768	0.50	.47034D+07435
0.01	.49017D+07234	0.51	.12989D+07436
0.02	.43739D+07272	0.52	.32848D+07436
0.03	.14073D+07295	0.53	.76120D+07436
0.04	.14372D+07311	0.54	.16168D+07437
0.05	.37925D+07323	0.55	.31481D+07437
0.06	.51373D+07333	0.56	.51184D+07437
0.07	.17827D+07342	0.57	.91886D+07437
0.08	.41622D+07349	0.58	.13763D+07438
0.09	.12228D+07356	0.59	.18868D+07438
0.10	.69454D+07361	0.60	.23649D+07438
0.11	.10370D+07367	0.61	.27070D+07438
0.12	.51107D+07371	0.62	.28255D+07438
0.13	.98874D+07375	0.63	.26848D+07438
0.14	.85949D+07379	0.64	.23175D+07438
0.15	.37370D+07383	0.65	.18132D+07438
0.16	.88625D+07386	0.66	.12823D+07438
0.17	.12306D+07390	0.67	.81721D+07437
0.18	.10612D+07393	0.68	.46771D+07437
0.19	.59692D+07395	0.69	.23945D+07437
0.20	.22838D+07398	0.70	.10918D+07437
0.21	.61582D+07400	0.71	.44111D+07436
0.22	.12067D+07403	0.72	.15704D+07436
0.23	.17645D+07405	0.73	.48952D+07435
0.24	.19704D+07407	0.74	.13265D+07435
0.25	.17146D+07409	0.75	.30993D+07434
0.26	.11835D+07411	0.76	.61864D+07433
0.27	.65833D+07412	0.77	.10439D+07433
0.28	.29926D+07414	0.78	.14714D+07432
0.29	.11256D+07416	0.79	.17087D+07431
0.30	.35426D+07417	0.80	.16089D+07430
0.31	.94224D+07418	0.81	.12060D+07429
0.32	.21371D+07420	0.82	.70437D+07427
0.33	.41670D+07421	0.83	.31254D+07426
0.34	.70358D+07422	0.84	.10226D+07425
0.35	.10356D+07424	0.85	.23804D+07423
0.36	.13366D+07425	0.86	.37757D+07421
0.37	.15210D+07426	0.87	.38710D+07419
0.38	.15335D+07427	0.88	.24025D+07417
0.39	.13759D+07428	0.89	.83115D+07414
0.40	.11031D+07429	0.90	.14414D+07412
0.41	.79308D+07429	0.91	.10901D+07409
0.42	.51301D+07430	0.92	.29771D+07405
0.43	.29945D+07431	0.93	.22562D+07401
0.44	.15814D+07432	0.94	.32297D+07396
0.45	.75743D+07432	0.95	.48225D+07390
0.46	.32969D+07433	0.96	.27921D+07383
0.47	.13066D+07434	0.97	.99331D+07373
0.48	.47224D+07434	0.98	.36162D+07360
0.49	.15587D+07435	0.99	.35433D+07337

TABELA 4

AREA SOB L^* (TETA) = 0.823331208E-01

LI	LS	AREA PARCIAL	PROP. DE AREA
0.62	0.63	0.99491570E-02	0.12084030E-00
0.61	0.63	0.19549720E-01	0.23744870E-00
0.61	0.64	0.28951700E-01	0.35164110E-00
0.60	0.64	0.37414080E-01	0.45423200E-00
0.60	0.65	0.45500070E-01	0.55263390E-00
0.59	0.65	0.52323480E-01	0.63550950E-00
0.59	0.66	0.58637900E-01	0.71220300E-00
0.58	0.66	0.63676910E-01	0.77340580E-00
0.58	0.67	0.68142580E-01	0.82764480E-00
0.57	0.67	0.71720570E-01	0.87110230E-00
0.57	0.68	0.74572220E-01	0.90573790E-00
0.56	0.68	0.76857720E-01	0.93349710E-00
0.56	0.69	0.78496430E-01	0.95340040E-00

INTERVALO DE CREDIBILIDADE ----> (0.56 , 0.69)

ANEXO "B"

PROGRAMA 1

```
      INTEGER X(232),N(232)
      REAL Y(101),TETA,XX(101)
      REAL*8 DL(101),S,DMAX
C
C     LEITURA DOS DADOS
C
      DO 1 I=1,232
    1 READ(5,10) N(I),X(I)
   10 FORMAT(2I5)
C
C     INCREMENTA VALOR DO PARAMETRO TETA
C
      DO 2 J=1,101
      S=0
      TETA=(K-1)/100.
      XX(J)=TETA
C
C     CALCULA A FUNÇAO DE VEROSSIMILHANÇA NO PONTO TETA
C
      DO 3 I=1,232
    3 S=S+DLOG(DELE(3.**X(I)*TETA+2.**N(I)*(1-TETA)))
      DL(J)=DEXP(S)
C
C     IMPRESSAO DO PAR TETA, VALOR DA FUNÇAO
C
    2 WRITE(6,20) TETA,DL(J)
   20 FORMAT(1X,F10.2,D20.5)
C
C     CALCULA O MAXIMO VALOR DA FUNÇAO DE VEROSSIMILHANÇA
C
      DMAX=DL(1)
      DO 5 I=2,101
      IF(DMAX.LT.DL(I)) DMAX=DL(I)
    5 CONTINUE
C
C     PADRONIZA A FUNÇAO PELO MAXIMO
C     CONVERTE OS VALORES DA FUNÇAO PARA PRECISAO SIMPLES
C
      DO 7 I=1,101
      DL(I)=DL(I)/DMAX
      IF(DL(I).LT.0.1D-36) GO TO 6
      Y(I)=SNGL(DL(I))
      GO TO 8
    6 Y(I)=0.1E-36
    8 WRITE(1,30) XX(I),Y(I)
   30 FORMAT(1X,F10.2,F20.5)
    7 CONTINUE
C
C     TRACA O GRAFICO DA FUNÇAO DE VEROSSIMILHANÇA PADRONIZADA
C
      CALL PLOT(101,XX,Y)
      STOP
      END
```

PROGRAMA 2

```
      DIMENSION X(101),Y(101)
C
C     LEITURA DOS DADOS
C
      READ(1,10) (X(I),Y(I),I=1,100)
   10 FORMAT(1X,F10.2,F20.5)
C
C     CALCULO DA AREA SOB A FUNÇAO DE VEROSSIMILHANÇA
C
      AREA=(Y(1)+Y(100))/2.
      DO 1 I=2,99
      AREA=AREA+Y(I)
    1 CONTINUE
      AREA=AREA/100.
      WRITE(6,20) AREA
   20 FORMAT(1H1,/,32X,'TABELA 4',/,32X,'--------',////,25X,'*',/,15X,
     *'AREA SOB L (TETA)    =    ',E15.9,////,14X,51('-'),/,16X,'LI',7X,
     *'LS',7X,'AREA PARCIAL',6X,'PROP. DE AREA',/,14X,51('-'))
C
C     CALCULO DAS AREAS PARCIAIS E INTERVALO DE CREDIBILIDADE
C
      LI=63
      LS=64
      LL=LI
      ALFA=0.
    2 ALFA=ALFA+(Y(LL)+Y(LL+1))/200.
      PROP=ALFA/AREA
      WRITE(6,30) X(LI),X(LS),ALFA,PROP
      IF(PROP.GT.0.95) GO TO 4
      IF(Y(LI-1).GT.Y(LS+1)) GO TO 3
      LL=LS
      LS=LS+1
      GO TO 2
    3 LI=LI-1
      LL=LI
      GO TO 2
    4 WRITE(6,30) X(LI),X(LS)
   30 FORMAT(14X,51('-'),///,14X,'INTERVALO DE CREDIBILIDADE    ---->
     * )',F4.2,' , ',F4.2,')')
      STOP
      END
```