

TESTES EXATOS PARA TABELAS 2X2 :

BAYES X FISHER

TELBA ZALKIND IRONY

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO : ESTATÍSTICA

ORIENTADOR:

Prof. Dr. Carlos Alberto de Bragança Pereira

- SÃO PAULO, AGOSTO DE 1984 -

Ao Ilan

pelo amor e

incentivo que

sempre tive;

com amor

AGRADEÇO

Ao Carlinhos pela amizade, pelo entusiasmo e pela orientação brilhante sem a qual este trabalho não teria sido possível.

À Itaudata Ltda. e em especial a Luciano Bacchi cujo discernimento permitiu que eu seguisse o programa de mestrado. Que o crédito dado à Universidade pela liderança de uma empresa bem sucedida sirva de exemplo para governos de instituições não tão bem sucedidas assim.

À Rosa Maria Vega Perez que com simpatia e perseverança elaborou os gráficos comparativos deste trabalho no PDP11/60.

Ao Prof. Dr. Carlos Alberto Barbosa Dantas pela compreensão e apoio que me foram dados quando iniciei o programa de mestrado.

Ao Galves, cujos conselhos, embora nem sempre seguidos, foram sempre levados em consideração.

À Rosa Nilde Aparecida Rubio pela paciência com a qual digitou os originais.

ÍNDICE

I.	INTRODUÇÃO	1
II.	O TESTE EXATO DE FISHER	
II.1.	Descrição	8
II.2.	A Lógica do Teste Exato de Fisher	18
II.2.1.	A Distribuição de Probabilidades Utilizada no Teste	19
II.2.2.	O Critério de Decisão Adotado por Fisher	33
III.	DISCUSSÃO DA LÓGICA UTILIZADA NO TESTE EXATO DE FISHER	
III.1.	A Distribuição de Probabilidades do Teste	39
III.2.	O Critério de Decisão adotado no Teste	47
III.2.1.	Nível de Significância, Tamanho do Teste e Nível Descritivo	47
III.2.2.	O Uso do Nível Descritivo: $P_t(x N, m, t)$	54
III.2.3.	A Pré-especificação do Nível de Significância	56
III.3.	O Desprezo pela Informação a Priori	60
IV.	UMA ALTERNATIVA BAYESIANA PARA O TESTE EXATO DE FISHER	
IV.1.	Introdução	61
IV.2.	O Teste Bayesiano	63
IV.2.1.	A Técnica Bayesiana de Eliminação de Parâmetros Nuisance	67
IV.2.2.	Uma Adaptação da Técnica Bayesiana de Eliminação de Parâmetros Nuisance aos Testes de Hipóteses	71

IV.3. Exemplo: A Tabela 2x2	77
IV.3.1. Cálculo da Função Preditiva sob H_1	81
IV.3.2. Cálculo da Função Preditiva sob H_0	83
IV.3.3. Comparação da Distribuição de Probabilidades Adotada no Teste Exato de Fisher com a Função Preditiva sob H_0	84
IV.3.4. O Escore a Favor de H_0 a Posteriori $\epsilon(x,y)$	85
IV.4. O Critério de Decisão Adotado no Teste Bayesiano	86
IV.4.1. Erros de 1a. e 2a. Espécies	86
IV.4.2. O Teste Bayesiano minimiza a Combinação Linear a $\alpha(\delta_B) + b \beta(\delta_B)$	88
IV.4.3. A Escolha da Constante c	89

V. CONFRONTO: TESTE EXATO DE FISHER X TESTE BAYESIANO

V.1. Introdução	92
V.2. Os Procedimentos de Teste	96
V.2.1. A Tábua de Decisão Bayesiana	97
V.2.2. A Tábua de Decisão de Fisher	99
V.3. A Simulação	103
V.3.1. Exemplo para o caso em que $m=n=5$	103
V.3.2. Exemplo para o caso em que $m=8$ e $n=14$	109
V.3.3. Exemplo para o caso $m=5$ e $n=3$ utilizando para o Teste Exato de Fisher o valor ALFABT	113
V.3.4. A Simulação para Tabelas de Contingência em que $m=n$	116
V.4. Conclusões	119

APÊNDICE A	122
APÊNDICE B	125
APÊNDICE C	127
APÊNDICE D	129
REFERÊNCIAS BIBLIOGRÁFICAS	131

CAPÍTULO I

INTRODUÇÃO

Uma pessoa que inicia estudos no campo da estatística não tarda a defrontar-se com dois modos de pensar que, a princípio, confundem-se em determinados pontos e divergem em outros: a Perspectiva Clássica e a Perspectiva Bayesiana.

Surge então, uma série de dúvidas. Qual das metodologias deve ser utilizada? Existem diferenças nos resultados obtidos? Em quais circunstâncias as duas filosofias conduzem a inferências semelhantes? Em algumas situações as diferenças metodológicas não produzem resultados diferentes do ponto de vista técnico, como era de se esperar. Este fato traz novas dúvidas ao iniciante.

Caso os resultados analíticos sejam distintos, qual deles é o melhor? Será que o fato de uma metodologia revelar-se superior em uma determinada situação indica que ela seja superior em todas as situações? Em outras palavras, existe uma linha filosófica mais convincente?

No presente trabalho procurou-se fugir das discussões intermináveis, embora por vezes fascinantes, sobre a coerência das filosofias Clássica e Bayesiana. Em lugar disso, escolheu-se um problema elementar onde analisa-se o funcionamento da estatística Bayesiana em confronto com a Clássica.

Aplicou-se o método Bayesiano, voltado aos testes de hipóteses, na análise de tabelas de contingência 2×2 . O tratamento Clássico deste problema, o Teste Exato de Fisher, tem sido largamente discutido e refutado na literatura estatística, e até o momento não foi apresentada uma solução alternativa de perspectiva Clássica que o supere (1). Assim, o ponto básico deste trabalho é comparar o Teste Exato de Fisher com o alternativo Bayesiano proposto.

Quando se estuda um fenômeno sob o ponto de vista estatístico, procura-se estabelecer um modelo e uma assertiva que o caracterize. Os objetivos do estatístico que faz a análise são:

1^a) Propor uma assertiva que tenha grande chance de ser verdadeira.

(1) YATES, F. Tests of significance for 2×2 contingency tables. J.R. Statist. Soc. A., 147 (3), 1984. (To Appear)

2^a) Acertar a chance de que a assertiva seja verdadeira.

As discussões sobre as linhas filosóficas que devem ser adotadas interessam aos estatísticos. Aos usuários das técnicas estatísticas interessa que sejam atingidos os objetivos acima, sobretudo o segundo. Assim, para estes, um bom estatístico é aquele que dá 30% de probabilidade para a veracidade de uma assertiva antes de que ela se mostre verdadeira em 30% das vezes. Em outras palavras, um bom estatístico não é somente aquele que acerta mais, mas também aquele que, além de acertar mais, acerta quantas vezes vai errar. Esta é uma qualidade objetiva que não depende de discussões filosóficas (2).

Considerando os pontos de vista acima levantados, concluímos neste trabalho, que a solução Bayesiana mostra-se objetivamente superior à Clássica para o caso específico das tabelas de contingência 2x2.

Ressalte-se, entretanto, que não existe a intenção de inferir a partir disso, que a metodologia Bayesiana supera a Clássica de maneira geral. Isto seria uma indução errônea, pois estaríamos concluindo que uma premissa é sempre verdadeira a partir de sua veracidade para um caso particular.

O capítulo II concentra-se na descrição e fundamentação lógica da técnica de análise Clássica para tabelas de contin-

(2) DAWID, A.P. The well-calibrated bayesian. JASA, 77 (379):605-613, 1982

gência 2×2 , o Teste Exato de Fisher. O teste é introduzido como um exemplo em um artigo de Fisher (3), que versava sobre o conceito de Quantidade de Informação, assunto que historicamente lhe deu origem. As características básicas do teste Exato são:

- 1ª) A distribuição de probabilidades condicional utilizada.
- 2ª) O critério de decisão adotado.

Inicialmente estas características foram aceitas pela comunidade estatística devido à falta de uma solução alternativa a este tipo de problema. Mais tarde, o teste Exato foi fundamentado sobre uma lógica gerada por uma teoria muito particular: a Inferência Parcial (4). A distribuição de probabilidades condicional utilizada no teste surge do uso do Método da Redução, explicado através da Inferência Parcial. O critério de decisão adotado decorre da determinação da probabilidade "exata" (condicionada nas marginais) de obtenção de cada uma das possíveis tabelas de frequências, valendo-se da pré-especificação do Nível de Significância α . Ainda no capítulo II mostramos que se pode aplicar o mesmo teste tanto para a tabela originada a partir da distribuição Multinomial como para a tabela originada a partir de duas distribuições

(3) FISHER, R.A. The logic of inductive inference. J.R. Statist. Soc. A, 98(1):39-54, 1935.

(4) MARIOTTO, A.B. Inferência Parcial. São Paulo, 1983
107p. Dissertação (Mestrado) - IME-USP

Binomiais.

No capítulo III questionamos tanto o uso da distribuição de probabilidades condicional quanto o critério de decisão adotado por Fisher em seu teste Exato. A distribuição de probabilidades envolve uma restrição do espaço amostral obtida através do Método da Redução por condicionamento. No caso do teste em questão, este método deve ser aplicado aliado ao conceito de G-ancilaridade que leva em conta a Informação de Fisher. Isto gera uma violação do Princípio da Verossimilhança além de uma certa perda de informação. A validade do Método da Redução é discutida por não ser uma técnica bem definida uma vez que não temos uma regra geral para o encontro da parametrização adequada ao método.

Além disso, este método conduz ao chamado "Paradoxo de Borel" (5),(6). As definições de Nível de Significância, Tamanho do Teste e Nível Descritivo são apresentadas na sequência do capítulo III com o objetivo de facilitar a leitura do texto. Procura-se uma medida que reflita a segurança com a qual uma decisão é tomada. A validade da medida $P_t(x|N,m,t)$, que leva em consideração todas as amostras que poderiam ter

(5) PEREIRA, C.A.B. Stopping rules and conditional inference in 2x2 contingency tables. São Paulo, IME-USP, 1983 7p. (Relatório Técnico, RT-MAE-8301)

(6) PEREIRA, C.A.B. & LINDLEY, D.V. Examples questioning the use of partial likelihood. São Paulo, IME-USP, 1983 9p. (Relatório Técnico, RT-MAE-8306).

sido selecionadas mas não o foram é discutida. A pré-especificação do Nível de Significância α , independentemente do tamanho da amostra, é criticada. Ressaltamos o fato de o procedimento de Fisher menosprezar eventuais informações prece-
dentes que o pesquisador tenha a respeito do experimento.

No capítulo IV apresentamos uma alternativa para o Teste Exato de Fisher. Tal alternativa faz uso das técnicas Bayesianas para testes de hipóteses assumindo conhecimentos pre-
cedentes tanto para as chances das hipóteses em teste quanto para atribuir distribuições de probabilidades aos parâmetros em exame. O critério de decisão é baseado no Fator de Bayes para hipóteses discretas.

As distribuições de probabilidades em jogo são combina-
ções da verossimilhança do modelo com as distribuições de probabilidades conferidas a priori aos parâmetros. O critério de decisão para aceitar ou rejeitar a hipótese nula é baseado na Razão de Probabilidades Complementares ou Escore a Favor do H_0 . O Escore a Priori, $E = \xi / 1 - \xi$, é a razão entre os pesos ξ e $(1 - \xi)$ dados de início à H_0 e H_1 respectivamente. O Escore a Posteriori, $E(x,y) = E \cdot f_0(x,y)/f_1(x,y)$ corresponde ao produto do Escore a Priori pela razão das Funções Preditivas sob H_0 e H_1 . Para o cálculo das Funções Preditivas, $f_0(x,y)$ condicionada à hipótese H_0 e $f_1(x,y)$ condicionada à hipótese H_1 , fizemos uso da Integral de Linha, onde os parâmetros em jogo, p e q , pertencem aos espaços paramétricos determinados pelas hipóteses condicionantes. Quando $E(x,y) < c$, onde c é uma constante determinada teoricamente, a decisão é

contrária a H_0 . Se $C(x,y) \geq c$, a decisão favorece H_0 .

Finalmente, no capítulo V comparamos o Teste Exato de Fisher e o Teste Bayesiano através de amostras simuladas. Para que os testes sejam confrontados, eles devem ser submetidos a condições semelhantes, as quais são discutidas na sequência do capítulo. Para alguns tamanhos de amostras, são construídos exemplos de Tábuas de Decisão para ambos os testes. Com base nas amostras simuladas, provenientes de populações cujos parâmetros são conhecidos, são calculadas as probabilidades dos erros de 1a. e 2a. ordem, ALFA e BETA, para os dois testes. Para todos os tamanhos de amostras geradas, o Teste Bayesiano apresentou um poder maior do que o poder do seu concorrente. Além disso, a soma dos erros de 1a. e 2a. espécies é significativamente menor quando utilizamos o Teste Bayesiano.

Enfim, um leitor que esteja interessado somente no sentido prático do trabalho, pode voltar-se apenas para a leitura do capítulo V, no qual constatará de forma objetiva que o Teste Bayesiano apresenta qualidade superior ao Teste Exato de Fisher se forem levados em conta os critérios tradicionalmente aceitos pela comunidade estatística: "o melhor teste é aquele que, além de errar menos, acerta quantas vezes vai errar".

CAPÍTULO II

O TESTE EXATO DE FISHER

II.1 - Descrição

O Teste Exato de Fisher foi apresentado pela primeira vez como um método de análise para tabelas de contingência 2x2 no artigo The Logic of Inductive Inference (1), em 1935. O artigo versava sobre um conceito muito importante e polêmico: Quantidade de Informação. Note que a finalidade da Inferência Estatística é justamente extrair toda a Informação a respeito de um parâmetro θ contida nos dados x .

(1) FISHER, R.A. The logic of inductive inference. J.R. Statist. Soc. A, 98(1):39-54, 1935

Ao introduzir o conceito de informação, Fisher tentou justificá-lo baseando-se em argumentos intuitivos. Para isto estabeleceu quatro postulados:

- 1ª) Informação Nula: Quando as probabilidades de observar-se os dados x independem do parâmetro θ , os dados x não fornecem informação a respeito de θ .
- 2ª) Informação Total: Existem determinadas estatísticas tais que, quando seus valores são fornecidos, as distribuições de todas as outras estatísticas passam a não depender do valor do parâmetro de interesse θ . Estatísticas com esta propriedade extraem toda a informação a respeito de θ contida nos dados e são chamadas Suficientes.
- 3ª) A informação a respeito de θ extraída dos dados através de uma estatística nunca pode superar a informação presente nos dados.
- 4ª) Observações estatisticamente independentes fornecem quantidades de informação que são aditivas.

Definindo quantidade de informação, Fisher procurava obter um critério para julgar as qualidades de um estimador. Faz parte do "senso comum" pensar que o melhor estimador deve usar toda a informação a respeito do valor do parâmetro contida na amostra e, mais do que isso, somente esta informação. Analisando-se o problema com cuidado, surge o desejo de eliminar toda informação "inútil" com respeito à estimação do parâmetro de interesse a qual, na melhor das hipóteses, será inócua.

A esta informação adicional, Fisher associou o conceito

de Estatística Ancilar e afirmou que tal estatística nada informa a respeito do valor do parâmetro de interesse. Contudo admitiu que esta estatística fornece informação a respeito da "precisão" da estimativa do parâmetro θ .

Para ilustrar o uso da estatística ancilar, Fisher considerou um estudo sobre gêmeos de condenados, desejando testar se a probabilidade de condenação de monozigóticos é a mesma de dizigóticos.

A análise foi feita a partir da seguinte tabela de frequências:

	Condenados	Absolvidos	Total
Monozigóticos	10	3	13
Dizigóticos	2	15	17
Total	12	18	30

Tabela I

Chamando de p a probabilidade de condenação de gêmeos monozigóticos e de q a probabilidade de condenação de gêmeos dizigóticos, surge o interesse de testar as seguintes hipóteses:

$$H_0: p = q$$

$$H_1: p > q$$

O argumento utilizado por Fisher (2) foi o seguinte:

(2) FISHER, R.A. The logic of inductive inference.

J.R.

Statist. Soc. A, 98(1):48-49, 1935

" Se admitirmos que as freqüências marginais não forne-
cem informação sobre a proporcionalidade das freqüên-
cias no corpo da tabela, poderemos tratá-las como uma
estatística ancilar. Desta forma, poderemos nos limitar
ao estudo das formas distintas através das quais a ta-
bela pode ser preenchida, dadas estas freqüências mar-
ginais. Estas formas estarão completamente especifica-
das fixando-se uma casela, a de condenados dizigóticos
por exemplo, e preenchendo-a com seus 13 valores possí-
veis (de 0 à 12)". (o grifo é nosso.)

Sejam os gêmeos monozigóticos, que têm a probabilidade p
de serem condenados. Como existem 13 gêmeos monozigóticos,
dos quais, no máximo 12 podem ser condenados, a probabilidade
de ocorrência de $(12-x)$ condenados e $(x+1)$ absolvidos é dada
por:

$$P_m(12-x) = \binom{13}{12-x} p^{(12-x)} (1-p)^{(x+1)} \quad x = 0, 1, 2, \dots, 12$$

Considerando H_0 verdadeira, ou seja, $p=q$, a probabilidade
de que, dos 17 gêmeos dizigóticos, x sejam condenados e
 $(17-x)$ sejam absolvidos é dada por:

$$P_d(x) = \binom{17}{x} p^x (1-p)^{(17-x)} \quad x = 0, 1, 2, \dots, 12$$

Assim, a probabilidade da ocorrência simultânea dos dois
eventos pode ser escrita como:

$$P_{md}(x) = \binom{13}{12-x} \binom{17}{x} p^{12} (1-p)^{18} \quad x = 0, 1, 2, \dots, 12$$

Note que as potências de p e de $(1-p)$ independem de x e,

portanto, representam um fator constante para as 13 possibilidades. Desta forma, a probabilidade de ocorrer algum valor de x é proporcional a:

$$\binom{13}{12-x} \binom{17}{x}$$

Para determinar-se a constante de proporcionalidade basta recorrer à normalização:

$$\sum_{x=0}^{12} k \cdot \binom{13}{12-x} \binom{17}{x} = 1$$

A expressão acima é facilmente reconhecível como a soma das probabilidades de uma distribuição Hipergeométrica onde a constante k deve valer

$$\frac{1}{\binom{30}{12}}$$

Portanto, a probabilidade de que se observe a tabela determinada pelo valor de x , dado que H_0 é verdadeira pode ser escrita como:

$$Pf(12-x) = \frac{\binom{13}{12-x} \binom{17}{x}}{\binom{30}{12}} \quad x = 0, 1, 2, \dots, 12$$

Com esta argumentação, Fisher conseguiu determinar exatamente as probabilidades de cada uma das possíveis configurações da tabela, sujeita às frequências marginais obtidas. Nenhum processo de estimação ou de convergência foi envolvido. Daí a denominação "Teste Exato de Fisher".

Dadas as probabilidades acima, intui-se o procedimento de

teste. Como H_1 indica um teste unicaudal, as configurações que favoreceriam H_1 com intensidade maior ou igual à da tabela apresentada seriam aquelas para as quais $x = 0, 1, 2$. Para obter-se o nível descritivo do conjunto de dados é suficiente calcular quão prováveis seriam estas configurações se H_0 fosse verdadeira. Tem-se, então:

$$P_t(10) = \sum_{x=0}^2 P_f(12-x) = \frac{\binom{13}{12} \binom{17}{0}}{\binom{30}{12}} + \frac{\binom{13}{11} \binom{17}{1}}{\binom{30}{12}} + \frac{\binom{13}{10} \binom{17}{2}}{\binom{30}{12}} = 4,65 \times 10^{-4}$$

A probabilidade de obtenção de uma entre as três configurações citadas é $4,65 \times 10^{-4}$, ou seja, somente uma de cada 2150 tabelas apresentará alguma dentre as três configurações extremas se $p=q$. Logo, considerando-se a seguinte região crítica para o teste:

$$RC = \{x; x \leq 2\}$$

a probabilidade de rejeitar-se H_0 , dado $p=q$ é dada por $Pr(\text{rejeitar-se } H_0 | p=q) = 4,65 \times 10^{-4}$. Note que esta conclusão é sujeita aos valores fixados das marginais.

O Teste Exato de Fisher pode ser utilizado de modo geral nos casos em que se deseja verificar a independência entre as características A e B dos indivíduos de uma população.

Para isso esquematiza-se a seguinte tabela:

	A	A ^c	
B	P	R	P+R
B ^c	Q	S	Q+S
	P+Q	R+S	1

Tabela II

O interesse está em avaliar se o fato de o indivíduo possuir a característica B(ou B^c) não influencia a probabilidade de que ele possua a característica A(ou A^c) e vice-versa.

A hipótese para teste é:

$$H_0: \Pr(A|B) = \Pr(A|B^c)$$

$$(\text{ou } \Pr(B|A) = \Pr(B|A^c))$$

e pode ser escrita como:

$$H_0: (P/P+R) = (Q/Q+S) \quad \text{ou} \quad H_0: (P/P+Q) = (R/R+S)$$

As frequências amostrais obtidas em um experimento cujo objetivo é o teste da hipótese acima podem ser dispostas no seguinte formato:

	A	A ^c	
B	x	m-x	m
B ^c	y	n-y	n
	t	N-t	N

Tabela III

Aqui, consideram-se dois experimentos que podem gerar a tabela acima:

- 1) **Multinomial:** retira-se, com reposição, uma amostra de N elementos de uma população, classificando-os nas categorias AB , AB^c , A^cB , A^cB^c . A disposição dos elementos da amostra nas 4 categorias segue uma distribuição Multinomial com parâmetros (N, P, Q, R) que é dada por:

$$P(x, y, m | N, P, Q, R) = \frac{N!}{x!(m-x)!y!(n-y)!} p^x R^{(m-x)} Q^y S^{(n-y)}$$

onde $n = N - m$

- 2) **Binomial:** Existem m indivíduos com a característica B e n indivíduos com a característica B^c . Para cada uma destas duas categorias observa-se o número de indivíduos com a característica A , x e y respectivamente. Estes números seguirão distribuições Binomiais, isto é:

$$x \sim \text{Bin}(m; P/P+R)$$

$$y \sim \text{Bin}(n; Q/Q+S)$$

A distribuição conjunta das variáveis aleatórias (x, y) é dada por:

$$P(x, y | N, m, P, Q, R) = \binom{m}{x} \left(\frac{P}{P+R} \right)^x \left(\frac{R}{P+R} \right)^{(m-x)} \binom{n}{y} \left(\frac{Q}{Q+S} \right)^y \left(\frac{S}{Q+S} \right)^{(n-y)}$$

Embora o modelo Multinomial seja diferente do modelo Binomial, pode-se utilizar o mesmo teste para ambos. As razões para isto são descritas abaixo, onde ao longo da discussão, o valor de N é considerado conhecido.

O modelo Multinomial pode ser escrito como:

$$P(x, y, m | N, P, Q, R) = P(x, y | N, m, P, Q, R) \cdot P(m | N, P, Q, R)$$

onde m é a variável aleatória que indica o número de indivíduos da amostra que possuem a característica B , tendo distribuição Binomial com parâmetros $(N, (P+R))$. Tal distribuição é dada por:

$$P(m | N, P, Q, R) = \binom{N}{m} (P+R)^m (1-P-R)^{(N-m)}$$

A distribuição de $(m | N, P, Q, R)$ não depende dos valores em teste, e sim de $(P+R)$. Logo, segundo o 1º postulado de Fisher, esta distribuição não fornece informação sobre (P, Q, R, S) . Neste caso, podemos abandoná-la, utilizando para o teste apenas a distribuição de $(x, y | N, m, P, Q, R)$, a qual, conforme foi visto anteriormente, é a distribuição do modelo Binomial.

O teste de Fisher é o mesmo para os dois modelos e usa o argumento básico de que as estatísticas marginais (m, t) não contêm informação sobre as probabilidades em questão.

Assim, a probabilidade de se obter a tabela descrita é dada pela distribuição Hipergeométrica, isto é:

$$P(x|N, m, t) = Pf(x|N, m, t) = \frac{\binom{m}{x} \binom{n}{y}}{\binom{N}{t}} \quad \text{onde } n = N - m \text{ e } y = t - x$$

O objetivo é testar $H_0: Pr(A|B) = Pr(A|B^c)$, ou seja, $H_0: p=q$ contra $H_1: p>q$.

Os casos que favoreceriam a rejeição de H_0 com mais ênfase do que o apresentado na tabela III, mantidos os totais marginais, poderiam ser esquematizados como segue (supondo, sem perda de generalidade, que $m>t$):

	A	A ^c	
B	x+1	m-x-1	m
B ^c	y-1	n-y+1	n
	t	N-t	N

Tabela IV

	A	A ^c	
B	x+2	m-x-2	m
B ^c	y-2	n-y+2	n
	t	N-t	N

Tabela V

	A	A ^c	
B	t	m-t	m
B ^c	o	n	n
	t	N-t	N

Tabela VI

Assim, o **Nível Descritivo** deste teste será a soma das probabilidades de obtenção destas tabelas, condicionadas ao valor de t, supondo H_0 verdadeira.

$$P_t(x|N, m, t) = \sum_{r=x}^t P_f(r|N, m, t) = \sum_{r=x}^t \frac{\binom{m}{r} \binom{n}{t-r}}{\binom{N}{t}}$$

Se esta soma for alta, os dados estarão fornecendo evidências de que H_0 é verdadeira e portanto H_0 não será rejeitada.

Se a soma for baixa, H_0 será rejeitada. A linha divisória entre as somas altas e baixas é o **Nível de Significância** do teste.

II.2 - A Lógica do Teste Exato de Fisher

O Teste Exato de Fisher é fundamentado sobre duas proposições básicas. A primeira diz respeito à distribuição de probabilidades utilizada no teste e a segunda diz respeito ao critério de decisão adotado. O objetivo aqui é entender e discutir a lógica sobre a qual estão baseadas estas duas pro-

posições. Parte da discussão abaixo é baseada no trabalho de Basu (3).

II.2.1 - A Distribuição de Probabilidades Utilizada no Teste

Como foi visto na seção anterior, a distribuição de probabilidades utilizada no Teste Exato de Fisher é a Hipergeométrica. Fisher chegou a esta distribuição servindo-se do argumento de que as estatísticas marginais (m, t) não conteriam informação sobre a hipótese em questão. Desta forma, não seria necessário levar-se em consideração a distribuição marginal dos dados (m, t) os quais poderiam ser considerados fixados para a obtenção da distribuição de probabilidades condicional associada ao corpo da tabela.

O objetivo desta seção é chegar ao âmago desta questão a qual tem origem num assunto bastante geral conhecido como Inferência Parcial.

A Inferência Parcial tenta resolver um problema muito importante que surge na elaboração de modelos cuja finalidade é o estudo de determinado estado desconhecido da natureza. Para que o cientista obtenha informação a respeito deste estado da natureza, o parâmetro θ , é necessário que ele planeje um experimento cujo resultado, x , esteja relacionado com o valor

(3) BASU, D. On the elimination of nuisance parameters.
JASA, 72:355-366, 1977.

desconhecido θ . A relação entre o resultado do experimento, x , e o parâmetro de interesse, θ , é descrita pela família de distribuições de probabilidades de x , que deve ser indexada pelo parâmetro θ : $P(x|\theta)$

O que ocorre freqüentemente, porém, é que não se consegue planejar experimentos de forma que o resultado x esteja ligado exclusivamente ao parâmetro de interesse θ . Em geral, x está relacionado a uma série de características (incluindo θ), resumidas aqui por λ . Dessa forma o modelo passa a ser denotado por $P(x|\lambda)$. Agora θ , o parâmetro de interesse, é parte de λ , ou seja, $\theta = \theta(\lambda)$. Existe, em contrapartida, um parâmetro complementar a θ , $\phi(\lambda)$, que resume as características do modelo que não interessam ao pesquisador. Assim, podemos montar uma relação biunívoca entre λ e o par $(\theta(\lambda), \phi(\lambda))$. Com esta representação, ϕ é conhecido como **Parâmetro Nuisance** pois além de ser uma característica do modelo que não interessa ao cientista, pode atrapalhá-lo na realização de inferências sobre o parâmetro de interesse, θ .

A relação de x com θ e ϕ é traduzida pela 'distribuição de probabilidades de x que passa a ser indexada por θ e ϕ e que será denotada por $P(x|\theta, \phi) = P(x|\lambda)$.

Como o problema é encontrar um método de inferência sobre θ , deseja-se eliminar do estudo o parâmetro nuisance ϕ . Uma das técnicas utilizadas para isso é o **Método da Redução**, ou seja, a elaboração de um novo modelo a partir do primeiro, de forma que a família de distribuições associada a este modelo dependa apenas de θ . Este novo modelo deve caracterizar-se

por duas propriedades básicas:

1a.) Simplicidade: o modelo reduzido deve possuir uma distribuição que dependa somente de θ .

2a.) Informatividade: o modelo reduzido não pode fornecer menos informação sobre θ do que o modelo original.

A segunda propriedade é bastante discutível pois o conceito de informação é polêmico. Um modelo considerado "informativo" por um estatístico pode ser considerado "não-informativo" por outro. Por outras vezes, pode-se considerar que a simplificação obtida através da redução de um modelo compensa uma "pequena" perda de informação. Em resumo, existem várias definições de informação que obedecem aos quatro postulados de Fisher citados no início do trabalho.

Como já foi visto, a relação dos dados x com os parâmetros θ e ϕ é traduzida pela distribuição de probabilidades $P(x|\theta, \phi)$. Uma vez observado o resultado x , $P(x|\theta, \phi)$ será função apenas dos parâmetros desconhecidos θ e ϕ . Neste caso, $P(x|\theta, \phi)$ será chamada de verossimilhança (ou função de verossimilhança) e denotada por $L(\theta, \phi|x)$. A verossimilhança transporta toda a informação contida no resultado x a respeito dos parâmetros θ e ϕ . Para que se retire alguma informação de x a respeito de θ , deve-se estar certo de que, ao variar-se θ , a verossimilhança $L(\theta, \phi|x)$ também varia. Se isto não ocorre, ou seja, se para qualquer valor de θ obtém-se a mesma verossimilhança, não deve haver relacionamento entre θ e x , e nenhuma informação a respeito de θ poderá ser retirada de x (1º postulado de Fisher acerca de informação).

Deste modo, para inferências sobre θ baseadas nos dados x , pode-se abandonar todos os fatores de $L(\theta, \phi | x)$ que não se alteram ao variar-se θ , considerando-se apenas os fatores que são função de θ .

O segredo do Método da Redução consiste precisamente na fatoração da função de verossimilhança em fatores que dependam apenas de θ e fatores que dependam apenas de ϕ . Deste modo, pode-se abandonar os fatores dependentes de ϕ , uma vez que eles não carregam informação sobre θ . Com os fatores dependentes apenas de θ , pode-se usar as técnicas usuais de inferência. Para que este método seja coerente, é necessário que não exista ligação entre θ e ϕ . Com o objetivo de exprimir a ausência de ligação entre θ e ϕ , estabeleceu-se o conceito de Variação Independente apresentado na sequência (Definição II.1).

Para que se consiga a fatoração desejada da função de verossimilhança, é necessário que se enfrente dois problemas:

O primeiro diz respeito à escolha do parâmetro de interesse θ . O teste $H_0: \theta = \mu_1 - \mu_2 = 0$ é equivalente ao teste $H_0: \theta = \mu_1/\mu_2 = 1$. Ligada à escolha do parâmetro de interesse, está a escolha do parâmetro nuisance ϕ e uma escolha adequada do par (θ, ϕ) poderá facilitar bastante o problema da fatoração da função de verossimilhança. Esta escolha é arbitrária e não existe uma receita que indique qual delas levará à solução mais simples para o problema. Muitas vezes não se consegue a fatoração da função de verossimilhança por falta de uma reparametrização adequada.

O segundo problema está ligado à determinação de uma estatística $t=t(x)$, função dos dados provenientes do experimento, de forma que se consiga fatorar $L(\theta, \phi|x)$.

Com a finalidade de solucionar estes problemas, seguem as definições:

Definição II.1: Os parâmetros $\theta: \Lambda \longrightarrow \Theta$ e $\phi: \Lambda \longrightarrow \Phi$ são de **Variação Independente** se o campo de variação de θ é o mesmo para qualquer valor de ϕ e vice-versa.

Seja $t=t(x)$ uma estatística, função dos dados x . Sendo t uma variável aleatória, pode-se escrever:

$$P(x|\theta, \phi) = \sum_t P(x|t, \theta, \phi) \cdot P(t|\theta, \phi)$$

Porém, como t é função de x , cada valor de x determina um único valor de t , e a somatória fica reduzida a uma única parcela:

$$P(x|\theta, \phi) = P_1(t|\theta, \phi) \cdot P_2(x|t, \theta, \phi)$$

Definição II.2: Se θ e ϕ são parâmetros de Variação Independente e se $P_2(x|t, \theta, \phi) = P_2(x|t, \theta)$, ou seja, se $P_2(x|t, \theta, \phi)$ não depende do valor de ϕ , então t é **Especificamente Suficiente** para inferir sobre ϕ a partir de x . Neste caso, pode-se escrever:

$$P(x|\theta, \phi) = P_1(t|\theta, \phi) \cdot P_2(x|t, \theta)$$

Definição II.3: Se θ e ϕ são parâmetros de Variação Independente e se $P_1(t|\theta, \phi) = P_1(t|\phi)$, ou seja, se $P_1(t|\theta, \phi)$ não depende do valor de θ , então t é Especificamente Ancilar para inferir sobre θ a partir de x .

Neste caso, escreve-se:

$$P(x|\theta, \phi) = P_1(t|\phi) \cdot P_2(x|t, \theta, \phi)$$

Dadas as definições acima, o problema fica reduzido à procura de uma estatística $t=t(x)$ que seja ao mesmo tempo especificamente suficiente para inferir sobre ϕ a partir de x e especificamente ancilar para inferir sobre θ a partir de x . Assim, o modelo se escreverá como:

$$P(x|\theta, \phi) = P_1(t|\phi) \cdot P_2(x|t, \theta)$$

Tomando-se $P(x|\theta, \phi)$ como função de θ e ϕ , isto é, a verossimilhança $L(\theta, \phi|x)$, e servindo-se da estatística t com as propriedades acima, alcança-se o objetivo inicial. Através da estatística t , fatora-se a verossimilhança em uma parte que depende apenas de ϕ e outra que depende apenas de θ . Neste caso diz-se que a estatística t define um corte de Barndorff - Nielsen para o modelo $P(x|\theta, \phi)$.

É importante ressaltar a dificuldade existente no encon-

tro de uma reparametrização e de uma estatística que defina um corte de Barndorff - Nielsen para o modelo.

Vejamos como a inferência parcial se aplica no caso do Teste Exato de Fisher.

Retomando a tabela 2x2, vem:

apud new apud

	A	A ^c	Total
<i>Marcas</i> B	x	m-x	m
<i>new</i> B ^c	y	n-y	n
Total	t	N-t	N

Tabela VII.

Utilizando-se o modelo Binomial, isto é, considerando-se que:

$$x \sim \text{Bin}(m; p)$$

$$y \sim \text{Bin}(n; q)$$

o interesse está em testar:

$$H_0: p=q$$

x

$$H_1: p>q$$

A distribuição dos dados é escrita como:

$$P(x, y | N, m, p, q) = \binom{m}{x} p^x (1-p)^{m-x} \binom{n}{y} q^y (1-q)^{n-y}$$

Através da reparametrização $\theta = p(1-q)/q(1-p)$ e $\phi = q$, as hipóteses em teste tomam a seguinte forma:

$$H_0: \theta = 1$$

x

$$H_1: \theta > 1$$

A próxima etapa é encontrar um modelo reduzido que dependa apenas de θ , eliminando-se assim o parâmetro nuisance ϕ .

Substituindo n por $N-m$ e y por $t-x$, chega-se à seguinte expressão:

$$P(x, t | N, m, \theta, \phi) = \frac{\binom{m}{x} \binom{N-m}{t-x} \theta^x \phi^t (1-\phi)^{N-t}}{(1-\phi+\theta\phi)^m}$$

Agora segue-se a busca de uma fatoração tal que se possa escrever:

$$P(x, t | N, m, \theta, \phi) = P_1(t | N, m, \phi) \cdot P_2(x | t, N, m, \theta)$$

Para que se obtenha P_1 , basta que se some $P(x, t | N, m, \theta, \phi)$ para todos os valores possíveis de x :

$$\begin{aligned} \sum_{k=0}^t P(k, t | N, m, \theta, \phi) &= \frac{\phi^t (1-\phi)^{N-t}}{(1-\phi+\theta\phi)^m} \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k \\ &= P_1(t | N, m, \theta, \phi) \end{aligned}$$

Tendo $P_1(t | N, m, \theta, \phi)$, a obtenção do segundo fator é imediata:

$$\begin{aligned}
 P_2(x|t, N, m, \theta, \phi) &= \frac{P(x, t|N, m, \theta, \phi)}{P_1(t|N, m, \theta, \phi)} = \\
 &= \frac{\binom{m}{x} \binom{N-m}{t-x} \theta^x}{\sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k} = P_2(x|t, N, m, \theta)
 \end{aligned}$$

A fatoração obtida é, então:

$$P(x, t|N, m, \theta, \phi) = P_1(t|N, m, \theta, \phi) \cdot P_2(x|t, N, m, \theta) \text{ onde}$$

$$P_1(t|N, m, \theta, \phi) = \frac{\phi^t (1-\phi)^{N-t}}{(1-\phi+\theta\phi)^m} \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k$$

$$P_2(x|t, N, m, \theta) = \frac{\binom{m}{x} \binom{N-m}{t-x} \theta^x}{\sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k}$$

Desta forma chegou-se a um modelo reduzido pois $P_2(x|t, N, m, \theta)$ depende apenas do parâmetro de interesse θ . Sob

$$H_0, \theta=1 \text{ é } P_2(x|t, N, m, 1) = \frac{\binom{m}{x} \binom{N-m}{t-x}}{\sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k}}$$

$P_2(x|t, N, m, 1)$ é uma distribuição de probabilidades e portanto $\sum_{k=0}^t P_2(k|t, N, m, 1) = 1$. Impondo esta condição, reconhece-se que $(x|t, N, m, 1)$ deve seguir uma distribuição Hipergeométrica.

$$\text{Logo, } P(x|t, N, m) = \frac{\binom{m}{x} \binom{N-m}{t-x}}{\binom{N}{t}}$$

e tem-se, então, a distribuição do Teste de Fisher apresentada na seção precedente.

Observe, porém, que o modelo $P_1(t|N, m, \theta, \emptyset)$ depende de θ , ou seja, carrega alguma informação sobre θ que está sendo desprezada quando se considera apenas o modelo $P_2(x|t, N, m, \theta)$. Em resumo, a estatística t não define um corte de Barndorff - Nielsen para o modelo.

Como então, se justifica o uso do modelo reduzido? Quanta informação está sendo perdida ao desprezar-se $P_1(t|N, m, \theta, \emptyset)$? Quando um fator do modelo, embora dependendo de θ , pode ser considerado não informativo com respeito a θ e conseqüentemente eliminado da análise?

Alguns estatísticos consideram muito restritiva a condição de que o modelo não informativo não dependa do valor de θ e tentaram relaxá-la visando justificar, entre outras coisas, o Teste Exato de Fisher. A partir daí, surgiu o conceito de G-ancilaridade, desenvolvido por Godambe (4).

Definição II.4: A estatística t é G-ancilar com respeito a θ se:

(4) GODAMBE, V.R. On Sufficiency and Ancilarity in presence of a nuisance parameter. *Biometrika*, 67:155-162, 1980

- (a) A distribuição condicional de x dado t depende somente de θ , isto é:

$$P(x|\theta, \phi) = P_1(t|\theta, \phi) \cdot P_2(x|t, \theta)$$

- (b) Para cada $\theta \in \Theta$, a classe das distribuições marginais de t é completa.

A condição (a) indica um modelo mais simples sobre o qual pode-se fazer inferências a respeito de θ . Basta que se utilize $P_2(x|t, \theta)$. Porém o que nos garante que o modelo $P_1(t|\theta, \phi)$ é não informativo com respeito a θ ?

Godambe desenvolveu uma teoria que mostra que a distribuição condicional $P_2(x|t, \theta)$ e a distribuição $P(x|\theta, \phi)$ contêm a mesma informação sobre o parâmetro θ se e somente se a estatística t for G-ancilar com respeito a θ . Como consequência conclui-se que se t é G-ancilar com respeito a θ , $P_1(t|\theta, \phi)$ não deve conter informação sobre θ , apesar de depender de θ . Ressalte-se que Godambe serviu-se da Informação de Fisher para demonstrar sua teoria. A Informação de Fisher, definida por:

$$I(\theta) = E \left[\frac{\partial}{\partial \theta} \ln P(x|\theta, \phi) \right]^2 = - E \left[\frac{\partial^2}{\partial \theta^2} \ln P(x|\theta, \phi) \right]$$

é um conceito bastante discutível, principalmente se considerarmos que trata-se de uma média para todas as amostras possíveis.

A partir da definição de G-ancilaridade, torna-se interessante generalizar o Teorema de Lehmann - Scheffé.

Seja a estatística t G-ancilar.

Considerando $\theta_0 \in \Theta$ fixo, tem-se pela condição (a) da Definição II.4 que:

$$P(x|\theta_0, \phi) = P_1(t|\theta_0, \phi) \cdot P_2(x|t, \theta_0)$$

Logo t é uma estatística suficiente para os dados x com respeito a ϕ . Além disso, a condição (b) afirma que a família $P_{t, \theta_0} = \{P(t|\theta_0, \phi), \phi \in \Phi\}$ é completa. Estas são as condições do Teorema de Lehmann - Scheffé e portanto t é a única Estatística Suficiente Mínima para a família $P_{\theta_0} = \{P(x|\theta_0, \phi), \phi \in \Phi\}$, $\theta_0 \in \Theta$.

Conseqüentemente segue o teorema:

Teorema II.1: Se a estatística t é G-ancilar com respeito a θ , então para $\theta_0 \in \Theta$, fixado e arbitrário, t é a única Estatística Suficiente Mínima para a família:

$$P_{\theta_0} = \{P(x|\theta_0, \phi), \phi \in \Phi\}$$

Neste caso t será chamada de Estatística Supra-Ancilar com respeito a θ .

Este teorema auxilia a justificar o abandono do fator $P_1(t|\theta, \phi)$, apesar do mesmo depender de θ . Se t é uma estatística suficiente mínima, t é função de todas as outras estatísticas suficientes para inferir sobre ϕ . Conseqüentemente,

é a estatística mais resumida que mantém a propriedade de suficiência. Logo, se t contiver alguma informação sobre θ , esta será mínima se comparada à informação contida nas demais estatísticas suficientes com respeito à ϕ . Assim, se concordarmos em adotar um modelo reduzido do tipo $P_2(x|t, \theta)$, é melhor que utilizemos a estatística G-ancilar t . Neste caso, se houver perda de informação sobre θ ao descartarmos o fator $P_1(t|\theta, \phi)$, tal perda será a menor possível.

Neste ponto, estamos em condições de justificar o abandono do fator $P_1(t|N, m, \theta, \phi)$ na determinação da distribuição utilizada no Teste Exato de Fisher. Retomando a fatoração do modelo, vem:

$$P(x, t|N, m, \theta, \phi) = P_1(t|N, m, \theta, \phi) \cdot P_2(x|t, N, m, \theta)$$

Se conseguirmos provar que t é G-ancilar com respeito a θ , provaremos que, segundo Godambe, $P_1(t|N, m, \theta, \phi)$ não transporta informação sobre θ e o Teste de Fisher estará justificado. Para tanto, basta provar que para $\theta \in \Theta$ fixado, a família

$$P_\theta = \{ P_1(t|N, m, \theta, \phi), 0 < \phi < 1 \} \text{ é completa.}$$

Temos que:

$$P_1(t|N, m, \theta, \phi) = \frac{\phi^t (1-\phi)^{N-t}}{(1-\phi+\theta\phi)^m} \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k$$

Devemos mostrar que, sendo $u(t)$ uma função qualquer de t , vale a implicação:

$$E \{ u(t) | N, m, \theta, \phi \} = 0 \Rightarrow u(t) \equiv 0$$

Demonstração:

Para efeito de simplificação, os condicionantes serão eliminados da notação.

$$E \{ u(t) \} = \sum_{t=0}^N \frac{u(t) \cdot \phi^t (1-\phi)^{N-t}}{(1-\phi+\theta\phi)^m} \cdot \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k =$$

$$= \frac{1}{(1-\phi+\theta\phi)^m} \sum_{t=0}^N u(t) \phi^t (1-\phi)^{N-t} \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k$$

$$\therefore E \{ u(t) \} = 0 \Rightarrow \sum_{t=0}^N u(t) \phi^t (1-\phi)^{N-t} \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k = 0$$

$$\Rightarrow u(t) \equiv 0$$

A última implicação tem a seguinte justificativa: A expressão é um polinômio de grau N em ϕ , com coeficientes

$$c(t) \cdot u(t), \text{ onde } c(t) = \sum_{k=0}^t \binom{m}{k} \binom{N-m}{t-k} \theta^k > 0$$

pois $\theta > 0$. Este polinômio deve anular-se para qualquer valor de $\phi \in (0, 1)$ e portanto deve ter mais do que N raízes. Um polinômio de grau N só possui mais do que N raízes se for identicamente nulo, portanto $u(t) \equiv 0$.

Logo, a família $P_{\theta_0} = \{P_1(t|N, m, \theta_0, \phi), 0 < \phi < 1\}$ é completa e a estatística t é G-ancilar com respeito à θ . Justifica-se, desta forma, o Teste Exato de Fisher.

Devemos ressaltar aqui que o critério de Godambe considera o modelo como um todo não levando em conta a informação contida em um ponto amostral específico. É evidente que pontos amostrais distintos podem carregar informações diferentes sobre o valor de θ . Se, por exemplo, $t=0$, a função $P_2(0|N, m, \theta, \phi)$ carrega toda a informação que a verossimilhança possui sobre θ pois nesse caso:

$$P_1(x|t=0, N, m, \theta, \phi) = \begin{cases} 1 & \text{se } x = 0 \\ 0 & \text{se } x \neq 0 \end{cases} \quad \text{qualquer que seja } \phi$$

No entanto, a situação pode ser invertida para outros valores de t . Voltaremos a analisar este ponto na sequência do trabalho.

II.2.2 - O Critério de Decisão Adotado por Fisher

O critério de decisão adotado no Teste Exato de Fisher decorre do conhecimento da probabilidade exata de obtenção de cada uma das possíveis tabelas de freqüências, considerando os totais marginais fixos e H_0 verdadeira. Para uma determinada tabela de freqüências resultante de um experimento, calcula-se esta probabilidade. A primeira idéia que nos ocorre é a de que devemos aceitar H_0 se esta probabilidade for alta (o

resultado obtido é "verossímil" sob H_0) e rejeitar H_0 se esta probabilidade for baixa (o resultado obtido é "inverossímil" sob H_0).

Contudo, esta idéia merece restrições. Suponha que o resultado observado tenha alta probabilidade sob H_0 mas tenha uma probabilidade ainda mais alta sob H_1 . Isto ocorre quando, por exemplo, deseja-se testar se os dados obtidos através da realização de um experimento provêm de uma distribuição $N(\mu, \sigma_1^2)$ ou de uma distribuição $N(\mu, \sigma_2^2)$ com $\sigma_1^2 > \sigma_2^2$. Aqui, $N(\mu, \sigma^2)$ representa a distribuição Normal com média μ e variância σ^2 . O teste seria:

$$H_0: \sigma^2 = \sigma_1^2$$

x

$$H_1: \sigma^2 = \sigma_2^2$$

As duas distribuições podem ser esquematizadas como se segue:

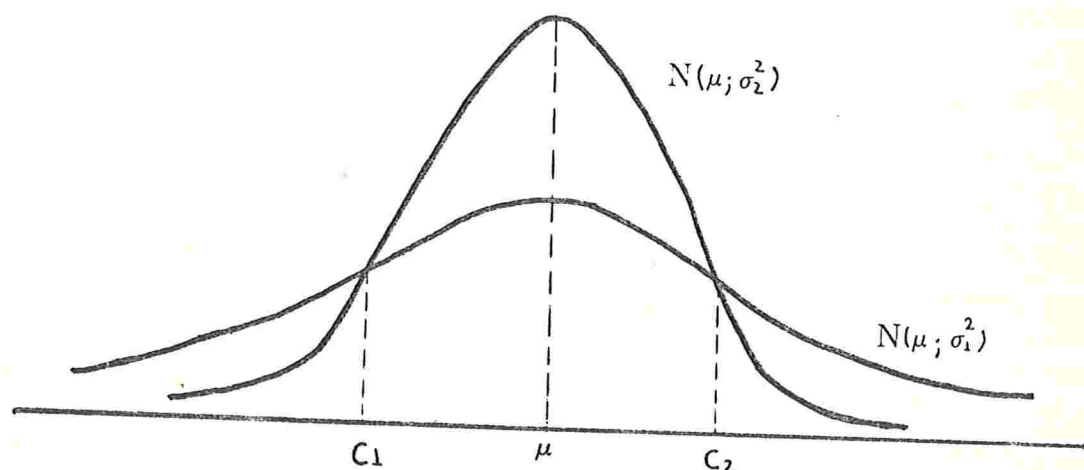


Figura II.1

O resultado do experimento produzirá uma média amostral que, com alta probabilidade, assumirá um valor no intervalo $(C1, C2)$. Neste caso, se adotarmos nossa idéia inicial, aceitaremos H_0 em favor de H_1 . Contudo, este evento é mais provável sob H_1 do que sob H_0 e possivelmente estaremos cometendo um erro. Conclui-se, então, que nossa idéia inicial deve ser submetida a uma análise mais profunda. Não basta apenas observar quão provável é o resultado do experimento sob H_0 ; H_1 também deve ser analisada. Segue daí que um procedimento de teste coerente deveria comparar as verossimilhanças sob H_0 e sob H_1 assim como no exemplo abaixo:

$\text{Pr}(\text{resultado} | H_0) > \text{Pr}(\text{resultado} | H_1) \Rightarrow$ não rejeito H_0
 $\text{Pr}(\text{resultado} | H_0) \leq \text{Pr}(\text{resultado} | H_1) \Rightarrow$ rejeito H_0

Para uma melhor formalização de tal procedimento, consideramos o conceito de Razão de Teste:

Definição II.5: Seja um experimento realizado com o intuito de testar a hipótese H_0 contra a alternativa H_1 . Seja x o resultado do experimento. Define-se Razão de Teste como sendo o quociente das probabilidades de obtenção de x sob H_0 e sob H_1 como segue:

$$R(x) = L(H_0 | x) / L(H_1 | x)$$

A estatística $R(x)$ pode ser encarada como uma medida da vantagem que H_0 leva sobre H_1 dado o resultado do experimento.

to, x.

Um critério de decisão poderia ser o de rejeitar H_0 se a Razão de Teste for menor do que uma constante arbitrária K e não rejeitar H_0 em caso contrário.

A escolha de K pode estar ligada ao crédito que se dá à veracidade de H_0 , a priori. Pode também ser tal que a probabilidade de que se cometa o erro de tipo I, ou seja, $\Pr(\text{rejeitar } H_0 | H_0 \text{ verdadeira})$ seja α .

Definição II.6: Critério da Razão de Teste.

Rejeitamos H_0 se e somente se $R(x) < K$.

Não rejeitamos H_0 se e somente se $R(x) \geq K$.

Teorema II.2: Suponha um estudo baseado em uma estatística $t=t(x)$, função do resultado do experimento. Suponha também, que para o problema em questão, a Razão de Teste $R(x)$ seja uma função monotônica crescente na estatística t , isto é, se $t(x) > t(x')$ então $R(x) > R(x')$. Além disso, $R(x)=K$ no conjunto $\{x: t(x)=t_0\}$.

Neste caso, o critério de decisão da Definição II.6 é equivalente ao critério:

Rejeitamos H_0 se e somente se $t(x) < t_0$.

Não rejeitamos H_0 se e somente se $t(x) \geq t_0$.

A demonstração deste resultado é imediata, e o teorema é análogo para o caso de $R(x)$ ser uma função monotonicamente decrescente na estatística $t(x)$. Neste caso, as desigualdades

do critério de decisão devem ser invertidas.

Assim, segundo o teorema acima, em lugar de fixarmos a constante K como limite dos valores de $R(x)$ podemos fixar t_0 como limite dos valores da estatística $t(x)$. Como os testes, em geral, são baseados no erro de tipo I, e como a distribuição de probabilidades da estatística $t(x)$ é conhecida, costuma-se fixar t_0 de forma que o Nível de Significância do teste, isto é, $\Pr(t(x) < t_0 | H_0 \text{ verdadeira})$ seja α_0 (ou mais próximo possível de α_0).

Retomando-se o Teste Exato de Fisher, tem-se as hipóteses:

$$H_0: p=q$$

x

$$H_1: p > q$$

Considerando-se a reparametrização $\theta = p(1-q)/q(1-p)$ e $\phi = q$, as hipóteses podem ser reescritas como segue:

$$H_0: \theta = 1$$

x

$$H_1: \theta > 1$$

É oportuno lembrar que existe uma correspondência um a um entre (θ, ϕ) e (p, q) .

Neste caso, a Razão de Teste será:

$$R(x) = \frac{L(N, m, \theta = 1, \phi, t|x)}{L(N, m, \theta > 1, \phi, t|x)} \quad \text{pois } t \text{ é fixado}$$

Logo

$$R(x) = \frac{[1 + \phi(\theta - 1)]^m}{\theta^x} \quad \text{para } \theta > 1 \text{ e } 0 \leq \phi \leq 1$$

A estatística do teste é x , com m e t constantes. Sendo $\theta > 1$, a Razão de Teste é monotonicamente decrescente em x . Desta forma, pode-se determinar um critério de decisão para o teste encontrando um valor de x , x_0 , abaixo do qual deixa-se de rejeitar H_0 . Enfim, H_0 será rejeitada sempre que for obtida uma tabela cujo corpo contenha um valor de x maior do que x_0 .

A determinação do valor x_0 é baseada na escolha do Nível de Significância α_0 do teste. Para isso, basta que se encontre o valor x_0 de modo que a soma das probabilidades de obtenção de todas as tabelas tais que $x \geq x_0$ seja α_0 , supondo $p=q$. O cálculo destas probabilidades no Teste Exato de Fisher é feito através da distribuição Hipergeométrica, conforme foi discutido na seção II.2.1.

Justifica-se, assim, o procedimento de decisão utilizado no Teste Exato de Fisher.

CAPÍTULO III

DISCUSSÃO DA LÓGICA UTILIZADA NO TESTE EXATO DE FISHER

III.1 - A Distribuição de Probabilidades do Teste

No capítulo anterior tentou-se justificar o uso da distribuição Hipergeométrica no Teste Exato de Fisher fazendo-se uso de Método da Redução o qual fatora a função de verossimilhança associada ao teste em dois fatores, um dependendo do parâmetro de interesse θ e outro dependendo do parâmetro nuisance ϕ . Dessa forma foi obtida a seguinte expressão:

$$P(x, t | N, m, \theta, \phi) = P_1(t | N, m, \theta, \phi) \cdot P_2(x | t, N, m, \theta)$$

Como não conseguimos que um fator dependesse apenas de θ e outro apenas de ϕ , isto é, como a estatística t não define um corte de Barndorff-Nielsen para o modelo, fomos obrigados a recorrer ao conceito de G-ancilaridade para justificar o

abandono do fator $P_1(t|N, m, \theta, \phi)$. Segundo este conceito, $P(x, t|N, m, \theta, \phi)$ e $P_2(x|t, N, m, \theta)$ contêm a mesma informação a respeito do parâmetro θ .

Contudo, ao abandonar-se o fator $P_1(t|N, m, \theta, \phi)$, abandona-se um fator que depende do parâmetro de interesse θ . Esta dependência significa que a probabilidade de observarmos um evento associado à estatística t varia à medida que θ varia. O desprezo de tal fator em inferências sobre θ contraria o Princípio da Verossimilhança, uma vez que o fator abandonado representa uma relação entre a estatística t e o parâmetro θ . Desta forma, de acordo com o primeiro postulado do próprio Fisher, a estatística t deve fornecer informação sobre θ .

O aparente paradoxo que se apresenta aqui pode ser explicado se levarmos em conta que Godambe valeu-se do conceito de Informação de Fisher para justificar a ausência de Informação a respeito de θ em $P_1(t|N, m, \theta, \phi)$. A Informação de Fisher, definida através da expressão:

$$I(\theta) = E \left[\frac{\partial}{\partial \theta} \ln P(x|N, m, \theta, \phi) \right]^2$$

é uma média cujo cálculo leva em conta todas as amostras possíveis.

O exemplo seguinte, dado por Berkson (1) mostra que

(1) BERKSON, J. Do the marginal totals of the 2x2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, 2 : 43-44, 1978

determinadas amostras contêm informação relevante nas frequências marginais das tabelas 2x2.

	A	A ^c	Total
B	x	m-x	m
B ^c	y	n-y	n
Total	t	N-t	N

Na tabela ao lado
N e m são fixados
previamente.

$$p = \Pr(A|B) \quad \hat{p} = x/m$$

$$q = \Pr(A|B^c) \quad \hat{q} = y/n$$

$$H_0: p = q$$

$$H_1: p > q$$

Existem (N+1) totais possíveis para o valor de t, (0, 1, 2, ..., N). Correspondendo a cada valor de t, existe um determinado número de tabelas para as quais $x \leq m$, $y \leq n$ e $x+y=t$. Para efeito do teste só são consideradas as tabelas que favoreceriam a rejeição de H_0 com ênfase maior ou igual ao da tabela obtida. Desta forma, calcula-se $P_t(x|N, m, t)$ definido no capítulo anterior por:

$$P_t(x|N, m, t) = \sum_{r=x}^t P_f(r|N, m, t) = \sum_{r=x}^t \frac{\binom{m}{r} \binom{n}{t-r}}{\binom{N}{t}}$$

Se $P_t(x|N, m, t) \leq \alpha$ (onde α é um valor previamente estipulado pelo pesquisador), p é considerado significativamente maior do que q ao nível α . Uma tabela assim é dita significativa.

Para diferentes totais N e t , a proporção de tabelas significantes pode variar de 0 a 1. Pode-se chamar esta proporção de "probabilidade de p ser significativamente maior do que q dados os totais marginais" e denotá-la Ps . Considerando $\alpha=0,05$, temos o seguinte quadro (Berkson - 1976):

t	x	Pt	x	Pt	x	Pt	Ps
0	0	1,00					0
1	1	0,50					0
2	1	0,78	2	0,22			0
3	2	0,50	3	0,08			0
4	2	0,74	3	0,26	4	0,02	1/3
5	3	0,50	4	0,10	5	0,004	1/3
6	3	0,74	4	0,26	5	0,02	1/3
7	4	0,50	5	0,08			0
8	4	0,78	5	0,22			0
9	5	0,50					0
10	5	1,00					0

$\alpha=0,05$ $N=10$ $m=5$

Ps =proporção de tabelas significantes dados os totais marginais

É fácil perceber que a probabilidade de rejeição de H_0 varia de acordo com os totais marginais. Nos casos extremos, isto é, nos casos em que $t=0$ e $t=10$, o corpo da tabela é de-

terminado exatamente e quando isto ocorre, nunca rejeitamos H_0 .

Examinando o quadro acima fica claro que à medida que t se aproxima de 0 ou de N , a diferença entre x e y tende a diminuir, dificultando a rejeição da hipótese nula. Conclui-se então, que a informação que podemos extrair de t depende do valor de t . A amostra efetivamente observada é que deveria ter o maior peso em inferências sobre θ . Os outros pontos amostrais que poderiam ter ocorrido, caso tenham que ser levados em consideração, deveriam ter um peso insignificante em relação ao observado.

Neste contexto somos induzidos a colocar em dúvida o conceito de Informação de Fisher, que nada mais é do que uma medida que leva em conta todas as amostras que poderiam ter sido sorteadas e não dá importância especial à amostra efetivamente observada.

Tentando entender o problema, Basu (2) considerou o seguinte exemplo, em que $m=n=3$ e $(x,y) = (3,0)$

	A	A ^c	Total
B	3	0	3
B ^c	0	3	3
Total	3	3	6

(2) BASU, D. Discussion of Joseph Berkson's paper "In dispraise of the exact test". Journal of Statistical Planning and Inference, 3: 189-192, 1979

Os totais m e n são fixados e o vetor amostral (x,y) varia sobre um espaço amostral de $2^4 = 16$ pontos. A amostra $(x,y) = (3,0)$ compõe o caso mais extremo em que $p \gg q$. Portanto, estes dados deveriam fornecer uma forte evidência contra a hipótese nula.

Considerando-se $t=3$ fixo, reduz-se o espaço amostral de 16 para 4 pontos, representados no conjunto $\{ (0,3), (1,2), (2,1), (3,0) \}$. Sob a hipótese nula, a probabilidade de que o par $(3,0)$ seja observado no conjunto acima

é dada por $\frac{\binom{3}{3} \binom{3}{0}}{\binom{6}{3}} = 0,05$. Este é o caso mais extremo e $P_t(3)=0,05$.

A um nível de significância $\alpha=0,045$ por exemplo, nem mesmo esta amostra levaria à rejeição de H_0 .

Tomando o espaço amostral de 16 pontos como conjunto de referência, a probabilidade de se obter o valor $(3,0)$ sob H_0 seria dada por:

$$p^3 (1-p)^3 \quad \text{sendo} \quad H_0: p=q$$

Como $p^3 (1-p)^3 \leq 1/64$, poder-se-ia dizer que os dados $(3,0)$ são altamente significantes ao nível $1/64 = 0,0156$.

Conclui-se que a redução do espaço amostral provocada pelo condicionamento nos totais marginais pode gerar resultados diferentes dos obtidos caso tal condicionamento não tivesse sido realizado. O Teste de Fisher não rejeitaria a hipótese nula com a amostra $(3,0)$. O teste que utiliza o espaço amostral completo (16 pontos) rejeitaria H_0 . Aliás, para $m=n=3$, o Teste de Fisher jamais rejeitaria H_0 a um nível de signifi-

cância $\alpha=0,045$, mesmo que as evidências favorecessem a hipótese alternativa.

Um outro problema que deve ser ressaltado no uso da Inferência Parcial é a determinação da reparametrização conveniente. Tal reparametrização deve fornecer a fatoração da função de verossimilhança de forma que se possa valer da teoria de Godambe.

No caso da tabela de contingência 2×2 em que foram consideradas distribuições Binomiais, ou seja, $x \sim \text{Bin}(m, p)$ e $y \sim \text{Bin}(n, q)$, fez-se uso da reparametrização $\theta = p(1-q)/q(1-p)$ e $\phi = q$ recaindo-se no teste de $H_0: \theta=1$ contra $H_1: \theta > 1$.

A distribuição

$$P(x|t, \theta, \phi) = \frac{\binom{m}{x} \binom{n}{y} \theta^x}{\sum_{k=0}^t \binom{m}{k} \binom{n}{t-k} \theta^k}$$

foi adotada para a realização de inferências sobre θ porque não depende do parâmetro nuisance ϕ e porque t é suficiente e completa com respeito a ϕ para cada valor de θ .

Suponha, porém, que a mesma tabela de contingência 2×2 tenha sido obtida de maneira diversa, ou seja, x e y têm distribuição Binomial Negativa. Desta forma, em lugar de se considerar m e n fixados, $(m-x)$ e $(n-y)$ é que são conhecidos previamente. A distribuição dos dados seria escrita como:

$$P(x, y | (m-x), (n-y), p, q) = \binom{m-1}{x-1} \binom{n-1}{y-1} p^x (1-p)^{(m-x)} q^y (1-q)^{(n-y)}$$

Em termos de θ e ϕ :

$$P(x, t | (m-x), (n-y), \theta, \phi) = \frac{\binom{m-1}{x-1} \binom{n-1}{y-1} \theta^x \phi^t (1-\phi)^{(m+n)}}{(1-\phi+\theta\phi)^t}$$

onde $t=x+y$

Neste caso, (x, t) e $(m-x, n-y)$ são fixados previamente.

Agora não se consegue repetir o procedimento utilizado no caso Binomial Positivo pois a estatística suficiente mínima com respeito a ϕ dado θ é a amostra toda, isto é, (x, t) .

Contudo, se for considerada uma outra reparametrização, $\xi = p/q$ e $\phi = q$, as hipóteses em confronto serão $H_0: \xi=1$ e $H_1: \xi > 1$. Neste caso, ϕ continua sendo o parâmetro nuisance e a distribuição dos dados passa a ser escrita como:

$$P(x, t | (m-x), (n-y), \xi, \phi) = \binom{m-1}{x-1} \binom{n-1}{y-1} \xi^x (1-\xi\phi)^{(m-x)} \phi^t (1-\phi)^{(n-t+x)}$$

Para que se obtenha $P_1(t | (m-x), (n-y), \xi, \phi)$ basta que se some a expressão acima para todos os valores possíveis de x .

$$P_1(t | (m-x), (n-y), \xi, \phi) = (1-\xi\phi)^{(m-x)} \phi^t (1-\phi)^{(n-y)} \sum_{k=0}^t \binom{k+m-x-1}{k-1} \binom{t-k+n-y-1}{t-k-1} \xi^k$$

A obtenção do segundo fator é imediata:

$$P_2(x | t, (m-x), (n-y), \xi, \phi) = \frac{\binom{m-1}{x-1} \binom{n-1}{y-1} \xi^x}{\sum_{k=0}^t \binom{m-1}{k-1} \binom{n-1}{t-k-1} \xi^k}$$

$$= P_2(x | t, (m-x), (n-y), \xi)$$

O segundo fator não depende do parâmetro nuisance ϕ . Agora t é uma estatística suficiente mínima com respeito à ϕ dado ξ e podemos nos valer do argumento da verossimilhança parcial. Se esta nova reparametrização tivesse sido utilizada no caso Binomial Positivo, a estatística suficiente mínima com respeito à ϕ dado ξ seriam os dados (x, t) e não poderíamos utilizar o Método da Redução.

O que nos parece estranho aqui é que devemos considerar parametrizações distintas para tabelas iguais, somente por terem sido obtidas a partir de espaços amostrais diferentes. Não parece coerente analisar ξ se os dados vierem de uma distribuição Binomial Negativa e θ caso os dados venham de uma distribuição Binomial Positiva pois além de a tabela resultante ser a mesma, as verossimilhanças dos dois modelos são proporcionais.

Este fato, provavelmente ligado ao chamado Paradoxo de Borel, traz novas dúvidas a respeito do argumento da inferência parcial e coloca em xeque a distribuição de probabilidades utilizada no Teste Exato de Fisher.

III.2 - O Critério de Decisão adotado no Teste

III.2.1 - Nível de Significância, Tamanho do Teste e Nível Descritivo

Um teste de hipótese é um problema de decisão onde a questão é localizar um determinado parâmetro θ em alguma re-

gião do espaço paramétrico Ω . Para tomar esta decisão, o estatístico baseia-se em um experimento ao qual está associado um espaço amostral X .

O espaço paramétrico é dividido em duas regiões, Ω_0 e Ω_1 de forma que $\Omega_0 \cup \Omega_1 = \Omega$ e $\Omega_0 \cap \Omega_1 = \Phi$, e o problema é colocado da seguinte forma:

$$H_0: \theta \in \Omega_0$$

$$H_1: \theta \in \Omega_1$$

Um Procedimento de Teste δ é uma regra de decisão que nos leva a aceitar H_0 ou H_1 . O procedimento nada mais é do que uma partição do espaço amostral X em dois subconjuntos:

- Região Crítica (RC): Subconjunto contendo todos os pontos amostrais que, se obtidos quando o experimento for realizado, levarão à rejeição de H_0 .
- Região de Aceitação (RA): Subconjunto contendo todos os pontos amostrais que, se observados, levarão à aceitação de H_0 .

Em suma, um procedimento de teste δ está definido para um dado experimento quando a região crítica relativa ao espaço amostral associado ao experimento estiver determinada. Em vista disso, usaremos a notação RC ou δ indiscriminadamente para denotar um procedimento de teste.

No caso da tabela de contingência 2×2 , com valores de N e m fixados o procedimento de teste estará especificado assim que for determinado o conjunto de todas as tabelas (todos os pares (x, t)) que, se obtidas, levarão à rejeição de H_0 .

A caracterização de um procedimento de teste se dá atra-

vés de sua Função Poder. A função poder específica, para cada valor de $\theta \in \Omega$, a probabilidade de que o procedimento leve à rejeição da hipótese nula.

$$\Pi(\theta, \delta) = \Pr(\text{rejeitar } H_0 | \theta, \delta), \theta \in \Omega$$

Associa-se a um procedimento de teste dois tipos de erro:

- Erro de 1a. Ordem: rejeitar H_0 quando H_0 é verdadeira.
- Erro de 2a. Ordem: não rejeitar H_0 quando H_0 é falsa.

Denota-se a probabilidade de cometer estes erros por:

- $\alpha(\delta)$ = Probabilidade de se cometer um erro de 1a. ordem.
- $\beta(\delta)$ = Probabilidade de se cometer um erro de 2a. ordem.

É fácil construir um procedimento δ para o qual $\alpha(\delta) = 0$ bastando para isto fixar uma RC vazia. Contudo, para este procedimento, $\beta(\delta) = 1$. Como existe um compromisso entre $\alpha(\delta)$ e $\beta(\delta)$, não é possível construir um teste para o qual $\alpha(\delta)$ e $\beta(\delta)$ sejam arbitrariamente pequenos. Assim, costuma-se fixar um limite superior para a probabilidade do erro de 1a. espécie, α_0 , de forma que $\alpha(\delta) \leq \alpha_0$. Este limite é conhecido como **Nível de Significância** do teste.

Considera-se como candidatos à RC todos os subconjuntos do espaço amostral cujos valores de $\Pi(\theta, \delta)$, para $\theta \in \Omega_0$, forem no máximo iguais à α_0 . Desta forma, determina-se a classe A_1 , composta pelos subconjuntos que possuem a propriedade acima:

$$A_1 = \{ \delta : \Pi(\theta, \delta) \leq \alpha_0, \theta \in \Omega_0 \}$$

Da classe A1, seleciona-se o subconjunto do espaço amostral que forneça o poder máximo para $\theta \in \Omega_1$. Este subconjunto constituirá a RC do procedimento de teste δ . O procedimento assim determinado fornecerá um valor $\beta(\delta)$ mínimo dentre aqueles cujo $\alpha(\delta) \leq \alpha_0$:

$\beta(\delta)$ é mínimo $\langle == \rangle$

$P(\text{não rejeitar } H_0 | H_1 \text{ é verdadeira})$ é mínima $\langle == \rangle$

$P(x \in RA | H_1 \text{ é verdadeira})$ é mínima $\langle == \rangle$

$1 - P(x \in RC | H_1 \text{ é verdadeira})$ é mínima $\langle == \rangle$

$P(x \in RC | H_1 \text{ é verdadeira})$ é máxima $\langle == \rangle$

$\Pi(\theta, \delta)$ é máximo para $\theta \in \Omega_1$

O Lema de Neyman-Pearson fornece uma técnica útil na busca de uma RC para a determinação de um procedimento de teste.

Para a construção de seu Teste Exato, Fisher serviu-se do critério que fixa um nível de significância α_0 e, a partir daí, busca a RC(δ) de $\alpha(\delta) \leq \alpha_0$ que tenha o mínimo $\beta(\delta)$.

Vale a pena notar que para alguns tipos de experimento, $\alpha(\delta) < \alpha_0$. Isto ocorre para os testes aos quais estão associadas distribuições de probabilidades discretas. Chamaremos a probabilidade $\alpha(\delta)$ de **Tamanho do Teste**.

O nível de significância é fixado a priori pelo estatístico com o propósito de determinar a RC do teste. Uma vez determinado este subconjunto, calcula-se o tamanho do teste, ou seja, a probabilidade de que o nosso experimento produza uma

amostra que pertença à RC, se H_0 for verdadeira. O nível de significância de um teste δ , α_0 , é uma medida usada na busca do procedimento de teste δ , ou seja, da regra de aceitação/rejeição. O tamanho, $\alpha(\delta)$ é uma característica própria do teste, que só pode ser calculada após a determinação de δ . O valor de $\alpha(\delta)$ dá uma idéia da freqüência de rejeições errôneas de H_0 para uma grande seqüência de amostras aleatórias.

Berkson (3) chamou o tamanho do teste, $\alpha(\delta)$, de α -efetivo, e o nível de significância através do qual a RC do teste foi determinada, α_0 , de α -nominal. Sua crítica ao Teste Exato de Fisher consiste basicamente no fato de o α -efetivo ser sempre inferior ao α -nominal. Segundo Berkson, o estatístico que utilizasse o procedimento do teste δ determinado por um nível de significância α_0 , estaria errando ao afirmar que os dados que obteve são significantes ao nível α_0 . O risco, na realidade, seria inferior a α_0 .

Esta crítica é infundada pois a discrepância entre $\alpha(\delta)$ e α_0 vem do fato de a distribuição utilizada no teste ser a Hipergeométrica, isto é, uma distribuição discreta. Além disso, nem o nível de significância nem o tamanho do teste fornecem uma medida da força com a qual os dados obtidos como

(3) BERKSON, J. In dispraise of the exact test. Do the marginal totals of the 2x2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, 2 :27-42, 1978

resultado de um experimento estão favorecendo H_0 ou H_1 .

Se um estatístico, ao testar uma hipótese, afirma que a rejeitou com uma significância $\alpha_0=0,05$, como se pode saber com que força esta hipótese foi rejeitada? Como se pode saber se a amostra obtida como resultado do experimento está próxima ou distante da fronteira entre a região crítica e a região de aceitação?

Necessita-se de uma medida associada à amostra obtida, para exprimir a segurança com a qual foi tomada a decisão. Afinal, o que nos deveria fazer tomar uma decisão são os dados obtidos e não o procedimento de teste. Tanto o nível de significância quanto o tamanho do teste são medidas associadas ao experimento e ao procedimento de teste, e não ao resultado obtido.

Uma medida associada à amostra obtida, (x,t) é o $P_t(x|N,m,t)$, já definido anteriormente como:

$$P_t(x|N,m,t) = \sum_{r=x}^t \frac{\binom{m}{r} \binom{N-m}{t-r}}{\binom{N}{t}}$$

$P_t(x|N,m,t)$ fornece a soma da probabilidade de obtenção do valor (x,t) sob H_0 , com as probabilidades de obtenção das tabelas mais extremas, também sob H_0 .

Uma vez realizado o experimento e obtido o resultado (x,t) , chamamos $P_t(x|N,m,t)$ de **Nível Descritivo** do teste.

Inclui-se na RC do Teste Exato de Fisher todas as amostras (x,t) para as quais o nível descritivo é inferior ao nível de significância, ou seja, $P_t(x|N,m,t) \leq \alpha_0$.

Definição III.1: Define-se como amostra limite da RC, a amostra (x_0, t_0) tal que

$$P_t(x_0|N, m, t_0) \leq \alpha_0 \text{ e } P_t(x_0-1|N, m, t_0) > \alpha_0$$

Logo, o tamanho do teste será igual ao nível descritivo se obtivermos como resultado do nosso experimento a amostra limite (x_0, t_0) . Neste caso escreveremos:

$$P_t(x_0|N, m, t_0) = \alpha(\delta)$$

Suponha que ao realizarmos o experimento, obtivemos como resultado a amostra (x, t) . Qual é o significado do nível descritivo $P_t(x|N, m, t)$?

$P_t(x|N, m, t)$ representa o mínimo valor que poderia ser assumido por α_0 , o nível de significância, para que a amostra (x, t) conduzisse à rejeição da hipótese nula. Devido a isto, o procedimento de teste elaborado por Fisher nos leva a aceitar H_0 se $P_t(x|N, m, t) > \alpha_0$ e a rejeitar H_0 se $P_t(x|N, m, t) \leq \alpha_0$. Quando rejeitamos H_0 com os dados (x, t) , estamos afirmando que rejeitamos H_0 enquanto o tamanho do teste ultrapassar $P_t(x|N, m, t)$, ou seja, enquanto a frequência relativa de rejeições errôneas de H_0 , em uma longa seqüência de experimentos for no mínimo igual à $P_t(x|N, m, t)$.

Poderíamos, então, utilizar como medida da força com a qual tomamos uma decisão baseada na amostra (x, t) , a distância entre $P_t(x|N, m, t)$ e $\alpha(\delta)$. Quanto maior for esta distância, maiores serão as evidências de que a decisão tomada foi a correta.

III.2.2 - O Uso do Nível Descritivo: $Pt(x|N,m,t)$

Quando rejeitamos ou aceitamos uma hipótese, queremos saber com que segurança estamos tomando a decisão. A metodologia de fixar um α_0 , encontrar uma RC e em seguida calcular $\alpha(\delta)$ para que posteriormente seja comparado com o $Pt(x|N,m,t)$ pode não ser a ideal.

O valor $Pt(x|N,m,t)$ não reflete apenas a força da amostra (x,t) . Ele inclui também todas as amostras mais extremas do que (x,t) , que poderiam ter sido observadas mas não o foram. Na realidade, só interessaria ao pesquisador a probabilidade de obtenção da amostra (x,t) sob H_0 . Sendo esta probabilidade "grande", a tendência será o favorecimento da hipótese nula. Não nos interessa a probabilidade de obtenção de amostras que poderiam ter sido selecionadas mas não o foram.

O exemplo dado a seguir esclarecerá este ponto de vista.

Considere duas funções densidade de probabilidade (fdp) $h_0(x)$ e $h_1(x)$ definidas como se segue:

$$h_0(x) = \begin{cases} 1 & \text{se } 0 \leq x \leq 1 \\ 0 & \text{cc} \end{cases}$$
$$h_1(x) = \begin{cases} 1/2 & \text{se } 0 \leq x \leq 10/20 \\ 5 & \text{se } 10/20 < x \leq 11/20 \\ 2 & \text{se } 20/20 \leq x \leq 25/20 \end{cases}$$

Faz-se uma única observação de uma variável aleatória x , cuja fdp, $h(x)$, é $h_0(x)$ ou $h_1(x)$. As seguintes hipóteses devem

ser testadas:

$$H_0: h(x) = h_0(x)$$

$$H_1: h(x) = h_1(x)$$

As funções $h_0(x)$ e $h_1(x)$ são esboçadas abaixo:

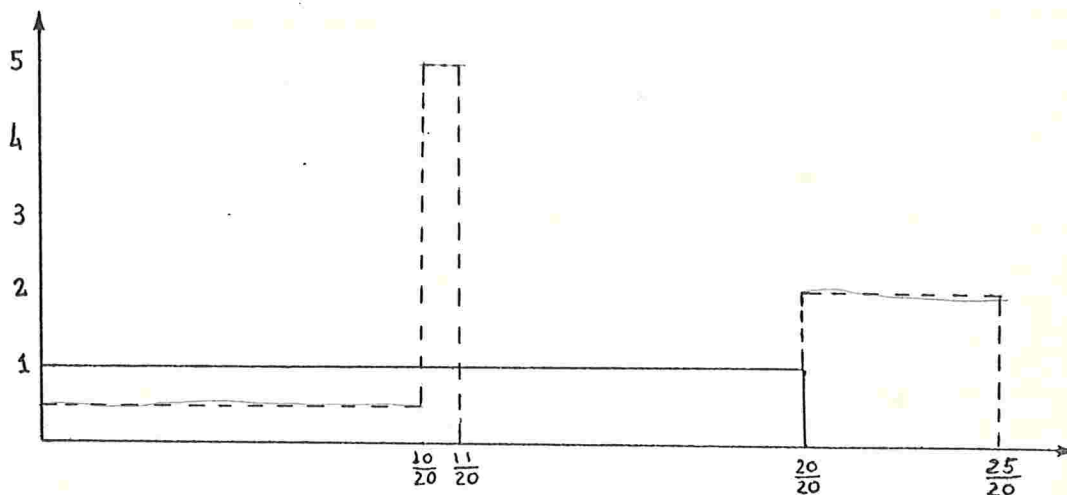


Figura III.1

Determina-se um nível de significância $\alpha_0 = 0,05$. Como as distribuições associadas ao teste são contínuas, o nível de significância coincidirá com o tamanho do teste $\alpha(\delta)$. A RC do teste é evidente:

$$RC = \{x: 10/20 \leq x \leq 11/20 \text{ ou } 20/20 \leq x \leq 25/20\}$$

$$\alpha(\delta) = P(\text{rejeitar } H_0 | H_0 \text{ verdadeira})$$

$$= P(x \in RC | h(x) = h_0(x)) = 0,05$$

Suponha que ao realizar o experimento, o cientista obtenha como resultado um valor de x tal que $20/20 \leq x \leq 25/20$ e conseqüentemente rejeite H_0 . Ao afirmar que rejeitou H_0 a um nível de significância $\alpha_0 = 5\%$, não estará fornecendo uma boa medida da força com a qual H_0 foi rejeitada. Se $20/20 \leq x \leq 25/20$, com certeza não foi cometido erro algum ao se rejeitar H_0 . Para esta amostra, a confiança é de 100%.

O nível de significância de 5% é uma característica do procedimento de teste δ . Ao considerarmos todos os resultados possíveis, o procedimento definido através da RC acima levará a uma rejeição errônea de H_0 em 5% das vezes.

Uma vez obtida a amostra, não há interesse no poder médio do teste, mas sim no poder do teste frente à amostra obtida. A confiabilidade do teste deve ser função da amostra efetivamente observada e mais do que isso, somente dela.

Uma medida adequada da confiabilidade do teste poderia ser a relação entre as funções densidade de probabilidade sob H_0 e sob H_1 para a amostra observada. Se esta relação for alta, devemos aceitar H_0 . Caso contrário H_0 deve ser rejeitada.

III.2.3- A Pré-especificação do Nível de Significância

Conforme já havíamos mencionado anteriormente, o procedimento de Fisher é determinado especificando-se um nível de significância α_0 e buscando-se o teste que, tenha $\beta(\delta)$ mínimo entre todos os procedimentos para os quais $\alpha(\delta) \leq \alpha_0$.

Tradicionalmente escolhe-se o nível de significância α_0 como sendo 0,10, 0,05 ou 0,01, dependendo da seriedade das conseqüências dos erros de 1a. e 2a. espécies. Para um determinado tamanho de amostra N , existe um compromisso entre os dois tipos de erro e ao tentarmos diminuir o $\alpha(\delta)$, poderemos aumentar o $\beta(\delta)$. Normalmente fixa-se α_0 como sendo 0,05. Se as conseqüências do erro de 1a. espécie não forem muito graves, escolhe-se $\alpha_0=0,10$. Por outro lado, toma-se $\alpha_0=0,01$

quando o cientista não quer rejeitar a hipótese nula a menos que a amostra forneça fortes evidências de que H_0 seja falsa.

Contudo, à medida que o tamanho da amostra aumenta, o compromisso entre $\alpha(\delta)$ e $\beta(\delta)$ faz com que a especificação de $\alpha_0=0,01$, por exemplo, leve a um procedimento que rejeita H_0 mesmo que se observe amostras que forneçam fortes indícios de que H_0 seja verdadeira.

O exemplo abaixo (4) esclarece a situação:

Toma-se uma amostra casual simples de tamanho n de uma distribuição de Bernoulli para a qual o valor do parâmetro p é desconhecido. As hipóteses em teste são:

$$H_0: p=0,1$$

$$H_1: p=0,5$$

A distribuição conjunta de x_1, \dots, x_n quando $p=0,1$ é dada por:

$$h_0(x) = \binom{n}{\sum_{i=1}^n x_i} (0,1)^{\sum_{i=1}^n x_i} (0,9)^{n - \sum_{i=1}^n x_i}$$

A distribuição conjunta de x_1, \dots, x_n quando $p=0,5$ é dada por:

(4) Exemplo semelhante é encontrado no livro:

DE GROOT, M.H. *Probability and Statistics*. London, Addison-Wesley, 1975. p.380-381.

$$h_1(x) = (0,5)^n$$

Para que se determine o procedimento de teste utiliza-se o Lema de Neyman-Pearson. Dessa forma, rejeita-se H_0 quando $h_1(x) > K \cdot h_0(x)$. K é determinado de forma que:

$\alpha(\delta) = \Pr(\text{rejeitar } H_0 | p=0,1) \leq 0,01$. Assim, conclui-se que entre todos os procedimentos para os quais $\alpha(\delta) \leq 0,01$, o valor $\beta(\delta)$ será mínimo para o procedimento δ^* que rejeita H_0 quando

$$\sum_{i=1}^n x_i > K'$$

onde $K' = \frac{\ln K + n \cdot \ln 9/5}{\ln 9}$

Deve-se encontrar um valor de K' de modo que

$$\Pr\left(\sum_{i=1}^n x_i > K' \mid p=0,1\right) = 0,01$$

Como a distribuição de Bernoulli é discreta, nem sempre conseguiremos encontrar um valor para K' de forma que a equação acima seja satisfeita. Contudo, tal fato não interferirá significativamente em nosso exemplo, pois determinaremos o valor de K' que torne $\alpha(\delta^*)$ próximo de 0,01.

Para cada valor de n , pode-se calcular

$$\beta(\delta^*) = \Pr\left(\sum_{i=1}^n x_i < K' \mid p=0,5\right)$$

Constrói-se, então, a seguinte tabela:

n	$\alpha(\delta^*)$	$\beta(\delta^*)$	κ'	κ
2	0,01	0,75	1	2,78
25	0,0095	$7,3 \times 10^{-3}$	6	0,2206
49	0,00800	$1,92 \times 10^{-5}$	10	0,00108

Se $n=2$, o teste δ^* dificilmente rejeita H_0 . H_0 só será rejeitada se os valores observados forem no mínimo 2,78 vezes mais prováveis sob H_1 do que sob H_0 .

Contudo, para $n=49$, H_0 será rejeitada ainda que a amostra obtida seja centenas de vezes mais provável sob H_0 do que sob H_1 pois rejeitaremos H_0 sempre que $h_1(x) > 0,00108 h_0(x)$. Isto ocorre porque o valor de $\beta(\delta^*)$ que pode ser atingido quando $n=49$ ($1,92 \times 10^{-5}$) é extremamente pequeno em relação à $\alpha(\delta^*) = 0,008$. Neste caso, δ^* torna-se bem mais cauteloso com respeito ao erro de tipo II.

Desta forma, um valor de α_0 que é apropriado para um determinado tamanho de amostra n pode ser excessivamente grande à medida que n aumenta.

Seria bem mais razoável valer-se de um critério que se preocupasse com a minimização simultânea das probabilidades dos dois tipos de erro. Em outras palavras, minimizar uma combinação linear $a\alpha(\delta) + b\beta(\delta)$ seria um procedimento mais adequado. As constantes a e b devem ser escolhidas de acordo com a gravidade dos erros de tipo I e II. Quanto mais grave for o erro, maior deverá ser a constante ligada a ele.

III.3 - O Desprezo pela Informação a Priori

Um problema só pode ser não familiar uma única vez. Em vista disso, o cientista que se envolve com um problema costuma possuir algum tipo de conhecimento a respeito do mesmo. Este conhecimento acerca do assunto em questão deveria ser levado em conta ao aplicar-se qualquer método de inferência, especificamente o Teste Exato de Fisher.

No que concerne às hipóteses, isto poderia ser feito atribuindo-se chances às veracidades da hipótese nula e da hipótese alternativa.

O conhecimento, ou a falta deste, a respeito dos parâmetros em questão, poderia ser expresso através de uma distribuição de probabilidades conferida aos mesmos com antecedência. Enfim, o teste de Fisher seria enriquecido se fossem valorizadas informações prévias sobre os estados da natureza em exame.

CAPÍTULO IV

UMA ALTERNATIVA BAYESIANA PARA O TESTE EXATO DE FISHER

IV.1 - Introdução

A filosofia Bayesiana define probabilidade como a medida do conhecimento (ou desconhecimento) que existe em torno do objeto de estudo.

Um cientista realiza um experimento com a finalidade de adquirir conhecimento a respeito de um determinado estado da natureza (parâmetro). Normalmente, seu interesse em relação a este parâmetro é despertado por algum conhecimento já existente em torno do mesmo. Na pior das hipóteses, tal conhecimento precedente é a informação de que este parâmetro existe e pode assumir "valores" em um conjunto bem definido. Isto já é suficiente para que se possa atribuir uma distribuição de probabilidades para os valores deste parâmetro. Com a reali-

zação do experimento, é obtido um resultado que "aumentará" o conhecimento do pesquisador a respeito do estado da natureza. Tal resultado vai calibrar a informação inicial do cientista dando origem a uma nova distribuição de probabilidades para os valores do parâmetro em questão. A calibragem é feita através da função de verossimilhança associada ao experimento, segundo a fórmula de Bayes.

Esta metodologia pode ser estendida para um Teste de Hipóteses, que é uma situação em que o cientista deve decidir-se pela veracidade de uma entre duas hipóteses, H_0 e H_1 , exaustivas e mutuamente exclusivas. Segundo algum conhecimento precedente que o pesquisador tenha acerca dos objetos submetidos às hipóteses, ele atribui chances iniciais para as veracidades de H_0 e de H_1 . A seguir, realiza-se um experimento cujo resultado calibra a informação inicial produzindo novas chances para as hipóteses em questão. Para que o procedimento seja coerente, os parâmetros presentes nas hipóteses em teste devem relacionar-se com o experimento realizado. Tal relação é dada pela função de verossimilhança, associada ao modelo experimental.

Na próxima seção discutiremos este procedimento com maiores detalhes. Ali, trataremos apenas do caso onde os espaços paramétricos, associados às hipóteses em estudo, possuem dimensões diferentes.

IV.2 - O Teste Bayesiano

Suponha a existência de um experimento cujo resultado aleatório, x , esteja relacionado aos parâmetros p e q através da função de verossimilhança $L(p, q|x)$. O objetivo é decidir-se pela veracidade de uma das seguintes hipóteses.

$$H_0: p=h(q)$$

$$H_1: p \neq h(q)$$

onde h é uma função de q lisa ou lisa por partes. Por função lisa entende-se uma função contínua que possua derivadas contínuas no intervalo considerado. Por função lisa por partes entende-se uma função formada por um número finito de funções, cada uma das quais é lisa.

Associados às hipóteses em teste estão os respectivos espaços paramétricos que podem ser escritos como:

$$\Omega_0 = \{(p, q); p \in D_p; q \in D_q \text{ e } p=h(q)\}$$

onde D_p = domínio de variação do parâmetro p .

D_q = domínio de variação do parâmetro q .

$$\Omega_1 = \{(p, q); p \in D_p; q \in D_q \text{ e } p \neq h(q)\}$$

Os domínios D_p e D_q são subconjuntos da reta.

Seja o parâmetro u definido como segue:

$$\mu = \begin{cases} 0 & \text{se } H_0 \text{ é verdadeira, isto é, } (p, q) \in \Omega_0 \\ 1 & \text{se } H_1 \text{ é verdadeira, isto é, } (p, q) \in \Omega_1 \end{cases}$$

A incerteza inicial que existe em torno dos parâmetros p e q é representada por uma função densidade de probabilidade conjunta $g(p, q)$, definida no conjunto $D_p \times D_q$. Da mesma forma, são atribuídas chances iniciais aos valores de p , ou seja:

$$P(\mu=0) = \xi$$

$$P(\mu=1) = 1 - \xi$$

Caso haja um desconhecimento total a respeito de (p,q) , deve-se atribuir uma distribuição de probabilidades conjunta uniforme para o par (p,q) . Assim, os valores $(p,q) \in D_p \times D_q$ serão equiprováveis. A mesma atitude deve ser tomada em relação às hipóteses. Não havendo preferências, os valores de ξ e de $1-\xi$ devem ser iguais, isto é, $\xi = 1-\xi = 1/2$.

Tendo que tomar uma decisão antes de realizar o experimento, o pesquisador deve examinar o quociente $C = \xi / (1-\xi)$, o qual será denominado **Razão de Probabilidades Complementares a Priori** ou **Escore a Favor de H_0 a Priori**. Sua decisão será a favor da hipótese cuja chance for a mais alta.

Neste ponto origina-se uma discussão delicada. Na maior parte das vezes, a distribuição de probabilidades conjunta para o par (p,q) é representada por uma densidade $g(p,q)$ sobre um subconjunto do R^2 , no caso $D_p \times D_q$. Qualquer sub-espaço de dimensão menor, como por exemplo uma curva sobre o plano $D_p \times D_q$, terá chance nula de ocorrência. Isto inclui a curva $p=h(q)$. Neste caso, como pode-se atribuir massa positiva à $\{\mu=0\}$? Não é absurda a expressão $P(p=h(q)) > 0$?

A atribuição de probabilidade positiva à hipóteses que restringem a dimensão do domínio de variação dos parâmetros não é absurda. Este fato pode ser entendido através do seguinte exemplo:

Suponha uma urna contendo duas moedas. Desta urna são feitas duas retiradas sendo que a cada retirada a moeda é

lançada, o resultado do lançamento é registrado e, finalmente, a moeda é recolocada na urna. Sejam p e q as probabilidades de obtenção de cara no primeiro e segundo lançamentos respectivamente. O desconhecimento inicial a respeito dos parâmetros p e q é simbolizado pela distribuição uniforme sobre o quadrado unitário. Assim:

$$g(p,q) = \begin{cases} 1 & \text{se } 0 \leq p \leq 1 \text{ e } 0 \leq q \leq 1 \\ 0 & \text{cc} \end{cases}$$

De acordo com a distribuição uniforme acima, qualquer curva sobre o quadrado unitário terá massa nula. As hipóteses em teste são:

$$H_0: p=q$$

$$H_1: p \neq q$$

Os espaços paramétricos associados às hipóteses são:

$$\Omega_0 = \{(p,q); 0 \leq p \leq 1; 0 \leq q \leq 1 \text{ e } p=q\}$$

$$\Omega_1 = \{(p,q); 0 \leq p \leq 1; 0 \leq q \leq 1 \text{ e } p \neq q\}$$

e podem ser esquematizados segundo a **Figura IV.1**:

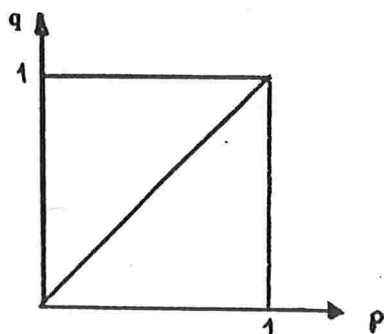


Figura IV.1: O conjunto Ω_1 é a região interna ao quadrado incluindo seus lados e excluindo a diagonal. O conjunto Ω_0 é a região formada pelos pontos interiores ao quadrado que pertencem ao segmento de reta $p=q$.

Embora a distribuição de probabilidades $g(p,q)$ confira massa nula ao segmento de reta $p=q$, qualquer pessoa concordaria em atribuir o valor $1/2$ à $P(\mu=0)$ pois esta é a chance de que a mesma moeda seja retirada nos dois sorteios. Justifica-se, então, a atribuição de uma probabilidade positiva a uma hipótese que restringe a dimensão do domínio de variação dos parâmetros em exame.

Dispondo de um experimento para aumentar seu conhecimento a respeito dos parâmetros analisados, o cientista terá interesse em comparar as probabilidades de que as hipóteses sejam verdadeiras depois de obtidos os resultados do experimento.

Chamaremos de **Escore a Favor de H_0 a Posteriori** (ou **Razão de Probabilidades Complementares a Posteriori**) ao quociente:

$$E(x) = P(\mu=0|x)/P(\mu=1|x)$$

Contudo, não existe uma relação direta entre o parâmetro μ (as hipóteses em teste) e o resultado x do experimento. A relação existente é indireta e se dá através dos parâmetros (p,q) e da função de verossimilhança $L(p,q|x)$ associada ao modelo experimental. Em consequência disso, é mais fácil que se obtenha a distribuição de probabilidades $P(\mu,p,q|x)$ em lugar da distribuição $P(\mu|x)$.

Através do raciocínio Bayesiano, deduz-se que a distribuição de probabilidades a posteriori conferida a um parâmetro é proporcional ao produto da função de verossimilhança pela distribuição a priori deste parâmetro, ou seja:

$$P(\mu, p, q|x) = \frac{P(\mu, p, q, x)}{P(x)} \propto L(\mu, p, q|x) P(p, q|\mu) P(\mu)$$

onde:

- $P(\mu, p, q|x)$ é a distribuição de probabilidades conjunta a posteriori de (μ, p, q) .
- $L(\mu, p, q|x)$ é a função de verossimilhança associada ao modelo condicionada em μ .
- $P(p, q|\mu)$ é a função densidade de probabilidade conjunta do par (p, q) dado μ .

Segundo a notação seguida até aqui:

$$P(p, q|\mu) = g(p, q|\mu)$$

- $P(\mu)$ é a distribuição de probabilidades a priori para μ

Na realidade, nosso interesse está concentrado nas chances atribuídas às hipóteses a posteriori, independentemente dos valores dos parâmetros p e q . Neste contexto pode-se considerar p e q como parâmetros nuisance e o objetivo imediato passa a ser sua eliminação.

Antes de prosseguirmos com a descrição do Teste Bayesiano, discutiremos na próxima seção uma técnica de eliminação de parâmetros nuisance.

IV.2.1 - A Técnica Bayesiana de Eliminação de Parâmetros Nuisance

A distribuição de probabilidades condicional utilizada no Teste Exato de Fisher surgiu da necessidade de eliminar-se o parâmetro nuisance do modelo. Tal distribuição origina-se a

partir do uso do Método da Redução que não é um método geral uma vez que não existe uma regra que determine a reparametrização ideal para cada modelo.

Em contrapartida, as técnicas Bayesianas, no que concernem à eliminação de parâmetros nuisance, são bastante gerais.

Assim como sob o ponto de vista Clássico, a perspectiva Bayesiana entende a função de verossimilhança $L(\lambda|x) = L(\theta, \phi|x)$ como sendo a relação existente entre os dados x e a quantidade desconhecida $\lambda = \lambda(\theta, \phi)$. A função de verossimilhança calibra a informação a priori sobre a quantidade λ , especificada em termos da distribuição a priori $\Pi(\lambda)$, transformando-a em informação a posteriori. Esta informação a posteriori é representada através da distribuição a posteriori de λ . Note que, sob o ponto de vista Bayesiano, θ e ϕ são (tecnicamente) variáveis aleatórias.

Para tanto, o modo de pensar Bayesiano baseia-se na fórmula de Bayes, dada por:

$$\Pi(\lambda|x) = \frac{\Pi(\lambda) L(\lambda|x)}{\int_{\lambda} \Pi(\lambda) L(\lambda|x) d\lambda} \propto \Pi(\lambda) L(\lambda|x)$$

Se o interesse concentra-se no subparâmetro $\theta(\lambda)$, basta que se calcule a distribuição marginal a posteriori de $\theta(\lambda)$, $\Pi(\theta|x)$, a partir de $\Pi(\lambda|x)$.

Para simplificar a determinação da distribuição a priori $\Pi(\lambda)$, escolhem-se os subparâmetros θ e ϕ de forma que:

$$\Pi(\lambda) = \Pi(\theta, \phi) = \Pi(\theta) \cdot \Pi(\phi|\theta)$$

onde:

$\Pi(\theta)$ é a distribuição marginal a priori de θ .

$\Pi(\phi|\theta)$ é a distribuição condicional de ϕ dado θ a priori.

O passo mais difícil é a seleção coerente das distribuições a priori acima. Uma vez selecionadas, obtém-se facilmente a distribuição a posteriori de θ :

$$\Pi(\theta, \phi|x) = \frac{\Pi(\theta, \phi) \cdot L(\theta, \phi|x)}{\int_{\lambda} \Pi(\lambda) L(\lambda|x) d\lambda}$$

$$\therefore \Pi(\theta|x) = \int_{\phi} \Pi(\theta, \phi|x) d\phi \propto \int_{\phi} \Pi(\theta, \phi) L(\theta, \phi|x) d\phi =$$

$$\int_{\phi} \Pi(\theta) \Pi(\phi|\theta) L(\theta, \phi|x) d\phi =$$

$$\Pi(\theta) \int_{\phi} \Pi(\phi|\theta) L(\theta, \phi|x) d\phi$$

Dessa forma as técnicas Bayesianas eliminam o parâmetro nuisance do modelo $L(\theta, \phi|x)$ através de integração sobre o mesmo (média). Uma vez eliminado ϕ , qualquer inferência sobre θ pode ser feita a partir da distribuição a posteriori $\Pi(\theta|x)$.

A discussão que surge em torno desta técnica é devida à subjetividade introduzida no modelo através das distribuições a priori, principalmente da distribuição $\Pi(\theta|\phi)$. Distribuições a priori diferentes poderão dar origem a modelos diferentes.

Enfim, a técnica Bayesiana de eliminação do parâmetro

nuisance é, em essência, bastante simples. Basta que se forneça uma distribuição precedente conjunta para os parâmetros de interesse, θ , e nuisance, ϕ . Baseando-se na função de verossimilhança, chega-se à distribuição a posteriori para θ e ϕ a qual, integrada sobre ϕ , resulta na distribuição a posteriori de θ . Neste caso, estar-se-á considerando uma espécie de média condicional do parâmetro nuisance dado o parâmetro de interesse.

Devido às dificuldades existentes na determinação de uma distribuição condicional de ϕ dado θ a priori, $\Pi(\phi|\theta)$, costuma-se considerar que, precedentemente, os parâmetros θ e ϕ são independentes. Esta independência (de parâmetros) seria a correspondência Bayesiana à afirmação Clássica de que os parâmetros θ e ϕ são de variação independente (ϕ não contém informação sobre θ e vice-versa). Contudo, em alguns casos é difícil concordar com a idéia de que θ e ϕ são independentes. Tal suposição é fortificada quando estão em jogo amostras grandes que fortalecem a verossimilhança tendendo a anular as informações a priori.

O modelo Bayesiano é mais rico do que o modelo condicional considerado no Teste Exato de Fisher, pois naquele nenhuma informação é desprezada. Assim, o modelo Bayesiano consegue extrair dos dados toda a informação contida neles. Além disso, permite a incorporação de informações precedentes que são ignoradas pela Estatística Clássica. Sendo o objetivo da Inferência Estatística extrair informações a respeito dos parâmetros em questão, seria um desperdício desprezar informa-

ções disponíveis.

IV.2.2 - Uma Adaptação da Técnica Bayesiana de Eliminação de Parâmetros Nuisance aos Testes de Hipóteses

Na construção do teste Bayesiano foi encontrado um obstáculo técnico na obtenção das probabilidades $P(\mu=0|x)$ e $P(\mu=1|x)$ a partir das distribuições $P(\mu=0, p, q, |x)$ e $P(\mu=1, p, q|x)$ respectivamente.

Inicialmente, tentaremos obter a probabilidade $P(\mu=0|x)$ a partir de $P(\mu=0, p, q|x)$. A primeira idéia que surge, inspirada no método descrito na seção anterior, é a integração de $P(\mu=0, p, q|x)$ nos parâmetros nuisance p e q . O ponto de partida é a expressão obtida anteriormente para $P(\mu=0, p, q|x)$:

$$\begin{aligned} P(\mu=0, p, q|x) &= \frac{L(\mu=0, p, q|x) \cdot P(p, q|\mu=0) \cdot P(\mu=0)}{P(x)} = \\ &= \frac{\xi \cdot L(\mu=0, p, q|x) \cdot P(p, q|\mu=0)}{P(x)} \end{aligned}$$

Conseqüentemente:

$$\begin{aligned} P(\mu=0|x) &= \int_* P(\mu=0, p, q|x) dpdq \\ &= \frac{\xi}{P(x)} \int_* L(\mu=0, p, q|x) P(p, q|\mu=0) dpdq \end{aligned}$$

onde o símbolo $*$ indica que a integral é calculada sobre os

valores possíveis para o par (p,q) .

O valor $\mu=0$ indica apenas o domínio de variação do par (p,q) , ou seja, indica que o par (p,q) pertence ao espaço paramétrico Ω_0 . Deste raciocínio decorrem as seguintes definições:

Definição IV.2.1: Define-se a função $L(\mu=0, p, q|x)$ como sendo a função de verossimilhança associada ao modelo experimental em estudo, $L(p, q|x)$, restrita ao conjunto Ω_0 ; trata-se, portanto, da função $L(p, q|x)$ onde o par (p, q) assume somente os valores para os quais $p=h(q)$.

Definição IV.2.2: Define-se a função $P(p, q|\mu=0)$ como sendo a função densidade de probabilidade conjunta para o par (p, q) , $g(p, q)$, restrita ao conjunto Ω_0 . Para que $P(p, q|\mu=0)$ continue sendo uma função densidade de probabilidade, precisamos normalizá-la e portanto definimos:

$$P(p, q|\mu=0) = \begin{cases} \frac{g(p, q)}{\int_{*} g(p, q) dp dq} & \text{para } (p, q) \in \Omega_0 \\ 0 & \text{cc} \end{cases}$$

o símbolo $*$ indica que a integral é calculada sobre os valores possíveis para o par (p, q) .

Enfim, a integral $\int_{*} L(\mu=0, p, q|x) \cdot P(p, q|\mu=0) dp dq$ nada mais

é do que a média da função de verossimilhança associada ao modelo experimental, ponderada pela distribuição a priori dos parâmetros p e q , condicionada à hipótese nula. Tal condicionamento traduz-se através do fato de que o par (p, q) percorre o espaço paramétrico Ω_0 . Esta média ponderada da verossimilhança, condicionada à hipótese nula será chamada "Função Preditiva sob a Hipótese H_0 ", e denotada por $f_0(x)$. É definida através das Integrais de Linha (1) (também ditas Integrais Curvilíneas ou Integrais de Contorno) como se segue:

Definição IV.2.3

$$f_0(x) = f(x|H_0) = \frac{\int_{\Omega_0} g(p, q) L(p, q|x) ds}{\int_{\Omega_0} g(p, q) ds}$$

onde s é o comprimento de arco definido pela função h e $\Omega_0 = \{(p, q); p \in D_p; q \in D_q \text{ e } p=h(q)\}$

Desta forma, obtém-se $P(\mu=0|x)$ a partir de $P(\mu=0, p, q|x)$ e pode-se escrever:

$$P(\mu=0|x) = \frac{\xi}{P(x)} \cdot f_0(x)$$

A obtenção de $P(\mu=1|x)$ a partir de $P(\mu=1, p, q|x)$ se faz de forma análoga:

$$P(\mu=1, p, q|x) = \frac{L(\mu=1, p, q|x) \cdot P(p, q|\mu=1) \cdot P(\mu=1)}{P(x)}$$

(1) Vide Apêndice A.

$$= \frac{(1 - \xi)}{P(x)} \cdot L(\mu=1, p, q|x) \cdot P(p, q|\mu=1)$$

Portanto:

$$P(\mu=1|x) = \int_* P(\mu=1, p, q|x) dpdq$$

$$= \frac{(1 - \xi)}{P(x)} \int_* L(u=1, p, q|x) P(p, q|u=1) dpdq$$

onde o símbolo * indica que a integral é calculada sobre os valores possíveis para o par (p,q).

Neste caso, os valores possíveis para (p,q) são indicados pelo espaço paramétrico Ω_1 uma vez que $\mu=1$.

Por analogia com o caso $\mu=0$, definimos para $\mu=1$ a "Função Preditiva sob a Hipótese H1":

Definição IV.2.4: A "Função Preditiva sob a Hipótese H1", denotada por $f_1(x)$ é dada por:

$$f_1(x) = f(x|H_1) = \frac{\int_{\Omega_1} g(p, q) L(p, q|x) dpdq}{\int_{\Omega_1} g(p, q) dpdq}$$

$$\text{onde } \Omega_1 = \{ (p, q); p \in Dp; q \in Dq; p \neq h(q) \}$$

Neste caso não se usa a Integral de Linha uma vez que o espaço Ω_1 é definido por uma região do plano e conseqüentemente a integral sobre Ω_1 é uma integral de volume. O volume sobre a curva $p=h(q)$ é nulo e portanto pode-se considerar os

limites de integração variando de modo contínuo sobre a região $D_{p \times D_q}$.

Assim:

$$P(\mu=1|x) = \frac{(1-\xi)}{P(x)} \cdot f_1(x)$$

O leitor pode estar surpreso pelo fato de não nos determos no cálculo do valor de $P(x)$. Desde que o valor de x é conhecido ao observar-se a amostra, o valor de $P(x)$ será aquele que faz com que $P(\mu=0|x) + P(\mu=1|x) = 1$. Assim, conhecidos $f_0(x)$ e $f_1(x)$, o valor de $P(x)$ é obtido trivialmente.

Neste ponto vale a pena ressaltar o seguinte lema:

Lema IV.2.1: A Função Preditiva sob a Hipótese H_i , com $i=0,1$, isto é, a função $f_i(x)$, é uma função densidade de probabilidade em x .

Prova:

1) $f_i(x)=f(x|H_i) \geq 0$ para $x \in X$ (espaço amostral do experimento)

Isto porque:

a) $g(p,q) \geq 0$ para $(p,q) \in \Omega_i$, $i=0,1$

pois $g(p,q)$ é uma função densidade de probabilidade.

b) $L(p,q|x) \geq 0$ para $x \in X$ e $(p,q) \in \Omega_i$, $i=0,1$

pois $L(p,q|x)$ é uma função densidade de probabilidade.

$$2) \sum_{x \in X} f_i(x) = \sum_{x \in X} \frac{\int_{\Omega_i} g(p,q) L(p,q|x) ds}{\int_{\Omega_i} g(p,q) ds} =$$

$$\begin{aligned}
 (2) \quad & \frac{\int_{\Omega_i} g(p,q) \sum_{x \in X} L(p,q|x) \, ds}{\int_{\Omega_i} g(p,q) \, ds} \\
 & = \frac{\int_{\Omega_i} g(p,q) \, ds}{\int_{\Omega_i} g(p,q) \, ds} = 1 \quad i=0,1
 \end{aligned}$$

onde ds indica que o par (p,q) percorre a curva ou a região determinada pelo espaço paramétrico Ω_i .

Estamos agora em condições de obter o Escore a Favor de H_0 a Posteriori, dado por:

$$\zeta(x) = P(\mu=0|x)/P(\mu=1|x) = \xi \cdot f_0(x) / (1-\xi) \cdot f_1(x)$$

O Escore a Posteriori corresponde ao produto do Escore a Priori pela razão das Funções Preditivas sob H_0 e H_1 respectivamente, ou seja, pela razão das verossimilhanças médias sob H_0 e sob H_1 .

Geralmente, quando $\zeta(x) < 1$, isto é, quando H_0 é menos provável do que H_1 depois de observado o resultado do experi-

(2) Aqui consideramos que as funções em jogo obedecem a condições de regularidade de forma que podemos permutar os símbolos \int e \sum pois a integral é calculada sobre o par (p,q) e a somatória é calculada sobre os valores de x .

mento, a decisão é contrária a H_0 . Em contrapartida, quando $\hat{C}(x) \geq 1$, a decisão favorece H_0 .

Contudo, outro critério de decisão pode ser definido, escolhendo-se uma constante c tal que, se $\hat{C}(x) < c$ rejeita-se H_0 ; caso contrário, $\hat{C}(x) \geq c$, aceita-se H_0 .

Antes de analisar os critérios de decisão com maiores detalhes, vamos exemplificar o cálculo do Escore a Favor de H_0 a posteriori, valendo-nos da tabela 2×2 .

IV.3 - Exemplo : A Tabela 2×2

As técnicas Bayesianas serão agora aplicadas ao problema da tabela 2×2 :

	A	A ^c	Total
B	x	m-x	m
B ^c	y	n-y	n
Total	t	N-t	N

Conforme discutiu-se anteriormente, pode-se utilizar o mesmo teste tanto para o modelo Multinomial quanto para o modelo Binomial. Considerando-se o modelo Binomial, o objetivo passa a ser a comparação das duas proporções, p e q , levando-se em conta a observação de $N=m+n$ variáveis aleatórias independentes de Bernoulli onde m têm p como parâmetro e n têm q como parâmetro. As hipóteses em teste são:

$H_0: p=q$

$H_1: p \neq q$

As chances inicialmente atribuídas à H_0 e H_1 são respectivamente ξ e $(1-\xi)$, isto é, $P(p=q)=\xi$ e $P(p \neq q)=1-\xi$. Para efeito de generalização considerou-se o teste bicaudal. Note que, neste caso, não existe interesse na determinação dos valores de p e q que serão tratados como informação supérflua. O escore a priori é $C = \xi / (1-\xi)$. O resultado do experimento é dado pelo par (x, y) e portanto o escore a posteriori é $C(x, y) = C \cdot f_0(x, y) / f_1(x, y)$. A verossimilhança associada ao modelo Binomial escreve-se como:

$$L(p, q | x, y) = \binom{m}{x} p^x (1-p)^{m-x} \binom{n}{y} q^y (1-q)^{n-y}$$

O primeiro obstáculo a ser vencido surge na determinação da distribuição conjunta a priori para os parâmetros p e q . Por motivos operacionais, um estatístico que enfoque o problema sob o ponto de vista Bayesiano, procura exprimir sua opinião precedente a respeito dos parâmetros em estudo através de uma distribuição que pertença à Classe das Distribuições Conjugadas em relação à função de verossimilhança considerada. Este procedimento favorece a obtenção de expressões analíticas.

Definição IV.3.1: Seja uma função de verossimilhança L , a qual depende de uma n -upla de parâmetros $(\theta_1, \theta_2, \dots, \theta_n)$. A esta n -upla é atribuída uma distribuição de probabili-

dades conjunta a priori, Π .

Uma família de distribuições de probabilidade é chamada de Classe Conjugada de Distribuições em relação a L se o fato de Π pertencer a esta classe implicar a pertinência da distribuição de probabilidades conjunta a posteriori da n -upla $(\theta_1, \theta_2, \dots, \theta_n)$ a esta classe.

Portanto procura-se atribuir ao par (p, q) uma distribuição de probabilidades conjunta a priori que pertença à Classe Conjugada de Distribuições em relação à verossimilhança adotada.

Sabe-se que a distribuição conjunta a posteriori de (p, q) é proporcional ao produto da distribuição conjunta a priori pela função de verossimilhança. Ao normalizar-se este produto, verifica-se que, se a distribuição a priori do par (p, q) for uma Beta, a distribuição a posteriori também o será. Desta discussão decorre o seguinte lema:

Lema IV.3.1: Considere a função de verossimilhança:

$$L(p, q | x, y) = \binom{m}{x} p^x (1-p)^{(m-x)} \binom{n}{y} q^y (1-q)^{(n-y)}$$

Assuma que os parâmetros p e q são "independentes" (o valor de p não depende do valor de q e vice-versa) e distribuídos segundo uma função densidade de probabilidade Beta:

$$p \sim B(a_1; b_1)$$

$$q \sim B(a_2; b_2)$$

isto é:

$$g_1(p) = \frac{1}{\beta(a_1, b_1)} p^{a_1-1} (1-p)^{b_1-1} \quad (3)$$

$$g_2(q) = \frac{1}{\beta(a_2, b_2)} q^{a_2-1} (1-q)^{b_2-1}$$

Conseqüentemente a distribuição conjunta $g(p, q)$ é dada por:

$$g(p, q) = g_1(p) \cdot g_2(q) = \frac{1}{\beta(a_1, b_1) \beta(a_2, b_2)} p^{a_1-1} (1-p)^{b_1-1} q^{a_2-1} (1-q)^{b_2-1}$$

Sob estas assertivas, a distribuição conjunta a posteriori do par (p, q) é dada através das condições:

$$p \amalg q | (x, y) \quad (4)$$

$$p | (x, y) \sim p | x \sim B(x+a_1; m-x+b_1)$$

$$q | (x, y) \sim q | y \sim B(y+a_2; n-y+b_2)$$

$$g(p, q | x, y) = \frac{p^{x+a_1-1} (1-p)^{m-x+b_1-1} q^{y+a_2-1} (1-q)^{n-y+b_2-1}}{\beta(x+a_1, m-x+b_1) \beta(y+a_2, n-y+b_2)}$$

Finalmente tem-se condições para enfrentar o terceiro aspecto do problema: as hipóteses em teste. O espaço paramétrico associado à hipótese alternativa é representado através do seguinte conjunto:

(3) A constante $\beta(a, b)$ é a função euleriana Beta avaliada no ponto (a, b) e definida por:

$$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad \text{para } a > 0 \text{ e } b > 0$$

(4) A notação \amalg indica independência

$$\Omega_1 = \{ (p, q); 0 \leq p \leq 1; 0 \leq q \leq 1; p \neq q \}$$

Quando a hipótese nula é verdadeira, o espaço paramétrico reduz-se ao conjunto:

$$\Omega_0 = \{ (p, q); 0 \leq p \leq 1; p = q \}$$

Tais conjuntos estão esquematizados na Figura IV.1

A diferença entre estes espaços paramétricos manifesta-se no cálculo das Funções Preditivas. Como o campo de variação dos parâmetros sob H_0 difere do campo de variação sob H_1 as médias das verossimilhanças sob H_0 e sob H_1 também diferem. Sob H_1 a média é tomada sobre todo o quadrado enquanto que sob H_0 esta média é tomada apenas sobre o segmento de reta.

IV.3.1 - Cálculo da Função Preditiva sob H_1

$$f_1(x, y) = f(x, y | H_1) = \frac{\int_{\Omega_1} g(p, q) L(p, q | x, y) dp dq}{\int_{\Omega_1} g(p, q) dp dq}$$

Calculando o numerador, vem:

$$\int_{\Omega_1} g(p, q) L(p, q | x, y) dp dq = \quad (5)$$

$$\int_0^1 \int_0^1 \frac{\binom{m}{x} \binom{n}{y} p^{x+a_1-1} (1-p)^{m-x+b_1-1} q^{y+a_1-1} (1-q)^{n-y+b_2-1}}{\beta(a_1, b_1) \beta(a_2, b_2)} dp dq$$

(5) A integral sobre Ω_1 é uma integral de volume. O volume sobre o segmento de reta $p=q$ é nulo e portanto pode-se considerar os limites de integração variando de modo contínuo no intervalo $[0; 1]$

$$= \frac{\binom{m}{x} \binom{n}{y} \beta(x+a_1, m-x+b_1) \beta(y+a_2, n-y+b_2)}{\beta(a_1, b_1) \beta(a_2, b_2)}$$

O cálculo do denominador é dado por:

$$\int_{\Omega_1} g(p, q) dp dq = \int_0^1 \int_0^1 g(p, q) dp dq = 1$$

Portanto:

$$f_1(x, y) = \frac{\binom{m}{x} \binom{n}{y} \beta(x+a_1, m-x+b_1) \beta(y+a_2, n-y+b_2)}{\beta(a_1, b_1) \beta(a_2, b_2)}$$

Para distribuições a priori simétricas, isto é, $a_1=a_2=a$ e $b_1=b_2=b$:

$$f_1(x, y) = \frac{\binom{m}{x} \binom{n}{y} \beta(x+a, m-x+b) \beta(y+a, n-y+b)}{\beta^2(a, b)}$$

Para prioris uniformes, ou seja $a=b=1$:

$$f_1(x, y) = \frac{\binom{m}{x} \binom{n}{y} \beta(x+1, m-x+1) \beta(y+1, n-y+1)}{\beta^2(1, 1)}$$

$$f_1(x, y) = \frac{1}{(m+1) \cdot (n+1)}$$

É interessante notar que $f_1(x, y)$ é uma constante, isto é, não varia com os pontos (x, y) . Este fato não deve surpreender o leitor pois a seleção de uma distribuição uniforme como priori para p (e para q) indica que não existe preferência por nenhum valor possível de p (e de q). Isto implica uma ausência de preferência sobre os valores de x e y que é descri-

ta através da equiprobabilidade destes valores. Finalmente, como x e y são condicionalmente independentes dado (p,q) e p e q são independentes, temos que x e y são (incondicionalmente) independentes. Esta última conclusão é confirmada pela função $f_1(x,y)$ descrita acima.

IV.3.2 - Cálculo da Função Preditiva sob H_0

$$f_0(x,y) = f(x,y|H_0) = \frac{\int_{\Omega_0} g(p,q) L(p,q|x,y) ds}{\int_{\Omega_0} g(p,q) ds}$$

Aqui, o comprimento de arco s é dado por meio das equações:

$$p=p \quad 0 \leq p \leq 1$$

$$q=h(p)=p$$

Calculando o numerador:

$$\begin{aligned} \int_{\Omega_0} g(p,q) L(p,q|x,y) ds &= \int_0^1 g(p,h(p)) \cdot L(p,h(p)|x,y) \sqrt{1 + \left(\frac{dh}{dp}\right)^2} dp = \\ &= \int_0^1 \frac{\binom{m}{x} \binom{n}{y} p^{(x+y+a_1+a_2-2)} (1-p)^{(N-(x+y)+b_1+b_2-2)} \sqrt{2}}{\beta(a_1,b_1) \beta(a_2,b_2)} dp = \\ &= \frac{\sqrt{2} \binom{m}{x} \binom{n}{y} \beta(t+a_1+a_2-1, N-t+b_1+b_2-1)}{\beta(a_1,b_1) \beta(a_2,b_2)} \end{aligned}$$

onde $t=(x+y)$

Calculando o denominador:

$$\int_{\Omega_0} g(p,q) ds = \int_0^1 g(p,h(p)) \cdot \sqrt{1 + \left(\frac{dh}{dp}\right)^2} dp =$$

$$\int_0^1 \frac{p^{(a_1+a_2-2)} (1-p)^{(b_1+b_2-2)} \sqrt{2}}{\beta(a_1, b_1) \beta(a_2, b_2)} dp =$$

$$\frac{\sqrt{2} \beta(a_1+a_2-1, b_1+b_2-1)}{\beta(a_1, b_1) \beta(a_2, b_2)}$$

Portanto:

$$f_o(x, y) = f(x, y | H_0) = \frac{\binom{m}{x} \binom{n}{y} \beta(t+a_1+a_2-1, N-t+b_1+b_2-1)}{\beta(a_1+a_2-1, b_1+b_2-1)}$$

Para distribuições a priori simétricas onde $a_1=a_2=a$ e $b_1=b_2=b$

$$f_o(x, y) = \frac{\binom{m}{x} \binom{n}{y} \beta(t+2a-1, N-t+2b-1)}{\beta(2a-1, 2b-1)}$$

No caso de distribuições uniformes, isto é, $a=b=1$

$$f_o(x, y) = \frac{\binom{m}{x} \binom{n}{y} \beta(t+1, N-t+1)}{\beta(1, 1)}$$

$$f_o(x, y) = \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{1}{(N+1)} \quad \text{onde } t=x+y$$

IV.3.3 - Comparação da Distribuição de Probabilidades Adotada no Teste Exato de Fisher com a Função Preditiva sob H_0

Tanto a distribuição de probabilidades adotada no Teste Exato de Fisher, $P_f(x|N, m, t)$, (que também pode ser escrita como $P_f(x|N, m, y)$), quanto a Função Preditiva sob H_0 , $f_o(x, y)$, fornecem uma medida da probabilidade de obtenção dos dados (x, y) supondo H_0 verdadeira.

No problema da tabela 2×2 , considerando distribuições a

priori uniformes para o teste Bayesiano, temos:

$$Pf(x|N, m, y) = Pf(x|N, m, t) = \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}}$$

$$fo(x, y) = fo(x, t) = \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{1}{(N+1)}$$

As duas funções diferem pelo fator $1/(N+1)$. Esta diferença decorre do fato de o Teste de Fisher considerar o total marginal t fixado. $(N+1)$ indica o número de valores distintos que t pode assumir $(0, 1, \dots, N)$. Como tomou-se distribuições a priori uniformes para os parâmetros p e q , os valores que t pode assumir são equiprováveis; assim cada um deles tem probabilidade de ocorrência $1/(N+1)$. Daí o aparecimento do fator $1/(N+1)$ em $fo(x, t)$, ou seja:

$$fo(x, t) = Pf(x|N, m, t) \cdot Pf(t|N, m) = Pf(x|N, m, t) \cdot 1/(N+1)$$

O fator $1/(N+1)$ é uma constante de proporcionalidade, ou seja, mantém-se constante a medida que o par (x, y) varia. Isto significa que, se fossem utilizadas da mesma forma, as funções $fo(x, t)$ e $Pf(x|N, m, t)$ levariam a inferências iguais a respeito da tabela 2×2 (Princípio da Verossimilhança). A diferença entre o Teste Exato de Fisher e o Teste Bayesiano é decorrente dos usos distintos que são feitos das funções $fo(x, t)$ e $Pf(x|N, m, t)$.

IV.3.4 - O Escore a Favor de H_0 a Posteriori - $E(x, y)$

Conforme foi visto na seção IV.2.2, o Escore a Favor de

Ho a Posteriori é dado por:

$$\epsilon(x) = P(\mu=0|x) / P(\mu=1|x) = \xi \cdot f_0(x) / (1-\xi) f_1(x)$$

No caso da tabela 2x2, teremos:

$$\epsilon(x,y) = \xi \cdot f_0(x,y) / (1-\xi) \cdot f_1(x,y) = \epsilon \cdot f_0(x,y) / f_1(x,y)$$

$$= \epsilon \cdot \frac{\beta(t+a_1+a_2-1, N-t+b_1+b_2-1) \beta(a_1, b_1) \beta(a_2, b_2)}{\beta(a_1+a_2-1, b_1+b_2-1) \beta(x+a_1, m-x+b_1) \beta(y+a_2, n-y+b_2)}$$

Para prioris simétricas, onde $a_1=a_2=a$ e $b_1=b_2=b$:

$$\epsilon(x,y) = \frac{\epsilon \cdot \beta(t+2a-1, N-t+2b-1) \cdot \beta^2(a, b)}{\beta(2a-1, 2b-1) \cdot \beta(x+a, m-x+b) \cdot \beta(y+a, n-y+b)}$$

Para prioris uniformes, isto é, $a=b=1$:

$$\epsilon(x,y) = \epsilon \cdot \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{(m+1)(n+1)}{(N+1)}$$

IV.4 - O Critério de Decisão Adotado no Teste Bayesiano

IV.4.1 - Erros de 1a. e 2a. Espécies

Como já foi mencionado em seção anterior, o critério de decisão adotado no teste Bayesiano baseia-se no Escore a Favor de Ho a Posteriori e, genericamente é dado por:

$$\epsilon(x) < c \Rightarrow \text{rejeita-se } H_0$$

$\bar{C}(x) \geq c \Rightarrow$ não se rejeita H_0

Assim, dada uma constante c , existe um conjunto $RC = \{x; \bar{C}(x) < c\}$ de pontos amostrais que levam à rejeição de H_0 . Este conjunto nada mais é do que a região crítica do teste. Portanto, no caso do teste Bayesiano, não é necessário que se forneça o nível de significância α para que o procedimento do teste δ_B , ou seja, a região crítica do teste seja determinada.

Uma vez obtida a região crítica do teste Bayesiano (δ_B), pode-se calcular facilmente as probabilidades dos erros de 1a. e 2a. espécies, isto é, $\alpha(\delta_B)$ e $\beta(\delta_B)$.

A probabilidade do erro de 1a. espécie, ou seja, o tamanho do teste é, por definição, a probabilidade de que H_0 seja rejeitada dado que H_0 é verdadeira. A probabilidade do erro de 2a. espécie é a probabilidade de que H_0 não seja rejeitada sendo que H_0 é falsa. Vimos na seção IV.2 que a Função Preditiva sob H_i , isto é, $f_i(x)$ é a média da função densidade de probabilidade da variável aleatória x condicionada à hipótese H_i , ou seja, calculada sobre todos os valores de p e q que obedecem à hipótese H_i . Conseqüentemente podemos definir:

Definição IV.4.1: O tamanho do teste δ_B ,

$\alpha(\delta_B) = \Pr(\text{rejeitar } H_0 | H_0 \text{ é verdadeira})$ é dado por:

$$\alpha(\delta_B) = \sum_{x \in RC} f_0(x)$$

Definição IV.4.2: A probabilidade de erro de 2a. espécie num teste δ_B , $\beta(\delta_B) = \Pr(\text{não rejeitar } H_0 | H_0 \text{ é falsa})$ é

dada por:

$$\beta(\delta_B) = 1 - \sum_{x \in R_C} f_1(x)$$

IV.4.2 - O Teste Bayesiano minimiza a Combinação Linear a $\alpha(\delta_B)$ + b $\beta(\delta_B)$

O Teste Bayesiano é um procedimento tal que rejeita H_0 se $\epsilon(x) < c$ e aceita H_0 se $\epsilon(x) \geq c$

Teorema IV.4.1: O procedimento Bayesiano δ_B minimiza a combinação linear $a\alpha(\delta) + b\beta(\delta)$, onde a e b são duas constantes fixadas.

$$\begin{aligned} \text{Demonstração (6): } a\alpha(\delta_B) + b\beta(\delta_B) &= a \sum_{x \in R_C} f_0(x) + b \left[1 - \sum_{x \in R_C} f_1(x) \right] = \\ &= b + \sum_{x \in R_C} [a \cdot f_0(x) - b \cdot f_1(x)] \end{aligned}$$

A combinação linear $a\alpha(\delta_B) + b\beta(\delta_B)$ será tanto menor quanto menor for a somatória $\sum_{x \in R_C} [a f_0(x) - b f_1(x)]$. Para que esta soma seja a menor possível, ela deve incluir todos os pontos x para os quais $a f_0(x) - b f_1(x) < 0$ e excluir qualquer ponto x para o qual $a f_0(x) - b f_1(x) > 0$. Será indi-

(6) Esta demonstração é baseada na prova citada por:
DE GROOT, M.H. Probability and Statistics. London,
Addison-Wesley, 1975. p.374-375

ferente incluir ou não os pontos para os quais $a f_0(x) - b f_1(x) = 0$. Logo, a combinação linear $a \alpha(\delta_B) + b \beta(\delta_B)$ será minimizada se a Região Crítica do teste δ_B contiver os pontos amostrais para os quais $a f_0(x) - b f_1(x) < 0$, ou seja, os pontos amostrais tais que: $f_0(x)/f_1(x) < b/a$.

O teste Bayesiano é aquele cuja Região Crítica contém os pontos amostrais para os quais:

$$\xi \cdot f_0(x) / (1 - \xi) \cdot f_1(x) < c \Rightarrow f_0(x) / f_1(x) < c \cdot (1 - \xi) / \xi.$$

Assim, basta escolher c de forma que $c = b \cdot \xi / a(1 - \xi)$.

Neste caso, o teste Bayesiano estará minimizando $a \alpha(\delta_B) + b \beta(\delta_B)$ quaisquer que sejam a e $b \in \mathbb{R}$.

IV.4.3 - A Escolha da Constante c

Como foi visto na seção anterior, a escolha da constante c está intimamente ligada aos valores de a e b , e deverá ser feita de acordo com a importância que dermos aos erros de 1ª. e 2ª. espécies.

Suponha que no processo de teste de hipóteses estejam envolvidas perdas, sendo w_0 a perda sofrida quando se comete o erro de tipo I e w_1 a perda sofrida quando se comete o erro de tipo II. A perda é nula quando a decisão correta é tomada. Suponha, ainda, que $P(\mu=0)=\xi$ e $P(\mu=1)=1-\xi$. Desta forma, a perda média para um procedimento de teste é dada por:

$$E[L(\delta)] = P(\mu=0) \cdot E[L(\delta)|\mu=0] + P(\mu=1) \cdot E[L(\delta)|\mu=1]$$

$$E[L(\delta)|\mu=0] = w_0 \alpha(\delta)$$

$$E[L(\delta)|\mu=1] = w_1 \cdot \beta(\delta)$$

Assim:

$$E[L(\delta)] = \xi \cdot w_0 \cdot \alpha(\delta) + (1 - \xi) \cdot w_1 \cdot \beta(\delta)$$

É evidente que o desejo de qualquer pesquisador seria a obtenção de um procedimento de teste que torne a perda média, $E[L(\delta)]$, mínima. O teste Bayesiano onde $c = (1 - \xi)w_1 / \xi w_0$ é um teste deste tipo.

Se for dada a mesma perda a ambos os erros ($w_0 = w_1$) e se $P(\mu=0) = P(\mu=1) = 1/2$, o valor da constante c deverá ser 1 e o teste Bayesiano estará minimizando a combinação linear $\alpha(\delta_B) + \beta(\delta_B)$.

Em suma, a constante c informa quantas vezes H_0 deverá ser mais provável do que H_1 , depois de observado o resultado do experimento, para que a decisão seja favorável à H_0 .

No próximo capítulo será feita uma comparação entre o teste Bayesiano desenvolvido neste capítulo e o Teste Exato de Fisher. Tal comparação dar-se-á através de amostras simuladas. Como os métodos clássicos não utilizam informações prévias, a comparação só terá sentido se forem utilizadas distribuições a priori não informativas. Caso contrário, poder-se-ia selecionar prioris favoráveis aos valores simulados de p e q e assim minimizar a probabilidade de erro no critério Bayesiano. Evidentemente isto nos levaria a super valorizar o método Bayesiano.

É muito importante que o leitor perceba que o teste aqui

construído tem um caráter geral. A técnica utilizada pode ser facilmente estendida para tabelas de contingência com dimensões e número de categorias maiores. Mesmo na teoria normal de comparação de médias (Análise de Variância) esta técnica pode ser utilizada de maneira direta. As únicas mudanças exigidas para este fim são com relação às distribuições de probabilidades a priori dadas às entidades envolvidas no processo.

CAPÍTULO V,

CONFRONTO: TESTE EXATO DE FISHER x TESTE BAYESIANO

V.1 - Introdução

O objetivo deste capítulo não é discutir a validade da filosofia Clássica ou da Bayesiana, embora, neste ponto, o leitor ainda possa estar em dúvida, sem saber se deve utilizar o Teste Exato de Fisher ou o Teste Bayesiano. Para solucionar esta questão vamos verificar objetivamente qual dos dois é o melhor teste.

Antes de mais nada, devemos definir um critério segundo o qual um teste será considerado melhor do que outro. A intuição nos diz que um teste é tanto melhor quanto menor for a probabilidade de que o pesquisador erre ao utilizá-lo.

O teste da hipótese nula, H_0 , contra a alternativa, H_1 , pressupõe duas atitudes: a decisão em favor de H_0 e a decisão

em favor de H_1 . Assim, os erros passíveis de serem cometidos num procedimento de teste de hipóteses são a decisão errônea em favor de H_1 (erro de 1a. espécie) e a decisão errônea em favor de H_0 (erro de 2a. espécie).

Uma medida tradicionalmente aceita para avaliação de testes de hipóteses é a função poder associada a eles. Como já foi visto em capítulo anterior, a função poder associada a um procedimento de teste δ é dada por:

$$\Pi(\mu, \delta) = \Pr(\text{rejeitar } H_0 | \mu, \delta) \quad \mu \in \Omega$$

Um procedimento de teste é bom quando a função poder associada a ele assume valores altos para $\mu \in \Omega_1$ e valores baixos para $\mu \in \Omega_0$.

Uma outra forma sob a qual pode-se avaliar a qualidade de um procedimento δ é através das probabilidades de que sejam cometidos os erros de 1a. e 2a. espécies. Seja:

$$\begin{aligned} \text{ALFA}(\delta) &= \Pr(\text{Rejeitar } H_0 | \mu \in \Omega_0, \delta) \\ \text{BETA}(\delta) &= \Pr(\text{Aceitar } H_0 | \mu \in \Omega_1, \delta) \end{aligned} \quad (1)$$

Diremos que o procedimento δ_1 é melhor do que o procedimento δ_2 se e somente se:

$$\text{ALFA}(\delta_1) + \text{BETA}(\delta_1) < \text{ALFA}(\delta_2) + \text{BETA}(\delta_2)$$

(1) Aqui, a notação difere da notação dos capítulos anteriores porque não é oportuno discutir se o ALFA considerado é o Tamanho do Teste ou o Nível de Significância do Teste.

Existem circunstâncias sob as quais um tipo de erro é mais grave de que outro, principalmente quando o experimento em questão envolve perdas (custo proveniente da tomada de uma decisão incorreta). Neste caso, o teste será tanto melhor quanto menor for a perda média (ou maior for o ganho médio) que se venha a sofrer ao aplicá-lo. Para exprimir as perdas relacionadas aos erros, costuma-se incluir coeficientes (w_0 e w_1) na combinação linear do ALFA e BETA. Generalizando, diremos que o procedimento δ_1 é melhor do que o procedimento δ_2 se e somente se:

$$w_0.ALFA(\delta_1) + w_1.BETA(\delta_1) < w_0.ALFA(\delta_2) + w_1.BETA(\delta_2)$$

Uma vez definido o critério para comparação dos testes, podemos confrontá-los. Para tanto, devemos, antes de mais nada, submetê-los às mesmas condições. Cabe aqui ressaltar que, sendo a metodologia envolvida no teste Bayesiano diferente daquela envolvida no Teste Exato de Fisher, a comparação não é óbvia.

O primeiro problema com o qual nos deparamos diz respeito às informações precedentes que temos acerca do fenômeno analisado. O teste Bayesiano incorpora os resultados experimentais ao estado de conhecimento inicial do fenômeno, representado por distribuições de probabilidades dadas aos parâmetros p e q e às veracidades das hipóteses H_0 e H_1 . Já o teste Exato não pressupõe informação precedente a respeito do experimento valendo-se apenas dos resultados experimentais. Para solucionar este problema, utilizaremos para o teste Bayesiano distribuições de probabilidades a priori não informativas.

Desta forma estaremos evitando a seleção de distribuições que ao fornecerem informação sobre os valores de p e q , minimizariam a probabilidade de erro do critério Bayesiano.

Assim, tomaremos como distribuições a priori:

$$g(p, q) = \begin{cases} 1 & 0 \leq p \leq 1 \text{ e } 0 \leq q \leq 1 \\ 0 & \text{cc} \end{cases}$$

$$P(u=0) = \xi = 1/2$$

$$P(u=1) = 1 - \xi = 1/2$$

O segundo problema vem do fato de o teste Bayesiano envolver perdas, w_0 e w_1 , associadas aos erros de 1a. e 2a. espécies respectivamente. Em contrapartida, o teste de Fisher não se vale das perdas para indicar preferências com relação aos erros. Para tornar os testes equivalentes sob este ponto de vista, assumiremos perdas iguais para os dois tipos de erros no teste Bayesiano ($w_0 = w_1$).

A solução dos dois problemas anteriores determina o valor da constante c no teste de Bayes: $c = w_1(1 - \xi)/w_0\xi = 1$

Tomando $c=1$ e empregando o teste Bayesiano, estaremos, em teoria, minimizando a combinação linear:

$$ALFA(\delta) + BETA(\delta)$$

Finalmente estamos em condições de enfrentar o terceiro obstáculo. O teste Bayesiano determina a região crítica, ou seja, o critério de aceitação/rejeição da hipótese H_0 , com base na constante c , cujo valor é definido a partir de premissas adotadas em relação ao experimento analisado. Dado o valor de c , a região crítica do teste de Bayes é constituída pelos valores de (x, y) tais que $C(x, y) < c$ enquanto que a

região de aceitação é dada pelos valores de (x,y) tais que $\epsilon(x,y) \geq c$. O ALFA e o BETA do procedimento são calculados a partir da região crítica do teste. Por outro lado, o teste Exato pressupõe o pré-estabelecimento de um valor para ALFA. Com base no valor de ALFA é construída a região crítica associada ao procedimento. Como solução para este problema foi adotado o seguinte esquema: estabelece-se a região crítica associada ao teste de Bayes tomando-se $c=1$; em seguida, calcula-se os valores ALFA (ALFABT) e BETA (BETABT) para o procedimento resultante. A partir do valor ALFABT determina-se a região crítica referente ao Teste Exato de Fisher. Com este esquema, pode-se dizer que os dois testes estão baseados no mesmo nível de significância devendo ter tamanhos próximos.

Uma vez de acordo com a solução adotada para os três problemas acima mencionados, diremos que os testes encontram-se sob as mesmas condições e podem ser comparados. O teste que efetivamente fornecer a menor probabilidade de que sejam cometidos erros, ou seja, o menor valor para $ALFA + BETA$, será o melhor.

V.2 - Os Procedimentos de Teste

Considere a tabela 2x2:

	A	A ^c	Total
B	x	m-x	m
B ^c	y	n-y	n
Total	t	N-t	N

Para cada tamanho de amostra, ou seja, para cada par (m,n) , o teste de Bayes e o teste Exato definem regiões críticas e regiões de aceitação distintas. Vamos esquematizá-las por meio de uma **Tábua de Decisão** na qual os valores de x estarão representados no eixo horizontal (colunas) e os valores de y no eixo vertical (linhas). O valor 1 indica que o par (x,y) pertence à região de aceitação (RA) e o valor 0 indica que o par (x,y) pertence à região crítica (RC).

	TÁBUA DE DECISÃO							
	0	1	2	3	.	.	.	m
0	1	1	0	0	.	.	.	0
1	1	1	1	0	.	.	.	0
2	0	1	1	1	.	.	.	0
3	0	0	1	1	.	.	.	0
.				
.				
.				
n								1

Figura V.1

V.2.1 - A Tábua de Decisão Bayesiana

O teste de Bayes utilizado para o confronto com o Teste Exato de Fisher foi construído de forma que a constante c tenha valor 1. O procedimento de teste, δ_B , é dado por:

$$(x,y) \in RC \Leftrightarrow C(x,y) = \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{(m+1)(n+1)}{(N+1)} < 1$$

$$(x, y) \in RA \Leftrightarrow \zeta(x, y) = \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{(m+1)(n+1)}{(N+1)} \gg 1$$

onde $t=x+y$ e $N=m+n$

As expressões acima equivalem a:

$$(x, y) \in RC \Leftrightarrow \frac{t!(N-t)!(m+1)!(n+1)!}{x!y!(m-x)!(n-y)!(N+1)!} < 1$$

$$(x, y) \in RA \Leftrightarrow \frac{t!(N-t)!(m+1)!(n+1)!}{x!y!(m-x)!(n-y)!(N+1)!} \gg 1$$

Apresentamos abaixo um exemplo de cálculo de Tábua de Decisão para o teste Bayesiano onde $m=n=5$

	MATBAYES					
	0	1	2	3	4	5
	↓	↓	↓	↓	↓	↓
0→	1	1	0	0	0	0
1→	1	1	1	0	0	0
2→	0	1	1	1	0	0
3→	0	0	1	1	1	0
4→	0	0	0	1	1	1
5→	0	0	0	0	1	1

Figura V.2

Legenda: corpo da tábua = 1 $\Rightarrow (x, y) \in RA$

corpo da tábua = 0 $\Rightarrow (x, y) \in RC$

Conhecendo-se a tábua acima, calcula-se o Alfa de Bayes Teórico (ALFABT) e o Beta de Bayes Teórico (BETABT):

$$ALFABT = \sum_{(x,y) \in RC} f_0(x, y) = \sum_{(x,y) \in RC} \frac{\binom{t}{x} \binom{N-t}{m-x}}{\binom{N}{m}} \cdot \frac{1}{(N+1)}$$

onde $t=x+y$

$$BETABT = \sum_{(x,y) \in RA} f_1(x, y) = \sum_{(x,y) \in RA} \frac{1}{(m+1)(n+1)}$$

Para o caso em que $m=n=5$ obtivemos:

ALFABT = 0,2251082251

BETABT = 0,4444444444

V.2.2 - A Tábua de Decisão de Fisher

Tendo em mãos o ALFABT, temos condições de montar a Tábua de Decisão associada ao Teste Exato de Fisher.

Neste ponto deve ser feita uma observação importante. O teste de Bayes foi idealizado como um teste bicaudal, isto é, testa-se $H_0: p=q$ contra $H_1: p \neq q$. Por outro lado, o teste Exato, da forma como foi construído, é unicaudal, testando $H_0: p=q$ contra $H_1: p > q$.

Para experimentos nos quais estão envolvidas distribuições de probabilidades simétricas, este fato não traria maiores problemas. Se um experimento mostra uma diferença significativa entre dois tratamentos (T_1 e T_2) a um nível de significância $ALFA=0,04$ e T_1 revelou ser superior a T_2 , pode-se afirmar que T_1 é significativamente melhor do que T_2 a um nível $(ALFA/2) = 0,02$ (pois T_1 poderia ser inferior a T_2).

Se uma distribuição de erros contínua é simétrica em torno do zero, desvios iguais em qualquer direção terão probabilidades iguais; por outro lado, se a distribuição de erros não for simétrica, estas probabilidades serão distintas. Assim, no caso de distribuições contínuas e simétricas, a regra para determinação da probabilidade de erro em uma cauda é tomar a metade da probabilidade bicaudal. Porém, no caso de

distribuições assimétricas podem existir dificuldades. Na tabela 2×2 com $m=n$ e $p=q$, existirão pares de pontos equidistantes do valor esperado de x , $e=E(x)=m.t/N$ ($p=t/N$). Cada elemento do par estará situado em uma das caudas e para um dado par, ambos os componentes possuirão a mesma probabilidade hipergeométrica. Conseqüentemente, para o caso em que $m=n$, é correto considerar o nível $(\alpha/2)$ para cada uma das caudas.

Se $m \neq n$, mas $2E(x)=2e$ for inteiro, ainda existirão pares de pontos equidistantes de e , mas existirão também alguns pontos desemparelhados na cauda mais comprida e neste caso a distribuição hipergeométrica é assimétrica. Como terceira alternativa temos a situação em que $m \neq n$ e $2E(x)=2e$ não é inteiro; com este arranjo não existirão pares de pontos equidistantes de e e a distribuição hipergeométrica também será assimétrica. Portanto, quando $m \neq n$, não é correto considerar o nível $(\alpha/2)$ para cada uma das caudas pois a probabilidade de que se cometa o erro de tipo I à direita de e não é igual à probabilidade de que se erre do lado esquerdo de e . Um valor à direita de e não é equiprovável a seu simétrico à esquerda de e .

Fisher nunca publicou o seu ponto de vista a respeito deste assunto, mas segundo Yates (2), referiu-se a este problema respondendo a uma carta de D.J.Finney, datada de maio

(2) YATES, F. Tests of significance for 2×2 contingency tables. J. R. Statist. Soc. A., 147(3), 1984. (To Appear)

de 1946. Em sua resposta, Fisher defende a solução de se considerar metade do nível de significância ($\alpha/2$), para cada extremidade, não porque haja correspondência entre as subdivisões discretas de cada cauda, mas porque, supostamente, o valor do nível de significância foi escolhido antes que o experimento fosse realizado. Isto equivale a dizer que se a probabilidade de obtenção de x for menor do que ($\alpha/2$), consideraremos $p=q$.

Resolvemos seguir o conselho de Fisher para comparar o teste Bayesiano com o teste Exato, principalmente para o caso em que $m=n$, quando tal procedimento não apresenta problemas.

Desta forma, para montar a Tábua de Decisão de Fisher, será utilizado o seguinte critério δ_F :

$$(x, y) \in RC \Leftrightarrow P_t(x|N, m, t) = \sum_{r=x}^t P_f(r|N, m, t) < \alpha/2$$

$$(x, y) \in RA \Leftrightarrow P_t(x|N, m, t) = \sum_{r=x}^t P_f(r|N, m, t) \geq \alpha/2$$

Este procedimento equivale a:

$$(x, y) \in RC \Leftrightarrow \sum_{r=x}^t \frac{\binom{m}{r} \binom{n}{t-r}}{\binom{N}{t}} < \alpha/2$$

$$(x, y) \in RA \Leftrightarrow \sum_{r=x}^t \frac{\binom{m}{r} \binom{n}{t-r}}{\binom{N}{t}} \geq \alpha/2$$

ou seja:

$$(x, y) \in RC \Leftrightarrow \sum_{r=x}^t \frac{m!n!t!(N-t)!}{r!(m-r)!(t-r)!(n-t+r)!N!} < \text{ALFABT}/2$$

$$(x, y) \in RA \Leftrightarrow \sum_{r=x}^t \frac{m!n!t!(N-t)!}{r!(m-r)!(t-r)!(n-t+r)!N!} \gg \text{ALFABT}/2$$

A Figura V.3 apresenta a Tábua de Decisão de Fisher para $m=n=5$.

	MATFISHER					
	0	1	2	3	4	5
	↓	↓	↓	↓	↓	↓
0→	1	1	1	0	0	0
1→	1	1	1	1	0	0
2→	1	1	1	1	1	0
3→	0	1	1	1	1	1
4→	0	0	1	1	1	1
5→	0	0	0	1	1	1

Figura V.3

Legenda: corpo da tábua = 1 ==> $(x, y) \in RA$

corpo da tábua = 0 ==> $(x, y) \in RC$

Uma vez definidos os procedimentos de teste δ_B e δ_F , podemos confrontá-los através de simulação. Cabe aqui salientar que a comparação deve se dar em separado para cada tamanho de amostra, ou seja, para cada valor do par (m, n) . O fato de um procedimento revelar-se melhor do que outro para um determinado valor de (m, n) não significa que ele será melhor para todos os pares.

V.3 - A Simulação

V.3.1 - Exemplo para o Caso em que $m=n=5$

. Cálculo dos ALFAS Simulados

Seja uma tabela de contingência 2×2 onde, conhecidamente, $p=q$. Se aplicarmos a ela um procedimento de teste δ e rejeitarmos $H_0: p=q$, estaremos cometendo um erro de 1a. espécie.

Para cada par (p,q) em que $p=q$, foram geradas 1000 tabelas de contingência 2×2 . A cada uma destas tabelas foi aplicado o procedimento de teste de Bayes (δ_B) e o procedimento de teste de Fisher (δ_F), computando-se o número total de rejeições de H_0 . Desta forma, o ALFA simulado é dado por:

$$\text{ALFAS}(p,q) = (\text{número de rejeições}) / 1000$$

A Tabela V.1 mostra os resultados obtidos para $m=n=5$.

TABALFASIM

M = 5		N = 5	
P=Q	SIMULADO BAYES ALFA(P,Q)	SIMULADO FISHER ALFA(P,Q)	
.100	.09200	.00900	
.200	.21100	.04600	
.300	.30500	.08200	
.400	.34000	.10800	
.500	.35400	.10800	
.600	.36600	.11900	
.700	.31000	.07100	
.800	.21600	.05000	
.900	.12400	.01300	
PARA CADA (P,Q) FORAM SIMULADAS 1000 TABELAS 2x2			

Tabela V.1

Analisando-se a tabela conclui-se que o ALFA simulado para o procedimento de Fisher ($ALFAS(p,q)$) é sempre menor do que o ALFA simulado para o procedimento de Bayes ($ALFABS(p,q)$). Esta diferença acentua-se para os valores médios de p e q (0,5 e 0,6) e diminui para os valores de p e q próximos de 0 e 1.

Além disso as curvas $ALFAS(p,q)$ variam paralelamente, ou seja, é mais provável que se cometa um erro de 1ª. espécie quando $p=q=0,5$ do que quando $p=q=0,1$, independentemente do procedimento de teste utilizado.

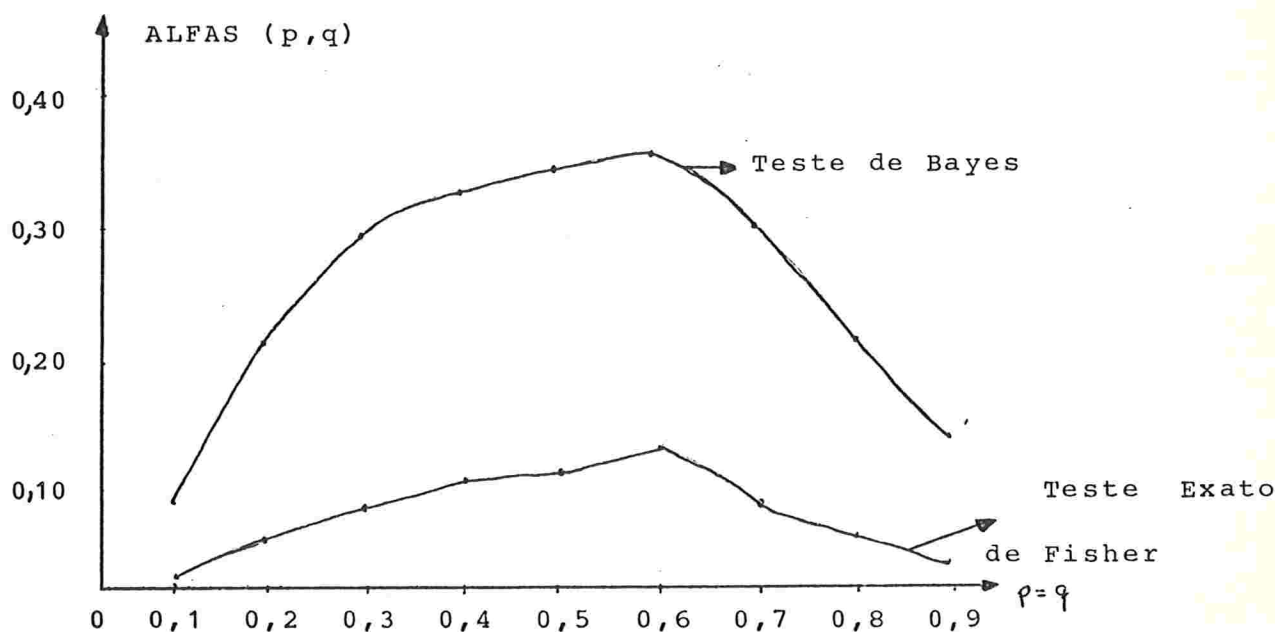


Figura V.4

Para que os procedimentos de teste sejam caracterizados para cada par (m,n) , calculamos o ALFA Simulado Médio, uma média dos $ALFAS(p,q)$ para todos os valores de p,q .

Alfa Médio de Bayes Simulado:

$$\text{ALFAMEBS} = \sum_{p,q=0,1}^{0,9} (\text{ALFABS}(p,q)) / 9$$

Alfa Médio de Fisher Simulado:

$$\text{ALFAMEFS} = \sum_{p,q=0,1}^{0,9} (\text{ALFAFS}(p,q)) / 9$$

Para $(m,n) = (5,5)$ obtivemos

$\text{ALFABT} = 0,2251082251$

$\text{ALFAMEBS} = 0,257555556$

$\text{ALFAMEFS} = 0,067333333$

O ALFAMEFS é bem menor do que o ALFAMEBS, ou seja, para $m=n=5$ o teste Exato erra bem menos em favor de H_1 do que o teste de Bayes. Tal feito pode ser explicado se examinarmos atentamente as Tábuas de Decisão associadas aos testes (Figuras V.1 e V.2) O procedimento Bayesiano possui uma região crítica bem maior do que a do procedimento de Fisher que, por sua vez é mais conservativo, rejeitando H_0 com maior dificuldade. O tamanho do teste Exato (alfa nominal) chega a ser muito menor do que o nível de significância escolhido preliminarmente ($\text{ALFABT} = 0,225108$) pois a distribuição sobre a qual baseia-se o teste Exato é discreta. (3)

(3) Este fato já foi discutido exhaustivamente no capítulo III e aqui está sendo comprovado através de simulação.

. Cálculo dos BETAS Simulados

Considere agora uma tabela de contingência 2×2 para a qual $p \neq q$. Se aplicarmos a ela um procedimento de teste δ e aceitarmos $H_0: p=q$, estaremos cometendo um erro de 2ª espécie.

Para uma série de pares (p, q) nos quais $p \neq q$, foram gerados 1000 tabelas 2×2 . A cada uma destas tabelas foi aplicado o procedimento de teste de Bayes (δ_B) e de Fisher (δ_F), computando-se, para cada um deles, o número total de aceitações de H_0 . O BETA simulado é dado por:

$$\text{BETAS}(p, q) = (\text{número de aceitações}) / 1000$$

A Tabela V.2 mostra os resultados obtidos para $m=n=5$.

No caso do cálculo dos BETAS a situação se inverte e o Beta simulado para o procedimento de Fisher ($\text{BETAFS}(p, q)$) é sempre maior do que o Beta para o procedimento Bayesiano ($\text{BETABS}(p, q)$). À medida que p aproxima-se de q , o $\text{BETA}(p, q)$ cresce em ambos os procedimentos, o que é muito natural pois se p é próximo de q , a probabilidade de que H_0 seja rejeitada deve diminuir.

O BETA SIMULADO MÉDIO é calculado de forma análoga ao ALFA SIMULADO MÉDIO:

Beta Médio de Bayes Simulado

$$\text{BETAMEBS} = \sum_{p=0,1}^{0,9} \sum_{\substack{q=0,2 \\ q > p}}^{0,9} (\text{BETABS}(p, q)) / 36$$

TABETASIM

		M = 5	N = 5
P	Q	SIMULADO BAYES BETA(P,Q)	SIMULADO FISHER BETA(P,Q)
.100	.200	.79200	.96200
.100	.300	.65500	.88200
.100	.400	.49100	.79600
.100	.500	.35100	.63400
.100	.600	.22100	.49100
.100	.700	.10500	.32800
.100	.800	.06200	.17700
.100	.900	.01600	.07000
.200	.300	.71700	.92100
.200	.400	.60300	.85600
.200	.500	.48100	.75500
.200	.600	.34000	.61600
.200	.700	.23400	.47800
.200	.800	.11500	.32200
.200	.900	.05700	.17100
.300	.400	.65500	.90000
.300	.500	.56400	.82200
.300	.600	.46300	.72400
.300	.700	.33900	.64600
.300	.800	.21800	.47600
.300	.900	.10900	.32300
.400	.500	.60800	.86500
.400	.600	.56800	.80300
.400	.700	.48100	.74200
.400	.800	.32500	.62000
.400	.900	.20800	.48100
.500	.600	.61800	.87500
.500	.700	.55400	.82800
.500	.800	.46600	.74500
.500	.900	.33300	.62800
.600	.700	.66400	.87400
.600	.800	.59100	.85300
.600	.900	.48500	.79900
.700	.800	.71500	.92800
.700	.900	.64000	.87900
.800	.900	.77700	.96300
PARA CADA (P,Q) FORAM SIMULADAS		1000	TABELAS 2x2

Tabela V.2

Beta Médio de Fisher Simulado

$$BETAMEFS = \sum_{p=0,1}^{0,9} \sum_{\substack{q=0,2 \\ q > p}}^{0,9} (BETA FS(p,q)) / 36$$

Para $m=n=5$, foram obtidos os seguintes valores:

$$BETABT = 0,4444444$$

$$BETAMEBS = 0,43391667$$

$$BETAMEFS = 0,67313889$$

O BETAMEFS é 55% maior que o BETAMEBS. Em outras palavras, o procedimento Exato comete o erro de tipo II um número maior de vezes do que o procedimento de Bayes.

. A Combinação Linear $ALFA(\delta) + BETA(\delta)$

$$ALFABT + BETABT = 0,669553$$

$$ALFAMEBS + BETAMEBS = 0,691472$$

$$ALFAMEFS + BETAMEFS = 0,740422$$

A proporção de erros referente ao teste Exato é 7% maior do que a referente ao teste de Bayes, sendo que a última é mais próxima da soma das probabilidades de erro teóricas. Isto significa que, para $m=n=5$, em cada 100 tabelas analisadas, o teste Exato decide em favor da hipótese falsa em 7 tabelas a mais do que o teste Bayesiano. Neste caso, a superioridade do teste de Bayes em relação ao teste Exato é evidente.

As funções poder dos testes são plotadas no Apêndice B. O poder do teste Bayesiano mostra-se maior do que o poder do Teste Exato de Fisher.

V.3.2 - Exemplo para o Caso em que $m=8$ e $n=14$

Embora neste exemplo, $m \neq n$, resolvemos agir de acordo com o ponto de vista de Fisher e considerar o valor $ALFABT/2$ para a construção da Tábua de Decisão do teste Exato. A Figura V.5 mostra as Tábuas de Decisão obtidas

Tábua de Decisão de Bayes

	MATBAYES								
	0	1	2	3	4	5	6	7	8
	↓	↓	↓	↓	↓	↓	↓	↓	↓
0→	1	1	0	0	0	0	0	0	0
1→	1	1	1	0	0	0	0	0	0
2→	1	1	1	1	0	0	0	0	0
3→	1	1	1	1	0	0	0	0	0
4→	0	1	1	1	1	0	0	0	0
5→	0	1	1	1	1	1	0	0	0
6→	0	0	1	1	1	1	0	0	0
7→	0	0	1	1	1	1	1	0	0
8→	0	0	0	1	1	1	1	0	0
9→	0	0	0	1	1	1	1	1	0
10→	0	0	0	0	1	1	1	1	0
11→	0	0	0	0	0	1	1	1	1
12→	0	0	0	0	0	1	1	1	1
13→	0	0	0	0	0	0	1	1	1
14→	0	0	0	0	0	0	0	1	1

Tábua de Decisão de Fisher

	MATFISHER								
	0	1	2	3	4	5	6	7	8
	↓	↓	↓	↓	↓	↓	↓	↓	↓
0→	1	1	1	0	0	0	0	0	0
1→	1	1	1	1	0	0	0	0	0
2→	1	1	1	1	1	0	0	0	0
3→	1	1	1	1	1	1	0	0	0
4→	1	1	1	1	1	1	0	0	0
5→	1	1	1	1	1	1	1	0	0
6→	0	1	1	1	1	1	1	0	0
7→	0	1	1	1	1	1	1	1	0
8→	0	0	1	1	1	1	1	1	0
9→	0	0	1	1	1	1	1	1	1
10→	0	0	0	1	1	1	1	1	1
11→	0	0	0	1	1	1	1	1	1
12→	0	0	0	0	1	1	1	1	1
13→	0	0	0	0	0	1	1	1	1
14→	0	0	0	0	0	0	1	1	1

Figura V.5-Observe que a região crítica associada ao procedimento de Fisher contém menos pontos do que a região crítica associada ao procedimento de Bayes.

$ALFABT = 0,1439684$

$BETABT = 0,422222$

. Cálculo dos ALFAS Simulados

TABALFASIM

M = 8		N = 14	
P=Q	SIMULADO BAYES ALFA(P,Q)	SIMULADO FISHER ALFA(P,Q)	
.100	.08500	.01000	
.200	.14300	.02500	
.300	.18700	.04300	
.400	.21100	.04800	
.500	.20300	.05700	
.600	.21300	.05200	
.700	.19400	.04500	
.800	.14500	.02700	
.900	.08500	.00800	
PARA CADA (P,Q) FORAM SIMULADAS 1000 TABELAS 2x2			

Tabelas V.3

Assim como para o exemplo anterior, aqui também o $ALFAS(p,q)$ é menor do que o $ALFABS(p,q)$ para qualquer par (p,q) onde $p=q$. As curvas $ALFAS(p,q)$ são paralelas, sendo mais provável que se cometa um erro de 1a. espécie para $p=q=0,5$ do que para $p=q=0,9$, independentemente do procedimento de teste utilizado

$ALFAMEBS = 0,1628888$

$ALFAMEFS = 0,035$

A probabilidade de que se cometa o erro de 1a. espécie é maior para o teste de Bayes, probabilidade esta que se aproxima bastante do $ALFABT(0,143968423)$

• Cálculo dos BETAS Simulados

TABETASIM

		M = 8	N = 14
P	Q	SIMULADO BAYES BETA(P,Q)	SIMULADO FISHER BETA(P,Q)
.100	.200	.79300	.93900
.100	.300	.59900	.83600
.100	.400	.38300	.63800
.100	.500	.24700	.45600
.100	.600	.11400	.29700
.100	.700	.03400	.15900
.100	.800	.00800	.04900
.100	.900	.00100	.00600
.200	.300	.77200	.92400
.200	.400	.59900	.84100
.200	.500	.41000	.70100
.200	.600	.26900	.53900
.200	.700	.11600	.35000
.200	.800	.04200	.16900
.200	.900	.00400	.04900
.300	.400	.75200	.92900
.300	.500	.62300	.85500
.300	.600	.45000	.71900
.300	.700	.27700	.53300
.300	.800	.12100	.31800
.300	.900	.03600	.15100
.400	.500	.74800	.92100
.400	.600	.62800	.86300
.400	.700	.45800	.69900
.400	.800	.26400	.52800
.400	.900	.09800	.27500
.500	.600	.75200	.91400
.500	.700	.61200	.83400
.500	.800	.44300	.70600
.500	.900	.22300	.49200
.600	.700	.74800	.93200
.600	.800	.63900	.86400
.600	.900	.41200	.72000
.700	.800	.78300	.94400
.700	.900	.62100	.90500
.800	.900	.84000	.97100
PARA CADA (P,Q) FORAM SIMULADAS		1000	TABELAS 2x2

Tabela V.4

O BETA simulado para o procedimento Bayesiano (BETABS(p,q)) é sempre menor do que o BETA simulado para o procedimento de Fisher (BETAFS)

BETABT = 0,4222222

BETAMEBS = 0,414416667

BETAMEFS = 0,61183333

O BETAMEBS aproxima-se do BETABT e é bem menor (cerca de 50%) do que o BETAMEFS.

. A Combinação Linear $ALFA(\delta) + BETA(\delta)$

ALFABT + BETABT = 0,5661906452

ALFAMEBS + BETAMEBS = 0,5773055556

ALFAMEFS + BETAMEFS = 0,646833333

A proporção média de erros é menor para o procedimento de Bayes do que para o de Fisher (cerca de 12%). Isto significa que se submetermos 100 tabelas aos dois testes, o segundo cometerá erros em 12 tabelas a mais do que o primeiro. Neste segundo exemplo o teste de Bayes voltou a revelar-se superior ao teste Exato de Fisher.

As funções poder dos testes são plotadas no Apêndice C.

O procedimento Bayesiano é uniformemente mais poderoso do que o de Fisher para $(p,q) \in \Omega_1$.

V.3.3 - Exemplo para o Caso $m=5$ e $n=3$ Utilizando para o

Teste Exato de Fisher o valor ALFABT

Para evitar discussões a respeito da utilização do valor (ALFABT/2) para a determinação do procedimento de teste de Fisher, vamos repetir a metodologia de simulação aplicada nos exemplos anteriores para comparar as duas Tábuas de Decisão, sendo que a de Fisher foi construída mediante o uso do valor ALFABT para cada cauda. É importante ressaltar que ao utilizarmos o valor ALFABT para cada extremidade, estamos tornando o teste Exato menos conservativo, ou seja, estamos aumentando sua probabilidade de rejeição de H_0 . Conseqüentemente estaremos diminuindo a probabilidade do erro de 2a. espécie. Vejamos se tal atitude fará com que o teste Exato supere o teste Bayesiano em qualidade.

. Tábuas de Decisão

Tábua de Decisão de Bayes

	MATBAYES					
	0	1	2	3	4	5
	↓	↓	↓	↓	↓	↓
0→	1	1	0	0	0	0
1→	1	1	1	1	0	0
2→	0	0	1	1	1	1
3→	0	0	0	0	1	1

Tábua de Decisão de Fisher

	MATFISHER					
	0	1	2	3	4	5
	↓	↓	↓	↓	↓	↓
0→	1	1	1	0	0	0
1→	1	1	1	1	1	0
2→	0	1	1	1	1	1
3→	0	0	0	1	1	1

ALFABT = 0,2222222

BETABT = 0,5

Figura V.6

Apesar de ter sido dobrado o nível de significância do teste Exato, sua região crítica contém menos pontos do que a região crítica associada ao procedimento Bayesiano.

• Cálculo dos ALFAS Simulados

TABALFASIM			
		M = 5	N = 3
P=Q	SIMULADO BAYES ALFA(P,Q)	SIMULADO FISHER ALFA(P,Q)	
.100	.06900	.01600	
.200	.20700	.07000	
.300	.32900	.13300	
.400	.32200	.13700	
.500	.34000	.15500	
.600	.35100	.13600	
.700	.29600	.10200	
.800	.21900	.08100	
.900	.09300	.02300	
PARA CADA (P,Q) FORAM SIMULADAS 1000 TABELAS 2x2			

Tabelas V.5

ALFAMEBS = 0,2473333

ALFAMEFS = 0,09477778

• Cálculo dos BETAS Simulados

BETAMEBS = 0,4853888889

BETAMEFS = 0,6793611111

TABETASIM

		M = 5	N = 3
P	Q	SIMULADO BAYES BETA(P,Q)	SIMULADO FISHER BETA(P,Q)
.100	.200	.76900	.94900
.100	.300	.62500	.86400
.100	.400	.45200	.76800
.100	.500	.34900	.61500
.100	.600	.24700	.49100
.100	.700	.16400	.35600
.100	.800	.09000	.24300
.100	.900	.04900	.14400
.200	.300	.67600	.88400
.200	.400	.59100	.82900
.200	.500	.49500	.74700
.200	.600	.38100	.60900
.200	.700	.28100	.49700
.200	.800	.19200	.40800
.200	.900	.13000	.27500
.300	.400	.62000	.84200
.300	.500	.57700	.81000
.300	.600	.50200	.71200
.300	.700	.40800	.61200
.300	.800	.32200	.53700
.300	.900	.24600	.40700
.400	.500	.64500	.85200
.400	.600	.56500	.80000
.400	.700	.53500	.75100
.400	.800	.44900	.66400
.400	.900	.39500	.53600
.500	.600	.64100	.84300
.500	.700	.60600	.82300
.500	.800	.58500	.76600
.500	.900	.50800	.65500
.600	.700	.67400	.85500
.600	.800	.67700	.83800
.600	.900	.68100	.78000
.700	.800	.72300	.89900
.700	.900	.76800	.86800
.800	.900	.85600	.92800
PARA CADA (P,Q) FORAM SIMULADAS 1000 TABELAS 2x2			

Tabela V.6

. A Combinação Linear $ALFA(\delta) + BETA(\delta)$

$$ALFABT + BETABT = 0,72222$$

$$ALFAMEBS + BETAMEBS = 0,7327222$$

$$ALFAMEFS + BETAMEFS = 0,77413889$$

Mesmo com o uso do ALFABT para cada uma das caudas como nível de significância para o teste Exato, o teste Bayesiano revelou-se superior em qualidade, na medida que minimizou $ALFA(\delta) + BETA(\delta)$. Além disso, os valores simulados de ALFA e BETA para o procedimento Bayesiano aproximaram-se bastante dos valores teóricos. Em contrapartida, no caso do teste Exato, os valores simulados mantêm maior distância dos valores teóricos. Tal fato constitui um argumento desfavorável à aplicação do teste de Fisher uma vez que os erros que são cometidos ao utilizá-lo são, na realidade, bastante diferentes dos erros que se acredita estar cometendo.

As funções poder dos testes para $m=5$ e $n=3$ são plotadas no Apêndice D.

V.3.4 - A Simulação para Tabelas de Contingência em que

$$m=n$$

A Tabela V.7 permite a comparação da combinação linear teórica e simulada para tabelas de contingência nas quais $m=n$. No exemplo são apresentados valores de m que vão de 1 à 20.

Para cada tamanho de amostra são construídas as Tábuas de

Decisão de Bayes e de Fisher. sendo que para esta última é utilizado o valor (ALFABT/2) como nível de significância para cada cauda.

Os ALFAS simulados correspondem à média dos ALFAS simulados para cada valor de p e q em que $p=q$ (ALFAMEBS e ALFAMEFS). Os BETAS simulados correspondem à média dos BETAS simulados para cada p e q, sendo $p \neq q$ (BETAMEBS e BETAMEFS). Para cada par (p,q) foram geradas 1000 tabelas de contingência.

TAB

		TEORICO BAYES			SIMULADO BAYES			SIMULADO FISHER		
M=N	ALFA	ALFA +		ALFA	ALFA +		ALFA	ALFA +		ALFA
		BETA	BETA		BETA	BETA		BETA	BETA	
1	.3333	.5000	.83333	.3622	.4836	.84578	.0000	1.0000	1.00000	
2	.4667	.3333	.80000	.5203	.3151	.83539	.0750	.7842	.85917	
3	.1286	.6250	.75357	.1478	.6222	.76997	.0153	.8870	.90233	
4	.1810	.5200	.70095	.1999	.5136	.71344	.0369	.7737	.81058	
5	.2251	.4444	.66955	.2474	.4361	.68358	.0677	.6719	.73956	
6	.2627	.3878	.65049	.2896	.3742	.66378	.0856	.5946	.68011	
7	.1647	.4688	.63344	.1839	.4766	.66053	.0337	.6970	.73069	
8	.1245	.4815	.60601	.1387	.4722	.61083	.0346	.6691	.70361	
9	.1448	.4400	.58477	.1578	.4242	.58194	.0370	.6326	.66964	
10	.1639	.4050	.56881	.1837	.3868	.57050	.0432	.5969	.64011	
11	.1818	.3750	.55681	.1992	.3535	.55269	.0498	.5633	.61306	
12	.1245	.4201	.54466	.1342	.4164	.55064	.0411	.5545	.59558	
13	.1125	.4184	.53087	.1266	.4114	.53792	.0442	.5376	.58178	
14	.1227	.3956	.51827	.1373	.3832	.52056	.0486	.5176	.56617	
15	.1328	.3750	.50777	.1440	.3613	.50531	.0503	.4958	.54614	
16	.1426	.3564	.49904	.1614	.3393	.50075	.0528	.4729	.52569	
17	.1145	.3765	.49102	.1279	.3552	.48306	.0456	.4867	.53225	
18	.1079	.3740	.48185	.1204	.3527	.47314	.0476	.4619	.50950	
19	.1028	.3700	.47284	.1218	.3511	.47286	.0456	.4584	.50394	
20	.0995	.3651	.46461	.1141	.3455	.45964	.0434	.4552	.49867	

Tabela V.7

Observação: O leitor deverá ter notado que no caso do teste de Bayes, o valor simulado do ALFA é sempre superior ao valor teórico, enquanto que o valor simulado do BETA é sempre inferior ao valor teórico. As somas dos valores simulados, embora próximas das somas teóricas, mostram-se superiores.

Quanto mais fina for a malha com a qual particionarmos o espaço paramétrico, mais próximos estarão os valores simulados dos teóricos. A malha sob a qual foi particionado o espaço paramétrico para que fosse gerada a **Tabela V.7** é de incremento $\Delta p=0,1$. Isto significa que a média dos ALFAS e BETAS simulados é calculada para valores de p e q que vão de 0,1 à 0,9 com incremento $\Delta p=0,1$.

Se diminuirmos o reticulado (o incremento), o valor simulado aproximar-se-á significativamente de teórico no caso do teste de Bayes. A **Tabela V.8** apresenta o cálculo dos ALFAS e BETAS simulados, feito com incremento $\Delta p=0,05$. Fizemos este cálculo para um número restrito de valores de m e n devido ao tempo excessivo de CPU consumido por este procedimento.

		TEORICO			SIMULADO			SIMULADO		
		BAYES			BAYES			FISHER		
M=N	ALFA	BETA	ALFA +	ALFA	BETA	ALFA +	ALFA	BETA	ALFA +	ALFA
			BETA			BETA			BETA	
1	.3333	.5000	.83333	.3500	.4889	.83894	.0000	1.0000	1.00000	
2	.4667	.3333	.80000	.4955	.3211	.81662	.0703	.7800	.85023	
3	.1286	.6250	.75357	.1342	.6242	.75837	.0153	.8813	.89656	
4	.1810	.5200	.70095	.1888	.5159	.70472	.0366	.7643	.80094	
5	.2251	.4444	.66955	.2372	.4357	.67289	.0619	.6676	.72949	

Tabela V.8

Esta tabela tem os valores simulados gerados para p e q variando de 0,1 à 0,9, com incremento $\Delta p = 0,05$. Note que para o caso do teste Bayesiano, a diferença entre os valores simulados e os teóricos é bastante reduzida.

V.4 - Conclusões

1º) O valor de $ALFA(\delta) + BETA(\delta)$ estimado a partir das tabelas simuladas correspondente ao teste Bayesiano, é menor do que o valor referente ao teste Exato para todos os tamanhos de amostra ($m=n$; $m=1, 2, \dots, 20$). Isto indica que o teste de Bayes supera o Exato em qualidade na medida que o pesquisador, ao fazer uso do primeiro, terá uma chance menor de cometer erros. Além disso, o valor de $ALFA(\delta) + BETA(\delta)$ calculado no teste de Bayes é mais próximo do valor teórico do que o calculado no teste de Fisher. Tal fato constitui mais um ponto a favor do teste Bayesiano que fornece uma segurança maior acerca da probabilidade de que se erre ao utilizá-lo.

2º) O Teste Exato de Fisher é bem mais conservativo, mesmo quando se considera, exageradamente, o valor $ALFABT$ como nível de significância para cada uma das caudas. Para $m=n=1$, sua região crítica é vazia. Neste caso o procedimento sempre aceitará H_0 , mesmo quando $p \neq q$.

($ALFAMEFS = 0$ e $BETAMEFS = 1$).

3^a) A diferença entre os dois procedimentos é mais acentuada para amostras pequenas (m e n pequenos). À medida que o tamanho da amostra cresce, os procedimentos aproximam-se entre si, tendendo ao procedimento gerado pelo Teste Qui-Quadrado para tabelas de contingência. Tal aproximação pode ser visualizada de forma extremamente elegante através dos gráficos dos Apêndices B e C. Nota-se facilmente que os gráficos relativos aos procedimentos de Fisher e de Bayes são bem mais parecidos quando $m=8$ e $n=14$ do que quando $m=n=5$. Paralelamente, o valor de $ALFA(\delta) + BETA(\delta)$ também diminui; quanto maior for o tamanho da amostra, menor será a probabilidade de que sejam cometidos erros.

4^a) Convém ressaltar que o teste Bayesiano parte de um conhecimento inicial a respeito dos valores dos parâmetros em jogo e das chances das hipóteses em teste. Este conhecimento inicial pode ter sido originado por experimentos anteriores ou mesmo pela intuição do especialista envolvido com o problema. A este conhecimento, traduzido por distribuições de probabilidades, são adicionados os resultados experimentais. Em outras palavras, os resultados experimentais calibram as distribuições de probabilidades iniciais, ou seja, o teste Bayesiano acumula informação. Em contrapartida, o Teste Exato de Fisher utiliza somente os resultados de um único experimento, não aproveitando possíveis informações obtidas em experimentos anteriores.

Na simulação aqui apresentada não incorporamos informações anteriores ao procedimento de Bayes para que ele fosse confrontado com o procedimento de Fisher sob condições semelhantes. Para tanto fizemos uso de distribuições de probabilidades a priori uniformes. Mesmo assim o teste de Bayes apresentou qualidade superior ao de Fisher.

O bom senso nos induz a crer que se incorporássemos informações prévias ao teste de Bayes, estaríamos contribuindo para aumentar ainda mais a sua eficiência.

Some-se a isso a vantagem de que o teste Bayesiano permite a inclusão de perdas relativas aos erros de 1a. e 2a. espécies, indicando assim a preferência do pesquisador em relação a estes erros.

APÊNDICE A

A Integral Curvilínea (1)

Por uma curva lisa C no plano pxq entende-se uma curva representada por meio das equações paramétricas:

$$p = \phi(t) \qquad h \leq t \leq k$$

$$q = \psi(t)$$

onde ϕ e ψ são funções contínuas e possuem derivadas contínuas no intervalo $h \leq t \leq k$. À curva C pode-se atribuir um sentido de percurso que é, em geral, o sentido dado pelos t crescentes. Se indicamos por A o ponto $[\phi(h), \psi(h)]$ e por B o ponto $[\phi(k), \psi(k)]$, então C pode ser vista como o caminho percorrido por um ponto que se desloca de modo contínuo de A até B .

(1) Uma descrição mais detalhada da Integral Curvilínea poderá ser encontrada em:

KAPLAN, W. Cálculo Integral Vetorial In: _____.

Cálculo Avançado. São Paulo, Edgard Blücher, 1977.

v.1, p.252-328.

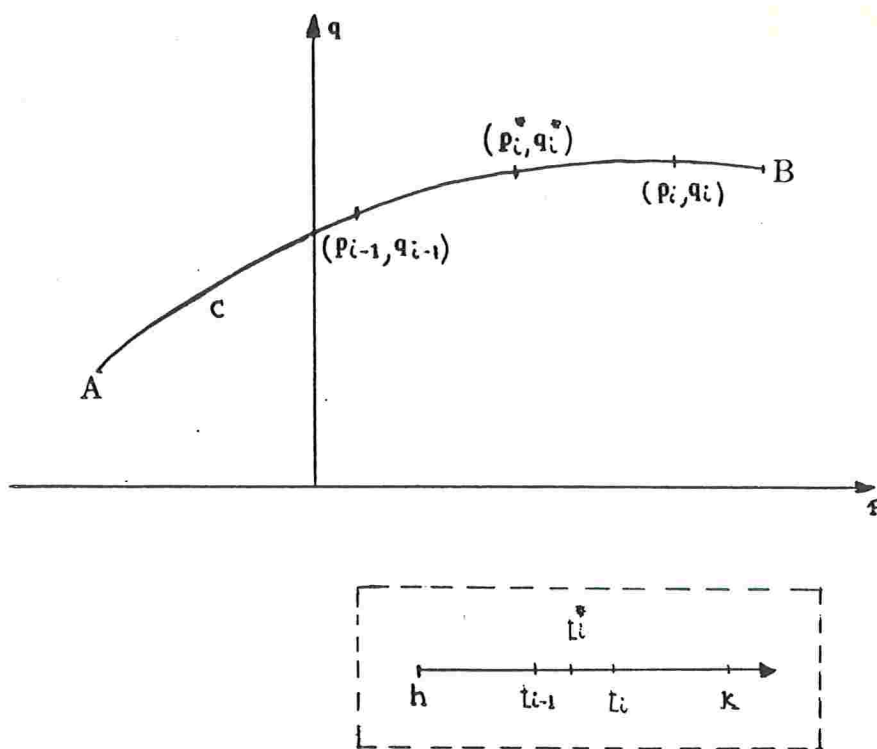


Figura A.1

Se C é uma curva lisa o comprimento de arco s está bem definido, ou seja, é a distância percorrida desde o ponto inicial ($t=h$) até um t genérico.

$$s = \int_h^t \sqrt{\left(\frac{dp}{dt}\right)^2 + \left(\frac{dq}{dt}\right)^2} dt$$

Se a orientação de C for dada pelo t crescente, então s também crescerá no sentido do movimento, variando seu valor de 0 até o comprimento L de C .

Subdivide-se C conforme a Figura A.1 e indica-se por Δs o acréscimo em s de t_{i-1} até t_i , isto é, a distância percorrida neste intervalo. Nestas condições, define-se Integral Curvilínea, também dita Integral de Linha ou Integral de Con-

torno:

$$\int_C f(p, q) ds = \lim_{\substack{n \rightarrow \infty \\ \max \Delta s \rightarrow 0}} \sum_{i=1}^n f(p_i^*, q_i^*) \Delta s$$

Se f for contínua em C , então esta integral existe e pode ser calculada através da fórmula:

$$\int_C f(p, q) ds = \int_h^k f[\phi(t), \psi(t)] \sqrt{\phi'^2(t) + \psi'^2(t)} dt$$

Deste modo a integral curvilínea reduz-se a uma integral comum.

Caso a curva C seja representada sob a forma $q=g(p)$, $a \leq p \leq b$, então poder-se-á tratar o próprio p como parâmetro de integração em substituição a t , ou seja, C é dada pelas equações:

$$p=p$$

$$q=g(p) \quad a \leq p \leq b$$

Assim:

$$\int_C f(p, q) ds = \int_a^b f(p, g(p)) \sqrt{1 + \left(\frac{dg}{dp}\right)^2} dp$$

Exemplificando, se a curva C representar um arame cuja densidade (massa por unidade de comprimento) varia ao longo de C , então a massa total do arame será dada por:

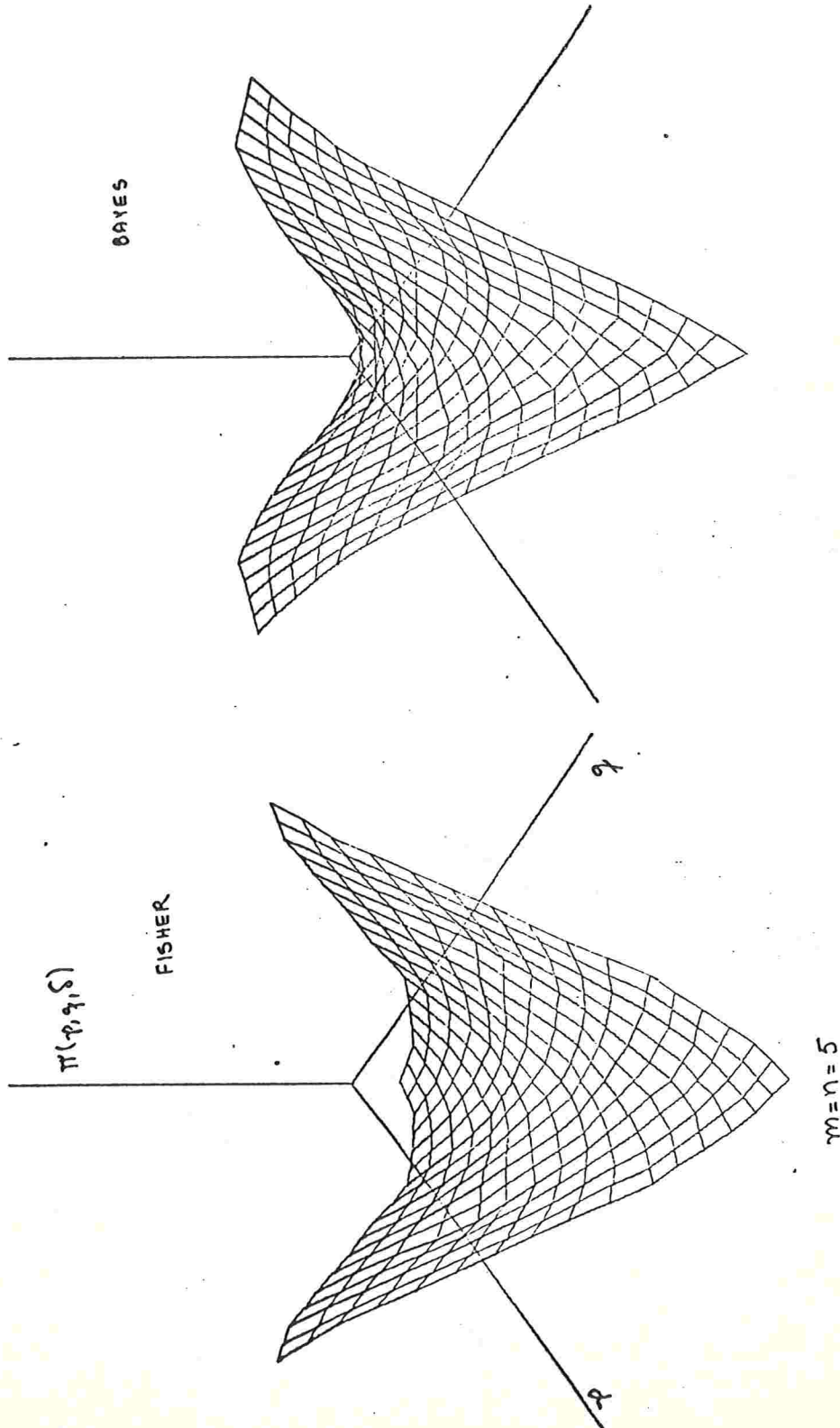
$$M = \int_C f(p, q) ds$$

onde $f(p, q)$ é a densidade no ponto (p, q) do arame.

APÊNDICE B

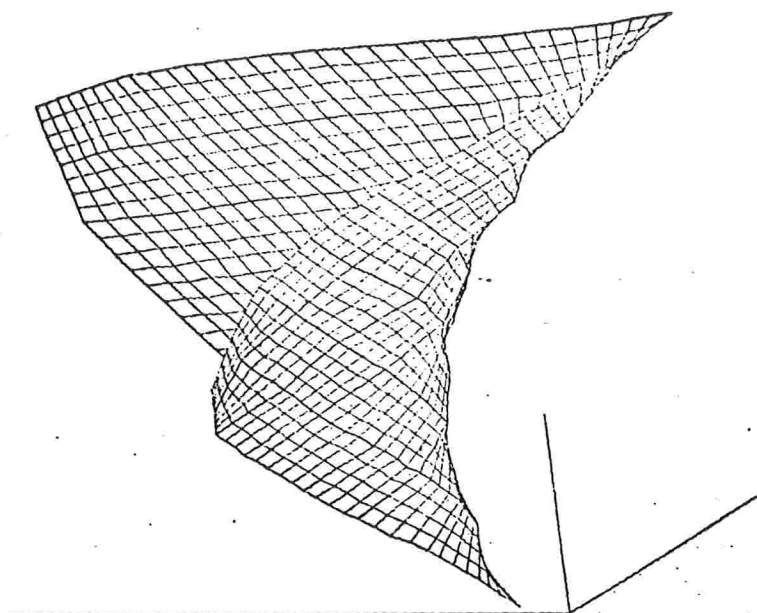
$$\Pi(p, q, \delta) = P(\text{Rejeitar } H_0 | p, q, \delta)$$

VISTA FRONTAL

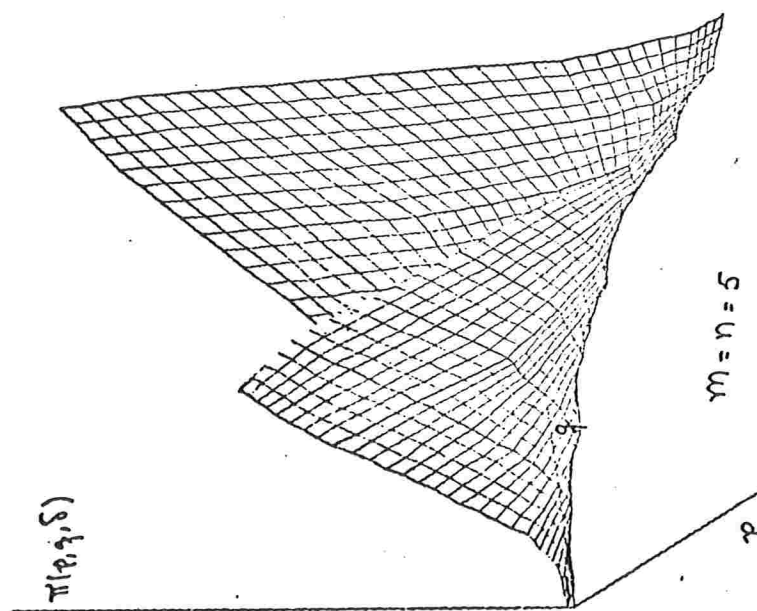


VISTA DE PERFIL

BAYES



FISHER

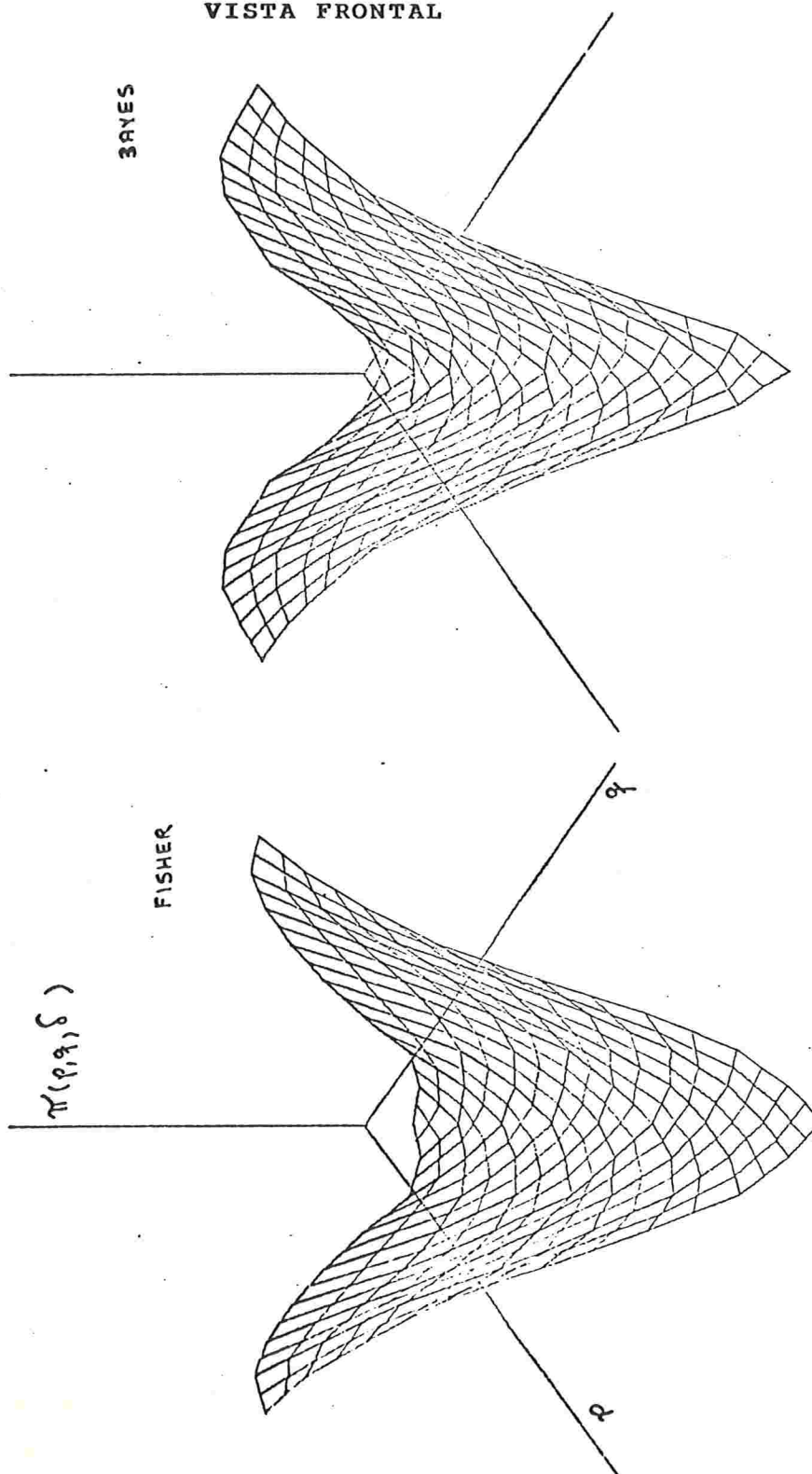


$\pi(\theta, \delta)$

APÊNDICE C

$$\Pi(p, q, \delta) = P(\text{Rejeitar } H_0 | p, q, \delta)$$

VISTA FRONTAL

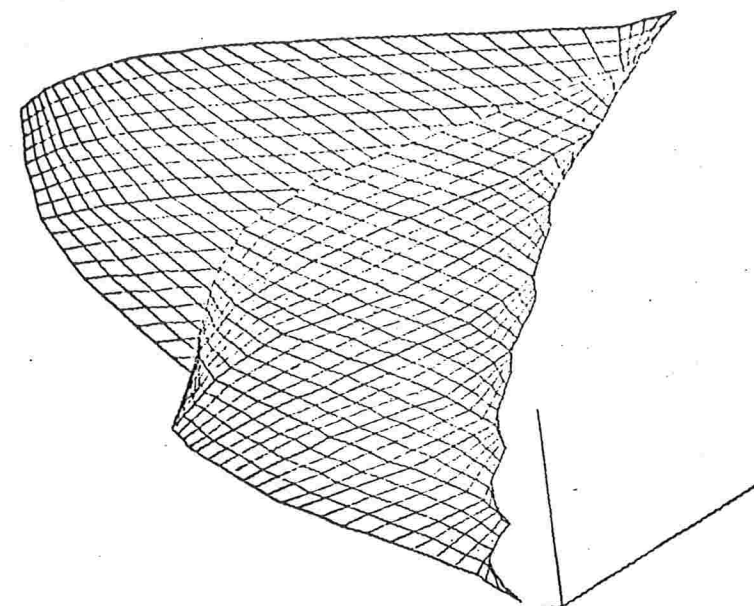


$$m = 8$$

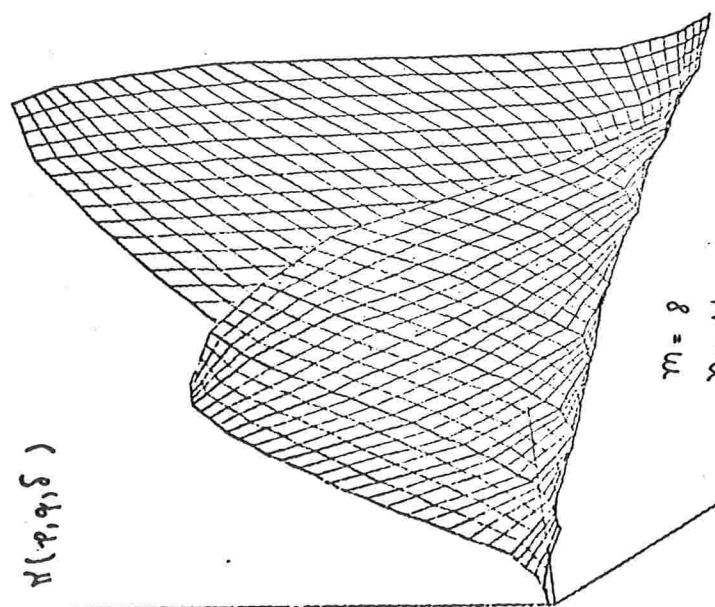
$$n = 14$$

VISTA DE PERFIL

3AYES



FISHER



$\pi(\varphi, \eta, \delta)$

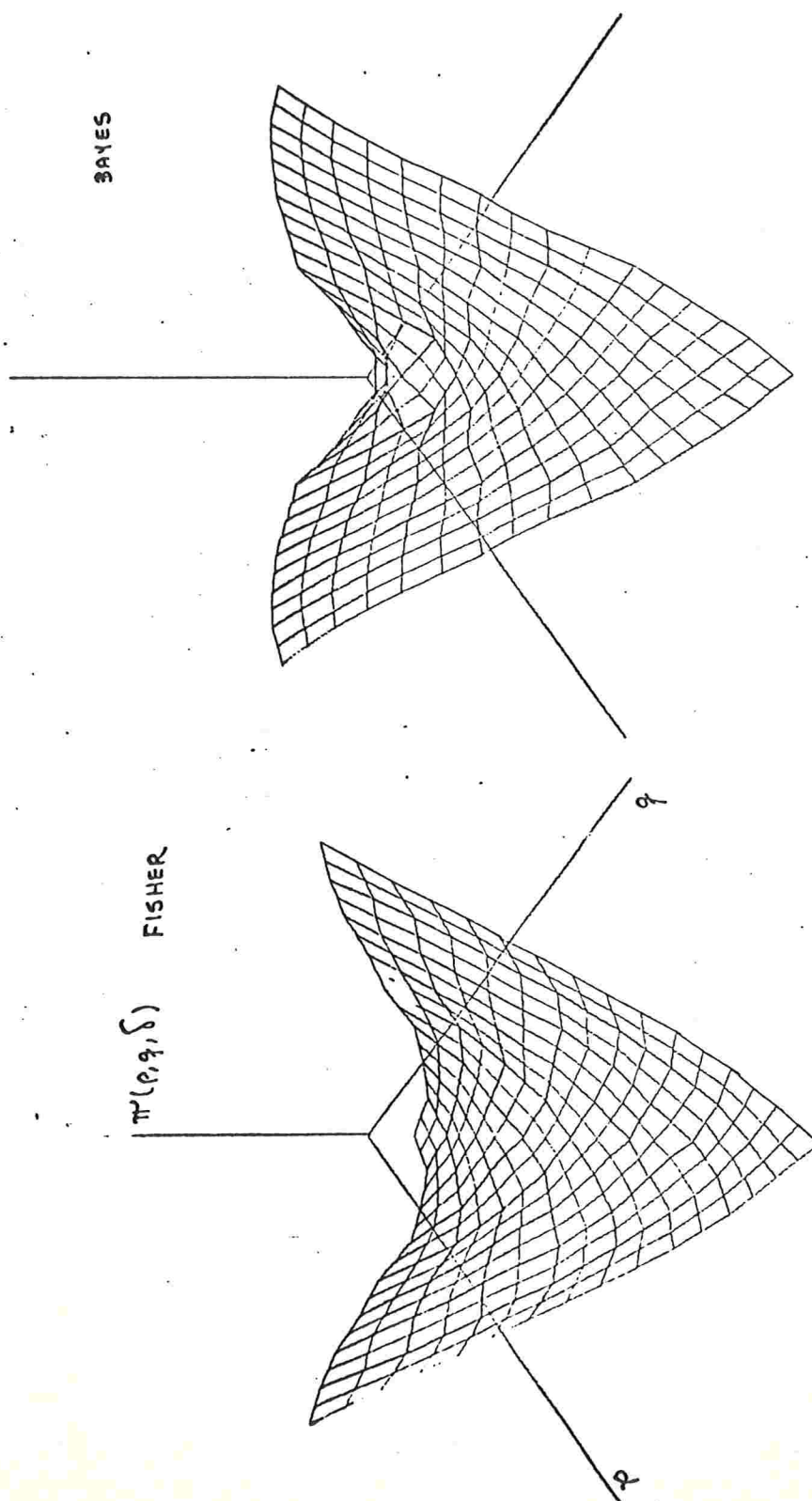
$m = 8$
 $n = 14$

p

APÊNDICE D

$$\Pi(p, q, \delta) = P(\text{Rejeitar } H_0 | p, q, \delta)$$

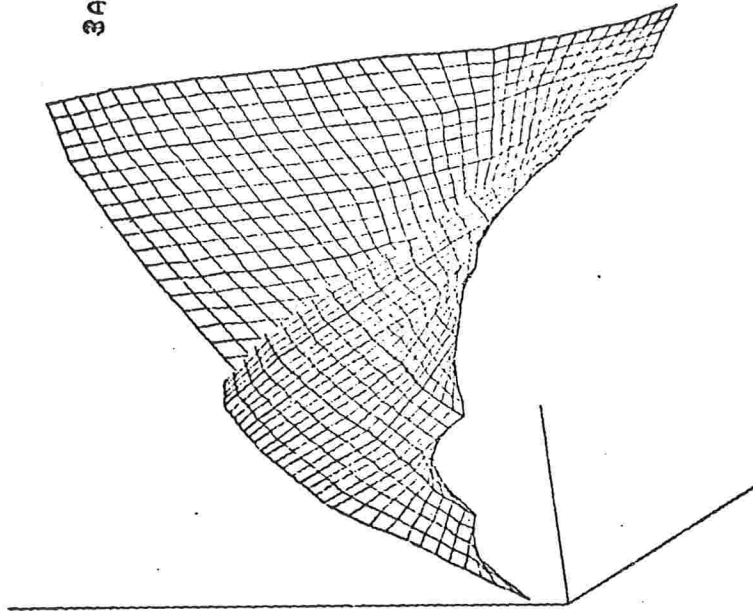
VISTA FRONTAL



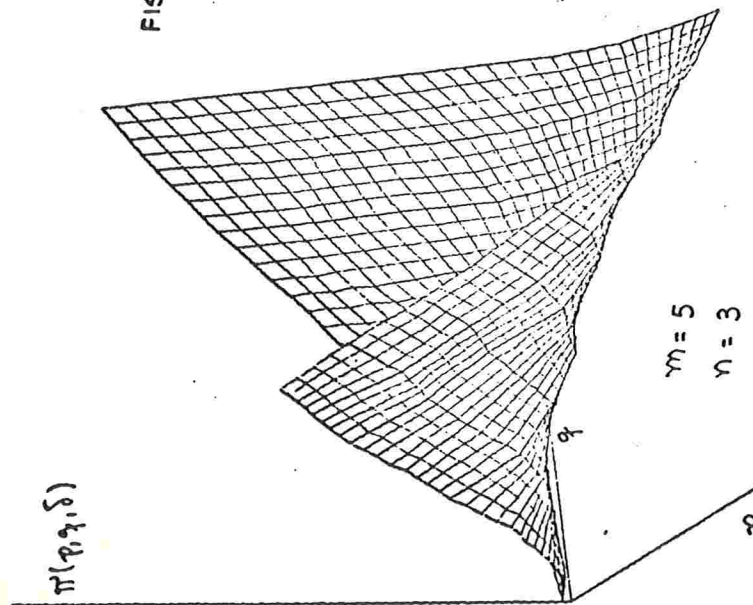
$m = 5$
 $n = 3$

VISTA DE PERFIL

3AYES



FISHER



REFERÊNCIAS BIBLIOGRÁFICAS

- AITCHISON, J. & BACON-SHONE, J. Bayesian risk ratio analysis. *American Statistician*, 35(4):254-257, 1981
- ALTHAM, P.M.E. Exact bayesian analysis of a 2x2 contingency table, and Fisher's "exact" significance test. *Journal of the Royal Statistical Society, B*, 31:261-269, 1969.
- BARNARD, G.A. In contradiction to J. Berkson's dispraise: conditional tests can be more efficient. *Journal of Statistical Planning and Inference*, 3:181-187, 1979.
- BASU, D. Discussion of Joseph Berkson's paper "In dispraise of the exact test". *Journal of Statistical Planning and Inference*, 3:189-192, 1979.
- BASU, D. On the elimination of Nuisance parameters. *JASA*, 72:355-366, 1977.
- BERKSON, J. "In dispraise of the exact test. Do the marginal totals of the 2x2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, 2:27-42, 1978.

BERKSON, J. Do the marginal totals of the 2x2 table contain relevant information respecting the table proportions? *Journal of Statistical Planning and Inference*, 2:43-44, 1978.

CORSTEN, L.C.A. & KROON, J.P.M. Comment on J. Berkson's paper "In dispraise of the exact test". *Journal of Statistical Planning and Inference*, 3:193-197, 1979.

DAWID, A.P. The well-calibrated bayesian. *JASA*, 77(379): 605-613, 1982.

DE GROOT, M. H. *Probability and Statistics*. London, Addison-Wesley, 1975

FISHER, R.A. The logic of inductive inference. *Journal of the Royal Statistical Society*, A, 98(1):39-54, 1935.

GODAMBE, V.R. On Sufficiency and Ancilearity in presence of a Nuisance parameter. *Biometrika*, 67:155-162, 1980.

KAPLAN, W. Cálculo integral vetorial. In: _____. *Cálculo Avançado*. São Paulo, Edgard Blücher, 1977. v.1.

KEMPTHORNE, O. In dispraise of the exact test: reactions. *Journal of Statistical Planning and Inference*, 3:199-213, 1979.

- MARIOTTO, A.B. *Inferência Parcial*. São Paulo, 1983. 107p.
Dissertação (Mestrado) - IME-USP.
- NOVICK, M.R. & GRIZZLE, J.E. A bayesian approach to the
analysis of data from clinical trials. *JASA*, 60(309):
81-96, 1965.
- PEREIRA, C.A.B. *Stopping rules and conditional inference in
2x2 contingency tables*. São Paulo, IME-USP, 1983. 7p.
(Relatório Técnico, RT-MAE-8301).
- PEREIRA, C.A.B. & LINDLEY, D.V. Examples questioning the use
of partial likelihood. São Paulo, IME-USP. 9p.
(Relatório Técnico, RT-MAE-8306).
- PEREIRA, C.A.B. & ROGATKO, A. The Hardy - Weinberg
equilibrium under a Bayesian perspective. São Paulo,
IME-USP, 1982. 16p. (Relatório Técnico, RT-MAE-8209).
- WEISBERG, H.I. Bayesian comparison of two ordered
multinomial populations. *Biometrics*, 28:859-867, 1972.
- YATES, F. Tests of significance for 2x2 contingency tables.
Journal of the Royal Statistical Society, A, 147(3), 1984.
(To Appear).