# A INFERÊNCIA FIDUCIAL E SUAS EXTENSÕES

*Marly Grasso Nunes*

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: INFERÊNCIA ESTATÍSTICA

ORIENTADOR: PROF. DR. CARLOS ALBERTO BRAGANÇA PEREIRA

- SÃO PAULO, JUNHO, 1985 -

# ÍNDICE

oOo

## AGRADEÇO

Ao Carlinhos, pela orientação nos acertos e pela experiência de cientista e de mestre, dando liberdade na escolha da direção seguida por esta obra.

Ao Julio, por sua boa vontade, trabalho e valiosos conselhos nos momentos de decisão.

Aos amigos Roberto, Enrique e Daniel, pelo incentivo fraterno, alicerce necessário para a conclusão deste trabalho.

À minha família, que ajudou a aliviar os momentos de tensão e de ansiedade, e pela compreensão e paciência em se ver privada das horas de convívio familiar.

Em especial à Antonia, que muito mais do que datilógrafa, se tornou minha amiga; pela ajuda, paciência e obstinação... com muito carinho.

# SÍNTESE

O objetivo deste trabalho é estudar o método de inferência fiducial e suas extensões, que são os modelos estrutural e funcional. No entanto não pretende esgotar o assunto, mas esclarecer e apresentar reunidos os principais resultados referentes a esse tema.

A introdução se ocupa em expor as principais tendências da Inferência Estatística com a finalidade de tentar clarificar as diferenças entre elas e mostrar de que forma se posiciona a Inferência Fiducial neste contexto.

O capítulo 2 descreve o método da inferência fiducial no caso uniparamétrico e chama a atenção para detalhes significativos desse método, como a propriedade de confiança e os subconjuntos relevantes.

O capítulo 3 discute as circunstâncias em que a inferência fiducial e a Bayesiana coincidem. Mostra a diferença lógica que há entre a metodologia dos intervalos de confiança e a fiducialista, ambas baseadas na propriedade de confiança; mostra também, de que forma atua a propriedade de confiança na verificação da validade de uma inferência fiducial.

O capítulo 4 apresenta vários exemplos no qual o método fiducial se aplica com sucesso.

No caso multiparamétrico a metodologia fiducialista é ainda pouco burilada. Algumas condições gerais foram estabelecidas mas os resultados necessitam de uma maior discussão. As principais condições são apresentadas no capítulo 5.

O capítulo 6 e 7 tratam das *extensões* do método fiducial: modelo estrutural e funcional dando uma visão geral dos resultados obtidos por essas extensões.

oOo

# CAPÍTULO 1

## Introdução

A inferência fiducial, introduzida por Ronald A. Fisher em 1930, foi alvo, desde o seu surgimento, de interpretações confusas e controvertidas devido à forma imprecisa como foi apresentada e desenvolvida. Mesmo no livro "Statistical Methods and Scientific Inference" de 1956, onde Fisher abordou o assunto de forma mais completa, muitas questões foram deixadas em aberto. Como consequência, este tema que suscitou polêmica durante muitos anos foi caindo aos poucos no esquecimento. Mesmo assim, alguns pesquisadores se dedicaram a esclarecer e estender esta idéia. O objetivo deste trabalho é estudar o método de inferência fiducial e suas extensões. Para ilustrar o contexto que motivou o desenvolvimento deste método, uma discussão das principais tendências de Inferência Estatística é apresentada nesta introdução. O método propriamente dito e suas extensões vêm descritos nos capítulos seguintes.

O passo decisivo para o desenvolvimento teórico da Inferência Estatística está associado à fundamentação do cálculo de probabilidades ocorrida no século XVIII. Apesar de alguns pesquisadores já terem utilizado modelos probabilísticos para a análise estatística no século XIX, foi no século XX que a Inferência Estatística como é conhecida hoje teve seu maior desenvolvimento. Esta consiste na obtenção e no desenvolvimento de técnicas que permitam, através das observações dos dados, induzir para o universo (população) avaliando ao mesmo tempo o grau de incerteza. Essa avaliação é feita em termos de probabilidade.

Apesar de o cálculo de probabilidades ser uma disciplina matemática bem fundamentada, o conceito de probabilidade, básico para a construção e interpretação dos resultados estatísticos, é controvertido e possui diversas interpretações. Essas diferentes interpretações foram em parte as responsáveis pelo surgimento e desenvolvimento de diferentes metodologias de inferência estatística, que podem ser essencialmente classificadas em dois grupos: o subjetivista e o objetivista. A descrição desses dois grupos que apresentaremos a seguir é importante para que se entenda em qual contexto surgiu o método fiducial. Alguns entes envolvidos na descrição são listados na sequência.

O presente trabalho está restrito ao contexto da inferência estatística paramétrica, ou seja, desejamos inferir sobre o valor de uma quantidade desconhecida, o *parâmetro* $\theta$, cujos valores possíveis estão restritos a um conjunto conhecido $\Theta$, o *espaço paramétrico*. O ponto de partida para o processo de inferência consiste em se realizar um experimento aleatório cujo resultado (*os dados*) $x = (x_1,...,x_n)$ correspondem a uma observação da *variável (vetor) aleatória* $X = (X_1,...,X_n)$. Os possíveis valores $x$, pertencem ao *espaço amostral* $\mathcal{X}$ e a distribuição de $X$ é caracterizada por uma função de probabilidade pertencente à família $\{f(x/\theta): \theta \in \Theta\}$. No caso discreto $f(x/\theta)$ é a função de probabilidade e no caso contínuo é a função de densidade de probabilidade. Note que uma outra classe de funções pode ser definida como $\{f(x/\theta), x \in \mathcal{X}\}$ e cujos elementos são funções definidas em $\Theta$. O elemento desta classe correspondente ao valor observado $x$, denotado

$$L(\theta;x) = f(x/\theta)$$

é denominado função de verossimilhança e cujo conceito foi introduzido por Fisher em 1920. Convém notar que a função de verossimilhança $L(\theta;x)$ não define necessariamente uma distribuição de probabilidades em $\Theta$, os elementos da classe $\{L(\theta;x);x \in X\}$ são apenas funções de $\theta$ para cada $x \in X$. Esta é de importância fundamental no contexto do método de inferência fiducial, pois é nela que está contida toda a informação que $x$ pode nos dar sobre $\theta$ (esse aspecto foi apontado por Fisher em 1921). Se a função puder ser fatorada em

$$L(\theta;x) = h(\theta,t(x))m(x)$$

onde $h$ e $m$ são funções quaisquer, a função $t(x)$ dos dados é denominada estatística suficiente com relação a $\theta$. Em outras palavras a estatística suficiente resume toda a informação sobre $\theta$ contida nos dados $x$.

A inferência Estatística consiste em tirar conclusões sobre o universo partindo de fatos observados na amostra, ou como no nosso caso, tirar conclusões sobre uma quantidade desconhecida $\theta$ partindo dos dados $x$. Essas conclusões podem estar expressas em termos de proposições do tipo "$\theta$ pertence ao intervado $(a,b)$" ou resumidamente "$\theta \in (a,b)$". Relativamente a essas proposições levanta-se contudo, uma dificuldade: não se pode dizer, contrariamente ao que sucede com as proposições "deduzidas", que são falsas ou verdadeiras. Tão pouco é de admitir-se um estado de completa ignorância. Existe, portanto, em relação a uma proposição "induzida" uma posição intermediária entre

falsa-verdadeira

posição de incerteza que leva a dizer que a proposição é mais ou menos provável.

Em resumo pode admitir-se que há um grau de incerteza em toda a inferência em contraste com o caráter de certeza absoluta que é atribuído às deduções matemáticas. A avaliação do grau de incerteza é medida em termos probabilísticos.

Tanto subjetivistas como objetivistas têm como ponto de partida os dados x e o modelo de probabilidades associado a eles, $f(x/\theta)$. No entanto, a forma como os utilizam está fundamentalmente relacionada ao conceito de probabilidade subjacente e consequentemente os processos de inferência resultantes serão diferentes.

A divisão em dois grupos subjetivistas e objetivistas, é bastante genérica e tem como referência as duas tendências mais conhecidas: a Bayesiana e a Clássica (a primeira é visualizada como subjetivista e a segunda como objetivista). No entanto, o aspecto que diferencia ambas, Bayesiana e Clássica, para além do subjetivismo e objetivismo está fundamentalmente associado à utilização ou não da probabilidade como medida de indução (probabilidade indutiva). Procuraremos esclarecer na exposição que segue o significado de probabilidade indutiva.

Sob a luz de certos fatos conhecidos, que chamaremos de A, existe uma posição de incerteza com relação à proposição "$\theta \in (a,b)$" que chamaremos de B. Para os subjetivistas o grau de incerteza que um indivíduo tem sobre a veracidade da afirmação contida em B pode ser expressa em termos numéricos. Esse valor numérico é a probabilidade. Então $p = P(B/A)$ é uma medida do grau de incerteza que uma pessoa pode ter sobre a ocorrência de B dado o conhecimento de A. Em casos extremos p=1 se B decorre de A e p=0 se A e B se contradizem mutuamente.

A interpretação subjetivista trata a probabilidade em termos de sentimentos de certeza e incerteza, com base na experiência pessoal do indivíduo (os "fatos conhecidos", A é a expressão dessa experiência). Ramsey e de Finetti, independentemente e simultaneamente, no período de 1925 a 1935 contribuíram para a construção de uma definição formal de probabilidade subjetiva com base em critérios de consistência e coerência, desenvolvida mais tarde por outros como Savage e Good.

A chave do método subjetivista é dizer que embora o verdadeiro valor de $\Theta$ seja desconhecido, os possíveis valores que ele pode assumir e que estão em Ⓗ podem ter diferentes graus de incerteza baseado em conhecimentos a priori de cunho pessoal. À essa incerteza associamos uma distribuição de probabilidades de natureza subjetiva no espaço paramétrico chamada distribuição a priori. Essas probabilidades são probabilidades condicionais (condicionadas pela experiência pessoal) (Lindley, 1965 e Good, 1950)(*).

Associando o modelo proposto, $f(x/\Theta)$, com a distribuição a priori de $\Theta$ e usando o teorema de Bayes chegamos a uma distribuição de probabilidades sobre o espaço paramétrico, dita distribuição a posteriori que é uma "calibragem" da nossa incerteza inicial sobre $\Theta$ já que considera a informação inicial (a priori) e aquela adicionada pelos dados.

---

(*) Lindley, D.V. (1965) - Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 1, Probability - Cambridge: C.U.P.

Good, I.J. (1950) - Probability and the Weighing of Evidence. London: Griffin

Com efeito, se denotarmos a distribuição a priori por $g(\theta)$ podemos obter a distribuição a posteriori de $\theta$ através da fórmula de Bayes:

$$g(\theta/x) = \frac{f(x/\theta)g(\theta)}{\int f(x/\theta)g(\theta)d\theta}$$

onde $\int_{\Theta}$ é a integral sobre o espaço paramétrico e $g(\theta/x)$ é a distribuição a posteriori de $\theta$. Esse mecanismo será ilustrado pelo seguinte exemplo:

Exemplo 1.1: Considere uma variável aleatória X com distribuição Uniforme em $(\theta-1/2;\ \theta+1/2)$ onde $\theta$ é desconhecido e pertencente ao espaço paramétrico $\Theta = [10,20]$. A função de verossimilhança é

$$L(\theta;x) = f(x/\theta) = \prod_{i=1}^{n} I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_i)$$

onde $$I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_i) = \begin{cases} 1 \text{ se } x_i \in (\theta-\frac{1}{2};\theta+\frac{1}{2}) \\ 0 \quad \text{caso contrário} \end{cases}$$

Assumindo como distribuição de probabilidades a priori uma distribuição Uniforme em [10,20] e usando a fórmula de Bayes obtemos

$$g(\theta/x) = \frac{\prod_{i=1}^{n} I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_i) \times [\frac{1}{10} \times I_{[10,20]}(\theta)]}{\int_{10}^{20} \prod_{i=1}^{n} I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_i) \times [\frac{1}{10} \times I_{[10,20]}(\theta)]d\theta}$$

$$= \frac{1}{y_1 - y_n + 1} \; I_{(y_n - 1/2;\; y_1 + 1/2)}(\Theta)$$

que corresponde a uma distribuição Uniforme em $(y_n - 1/2; y_1 + 1/2)$ onde $y_i$ é a i-ésima estatística de ordem.

Nossa informação inicial sobre $\Theta$ é modificada pelos dados x através dessa operação que funciona como uma forma de "atualização" dessa informação. A utilização formal do conhecimento a priori é um dos pontos mais polêmicos desse argumento.

Algumas das críticas feitas à utilização de uma distribuição a priori vêm na forma de perguntas como as que seguem:

1) mesmo aceitando o fato de que a "incerteza" pode ser descrita através de uma distribuição de probabilidades, como encontrar o modelo adequado para traduzir essa informação?

2) qual distribuição a priori deve ser utilizada quando não se tem informação? Nesses casos, alguns adotam a distribuição Uniforme quando o espaço paramétrico é um conjunto limitado, como no exemplo 1.1. Mas e se o espaço paramétrico não é um conjunto limitado?

A maior crítica está relacionada com o fato de a distribuição a priori ser uma medida subjetiva (personalizada): duas pessoas com informações diferentes poderão considerar diferentes distribuições a priori e consequentemente obter diferentes distribuições a posteriori. Assim, com os mesmos dados, dois estatísticos poderão fazer diferentes inferências sobre o mesmo parâmetro.

Mas, essa abordagem que utiliza a fórmula de Bayes não exige obrigatoriamente o uso do conceito subjetivista. É também compatível com o conceito lógico de probabilidade desenvolvido por Jeffreys (1961)[1]

Keynes em (1921)[2] motivado por princípios de lógica matemática, na qual a proposição A implica ou refuta outra proposição B, modificou e ampliou esse conceito medindo pela probabilidade a extensão em que A verifica B. Essa medida de probabilidade é uma probabilidade condicional e mede um grau de crença racional, que em distinção da idéia de probabilidade subjetiva é assumida como *única*. É uma quantidade de confiança que é adequado conferir a uma proposição B, à luz da informação ou conhecimento que obtemos de A. Há um e só um grau de implicação p = P(B/A) e é uma medida tão objetiva quanto o conceito frequentista que veremos adiante. Popper (1975)[3] considera o conceito lógico de probabilidade uma variante do conceito subjetivista interpretando os enunciados não psicologicamente,mas logicamente.

Nas palavras de Good (1965)[4]: A logical probability, or... "credibility" (in the terminology of early writers) is a rational intensity of conviction, implicit in the given information, and such that if a person does not agree with it he is wrong" (pág. 77).

As críticas ao conceito lógico de probabilidade ressaltam

---

(1) Jeffreys (1961) Theory of Probability, 3rd Edn. Oxford Clarendon Press.

(2) Keynes, J.M. (1921) Treatise on Probability. London: Mac Millan

(3) Popper, K.R. (1975) A lógica da pesquisa científica: tradução de L. Herberg e O. Silveira da Motta. São Paulo, Cultrix. Ed. da Universidade de São Paulo

(4) Good, I.J. (1965) The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Cambridge, Mass: M.I.T. Press.

em primeiro lugar uma insatisfação com a base de obtenção do valor numérico para a probabilidade. Perguntas do tipo: como medir o grau de crença racional? E como obter uma medida única (não pessoal)? - tornam esse conceito impopular entre os estatísticos modernos.

Existe uma variante entre os trabalhos de Keynes e Jeffreys. No contexto deste estudo citaremos apenas as idéias de Jeffreys no critério para escolha das distribuições que representam ignorância a priori sobre $\Theta$.

Jeffreys, utilizando o princípio da razão insuficiente de Bayes-Laplace[(1)] que conduz à distribuição uniforme para conjuntos fechados na reta e mantendo certas propriedades de invariância dos modelos propostos sugeriu que a distribuição a priori dependesse da natureza de $\Theta$. Para isso propôs os seguintes critérios:

a) $\Theta = (-\infty,\infty)$: distribuição uniforme contínua, assegurando invariância face a qualquer função linear de $\Theta$.

$$g(\Theta) \propto \text{constante} \qquad -\infty < \Theta < \infty$$

---

(1) O princípio da razão insuficiente diz que a probabilidade de dois eventos irá ser igual se nós não temos razão para acreditá-los diferentes. O primeiro a usar esse princípio foi Thomas Bayes em 1763 que postulou que o parâmetro da Binomial, p, tem distribuição Uniforme em (0,1) se nada é conhecido sobre ele, o que é conhecido como postulado de Bayes. Laplace em 1814 usou o princípio citado acima para outros parâmetros além daquele da Binomial. O princípio ficou conhecido como princípio da razão insuficiente de Bayes-Laplace.

b) $\Theta = (0,\infty)$: distribuição uniforme contínua para $\phi = \log\theta$, $\phi \in (-\infty,\infty)$, o que garante coerência com a) e invariância face a qualquer potência $\beta (\beta \neq 0)$ de $\theta$

$$g(\theta) \propto \frac{1}{\theta} \qquad 0 < \theta < \infty$$

Vale notar que $g(\theta)$ em a) e b) são distribuições de natureza imprópria, pois $\int g(\theta)d\theta = +\infty$ e podem ser visualizadas como "funções de ponderação" operando na função de verossimilhança $f(x/\theta)$, originando, após normalização, uma distribuição a posteriori. O exemplo abaixo ilustra este método.

Exemplo 1.2: Consideremos a mesma situação do exemplo 1.1, com exceção que aqui $\Theta = (-\infty,\infty)$ e a distribuição a priori é aquela proposta por Jeffreys

$$g(\theta) = 1 \qquad -\infty < \theta < \infty$$

obtemos como resultado

$$g(\theta/x) = \frac{f(x/\theta) \times 1}{\int_{-\infty}^{\infty} f(x/\theta) \times 1 \, d\theta} = \frac{\prod_{i=1}^{n} I_{(\theta-1/2;\theta+1/2)}(x_i)}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} I_{(\theta-1/2;\theta+1/2)}(x_i) d\theta} =$$

$$= I_{(y_n-1/2;y_1+1/2)}(\theta) \frac{1}{y_1 - y_n + 1}$$

Com a aplicação do princípio da razão insuficiente, $f(x/\theta)$ é proporcional a $g(\theta/x)$ o que revela completa "submersão" da informação a priori pelos dados amostrais.

Até este ponto descrevemos as abordagens subjetiva e objetiva que utilizam o argumento de Bayes o que é em geral conhecido como método Bayesiano. Até aqui a probabilidade é utilizada como medida indutiva (probabilidade indutiva) que gera distribuições a priori no espaço paramétrico, subjetivas ou objetivas (neste contexto, probabilidade indutiva é uma medida "induzida" no espaço paramétrico).

No entanto, nem todos são partidários da utilização de distribuições a priori, subjetivas ou não. É o caso da escola de Neyman-Pearson. Nesta escola isto é reflexo da visão de que probabilidade não pode ser utilizada no sentido indutivo descrito anteriormente. Para os seguidores da escola de N.P. a probabilidade de um evento é uma medida única, mas encarada apenas no sentido frequentista que descrevemos a seguir. Em um experimento repetido um grande número de vezes, imagina-se que um evento cuja probabilidade "verdadeira" é 0,6 ocorra aproximadamente sessenta vezes em cem. Segundo esta interpretação frequentista o enunciado "a probabilidade de obter 5 em um dado é 1/6" não é efetivamente uma afirmação acerca do próximo lançamento; é antes uma afirmação acerca de toda uma classe de lançamentos do qual o próximo é apenas um deles. Nos termos dessa concepção, os enunciados de probabilidade numérica (como o de obter 5 em um dado) só são admissíveis se pudermos oferecer uma interpretação frequentista para eles. E como o parâmetro $\theta$ é uma quantidade fixa não devemos associar qualquer probabilidade diferente de zero ou um à proposição "$\theta \in (a,b)$". Portanto, a probabilidade indutiva não é admitida nessa versão de N.P. da inferência, o que torna inaceitável qualquer distribuição de probabilidade sobre o espaço paramétrico.

Com base apenas na visão frequentista da probabilidade e nos dados x como única informação relevante que permite uma quantificação, além do modelo $f(x/\theta)$, as inferências mais comuns da inferência de N.P. são: estimação pontual, onde se encontra um particular valor $\hat{\theta}$ para $\theta$; estimação por intervalo, encontrando-se o intervalo $(\theta_*, \theta^*)$ contido no espaço paramétrico e associando-se a este um grau de confiança; testes de hipóteses, no sentido de "verificar" se $\theta \in \Theta_0 \subset \Theta$ e associando ao resultado os "erros" de inferência.

Dentro desse contexto nosso interesse está voltado para a teoria de estimação objetiva que se desenvolveu de 1930 a 1935 tendo dois eixos distintos: o de Neyman-Pearson e o de Fisher (responsável pelo método fiducial).

Neyman-Pearson desenvolveram a teoria de testes de hipóteses e de intervalos de confiança (estimação por intervalo) por volta de 1930. Como já mencionado, o uso lógico da probailidade é aqui inaceitável e portanto distribuições de probabilidades sobre o espaço paramétrico são impensáveis.

O processo de estimação por intervalos de confiança consiste em encontrar um intervalo do tipo $(A(X), B(X))$ que, pode ser considerado como um conjunto aleatório contido em $\Theta$. Para um dado valor $\theta_0 \in \Theta$ nós podemos calcular a probabilidade de $(A(X), B(X))$ conter $\theta_0$ como segue:

$$P(A(X) \leq \theta_0 \leq B(X)/\theta = \theta_0) = 1 - \alpha$$

para um $\alpha \in (0,1)$ conhecido e onde

$$P(A(X) \leq \Theta_0 | \Theta = \Theta_0) = \int_{\{x; A(x) \leq \Theta_0\}} f(x|\Theta_0) dx$$

Suponha que esta probabilidade $(1-\alpha)$ é a mesma para todo $\Theta_0 \in \Theta$, então

$P[(A(X),B(X)) \ni \Theta_0 | \Theta = \Theta_0) = 1-\alpha$, para todo $\Theta_0 \in \Theta$. O intervalo $(A(x),B(x))$ onde $x$ corresponde ao valor observado de $X$ é chamado intervalo de confiança para $\Theta$ com coeficiente de confiança $(1-\alpha)$.

<u>Exemplo 1.3</u>: Considere $X_1,\ldots,X_n$ variáveis aleatórias (v.a) independentes identicamente distribuídas (i.i.d.) com distribuição uniforme em $(\Theta-1/2, \Theta+1/2)$. Definimos $A(X) = Y_1 = \min\{X_1,\ldots,X_n\}$ e $B(X) = Y_n = \max\{X_1,\ldots,X_n\}$. Calculemos a probabilidade de $\Theta \in (Y_1,Y_n)$, ou melhor calculemos a probabilidade de $(Y_1,Y_n)$ conter $\Theta$. Então

$$P(Y_1 < \Theta < Y_n/\Theta) = \sum_{K=0}^{n-1} \binom{n}{K} [P(X < \Theta)]^K [1-P(X < \Theta)]^{n-K}$$

$$= \sum_{K=1}^{n-1} \binom{n}{K} [\Theta-\Theta+1/2]^K [1-\Theta+\Theta-1/2]^{n-K}$$

$$= \sum_{K=1}^{n-1} \binom{n}{K} [1/2]^n = \frac{2^{n-1}-1}{2^{n-1}}$$

para qualquer que seja $\Theta \in \Theta$. Dessa forma, $[y_1,y_n]$ é um intervalo de confiança para $\Theta$ com coeficiente de confiança $\frac{2^{n-1}-1}{2^{n-1}} = 1-\alpha$ ; aqui $y_1$ e $y_n$ são os valores observados de $Y_1$ e $Y_n$, respectivamente. Suponha agora que $n=2$, então $\alpha=1/2$ e $(1-\alpha) = 1/2$. Suponha também que os valores observados sejam $y_1 = 2{,}3$ e $y_n = 3{,}0$, então $(2{,}3;3{,}0)$ é um intervalo de confiança para $\Theta$ com coeficiente de confiança 1/2.

Mas, note que $3{,}0-2{,}3 = 0{,}7 > 1/2$ o que significa que $(2{,}3;3{,}0) \subset (\theta-1/2;\theta+1/2)$ contém $\theta$ *com certeza*. Pode-se então concluir que o coeficiente de confiança $(1-\alpha)$ associado ao intervalo de confiança, não é uma medida da probabilidade do intervalo amostral conter $\theta$. Ou seja, dizer que $(A(x),B(x))$ é um intervalo de confiança com coeficiente de confiança $(1-\alpha)$, significa dizer que quando $x$ é observado ou o intervalo contém $\theta$ ou $x$ é um evento com probabilidade $\alpha$. Se repetirmos esse procedimento várias vezes, ou seja, observarmos várias amostras $x$, espera-se que a frequência de vezes que o intervalo amostral gerado irá conter $\theta$ seja de $100(1-\alpha)\%$. Portanto, $(1-\alpha)$ é uma medida de erro do *procedimento* e não de um *específico* intervalo observado.

O uso exclusivo do conceito frequentista de probabilidade e principalmente a falta de um critério que permita medir a precisão final da inferência (a não ser em termos de repetições hipotéticas) são algumas das críticas à esse método. Uma outra crítica está relacionada com o fato de que esse procedimento não utiliza a informação a priori, mesmo nas circunstância em que esta é facilmente quantificável.

A maior divergência entre as escolas Bayesiana e a de Neyman-Pearson (Clássica) está no fato de que para os primeiros a regra de parada é irrelevante, efeito direto do princípio de verossimilhança(*), enquanto que para os seguidores de Neyman-Pearson isto é

---

(*) O princípio de verossimilhança diz que se duas amostras $x$ e $y$ têm verossimilhanças que são funções do mesmo parâmetro $\theta$ e estas são proporcionais, as duas contêm a mesma informação sobre o parâmetro $\theta$.

absurdo, pois estes não aceitam nenhum método inferencial que não considere a técnica de amostragem; segundo essa abordagem a função de verossimilhança é simplesmente um instrumento para a construção de métodos específicos de estimação e testes de hipótese.

As várias atitudes em face do conceito de probabilidade são evidentemente as causadoras básicas dos diferentes princípios de Inferência Estatística. Assim, entre a inferência Bayesiana e a Clássica, a diferença fundamental está no fato que a primeira usa a probabilidade no sentido indutivo (subjetiva ou objetiva) e a segunda não.

Quando se utiliza a probabilidade como grau de crença de uma proposição, "$\theta \in (a,b)$", este grau de crença, seja pessoal ou racional, exige que se encare o conceito de probabilidade além dos limites frequentistas.

R.A. Fisher ao propor a probabilidade fiducial, procurou unir o conceito frequentista de probabilidade com o uso da probabilidade no sentido indutivo. No entanto, Barnett (1973), por exemplo, chamou a atenção para o fato de que Fisher foi muito mais estimulado pela construção de um princípio de Inferência Estatística do que pela tentativa de propor uma nova visão de probabilidade.

Fisher, cuja contribuição para a moderna Estatística é uma das mais importantes, era adepto do conceito de probabilidade frequentista, no entanto a formulação proposta por Neyman-Pearson lhe parecia insuficiente: "wooden attitude". (Fisher, 1956, pg. 9). Nas palavras de Fisher (1956, pg. 44): "all expressions of incertain Knowledge must have the some logical form, namely that a statement of probability". No entanto para ele um método de inferência "do not generally lead to any probability statements about the real world, but to a rational and well-defined measure of reluctance to the acceptance of the hyphotesis they test".

Em outras palavras, Fisher acreditava que nossa incerteza sobre $\Theta$ poderia ser traduzida em termos probabilísticos sem a utilização de distribuições a priori (tendo como base somente os dados). Em primeiro lugar, porque não acreditava que pudéssemos expressar nossa crença racional em hipóteses competitivas através de valores numéricos, somente com base em conhecimentos anteriores, sem que esses valores numéricos fossem subjetivos (e para Fisher, que era objetivista, isto era inaceitável). Em segundo lugar, mesmo a proposta de distribuições à priori, segundo uma perspectiva objetiva de Jeffreys que é a escolha de distribuições a priori Uniformes (usando o princípio da razão de insuficiência de Bayes-Laplace) Fisher questionou apontando o que chamou de "inconsistência" do método Bayesiano.

Sobre essa inconsistência Fisher (1930) coloca: "se p é a probabilidade de sucesso de uma Binomial e esta é "completamente desconhecida", atribuir distribuição a priori para p, não significa "desconhecimento total", pois tomando ao pé da letra, se p é "totalmente desconhecida", $p^2$ também o é. Sendo assim, $p^2$ também deveria ter distribuição a priori Uniforme em (0,1). Mas se supusermos p com distribuição Uniforme, utilizando o cálculo de probabilidades, $p^2$ terá outra distribuição."

Nesse artigo de 1930, cujo título é "Inverse Probability", onde, por probabilidade inversa entende-se a distribuição a posteriori Bayesiana, Fisher combatia o método Bayesiano por considerá-lo inconsistente (devido ao uso de distribuições a priori) e definiu um método que procurou levar em conta os seguintes princípios:

1) Distribuições a priori são inaceitáveis
2) Nossa informação sobre $\Theta$, vinda dos dados, pode ser expressa em termos de probabilidade (frequentista)

Com base nessas idéias formulou o argumento fiducial.

As primeiras exposições do argumento fiducial foram vagas e informais baseadas em exemplos específicos, sem menção aos requisitos essenciais para a sua utilização. Na verdade, acredita-se que a metodologia fiducialista seja mais uma coleção de exemplos do que uma teoria. A abordagem mais completa sobre o assunto encontra-se em um livro de Fisher publicado em 1956 cujo título é "Statistical Methods and Scientific Inference". Neste livro encontra-se, pela primeira vez, os requisitos para o uso do método fiducial. Mas, mesmo assim algumas questõs permaneceram em aberto, como por exemplo: o método passo-a-passo ao qual Fisher se refere como uma forma "rigorosa" de obter inferências fiduciais para o caso multiparamétrico, nem sempre produz resultados únicos.

Em vista disto e da falta de consenso com relação à definição da probabilidade fiducial, não é de surpreender que este método tenha sido alvo de interpretações confusas e controvertidas, desde o seu surgimento em 1930. Este é talvez o capítulo de menor sucesso do trabalho de Fisher, devido à polêmica que suscitou e ainda hoje suscita.

Desde a morte de Fisher em 1962 o interesse sobre o método fiducial vem declinando. Alguns de seus seguidores procuraram esclarecer e extender essa idéia (por exemplo, Bunke (1975), Fraser (1961, 1964, 1968), Hacking (1965), Pedersen (1978), Seidenfeld (1979), Wilkinson (1977)).

Os modelos estrutural e funcional são extensões do método fiducial. São definidos em situações nas quais x pode ser expresso como uma função de $\Theta$ e e, onde e é uma variável aleatória erro, que é a expressão de "fontes indefinidas de variação". Por exemplo: Suponha que x pode ser expressa como $x = \Theta + e$; então e pode ser visualizado como a diferença entre o valor x (que é observável) e o parâmetro $\Theta$ (desconhecido). A variável aleatória e não é observável, portanto tem valor desconhecido, mas têm-se informação sobre o seu comportamento através de uma distribuição de probabilidades que independe de $\Theta$. Os modelos funcional e estrutural são compostos de duas partes:

1) A função que relaciona x com $\Theta$ e e.
2) A distribuição de probabilidades de e, que é independente de $\Theta$.

A base da inferência estrutural e funcional é reverter a relação inicial expressando $\Theta$ como função de x e e. Como o mecanismo probabilístico de e é conhecido e o valor de x é observado obtemos uma "distribuição de probabilidade" para $\Theta$.

As inferências fiducial e estrutural são descritas nos capítulos 6 e 7, somente nos aspectos mais significativos que as relacionam com a idéia que gerou o seu surgimento (a fiducial). Os detalhes específicos da inferência proposta para cada uma dessas extenções do argumento fiducial são omitidos deste trabalho.

## CAPÍTULO 2

# MÉTODO FIDUCIAL: CASO UNIPARAMÉTRICO

Para facilitar a compreensão do método fiducial e a consequente interpretação do conceito de probabilidade envolvida, passamos a expor o mecanismo no qual ele se fundamenta. Em seguida introduzimos o correspondente conceito de probabilidade e os requisitos necessários para sua utilização.

## 2.1. O mecanismo

O método fiducial consiste em encontrar uma distribuição de probabilidades sobre o espaço paramétrico, $\Theta$, baseado na informação fornecida pelos dados x e pelo modelo $f(x/\theta)$. O resultado, a "distribuição de probabilidades" para $\theta$ quando o valor de x é observado, recebe o nome de distribuição fiducial de $\theta$. O termo distribuição fiducial foi adotado por Fisher para diferenciar esta distribuição da distribuição a posteriori obtida pelo método Bayesiano.

Existem duas formas de se encontrar a distribuição fiducial de $\theta$: através de funções de distribuição monótonas e através de quantidades pivotais. Descrevemos a seguir estas duas formas.

### 2.1.1. O método da Função de Distribuição Monótona

Desejamos inferir sobre $\theta$ (parâmetro) que assume valores em $\Theta$ (espaço paramétrico) contido em $\mathbb{R}$.

O único ente que relaciona $\theta$ (desconhecido) com os dados x (observado) é a função $L(\theta;x)=f(x/\theta)$. Assim esperamos que ao se variar o valor de $\Theta$ o valor de $L(\Theta;x)$ se altere. Como consequência intuitiva, ao fatorarmos $f(x/\Theta)$, todo o fator que não se altera com a variação de $\Theta$, pode ser abandonado e somente aqueles fatores que são funções de $\Theta$ precisam ser considerados no processo de inferência.

Seja $t: \chi \to \tau$ uma estatística, onde $\tau = \{t; t(x)=t \text{ e } x \in \chi\}$. Como resultado $f(x/\Theta) = f(x/t,\Theta) f(t/\Theta)$. Além disso se t é uma estatística suficiente $f(x/t,\Theta) = f(x/t)$ e portanto, independente de $\Theta$. Dessa forma, para efeito de inferência, no lugar de trabalharmos com $f(x/\Theta)$ trabalharemos com $f(t/\Theta)$ ou com

$$P(t \leqq t_a/\Theta) = F(t_a/\Theta) = \int_{\{x;\, t(x) \leqq t_a\}} f(x/\Theta)\,dx \,, \qquad t_a \in \tau$$

De forma genérica $F(t/\Theta)$ indica a função de distribuição de probabilidades de t. Um valor conhecido de t, será denotado por $t_a$ e de $\Theta$ por $\Theta_a$.

O próximo passo para que possamos usar a inferência fiducial é considerar as seguintes condições:

a) $\tau \subset \mathbb{R}$, ou seja t tem a mesma dimensão de $\Theta$.

b) $F(t/\Theta)$ é contínua em t e tem mesmo domínio $\Theta$ para todo t.

c) $F(t/\Theta)$ é monótona (crescente ou decrescente) em $\Theta$, com

$$F(t/\Theta) \uparrow 1 \quad \text{se} \quad \Theta\uparrow \text{ e } F(t/\Theta) \downarrow 0 \quad \text{se} \quad \Theta\downarrow .$$

Nestas condições existe uma relação de ordem entre t e $\Theta$ e é possível encontrar através de $F(t/\Theta)$ uma distribuição de probabilidades fiducial para $\Theta$ quando $t=t_a$.

Descrevemos a seguir a construção da distribuição fiducial. Para cada valor de $\alpha \in (0,1)$ vamos definir:

i) $t_\alpha(\theta): \Theta \to \tau$ tal que para cada $\theta_a \in \Theta$, $t_\alpha(\theta_a)$ assume um valor $t \in \tau$, tal que $F(t/\theta_a) = \alpha$

ii) o conjunto de valores de $\theta \in \Theta$ tal que $t_\alpha(\theta_a) = t_a$ será denotado por $\theta_\alpha(t_a)$.

Se as condições

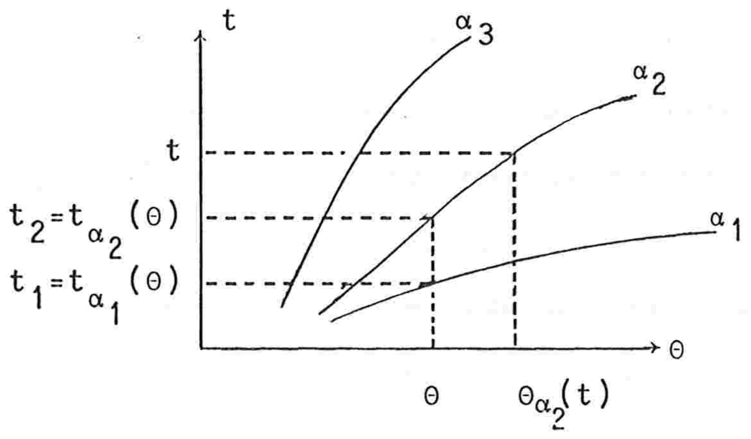
1) $\forall\ \alpha \in (0,1)$ e $t_a$ conhecido, $\theta_\alpha(t_a)$ é unitário e

2) $\forall\ \alpha \in (0,1)$, $t_\alpha(\theta)$ é uma função crescente (ou decrescente)

são satisfeitas, então $F(t/\theta)$ é contínua em t e monótona decrescente (crescente) em $\theta$. Na verdade, as condições 1) e 2) são uma outra forma de escrever as condições a), b) e c) iniciais.

Desejamos saber qual a probabilidade de $\theta \leq \theta_a$ para todo $\theta_a \in \Theta$. Na figura 2.1 representamos a função $t_\alpha(\theta)$ (crescente) para alguns "valores" de $\alpha \in (0,1)$.

Figura 2.1: Gráfico da função $t_\alpha(\theta)$ para $\alpha = \alpha_1, \alpha_2, \alpha_3$. $\tau = \Theta = (0,\infty)$.



Observe que $\alpha_1 < \alpha_2 \Longrightarrow t_{\alpha_1}(\Theta) = t_1 < t_2 = t_{\alpha_2}(\Theta)$ para todo $\Theta$ onde $F(t_1/\Theta) = \alpha_1$ e $F(t_2/\Theta) = \alpha_2$.

Podemos concluir que

$$(2.1.1) \quad \Theta \leqq \Theta_{\alpha_2}(t) \Longleftrightarrow t \geqq t_{\alpha_2}(\Theta)$$

Como t tem distribuição indexada pelo parâmetro $\Theta$, a afirmação $t \geqq t_{\alpha_2}(\Theta)$ nos dá como resultado que

$$P(t \geqq t_{\alpha_2}(\Theta)/\Theta) = 1 - F(t_{\alpha_2}(\Theta)/\Theta) = 1 - \alpha_2$$

Portanto, a afirmação $t \leqq t_{\alpha_2}(\Theta)$ é satisfeita com probabilidade $(1-\alpha_2)$ e como é válida a expressão em (2.1.1), então a primeira desigualdade de (2.1.1) é satisfeita com a mesma probabilidade $(1-\alpha_2)$. E assim

$$P(\Theta \leqq \Theta_{\alpha_2}(t)) = P(t \geqq t_{\alpha_2}(\Theta)/\Theta) = 1-\alpha_2 .$$

Note que (vide figura 1)

$$F(t_{\alpha_2}(\Theta)/\Theta) = F(t/\Theta_{\alpha_2}(t)) = \alpha_2$$

Então para todo $t \in \tau$ podemos definir

$$P(\Theta \leqq \Theta_{\alpha_2}(\Theta)/t) = G(\Theta_{\alpha_2}(\Theta)/t) = 1-F(t/\Theta_{\alpha_2}(t)) = 1 - \alpha_2$$

onde, $G(\Theta/t)$ é a função de distribuição fiducial de $\Theta$ para $t \in \tau$. Chamamos a atenção para o fato que $P(\cdot/t)$ não é uma probabilidade comum, pois t é a única v.a. Quando $t=t_a$

$$G(\Theta/t_a) = 1 - F(t_a/\Theta)$$

A construção anterior foi feita para o caso em que $t_\alpha(\Theta)$ é crescente; no caso em que $t_\alpha(\Theta)$ é decrescente, analogamente teremos

$$\Theta \leq \Theta_{\alpha_2}(t) \iff t \leq t_{\alpha_2}(\Theta)$$

e

$$G(\Theta/t) = F(t/\Theta)$$

Resumindo:

- $F(t/\Theta)$ decrescente em $\Theta \iff t_\alpha(\Theta)$ crescente e

  $G(\Theta/t_a) = 1 - F(t_a/\Theta)$ , quando $t = t_a$

- $F(t/\Theta)$ crescente em $\Theta \iff t_\alpha(\Theta)$ decrescente e

  $G(\Theta/t) = F(t/\Theta)$ , quando $t = t_a$ .

Se além disso $F(t/\Theta)$ é derivável com relação a $\Theta$, podemos obter a densidade fiducial de $\Theta$ que é, para os dois casos acima, igual a

$$g(\Theta/t_a) = \left|\frac{\partial F(t_a/\Theta)}{\partial \Theta}\right| .$$

Vejamos o seguinte exemplo:

<u>Exemplo 2.1.1</u>: Seja $X_1, \ldots, X_n$ v.a. mutuamente independentes com a mesma distribuição (i.i.d) exponencial com parâmetro $\Theta$ desconhecido, assumindo valores em $\Theta = (0,\infty)$. Uma estatística suficiente com relação a $\Theta$ é $t = n\overline{X}$. A distribuição de $t$ é Gamma com parâmetros $(n, 1/\Theta)$ e portanto

$$F(n\bar{x}/\Theta) = \int_0^{n\bar{x}} \frac{(1/\Theta)^n}{(n-1)!} s^{n-1} e^{-s/\Theta} ds$$

que é uma função monótona decrescente em $\Theta$. Então, a distribuição fiducial de $\Theta$ é

$$G(\Theta/n\bar{x}) = 1 - F(n\bar{x}/\Theta)$$

Derivando $F(n\bar{x}/\Theta)$ com relação a $\Theta$, através da regra da cadeia obtemos a densidade fiducial de $\Theta$.

$$g(\Theta/nx) = \left(\frac{n\bar{x}}{\Theta}\right)^{n-1} \frac{e^{-\frac{n\bar{x}}{\Theta}}}{(n-1)!} \left(\frac{1}{\Theta}\right) \quad , \quad \Theta \geq 0$$

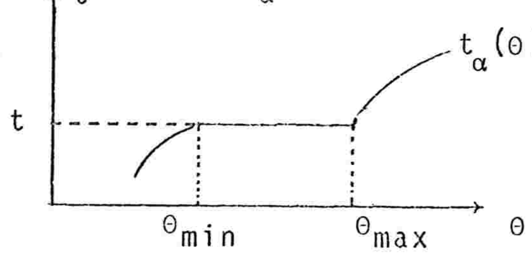
Quando substituimos $n\bar{x}$ pelo respectivo valor observado $n\bar{x}_a$, teremos a densidade fiducial de $\Theta$ quando $n\bar{x} = n\bar{x}_a$.

Como o caso mais comum é aquele em que $F(t/\Theta)$ é monótona decrescente em $\Theta$, toda referência subsequente será feita considerando este caso, a não ser quando mencionado explicitamente. As seguintes observações são de interesse:

I) É fundamental que t tenha a mesma dimensão de $\Theta$. No caso descrito t é suficiente completa.

II) A condição que estabelece que $\Theta_\alpha(t)$ é unitário pode ser "relaxada". Nestes casos teríamos uma relação entre t e $\Theta$ como indicado na figura abaixo.

Figura 2.2: Gráfico de uma função $t_\alpha(\Theta)$ tal que a condição $\Theta_\alpha(t)$ unitário não é válida



Naturalmente, continua válida a expressão (2.1.1). A única diferença nos casos em que $\Theta_\alpha(t)$ não é unitário é que substituímos $\Theta_\alpha(t)$ por $\Theta_{máx} = \underset{\Theta}{máx}\ \Theta_\alpha(t)$.

O raciocínio é análogo nos casos em que $t_\alpha(\Theta)$ assume o mesmo valor para diferentes valores de $\Theta$.

III) Nem sempre o espaço paramétrico $\Theta$, quando $t=t_a$, é o mesmo inicial, às vezes o domínio da distribuição de probabilidade fiducial depende do valor de t.

IV) Para variáveis aleatórias discretas não é aconselhável o uso do método fiducial, pois nestes casos a função $t_\alpha(\Theta)$ não é definida para *todo* $\alpha \in (0,1)$ e todo $\Theta \in \Theta$ e esta condição é fundamental (vide discussão cap. 3).

V) Nos casos em que $F(t/\Theta)$ não converge para zero quando $\Theta$ decresce para o limite inferior de $\Theta$ e/ou $F(t/\Theta)$ não converge à unidade quando $\Theta$ cresce para o limite superior de $\Theta$, a distribuição fiducial resultante é dita incompleta. Neste caso, também não se aconselha o uso do método fiducial. (Note que neste caso $P(\Theta \in \Theta) < 1$).

### 2.1.2 Quantidades Pivotais

Na seção anterior obtivemos a distribuição de probabilidade fiducial de $\Theta$ quando $t=t_a$, para situações em que a função de distribuição de uma estatística suficiente (completa) com respeito a $\Theta$ é monótona em $\Theta$.

Com o uso de quantidades pivotais, generalizamos o mecanismo das funções de distribuição monótonas para os casos em que t tem a mesma dimensão de $\Theta$.

Uma quantidade pinotal $\underline{e}(\theta;t)$ é uma função do parâmetro $\theta$ e da estatística suficiente t, cuja distribuição de probabilidades, P, independe do valor de $\theta$ ou seja

$$\forall\ \theta_1\ ,\ \theta_2 \in \Theta$$

$$P(\underline{e}(\theta_1;t) \in A/\theta_1) = P(\underline{e}(\theta_2;t) \in A/\theta_2)$$

onde A é um evento qualquer contido em $E$=conjuntos dos possíveis valores de $\underline{e}(\theta;t) = \underline{e}$.

As quantidades pivotais podem ser utilizadas para encontrar a distribuição fiducial de $\theta$ quando $t=t_a$, desde que as seguintes condições sejam satisfeitas: Para todo $t \in \tau$,

a') $\underline{e}(t;\theta)$ tem o mesmo domínio $\Theta$

b') $\underline{e}(t;\theta)$ é monótona ou inversível em $\theta$

c') $\underline{e}(t;\theta)$ é derivável em $\theta$.

Nestas condições a densidade fiducial de $\theta$ será

$$g(\theta/t) = p(\underline{e}(\theta;t)/\theta)\left|\frac{d\underline{e}(\theta;t)}{d\theta}\right|$$

onde $p(\underline{e}(\theta;t)/\theta)$ é a função densidade de probabilidade de $\underline{e}$.

O seguinte exemplo serve para ilustrar essas condições:

Exemplo 2.2.1: Sejam $X_1,\ldots,X_n$ v.a.i.i.d com distribuição normal com média $\theta$ e variância 1; neste caso se $\Theta=\mathbb{R}$. A estatística $\overline{X}$ é sufici ente para $\Theta$ $\underline{e}$ $e(\bar{x};\Theta) = \bar{x}-\Theta$ é uma quantidade pivotal, cuja distribuição é normal com média zero e variância $(1/n)$, para todo $\Theta \in \mathbb{R}$. As condições necessárias para a obtenção do método fiducial estão satisfeitas. Portanto, a função de densidade fiducial de $\Theta$ é

$$g(\Theta/\bar{x}) = p(\underline{e}(\bar{x};\Theta)/\Theta)\left|\frac{d\underline{e}}{\partial\Theta}\right| = \frac{\sqrt{n}}{\sqrt{2\pi}}\exp\{\frac{-n}{2}(\bar{x}-\Theta)^2\}\ ,\quad \Theta \in \mathbb{R}$$

que é a densidade de uma normal com média $\bar{x}$ e variância $(1/n)$. Quando $\bar{x} = \bar{x}_a$, $\Theta$ tem densidade fiducial normal com média $\bar{x}_a$ e variância $\frac{1}{n}$.

Note que $\underline{e}(t;\Theta) = F(t/\Theta)$ é uma quantidade pivotal com distribuição Uniforme em $(0,1)$ e as condições

a) $F(t/\Theta)$ tem o mesmo domínio $\Theta$, para todo $t$.

b) $F(t/\Theta)$ é monótona em $\Theta$.

c) $F(t/\Theta)$ é derivável em $\Theta$.

são praticamente as condições da seção anterior, necessitando-se acrescentar apenas que $F(t/\Theta)$ é contínua em $t$ e que $F(t/\Theta)\uparrow 1$ se $\Theta\uparrow$ e $F(t/\Theta)\downarrow 0$ se $\Theta\downarrow$.

A construção da densidade fiducial se dá de forma análoga à seção anterior. Assim, a equivalência (2.1.1) no caso da quantidade pivotal $\underline{e}(t;\Theta) = F(t/\Theta)$ com $F(t/\Theta)$ monótona seria

$$\Theta \leq \Theta_\alpha(t) \iff t \geq t_\alpha(\Theta) \iff F(t/\Theta) \geq F(t_\alpha(\Theta)/\Theta).$$

A segunda equivalência é imediata, pois $F(t/\Theta)$ é uma função de distribuição. A probabilidade fiducial associada a $\Theta \leq \Theta_\alpha(t)$ é a mesma que satisfaz $F(t/\Theta) \geq F(t_\alpha(\Theta)/\Theta)$. Como $F(t/\Theta)$ tem distribuição Uniforme em (0,1), para todo $\Theta$, a desigualdade $F(t/\Theta) \geq F(t_\alpha(\Theta)/\Theta)$ tem a probabilidade fiducial de $\Theta \leq \Theta_\alpha(t)$. A densidade fiducial de $\Theta$ será

$$g(\Theta/t) = p(F(t/\Theta)/\Theta) \left|\frac{\partial F(t/\Theta)}{\partial\Theta}\right| \quad , \quad \Theta \in \Theta$$

onde

$$p(F(t/\Theta)/\Theta) = \begin{cases} 1 \; , \; F(t/\Theta) \in (0,1) \\ 0 \; , \; \text{caso contrário} \end{cases}$$

então

$$g(\Theta/t) = \left| \frac{\partial F(t/\Theta)}{d\Theta} \right| \quad , \quad \Theta \in \Theta .$$

Nos casos em que existe $\underline{e}(\Theta;t)$ que não é a função de distribuição, mas que satisfaz as condições iniciais vale a relação

$$\Theta \leq \Theta_a \iff \underline{e}(t;\Theta) \leq \underline{e}(t;\Theta_a)$$

no caso em que $\underline{e}(\Theta;t)$ é monótona crescente em $\Theta$, para todo $t \in \tau$. E a probabilidade de $\underline{e}(t;\Theta) \leq \underline{e}(t;\Theta_a)$ é a probabilidade fiducial de $\Theta \leq \Theta_a$ , $\forall\ \Theta_a \in \Theta$ .

O raciocínio é análogo para $\underline{e}(\Theta;t)$ monótona decrescente em $\Theta$ e para $\underline{e}(\Theta;t)$ inversível em $\Theta$, para todo $t \in \tau$.

É importante observar que as equivalências

$$\Theta \leq \Theta_a \iff \underline{e}(t;\Theta) \leq \underline{e}(t,\Theta_a)$$

resultam na probabilidade fiducial de $\Theta \leq \Theta_a$ quando $t = t_a$.

Antes de prosseguirmos a discussão vejamos um exemplo em que a quantidade pivotal não seja uma função de distribuição.

Exemplo 2.2.2: Consideremos $X_1,\dots,X_n$ v.a. ii.d com distribuição Normal com média zero e variância $\sigma^2$ desconhecida. Neste caso $\Theta = (0,\infty)$. A estatística suficiente com respeito a $\sigma^2$ é $s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$. A quantidade pivotal $\underline{e}(s^2;\sigma^2) = \frac{(n-1)s^2}{\sigma^2}$ tem distribuição quiquadrado com (n-1) gℓ, denotada $\chi^2_{n-1}$.

Como $\underline{e}(s^2;\sigma^2)$ satisfaz as condições 1), 2) e 3), o resultado será

$$g(\sigma^2/s^2) = p\left\{\frac{(n-1)s^2}{\sigma^2}/\sigma^2\right\} \times \frac{\partial\left[\frac{(n-1)s^2}{\sigma^2}\right]}{\partial\sigma^2} =$$

$$= \frac{(1/2)^{\frac{(n-1)}{2}}}{\Gamma(\frac{n-1}{2})}\left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{(n-3)}{2}} \exp\left\{\frac{-(n-1)s^2}{\sigma^2}\right\}\left[\frac{(n-1)s^2}{(\sigma^2)^2}\right].$$

$$= \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})}\left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n-1}{2}} \exp\left\{\frac{-(n-1)s^2}{\sigma^2}\right\}\left(\frac{1}{\sigma^2}\right)$$

com $\sigma^2 \geqq 0$.

Até agora, descrevemos o método fiducial no caso em que t é suficiente completa com respeito a Θ. Discutiremos a seguir o método fiducial em casos mais gerais, mas antes faremos uma descrição do conceito de probabilidade fiducial.

## 2.2. Conjuntos de referência e o conceito Fisheriano de probabilidade

### 2.2.1 Propriedade de Confiança

Como já foi dito inicialmente o argumento fiducial foi apresentado de forma vaga e imprecisa. Por exemplo em Fisher (1935) colocou o argumento da seguinte forma:

"Suponha que uma amostra aleatória de tamanho n é retirada de uma população normal com média $\mu$ e $\sigma^2$ desconhecidos. Temos que $\bar{x}$ e $s^2$ são estimadores não viciados de $\mu$ e $\sigma^2$ respectivamente. Então a estatística t-Student é

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s}$$

que tem distribuição conhecida t-Student com (n-1)g.ℓ. Então a desigualdade $t > t_1$, que tem probabilidade conhecida, pode ser reescrita como

$$(1) \quad \mu < \bar{x} - s\, t_1/\sqrt{n}$$

e que, portanto, é satisfeita com a mesma probabilidade que a primeira desigualdade $t > t_1$. Para cada valor de $t_1$ teremos a probabilidade de fiducial de (1)."

Segundo a versão acima, o que significa a probabilidade fiducial do intervalo $\mu < \bar{x} - \frac{st_1}{\sqrt{n}}$ ? Segundo Fisher, a proposição $\mu < \bar{x} - \frac{st_1}{\sqrt{n}}$ tem chance de ser verdadeira com probabilidade igual a de $t < t_1$.

Fisher no seu artigo de 1930 encontrou a distribuição fiducial do coeficiente de correlação populacional, $\rho$, de uma normal bivariada com médias iguais a zero e variâncias iguais a um, tendo como base a distribuição do coeficiente de correlação amostral, r. O trecho abaixo reproduz a interpretação de probabilidade fiducial que Fisher dá para o exemplo referido acima.

"Whereas apart from any sampling for $\rho$, we know that if we take a number of samples of 4, from the same or from different populations, and for each calculate the fiducial 5 per cent value for $\rho$, then in 5 per centof cases the true value of $\rho$ will be less than the value we have found".

Note que Fisber usa a mesma linguagem frequentista dos intervalos de confiança e de fato durante muito tempo confundiram-se esses dois conceitos de probabilidade fiducial e de coeficiente de confiança. Muitos pesquisadores interpretaram os intervalos fiduciais $(-\infty,\theta_\alpha(t_a))$ com probabilidade fiducial $(1-\alpha)$ quando $t=t_a$, como intervalos de confiança com nível de confiança $(1-\alpha)$. Fisher era frequentista, mas havia uma diferença fundamental entre sua interpretação e a dos criadores do método dos intervalos de confiança, Neyman-Pearson; esta diferença estava no fato de que ele acreditava na probabilidade indutiva. A proposição "$\theta$ pertence ao intervalo $(-\infty,\theta_\alpha(t_a))$, quando $t=t_a$" tem a seguinte interpretação no contexo de intervalos de confiança: ou a proposição é verdadeira ou ela é falsa, mas nesse caso um evento com probabilidade $\alpha$ foi obtido, paralelamente a interpretação fiducialista é "a proposição é verdadeira com probabilidade $(1-\alpha)$".

A diferença entre o intervalo de confiança e o fiducial, no caso uniparamétrico não parece evidente, a não ser com relação à forma como os intervalos são obtidos e à interpretação de probabilidade indutiva associada ao caso fiducial. A diferença entre ambos fica mais clara no caso multiparamétrico é o caso do teste de Behrens discutido no cap. 5.

Voltamos a comparar os intervalos fiduciais e de confiança uniparamétricos no capítulo 3.

No momento cabe ressaltar que:

1) se observarmos várias amostras de tamanho n, a frequência com que um intervalo (a,b), gerado pela amostra, irá conter $\Theta$ é a probabilidade fiducial.

2) Essa frequência é usada como medida de chance da proposição "$\Theta$ pertence ao intervalo (a,b)" ser verdadeira.

3) A probabilidade fiducial é definida sobre o *conjunto de referência* (definido na sequência).

2.2.2. Conjuntos de Referência

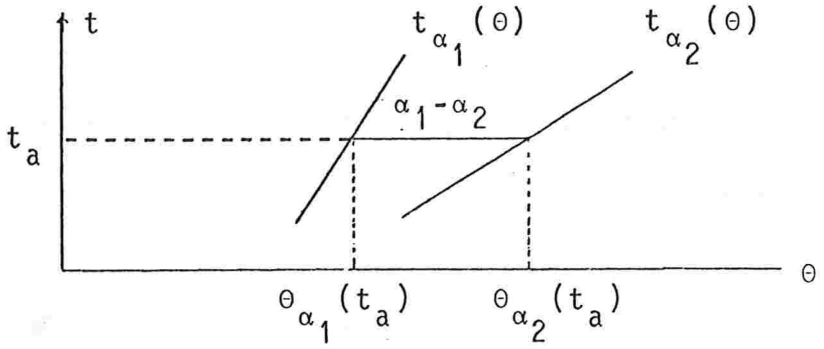
Muita confusão se originou do fato de o método fiducial favorecer a interpretação da densidade fiducial como uma densidade condicional em $t=t_a$. Esta é uma interpretação errada, pois $g(\Theta/t_a)$ não é uma densidade condicional em $t_a$, no sentido comum da palavra.

Vimos, na exposição inicial da Seção 2, que quando t tem distribuição $F(t/\Theta)$ satisfazendo determinadas condições, como monotonicidade em $\Theta$ e continuidade em t, as afirmações do tipo $\Theta \leq \Theta_\alpha(t_a)$ eram equivalentes a $t \geq t_a$ para todo $\alpha \in (0,1)$, onde $\Theta_\alpha(t_a)$ é o conjunto dos valores de $\Theta$ que satisfazem $F(t_a/\Theta) = \alpha$. A probabilidade fiducial de $\Theta \leq \Theta_\alpha(t_a)$ é a probabilidade de $t \geq t_a$.

Observe que as funções $t_\alpha(\Theta)$, $\alpha \in (0,1)$, dividem o plano $(t,\Theta)$ em regiões. Quando $t=t_a$, para cada $\alpha \in (0,1)$ existe um $\Theta_\alpha(t_a)$ que satisfaz $F(t_a/\Theta_\alpha(t_a)) = \alpha$. Se $\alpha_1 > \alpha_2$, existe $\Theta_{\alpha_1}(t_a)$ e $\Theta_{\alpha_2}(t_a)$

(Suponhamos $\Theta_{\alpha_1}(t_a) \neq \Theta_{\alpha_2}(t_a)$ , $\forall\ \alpha_1 \neq \alpha_2$). A probabilidade fiducial de $\Theta_{\alpha_1}(t_a) \leq \Theta \leq \Theta_{\alpha_2}(t_a)$ é $\alpha_1 - \alpha_2$ que é uma "projeção" no eixo $t=t_a$ como indicado na figura abaixo

Figura 2.3: Gráfico de duas funções $t_{\alpha_1}(\Theta)$ e $t_{\alpha_2}(\Theta)$. A probabilidade fiducial do intervalo $(\Theta_{\alpha_1}(t_a), \Theta_{\alpha_2}(t_a))$ é igual a $\alpha_1-\alpha_2$ .




Como $t_\alpha(\Theta)$ é definida para *todo* $\alpha \in (0,1)$, é possível encontrar a probabilidade fiducial de qualquer intervalo $(-\infty,\Theta_\alpha(t_a))$ para um dado $t_a$ e consequentemente, qualquer união ou intersecção de intervalos desse gênero o que resulta na distribuição fiducial de $\Theta$ "condicionado" em $t=t_a$.

Definimos, portanto as probabilidade fiduciais para $\Theta$ tendo como base o conjunto $\mathcal{D}$ dos pares $(t,\Theta)$ onde t tem distribuição $F(t/\Theta)$, que é chamado *conjunto de referência*. De forma genérica $\mathcal{D}$ é um subconjunto de $\tau \times \Theta$ .

<u>Exemplo 2.2.2.1</u>: Suponha $t=\bar{x}$ tendo distribuição normal com média $\Theta$ e variância $1/n$, onde $\tau = \mathbb{R}$ e $\Theta = \mathbb{R}$. O conjunto de referência $\mathcal{D}$ é

o conjunto dos pares $(\bar{x},\theta)$ tal que $\bar{x}$ tem distribuição normal com média $\theta$ e variância $1/n$ e portanto, neste caso, $\mathcal{D} = \tau \times \Theta$.

Exemplo 2.2.2.1: Suponha $t = Y_n = \max\{X_1,\ldots,X_n\}$, $\{X_i\}$ um conjunto de v.a. i.i.d. com distribuição uniforme em $(0,\theta)$. Temos $\tau = (0,\theta)$ e $\Theta = (0,\infty)$. A função de distribuição de $Y_n$ é dada por $F(Y_n/\theta) = [\frac{Y_n}{\theta}]^n$. O conjunto de referência neste caso é

$$\mathcal{D} = \{(y_n,\theta);\ F(y_n/\theta) = [\frac{Y_n}{\theta}]^n\ ;\ y_n \in (0,\theta)\ ,\ \theta > 0\}$$

Note que o conjunto de referência $\mathcal{D}$ é composto de *todos* os possíveis valores de $t$ e de $\theta$, com $t$ tendo distribuição $F(t/\theta)$. É sobre $\mathcal{D}$ que a distribuição fiducial é definida. Quando assumimos $t=t_a$ a densidade fiducial de $\theta$, $g(\theta/t=t_a)$, é uma "projeção" de $g(\theta/t)$ quando $t=t_a$. Ou seja, para *todo* $t=t_a$ existe $\theta_\alpha(t_a)$ tal que $\theta \leq \theta_\alpha(t_a)$ tem probabilidade fiducial $(1-\alpha)$.

É importante ressaltar que no conjunto de referência $\mathcal{D}$ a equivalência (2.1.1) deve ser válida para todo $t \in \tau$, todo $\theta \in \Theta$ e todo $\alpha \in (0,1)$. Se existe algum subconjunto do conjunto de referência onde, na equivalência (2.1.1), $\theta \leq \theta_\alpha(t)$ não resulta em probabilidade fiducial igual a $(1-\alpha)$ para algum $t \in \tau$, este subconjunto é chamado *subconjunto relevante*. Note que se para algum $\theta$ a relação (2.1.1) tem probabilidade fiducial diferente de $(1-\alpha)$, este $\theta$ é "diferente" dos demais e portanto, é "reconhecível". A existência de subconjuntos relevantes significa que há um "acréscimo" ou uma "perda" de informação sobre $\theta$, para além do que os dados fornecem. Isto contraria a imposição de Fisher da não existência de outra informação (a priori) que não seja aquela fornecida pelos dados. Na seção 3.2 enfocamos novamente o problema dos subconjuntos relevantes.

Nos casos que vimos até agora, onde t é uma estatística suficiente e completa com respeito a $\Theta$ e tem a mesma dimensão do parâmetro, aparentemente não temos problemas se as condições necessárias para o uso do método fiducial estiverem satisfeitas. Nos casos em que t não é suficiente completa e não tem a mesma dimensão que $\Theta$, a utilização do método fiducial deve ser cuidadosa, pois deve-se evitar a existência de subconjuntos relevantes. Como o método fiducial está baseado na utilização de estatísticas suficientes é natural que estudemos qual a implicação de usarmos essa "redução" no lugar dos dados originais.

## 2.3. Suficiência e Exaustividade

Como a idéia de Fisher era fazer indução sobre $\Theta$ baseado na informação contida nos dados x, era natural que pensasse em estatísticas suficientes como forma de resumir a informação sobre $\Theta$ contida em x. No entanto para estabelecer uma relação de ordem entre t e $\Theta$ é necessário que ambos tenham a mesma dimensão, permitindo estabelecer equivalências do tipo expresso em (2.1.1). Mas, é suficiente encontrarmos uma estatística t de mesma dimensão de $\Theta$ e fazermos nossa inferência com base em $F(t/\Theta)$? Os exemplos apresentados por Robinson (1975) mostram que podemos encontrar subconjuntos relevantes se não observarmos com cuidado o tipo de estatística t que estamos considerando.

A idéia é a seguinte: se existe uma estatística u tal que a distribuição condicional de t dado u é diferente da distribuição de t sem condicionamento, utilizar a distribuição de t sem condicionamento é perder informação sobre $\Theta$, pois nesse caso faz com que o conjunto dos pontos $(x,\Theta)$ onde x tem distribuição com parâmetro $\Theta$ e $u(x) = u(x_a)$, onde $x_a$ é a nossa mostra, seja um subconjunto relevante.

Suponha que não tenhamos uma estatística t unidimensional e completa com respeito à $\Theta$, mas exista uma transformação bijetora que transforma o conjunto das estatísticas de ordem, $(Y_1,\ldots,Y_n)$ em $t^* = (t,u)$ onde u é uma estatística (n-1) dimensional com distribuição independente de $\Theta$ (ou seja, é ancilar) e t é unidimensional e tem distribuição condicionada em $\Theta$ e u. Então

$$f(x_1,\ldots,x_n/\Theta) \propto f(y_1,\ldots,y_n/\Theta)$$

e

$$f(y_1,\ldots,y_n/\Theta) \propto f(t,u/\Theta) = f(t/\Theta,u)f(u).$$

(Fisher, 1941)

Nestas condições, t é dita suficiente exaustiva ou condicional com respeito a $\Theta$. A densidade fiducial de $\Theta$ será encontrada tomando como base

$$f(t/\Theta,u).$$

Assim temos a mesma idéia exposta na Seção 2.1, em que toda função que não varia com $\Theta$ pode ser eliminada,(que chamaremos de princípio de relevância ou condicionalidade), como é o caso de f(u).

Restringir nossa atenção a $f(t/\Theta,u)$ não acarreta perda de informação sobre $\Theta$ e nos permite estabelecer uma relação de ordem entre $t$ e $\Theta$ já que ambos têm mesma dimensão. O papel de $u$ é determinar o que Fisher chama de configuração da amostra. Lembremos que $f(t/\Theta,u)$, ou melhor,

$$F(t/\Theta,u) = \int_{-\infty}^{t} f(t/\Theta,u)dt$$

deve satisfazer as condições da seção 2.1, para todo $u \in \mathcal{U}$, onde

$$\mathcal{U} = \{u; u(x) = u \text{ com } x \in \chi\}.$$

Para ilustrar e proporcionar uma melhor compreensão, vejamos um exemplo

Exemplo 2.3.1: Sejam $X_1,\dots,X_n$ v.a.i.i.d com densidade

$$f(x_i/\Theta) = (\tfrac{1}{2}) \quad \exp\{-|x_i-\Theta|\} \ , \ x_i \in \mathbb{R}$$

onde $\Theta$ é desconhecido e $\textcircled{H} = R$.

A densidade conjunta de $(X_1,\dots,X_n)$ é

$$f(x_1,\dots,x_n/\Theta) = (\tfrac{1}{2})^n \exp\{-\sum_{i=1}^{n} |x_i-\Theta|\},\ x_i \in \mathbb{R},\ i=1,\dots,n;$$

$i=1,\dots,n$ ; $\Theta \in \mathbb{R}$ desconhecido.

Uma estatística suficiente com relação a $\Theta$, neste caso, é a mediana amostral. Consideremos, para simplificar $n=3$ e assim $t=y_2$ .

Suponhamos que não levássemos em conta que $y_2$ não é suficiente completa e encontrássemos a distribuição fiducial de $\Theta$ com base em

$$f(y_2/\Theta) = \frac{3}{4}\exp\{-2|y_2-\Theta|\}\{2-\exp[-|y_2-\Theta|]\}$$

o resultado seria

$$g(\Theta/y_2) = \frac{3}{4}\exp\{-2|y_2-\Theta|\}\{2-\exp[-|y_2-\Theta|]\} \quad (2.3.1)$$

Agora, levemos em conta que $y_2$ não é suficiente completa, mas que existe uma transformação bijetora de $(Y_1,Y_2,Y_3)$ em $(Y_2,Y_2-Y_1,Y_2-Y_3)$ tal que

$$f(y_1,y_2,y_3/\Theta) = f(y_2,y_2-y_1,y_2-y_3/\Theta) .$$

Note que $t=y_2$ é suficiente exaustiva e $u = (y_2-y_1,y_2-y_3)$ é ancilar com relação a $\Theta$. Portanto

$$f(y_2,y_2-y_1,y_2-y_3/\Theta) = f(y_2/\Theta,y_2-y_1,y_2-y_3)f(y_2-y_1,y_2-y_3)$$

o que nos permite restringir nossa atenção apenas a $f(y_2/\Theta,y_2-y_1,y_2-y_3)$ que é proporcional a

$$\exp\{-|y_2-\Theta|-|y_2-u_1-\Theta| - |y_2-u_2-\Theta|\}$$

onde $u_1 = y_2-y_1$ e $u_1= y_2-y_3$

A distribuição fiducial de $\Theta$ será proporcional a

$$\exp\{-|y_2-\Theta| - |y_2-u_1-\Theta| - |y_2-u_2-\Theta|\} .$$

Quando o valor observado é $(t_a,u_a)$ temos:

$$\begin{aligned} g(\Theta/t_a,u_a) &= g(\Theta/y_{2_a},u_{1_a},u_{2_a}) = \\ &= K\exp\{-|y_{2_a}-\Theta|-|y_{2_a}-u_{1_a}-\Theta|-|y_{2_a}-u_{2_a}-\Theta|\} \\ &= K\exp\{-\sum_{i=1}^{3}|x_{i_a}-\Theta|\} \end{aligned} \quad (2.3.2)$$

que é proporcional à verossimilhança do modelo e onde K é a constante de proporcionalidade que faz com que $\int g(\Theta/t_a,u_a)d\Theta = 1$.

A densidade fiducial obtida em (2.3.1) é evidentemente diferente de (2.3.2). E é fácil ver que

$$\int_{U} f(t/\Theta,u)f(u)du = f(t/\Theta) \ .$$

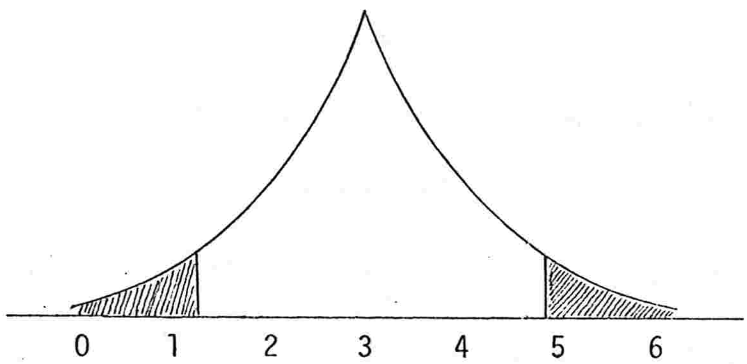
onde $U = \{u;u(x) = u,\ x \in \chi\}$.

Vejamos numericamente qual a importância de considerarmos o condicionamento relativamente a u. Suponhamos que tenhamos duas amostras de tamanho n=3, diferentes:

(2;3;6) e (2,7; 3 ; 3,5)

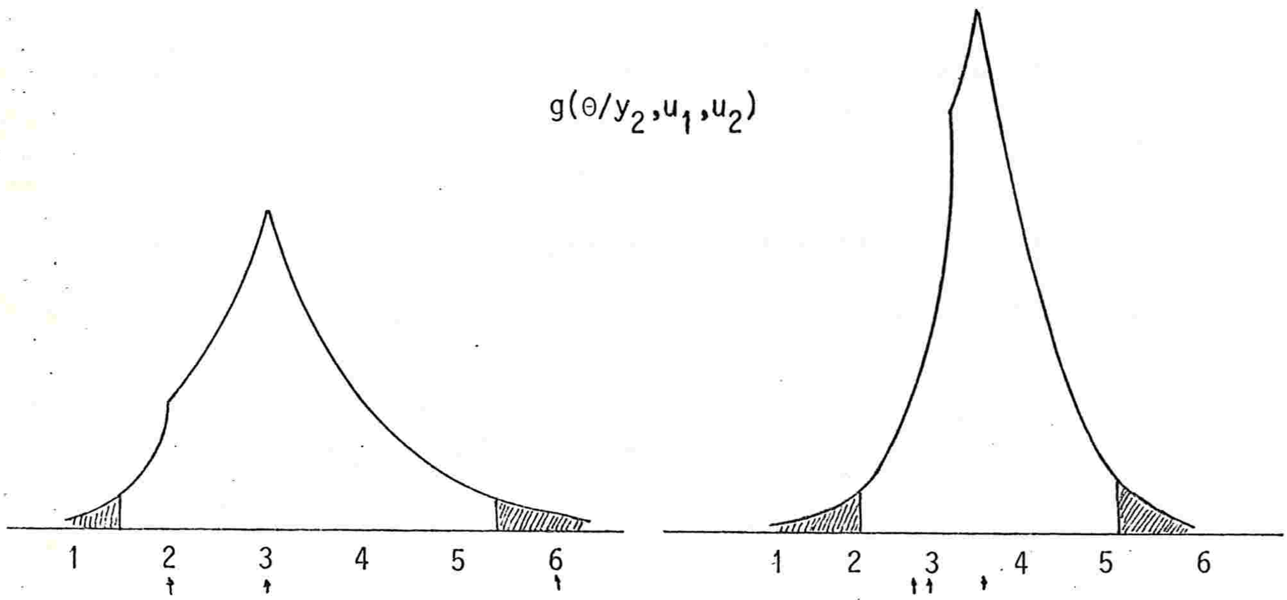
Para essas duas amostras o gráfico da densidade fiducial obtida somente com $y_2$ está representado na Figura 2.4.

Figura 2.4: Gráfico de $g(\Theta/y_2)$ com $y_2 = 3,0$




No entanto, no caso da distribuição (2.3.2), onde a configuração é levada em conta, teremos duas densidades distintas como representadas na Figura 2.5.

Figura 2.5: Gráfico de $g(\Theta/y_2,u,u_2)$ para duas amostras diferentes.




amostra (2;3;6) amostra (2,7; 3; 3,5)

Assim um intervalo do tipo (a,b) que contém $\Theta$ com probabilidade fiducial 0,95 terá seu comprimento determinado pela configuração u (a soma das probabilidades das áreas hachuradas corresponde a 0,05). Isto já não ocorre com $g(\Theta/y_2)$ onde qualquer que seja a amostra, um intervalo (a,b) com probabilidade 0,95 terá sempre o mesmo comprimento.

Essa situação leva à delicada discussão sobre estatísticas ancilares. Não há dúvida que toda a informação relevante está em $f(t/\Theta,u)$ e que se u é ancilar nenhuma informação com respeito a $\Theta$ pode ser extraída de u. No entanto essa estatística dá uma medida da "precisão" de t e abandoná-la significa perder essa medida. Consequentemente haverá perda de informação sobre $\Theta$, o que corresponde na inferência fiducial a gerar conjuntos de referência contendo subconjuntos relevantes. No exemplo acima é correta a densidade fiducial de $\Theta$ expressa em (2.3.2) e incorreta a de (2.3.1).

Dessa forma estamos diante do problema de encontrar estatísticas ancilares adequadas. Lembremos que em alguns problemas a classe das estatísticas ancilares não é um conjunto unitário, o que dificulta a procura da estatística mais apropriada. Fisher (1956) sugeriu que a estatística ancilar fosse encontrada por qualquer método adequado e posteriormente a distribuição fiducial fosse obtida baseando-se em $f(t/\Theta,u)$.

Não é do escopo deste trabalho discutir formas de obter estatísticas condicionalmente suficientes e ancilares. Algumas referências interessantes sobre esse tema são Buehler (1959, 1979), Dawid (1980), Fraser (1979)

Algumas observações importantes referentes ao condicionamento são:

I) Pelo Teorema de Basu: t suficiente completa e u ancilar $\Longleftrightarrow$ t e u são independentes. Nestes casos $f(t/\Theta,u) = f(t/\Theta)$. É o caso dos exemplos da seção 2.1.

II) A estatística ancilar quando existe pode não ser única. Se duas transformações bijetoras, $(t,u)$ e $(t^*,u^*)$ de $(Y_1,\ldots,Y_n)$ existem de tal forma que

$$f(t/\Theta,u) = f(t^*/\Theta,u^*)$$

onde $t$ e $t^*$ condicionalmente suficientes, $u$ e $u^*$ ancilares, então

$$g(\Theta/t,u) = g(\Theta/t^*,u^*).$$

Portanto, para encontrar a densidade fiducial, tanto podemos partir de $f(t/\Theta,u)$ como de $f(t^*/\Theta,u^*)$.

Tomando como base o exemplo 2.3.1 se, ao invés de termos usado $(y_2, y_2-y_1, y_2-y_3)$, tivéssemos usado $(x_1, x_2-x_1, x_3-x_1)$ chegaríamos ao mesmo resultado para a densidade fiducial de $\Theta$. Nos casos em que isto não ocorre não é aconselhável o uso do método fiducial.

III) É importante que o princípio de relevância seja levado em conta, ou seja, que consideremos apenas $f(t/\Theta,u)$ (somente a informação relevante é levada em conta). Um exemplo apresentado por Pedersen na discussão do artigo de Wilkinson (1977), mostra porque devemos nos restringir apenas à informação relevante, e é apresentado na sequência.

<u>Exemplo 2.3.2</u>: Considere a função densidade de probabilidade

$$f(x/\Theta) = \begin{cases} 0{,}8\phi(x-\Theta) & x < 0 \\ 0{,}2 & 0 \leq x \leq 1 \\ 0{,}8\phi((x-1)-\Theta) & x > 1 \end{cases} \tag{2.3.3}$$

onde $\phi(\cdot)$ é a densidade da normal padrão.

Observamos um único valor $x=x_a$. Se considerarmos apenas a monotonicidade em $\Theta$ da função $f(x/\Theta)$ a densidade fiducial resultante é

$$g(\Theta/x_a) = \begin{cases} 0{,}8\phi(\Theta-x_a) + 0{,}2 & x_a < 0 \\ 0{,}8\phi(\Theta-(x_a-1)) & x_a > 1 \\ \text{algo entre } 0{,}8\phi \text{ e } 0{,}8\phi+0{,}2 \text{ se } 0 \leq x_a \leq 1 \end{cases} \tag{2.3.4}$$

No entanto, respeitando o princípio de relevância a distribuição fiducial será

$$(2.3.5) \qquad g(\theta/x_a) = \begin{cases} \phi(\theta - x_a) & x_a < 0 \\ \phi(\theta-(x_a-1)) & x_a > 1 \\ \nexists & 0 \leq x_a \leq 1 \end{cases}$$

Note que se $0 \leq x \leq 1$ , nada se pode dizer sobre $\theta$. A conclusão é que (2.3.5) é a densidade fiducial correta.

IV) Nos casos em que u não é ancilar, encontrar a distribuição fiducial de $\theta$ com base em $f(t/\theta,u)$ é incorreto, pois estamos deixando de levar em consideração $f(u/\theta)$ e nestes casos u irá gerar subconjuntos relevantes.

Nas situações em que mais de um parâmetro está envolvido e estamos interessados em apenas um deles existem propostas para se encontrar estatísticas condicionalmente suficientes e ancilares. Veja Dawid (1980) e Mariotto (1982), por exemplo.

O exemplo do coeficiente de correlação populacional $\rho$, foi durante algum tempo, considerado inadequado, pois existem outros parâmetros envolvidos (se bem que Fisher considera o caso de médias zero e variâncias um).

Fisher em 1915 obteve a distribuição de r, coeficiente de correlação amostral de uma amostra de tamanho n dado $\rho$. Em outras palavras ele mostrou que a distribuição de r depende somente de $\rho$ (e não das médias e variâncias). Barnard (1963) introduziu uma definição formal de suficiência de uma estatística para um parâmetro na ausência de conhecimento dos demais parâmetros. Essa definição é a G-suficiência (Dawid, 1980). Nesta definição, $r$ é suficiente para $\rho$ e portanto a distribuição de $r$ dado $\rho$ pode ser utilizada para encontrar a distribuição fiducial de $\rho$ dado $r$ (que se encontra tabelada em muitos livros de estatística).

Este assunto voltará a ser considerado no caso multiparamétrico.

oOo

# CAPÍTULO 3

## DISCUSSÃO DA DISTRIBUIÇAO FIDUCIAL

Por ter surgido em um período em que Inferência Estatística estava iniciando e devido à forma imprecisa como foi apresentado o método fiducial gerou durante muito tempo várias polêmicas.

Através da distribuição fiducial podemos associar a um intervalo no espaço paramétrico a probabilidade deste conter o parâmetro. Esses intervalos são chamados intervalos fiduciais. A primeira dificuldade com relação ao método fiducial de Inferência surgiu quando os intervalos fiduciais foram confundidos com os intervalos de confiança. Na discussão do artigo de Neyman (1934), Fisher deixou claro, pela primeira vez, a diferença entre ambos. Apesar dos intervalos fiduciais possuirem a propriedade de confiança (seção 2.2.1) esses são essencialmente diferentes dos intervalos de confiança. A propriedade de confiança e a utilização desta na verificação da validade de uma distribuição fiducial uniparamétrica são o assunto da seção 3.2.

Outro ponto controvertido do método fiducial teve origem no fato de Fisher afirmar que a probabilidade fiducial é probabilidade como qualquer outra. O fato de a distribuição fiducial do parâmetro coincidir, em algumas situações, com a distribuição à posteriori Bayesiana, quando a priori utilizada é não informativa, contribuiu para que houvesse confusão. Kendall (1947/51) e Lindley (1957) observaram que é um erro dizer que a distribuição fiducial é uma distribuição de probabilidade no sentido comum. Lindley (1958) mostrou, para o caso uniparamétrico as condições em que ambas, distribuição fiducial e posteriori Bayesiana, coincidem. Somente nos modelos de locação e escala (ou no caso mais geral, em modelos invariantes) é que ocorre a coincidência mencionada; nesse caso a distribuição fiducial pode ser utilizada como distribuição a priori para novas observações sem resultados contraditórios. A discussão desse aspecto da distribuição fiducial está na seção 3.1.

Na seção 3.3 apresentamos algumas propriedades que auxiliam na identificação e interpretação de uma estatística ancilar u. Para que o modelo resultante $F(t/\theta,u)$ possa ser utilizado para encontrar a distribuição fiducial de $\theta$. Essas condições foram estabelecidas por Buehler (1979).

Finalmente apresentamos em 3.4 uma rápida discussão sobre a inadequação do método fiducial no caso em que $F(t/\theta)$ é a distribuição de uma variável aleatória discreta.

## 3.1. Comparação dos resultados Bayesianos e Fiduciais

Um aspecto interessante do método fiducial está relacionado com a comparação dos seus resultados com os resultados obtidos pelo método Bayesiano, já que em ambos encontramos uma distribuição de probabilidades para o parâmetro $\theta$ dadas as observações x. Em alguns casos os métodos fiducial e Bayesiano resultam na mesma distribuição para $\Theta$ dado x. Isto ocorre quando se consideram distribuições à priori não informativas (do tipo priori imprópria proposta por Jeffreys).

Como introdução consideremos dois exemplos relevantes

Exemplo 3.1.1: Sejam $X_1,...,X_n$ v.a. i.i.d. com distribuição $N(\mu,1)$, $\mu \in \mathbb{R}$. A distribuição de $\mu$ condicionada aos dados $x=(x_1,...,x_n)$ pode ser obtida através de:

(A) Método Fiducial - $\bar{X}$ é uma estatística suficiente para x com relação a $\mu$ tal que $\bar{X} \sim N(\mu,1/n)$. Consequentemente a distribuição de $\mu$ é $N(\bar{x},1/n)$ (vide exemplo 2.2.1)

(B) Método Bayesiano - A verossimilhança neste caso é

$$L(\mu;x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\frac{1}{2}[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2]\}$$

Assumindo para $\mu$ a distribuição a priori não informativa $g(\mu) = 1$, $\mu \in \mathbb{R}$, a distribuição a posteriori de $\mu$ será

$$g(\mu/x) \propto L(\mu;x)g(\mu) \propto \exp\{\frac{-n}{2}(\bar{x}-\mu)^2\}$$

que coincide com aquela obtida em (A)

Exemplo 3.1.2: Suponha $X_1,...,X_n$ v.a. i.i.d. com distribuição uniforme em $(0,\theta)$, $\theta \in (0,+\infty)$. Então a distribuição de $\theta$ dado $x=(x_1,...,x_n)$ pode ser obtida através de:

(A) Método Fiducial - $Y_n = \max\{X_1,...,X_n\}$ é uma estatística suficiente com distribuição $F(Y_n/\theta) = (\frac{y_n}{\theta})^n$. A distribuição fiducial de $\theta$ dado $y_n$ é

$$g(\theta/y_n) = (\frac{n}{\theta})(\frac{y_n}{\theta})^n I_{(y_n,+\infty)}(\theta)$$

(B) Método Bayesiano - A verossimilhança associada, neste caso, é

$$L(\theta;x) = (\frac{1}{\theta})^n \prod_{i=1}^{n} I_{(0,\theta)}(x_i) = (\frac{1}{\theta})^n I_{(y_n,+\infty)}(\theta)$$

Utilizando uma densidade de probabilidade a priori não informativa $g(\theta) = \frac{1}{\theta}$ para $\theta > 0$ , a posteriori será igual a

$$g(\theta/y_n) = (\frac{n}{\theta})(\frac{y_n}{\theta})^n I_{(y_n,+\infty)}(\theta)$$

que coincide com aquela obtida em (A).

O seguinte resultado devido a Lindley (1958) estabelece, no caso uniparamétrico, as condições nas quais os dois métodos coincidem, usando uma priori não informativa.

Teorema 3.1: A condição necessária e suficiente para que a distribuição fiducial de $\theta$ e a distribuição a posteriori de Bayes com priori não informativa, coincidam, é que exista uma função bijetora H da estatística suficiente t, $t_*=H(t)$, e uma função bijetora N do parâmetro $\theta, \theta_* = N(\theta)$, tal que $\theta_*$ seja um parâmetro de locação para $t_*$, ou seja, exista uma quantidade pivotal $\underline{e}(t;\theta) = t_*-\theta_*$ .

A demonstração desse teorema encontra-se no apêndice B.

A distribuição fiducial de $\theta$ é $g(\theta/t) = \left|\frac{\partial F(t/\theta)}{\partial \theta}\right|$ (com $F(t/\theta)$, função distribuição de t, obedecendo as restrições da seção 2.1). A distribuição a posteriori Bayesiana com priori não informativa, $n(\theta)$, é $g^*(\theta/t) \propto \frac{\partial F(t/\theta)}{\partial t} n(\theta)$ .

Se existe $\underline{e}(t;\theta) = H(t)-N(\theta)$, uma quantidade pivotal com função densidade de probabilidade

$$p(\underline{e}(t;\theta)) = p(H(t) - N(\theta)) \propto \frac{\partial F(t/\theta)}{\partial t}$$

então a densidade fiducial de $\theta$ é

$$g(\theta/t) = p(\underline{e}(t;\theta)) \left|\frac{\partial \underline{e}(t;\theta)}{\partial \theta}\right| \propto \frac{\partial F(t/\theta)}{\partial t} \left(\frac{\partial N(\theta)}{\partial \theta}\right)$$

Para que $g(\theta/t) = g^*(\theta/t)$ é necessário que $\frac{\partial N(\theta)}{\partial \theta} = n(\theta)$. Estão nestas condições os modelos de locação onde $H(t) = t$ , $N(\theta) = \theta$ e portanto, $n(\theta) = 1$, que é a priori não informativa neste caso.

No caso dos modelos de escala, onde $H(t) = \log t$ e $N(\theta) = \log(\theta)$, a priori não informativa é $n(\theta) = \frac{1}{\theta}$ $(= \frac{\partial \log\theta}{\partial \theta})$.

No caso uniparamétrico não existe outro modelo que esteja dentro das condições do Teorema 3.1, além dos modelos de locação e escala.

Além disso, Lindley (1958) mostrou, ainda no caso uniparamétrico, que as afirmaçãos de Fisher de que a "probabilidade fiducial" é "probabilidade" como qualquer outra (Fisher, 1956, pg. 51) e que a "distribuição fiducial nos dá informação da mesma forma que uma distribuição a priori" (Fisher, 1956, pg. 125) são afirmações falsas. Para tal conclusão, Lindley observou que

(1) Para que a distribuição fiducial possa ser considerada uma distribuição de probabilidades deve obedecer as regras básicas do cálculo de probabilidades.

(2) Se a distribuição fiducial nos dá informação da mesma forma que uma distribuição a priori, deveria poder ser utilizada como tal.

e assim, propôs o seguinte processo:

a) observamos duas amostras $x_1$ e $x_2$ (independentes)

b) usando a amostra $x=(x_1,x_2)$ obtemos a distribuição fiducial de $\Theta$, $g(\Theta/x) = g(\Theta/x_1,x_2)$.

c) usando apenas a amostra $x_1$ obtemos a distribuição fiducial de $\Theta$, $g(\Theta/x_1)$; usando $g(\Theta/x_1)$ como priori para $f(x_2/\Theta)$ encontramos a posteriori Bayesiana $g_1^*(\Theta/x_1,x_2)$.

d) realizamos o mesmo processo descrito em c) trocando a ordem de $x_1$ e $x_2$ e encontrando a posteriori Bayesiana $g_2^*(\Theta/x_1,x_2)$.

Concluiu que somente nos modelos de locação e escala é que vale a igualdade

$$g_1^*(\Theta/x_1,x_2) = g_2^*(\Theta/x_1,x_2) = g(\Theta/x_1,x_2)$$

Se a densidade fiducial de $\Theta$ fosse uma densidade condicional no sentido comum esperar-se-ia que

$$g_1^*(\Theta/x_1,x_2) \propto f(x_2/\Theta)g(\Theta/x_1) \propto f(x_1/\Theta)g(\Theta/x_2) \propto g_2^*(\Theta/x_1,x_2)$$

ou seja, a ordem das amostras não deveria ter efeito sobre o resultado final. O que implicaria também na igualdade acima com $g(\Theta/x_1,x_2)$. Como isto não ocorre em geral, podemos concluir que o método fiducial é inconsistente (no sentido de não obedecer os axiomas básicos do cálculo de probabilidades).

O seguinte exemplo foi utilizado por Lindley para ilustrar esse fato:

Exemplo 3.1.3: Seja X uma v.a. com densidade

$$f(x/\Theta) = \frac{\Theta^2}{\Theta+1}(x+1)\exp(-x\Theta) \quad \text{com} \quad x > 0 \; , \; \Theta \geqq 0$$

Consideremos duas observações independentes, $x_1$ e $x_2$, de X. A distribuição da estatística suficiente $z=x_1+x_2$ é representada pela densidade

$$f(z/\Theta) = \left(\frac{\Theta^2}{\Theta+1}\right)^2 (z+z^2+\frac{1}{6}z^3)\exp(-z\Theta).$$

A distribuição fiducial de $\Theta$ com base em $f(z/\Theta)$ é a densidade

$$(3.1.1) \quad g(\Theta/z) = \frac{\Theta^3}{(\Theta+1)^3}[(2z^2+\frac{4}{3}z^3+\frac{1}{6}z^4)+\Theta(z^2+z^3+\frac{1}{6}z^4)]\exp(-z\Theta)$$

Por outro lado a distribuição fiducial de $\Theta$ com base em $f(x_1/\Theta)$ é

$$g(\Theta/x_1) = \frac{\Theta}{(\Theta+1)^2}[1+(1+\Theta)(1+x_1)]x_1 \exp(-\Theta)$$

e a densidade a posteriori de Bayes utilizando $g(\Theta/x_1)$ como priori e $f(x_2/\Theta)$ como modelo é

$$(3.1.2) \quad g_1^*(\Theta/x_1,x_2) \propto f(x_2/\Theta)g(\Theta/x_1) \propto \frac{\Theta^3}{(\Theta+1)^3}x_1(1+x_2)[1+(1+\Theta)(1+x_1)]\exp\{-(x_1+x_2)\Theta\}$$

Trocando a ordem de $x_1$ e $x_2$ teremos de forma análoga

$$(3.1.3) \quad g_2^*(\Theta/x_1,x_2) \propto f(x_1/\Theta)g(\Theta/x_2) \propto \frac{\Theta^3}{(\Theta+1)^3}x_2(1+x_1)[1+(1+\Theta)(1+x_2)]\exp\{-(x_1+x_2)\Theta\}$$

Note que os resultados (3.1.2) e (3.1.3) são diferentes; a ordem das amostras tem efeito no resultado. Além disso, tanto $g_1^*(\theta/x_1,x_2)$ como $g_2^*(\theta/x_1,x_2)$ são diferentes de $g(\theta/z)$ expressa em (3.1.1).

Os resultados de Lindley (1958) podem ser estendidos ao caso multivariado. Nesse sentido Fraser (1961a) mostrou que para modelos com propriedades de invariância sob um grupo de transformações G, a distribuição fiducial e a distribuição a posteriori Bayesiana com priori não informativa coincidem. Ainda neste mesmo artigo, Fraser mostrou que a distribuição fiducial de $\theta$ condicionada na observação $x_1$ pode ser utilizada como priori para o modelo de uma nova observação $x_2$, produzindo resultados coerentes (no mesmo sentido que Lindley adotou na sua discussão).

A invariância é apresentada no capítulo 6, dentro do contexto dos modelos estruturais. A demonstração de alguns dos resultados citados acima encontram-se no apêndice B.

## 3.2. Propriedade de Confiança e Funções Paramétricas

### 3.2.1 Intervalos de Confiança e a Propriedade de Confiança dos Intervalos Fiduciais

Segundo a expressão (2.1.1) à afirmação $\theta \leq \theta_\alpha(t)$ associamos a probabilidade fiducial $(1-\alpha)$ que tem a interpretação apresentada na seção 2.2.1. Os intervalos fiduciais $(-\infty,\theta_\alpha(t))$ têm, pela própria construção, o que chamamos de propriedade de confiança, como é o caso dos intervalos de confiança, ou seja,

$$P[(-\infty,\Theta_\alpha(t)] \ni \Theta/\Theta_0) = \int\limits_{\{t;\Theta_0\in(-\infty,\Theta_\alpha(t))\}} f(t/\Theta_0)dt = 1-\alpha \quad , \forall\ \Theta_0 \in \Theta$$

onde $\Theta_\alpha(t) = A(t)$, tal que $(-\infty,A(t))$ é um intervalo de confiança com coeficiente de confiança $(1-\alpha)$. Ressaltamos, no entanto, que os métodos de intervalos de confiança e o fiducial podem produzir intervalos para $\Theta$ com a mesma propriedade de confiança, mas que são essencialmente diferentes, não apenas pela forma como são obtidos, mas principalmente pela interpretação que se faz da probabilidade associada ao intervalo (mesmo nos casos em que ambos coincidem numericamente).

Vejamos um exemplo para mostrar a diferença entre esses métodos e suas interpretações:

Exemplo 3.2.1: Consideremos o exemplo 1.2, com $n=2$; a distribuição de $X_1,X_2$ é uniforme em $(\Theta - \frac{1}{2}\ ;\Theta + \frac{1}{2})$. Vamos supor que a amostra observada é $(2,3;3,0)$.

(A) Intervalo de Confiança - Vimos no exemplo 1.2 que o intervalo de confiança $(Y_1,Y_2)$ tem coeficiente de confiança $1/2$. Para a amostra acima esse intervalo é $(2,3;3,0)$. Como já foi comentado, a probabilidade do intervalo conter $\Theta$ não é $\frac{1}{2}$; neste caso $(2,3;3,0)$ contém $\Theta$ com certeza.

(B) Intervalo Fiducial - Usando a técnica descrita na seção 2.3 temos $(Y_1,Y_2-Y_1) = (t,u)$, onde $u=Y_2-Y_1$ é uma estatística ancilar com relação a $\theta$, com distribuição uniforme em (0,1) e $t=Y_1$ é uma estatística com distribuição condicionada em u e $\theta$. Assim

$$f(y_1,y_2/\theta) \propto f(y_1,y_2-y_1/\theta) = f(t,u/\theta);$$

no entanto

$$f(y_1,y_2/\theta) \propto f(x_1,x_2/\theta) = I_{(\theta-\frac{1}{2};\ \theta+\frac{1}{2})}(x_1)\ I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_2)$$

e então

$$f(t,u/\theta) \propto I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_1)\ I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_2)\ .$$

Além disso $f(t,u/\theta) = f(t/y,\theta)f(u)$ , onde $f(u) = I_{(0,1)}(u)$. Para encontrar a distribuição fiducial de $\theta$ notemos que

$$f(t/u,\theta) \propto I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_1)\ I_{(\theta-\frac{1}{2};\theta+\frac{1}{2})}(x_2) =$$

$$= I_{(x_1-\frac{1}{2};x_1+\frac{1}{2})}(\theta)\ I_{(x_2-\frac{1}{2};x_2+\frac{1}{2})}(\theta) = I_A(\theta)$$

onde $A = (x_1 - \frac{1}{2};x_1 + \frac{1}{2}) \cap (x_2 - \frac{1}{2};x_2 + \frac{1}{2})$

Então a densidade fiducial de $\theta$ dado (t,u) é

$$g(\theta/t,u) = g(\theta/y_1,y_2) \propto I_A(\theta).$$

Para a amostra acima: $A = (1,8;2,8) \cap (2,5;3,5) = (2,5;2,8)$. A probabilidade fiducial associada ao intervalo (2,3;3,0) é igual a 1 (um).

Além de os resultados obtidos em cada um dos casos serem diferentes, a probabilidade associada a cada intervalo tem interpretação diferente: no caso dos intervalos de confiança o coeficiente de confiança não pode ser interpretado como a probabilidade de o intervalo conter $\Theta$, ao passo que no caso da probabilidade fiducial essa interpretação é válida.

A propriedade de confiança serve como indicador da validade de uma distribuição fiducial. Se um intervalo $(-\infty,\Theta_\alpha(t))$ não possue propriedade de confiança, ou seja, não tem probabilidade fiducial igual a $(1-\alpha)$ para qualquer que seja $t \in \tau$, isto indica a existência de subconjuntos relevantes (seção 2.2.2). Portanto, distribuições fiduciais sem a propriedade de confiança, em geral não são válidas. No caso multiparamétrico essa questão será analisada no problema de Beherens (capítulo 5).

Na próxima seção a relação entre a propriedade de confiança e a validade da distribuição fiducial será analisada.

## 3.2.2. Funções Paramétricas

Nesta seção mostramos empricamente que em qualquer modelo, inclusive nos modelos de locação e escala, não podemos encontrar a distribuição de uma função qualquer, $\ell(\Theta)$, tomando como base a distribuição fiducial de $\Theta$. Wilkinson (1977) mostrou que se tivermos a distribuição fiducial de $\Theta$, $g(\Theta/t)$, nem sempre podemos usar o cálculo de probabilidades para encontrar $g(\ell(\Theta)/t)$, onde $\ell(\Theta)$ é uma função paramétrica genérica.

Considere o exemplo 3.1.1, onde a distribuição fiducial de $\mu$ é $N(\bar{x},1/n)$. Suponha que estejamos interessados na distribuição de $\mu^2$. Utilizando o cálculo de probabilidades se $\mu \sim N(\bar{x},1/n)$, então $\mu^2$ tem distribuição quiquadrado não central com parâmetro de não centralidade $\bar{x}^2$ e n g.ℓ., $\chi_n^2(\bar{x}^2)$. Por outro lado, como $\bar{x}^2 \sim \chi_n^2(\mu^2)$ podemos encontrar a distribuição fiducial de $\mu^2$ com base nesta distribuição, obtendo um resultado diferente do anterior, como podemos ver através de um exemplo numérico devido a Wilkinson (1977): Se n=50 e $\bar{x}_a^2 = 100$ os intervalos fiduciais centrais com 95% de probabilidade descritos acima são (valores aproximados):

(1) $109 < \mu^2 < 196$ utilizando o fato de $\mu^2 \sim \chi_n^2(\bar{x}_a^2)$

(2) $21 < \mu^2 < 89$ utilizando o fato da distribuição de $\mu^2$ ser obtida de $\bar{x}^2 \sim \chi_n^2(\mu^2)$

Esse exemplo surgiu como uma reinterpretação da demonstração de Stein (1959). Esse autor produziu um exemplo no qual a distribuição fiducial resultante era claramente absurda se julgada pela sua propriedade de confiança: obteve a distribuição fiducial marginal de $\sum_{i=1}^{n} \theta_i^2$ da distribuição fiducial de $\theta_1,\ldots,\theta_n$ dado $x_1,\ldots,x_n$ onde $x_i \sim N(\theta_i,1)$ e mostrou que nesta distribuição marginal os intervalos fiduciais não possuiam propriedade de confiança.

Pedersen (1978) demonstrou que os intervalos gerados por $\mu^2 \sim \chi_n^2(\bar{x}_a^2)$ não têm a propriedade de confiança. O que significa que o intervalo obtido em (1) é incorreto. Através da demonstração que apresentaremos a seguir podemos entender melhor porque a distribuição de $\mu^2 \sim \chi_n^2(\bar{x}^2)$ não é válida.

*Demonstração da não existência da propriedade de confiança dos intervalos fiduciais para* $\mu^2 \sim \chi^2_n(\bar{x}^2)$ - A distribuição de probabilidades $u^2 \sim \chi^2_n(\bar{x}^2)$ foi obtida da distribuição fiducial $\mu \sim N(\bar{x},1/n)$. Então para um valor $\mu_1$ pertencente ao espaço paramétrico teremos

$$\mu \leq \mu_1^2 \iff -\mu_1 \leq \mu \leq \mu_1.$$

Para simplificar façamos n=1 e suponhamos $\mu_1$ um valor tal que fixando $\alpha \in (0,1)$ seja verdadeira a igualdade

$$(3.2.2.1) \quad \alpha = P(-\mu_1 \leq \mu \leq \mu_1/x) = \int_{-\mu_1}^{\mu_1} g(\mu/x)d\mu = \frac{1}{\sqrt{2\pi}} \int_{-\mu_1}^{\mu_1} \exp[-\frac{(\mu-x)^2}{2}]d\mu =$$

$$= \phi(\mu_1-x) - \phi(-\mu_1-x) = \alpha_1 - \alpha_2 .$$

onde $\phi(\cdot)$ é a função de distribuição da normal padrão. Então $\phi(\mu_1-x) = \phi(u_{\alpha_1}) = \alpha_1 \Rightarrow \mu_1 = u_{\alpha_1} + x$ e analogamente $-\mu_1 = u_{\alpha_2}+x$, com $\alpha=\alpha_1-\alpha_2$ , onde $\mu_1$ e $\alpha$ são valores conhecidos. Queremos mostrar que os intervalos do tipo $(-\mu_1,\mu_1)$ com probabilidade fiducial $\alpha$ não são intervalos de confiança com coeficiente de confiança $\alpha$, ou seja,

$$(3.2.2.2.) \quad P\{\mu \in (-\mu_1,\mu_1)/\mu\} = \int_{A(x)} f(x/\mu)dx \; > \alpha \;, \; \forall \mu \in \mathbb{R}$$

onde $A(x) = \{x \in \chi; -\mu_1 = u_{\alpha_2}(x) + x \text{ e } \mu_1 = u_{\alpha_1}(x) + x \;, \; \alpha_1(x) - \alpha_2(x) = \alpha\}$

Podemos reescrever a expressão acima da seguinte forma

$$(3.2.2.3) \quad P\{u_{\alpha_2}(x) \leq \mu - X \leq u_{\alpha_1}(x)/\mu\} = \int_{B(x)} \phi(\mu-x)dx \;, \; \forall \mu \in \mathbb{R}$$

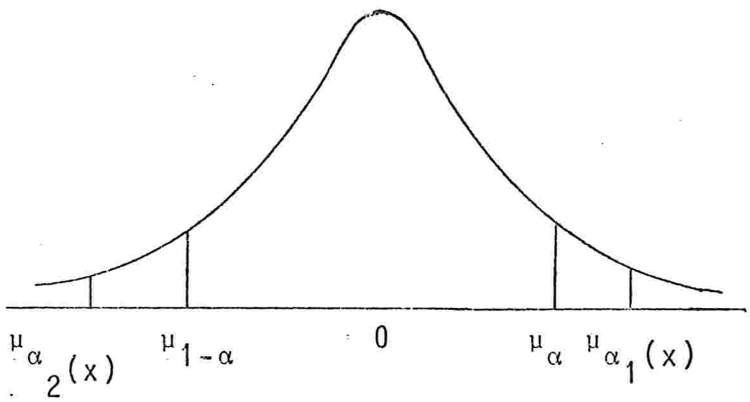
onde $B(x) = \{x \in \chi;\ \alpha_1(x)-\alpha_2(x) = \alpha\}$

Agora notemos que

(1) $u_{\alpha_1}(x) = \mu_1 - x$ e $u_{\alpha_2}(x) = -\mu_1 + x$ são funções decrescentes de x; além disso $u_{\alpha_1}(x) = -u_{\alpha_2}(x) - 2x$

(2) Como $\alpha_1(x) - \alpha_2(x) = \alpha$ então $\alpha_1(x) > \alpha$ e $\alpha_2(x) < 1-\alpha$, $\forall\, x \in \chi$ e a representação dos respectivos quantis $u_{\alpha_1}(x)$ e $u_{\alpha_2}(x)$ com relação a $u_\alpha$ e $u_{1-\alpha}$ está indicada na figura abaixo.

Figura 3.1: Gráfico da função densidade da normal padrão com os respectivos quantis $u_\alpha, u_{1-\alpha}, u_{\alpha_1}(x), u_{\alpha_2}(x)$ onde $\alpha_1(x) - \alpha_2(x) = \alpha$, $\forall\, x \in \chi$ .




Então $u_{\alpha_1}(x) > u_\alpha$ e $u_{\alpha_2}(x) < u_{1-\alpha}$ , $\forall\, x \in \chi$ .

(3) $-\mu_1 = u_{\alpha_2}(x) + x$ ; quando x decresce $(x \to -\infty)$ e $-\mu_1$ é um valor fixado, isto significa que $u_{\alpha_2}(x)$ cresce e como $u_{\alpha_2}(x) < u_{1-\alpha}$, $\forall\, x \in \chi$ , então $u_{\alpha_2}(x) \to u_{1-\alpha}$ quando $x \to -\infty$ .

O raciocínio é análogo para $u_{\alpha_1}(x)$. Então $u_{\alpha_1}(x) \to u_\alpha$ quando $x \to +\infty$.

(4) Por outro lado, como $u_{\alpha_1}(x) = -u_{\alpha_2}(x) - 2x$ então

(i) quando $x \to -\infty$, $u_{\alpha_2}(x) \to u_{1-\alpha}$ e $u_{\alpha_1}(x) \to u_{1-\alpha} - 2x$

(ii) quando $x \to +\infty$ , $u_{\alpha_1}(x) \to u_\alpha$ e $u_{\alpha_2}(x) \to -u_\alpha - 2x$

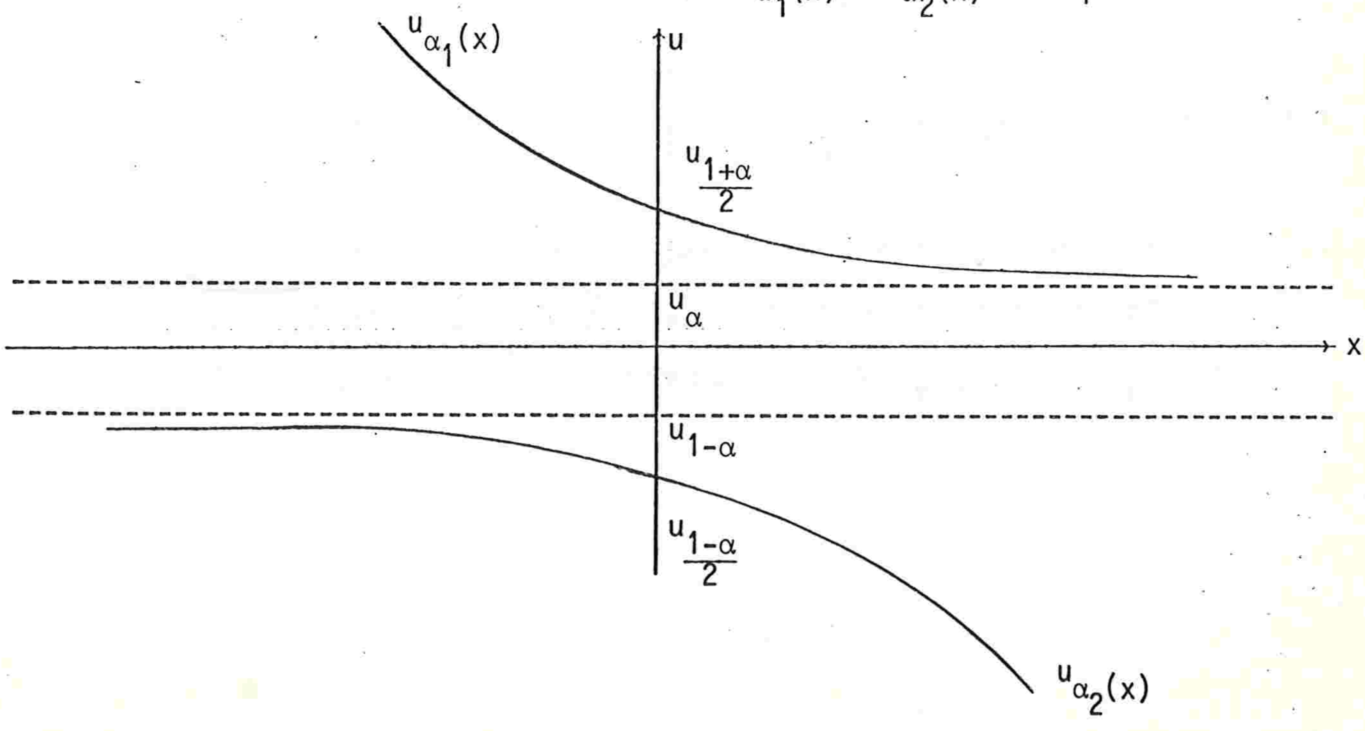
(5) Quando $x=0$, $u_{\alpha_1}(o) = \mu_1$ e $u_{\alpha_2}(o) = -\mu_1 \Longrightarrow -u_{\alpha_1}(o) = u_{\alpha_2}(o)$

o que significa que $\alpha_1(o) = 1-\alpha_2(o)$. E como
$\alpha_1(o) - \alpha_2(o) = \alpha$ então $\alpha_1(o) = \frac{1+\alpha}{2}$ e $\alpha_2(o) = \frac{1-\alpha}{2}$ .

(6) $\dfrac{\partial u_{\alpha_1}(x)}{\partial x} = \dfrac{\partial u_{\alpha_2}(x)}{\partial x} = -1$ e $u_{\alpha_1}(x) = -u_{\alpha_2}(x) - 2x \Rightarrow u_{\alpha_1}(x)$
é convexa e $u_{\alpha_2}(x)$ é côncava.

Com base nos ítens (1) a (6) acima podemos construir os gráficos das funções $u_{\alpha_1}(x) = \mu_1 + x$ e $u_{\alpha_2}(x) = -\mu_1 + x$ representados na Figura 3.2 (com $\mu_1$ e $\alpha$ conhecidos).

Figura 3.2: Gráfico das funções $u_{\alpha_1}(x)$ e $u_{\alpha_2}(x)$ com $\mu_1$ e $\alpha$ conhecidos



Com base nesse gráfico vamos provar que a expressão (3.2.2.2.) é válida. Devemos mostrar que a probabilidade da reta $u = \mu - X$ para $\forall\ \mu \in \mathbb{R}$, estar entre $u_{\alpha_2}(x)$ e $u_{\alpha_1}(x)$ é maior do que $\alpha$.

Caso 1: Façamos $u_{\frac{1-\alpha}{2}} \leqq \mu \leqq u_{\frac{1+\alpha}{2}}$. Pela figura 3.3 vemos que a reta $u=\mu-X$ está entre $u_{\alpha_1}(x)$ e $u_{\alpha_2}(x)$ para todo $x \in \chi$. Então concluimos que

que $P(u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)/\mu) = 1 > \alpha$, para $u_{\frac{1-\alpha}{2}} < \mu < u_{\frac{1+\alpha}{2}}$

Caso 2: Façamos $\mu > u_{\frac{1+\alpha}{2}}$. Pela figura 3.4 vemos que existe um g tal que quando $x > \mu-g$ então $u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)$

Figura 3.3: Gráfico das funções $u_{\alpha_1}(x)$ e $u_{\alpha_2}(x)$ e da reta $u = \mu-X$ quando $u_{\frac{1-\alpha}{2}} \leqq \mu \leqq u_{\frac{1+\alpha}{2}}$

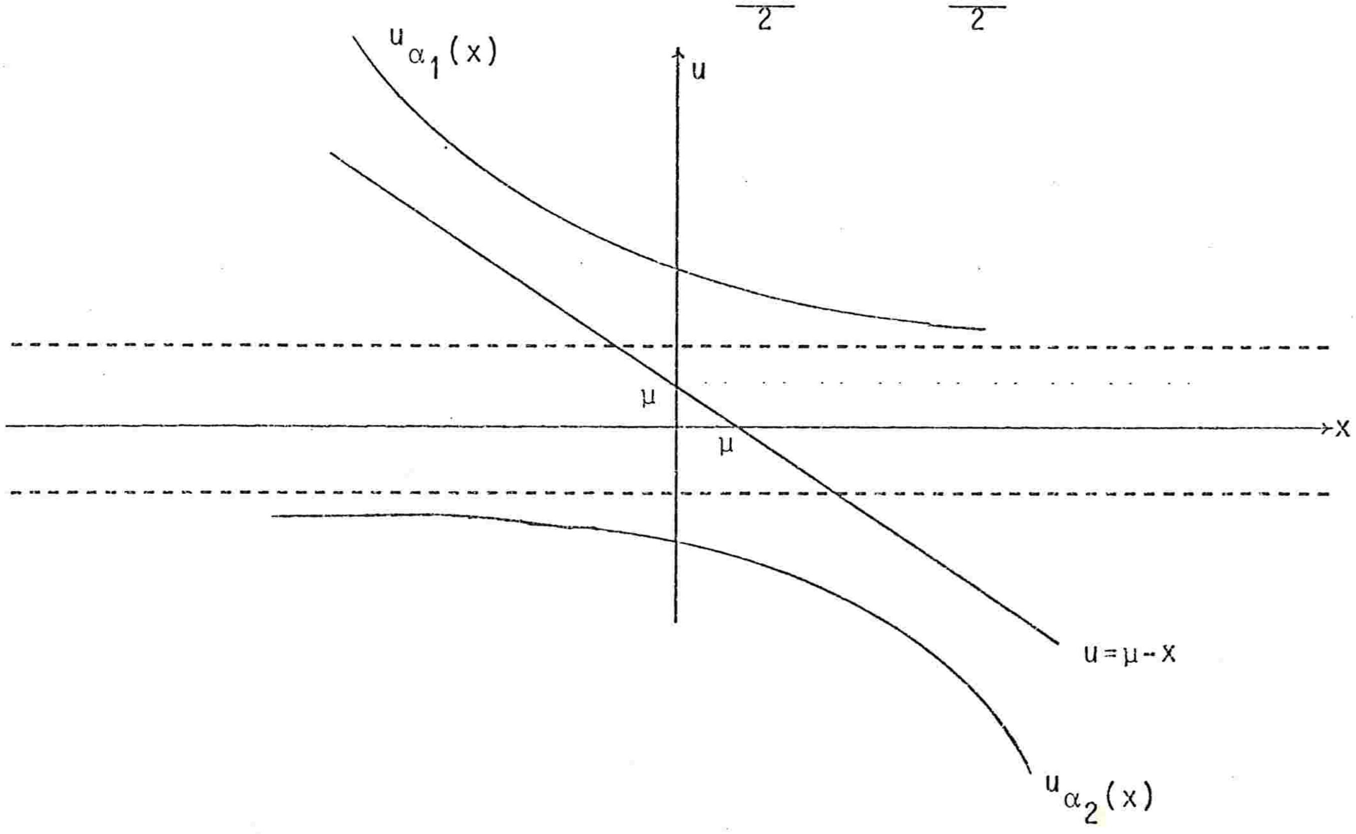
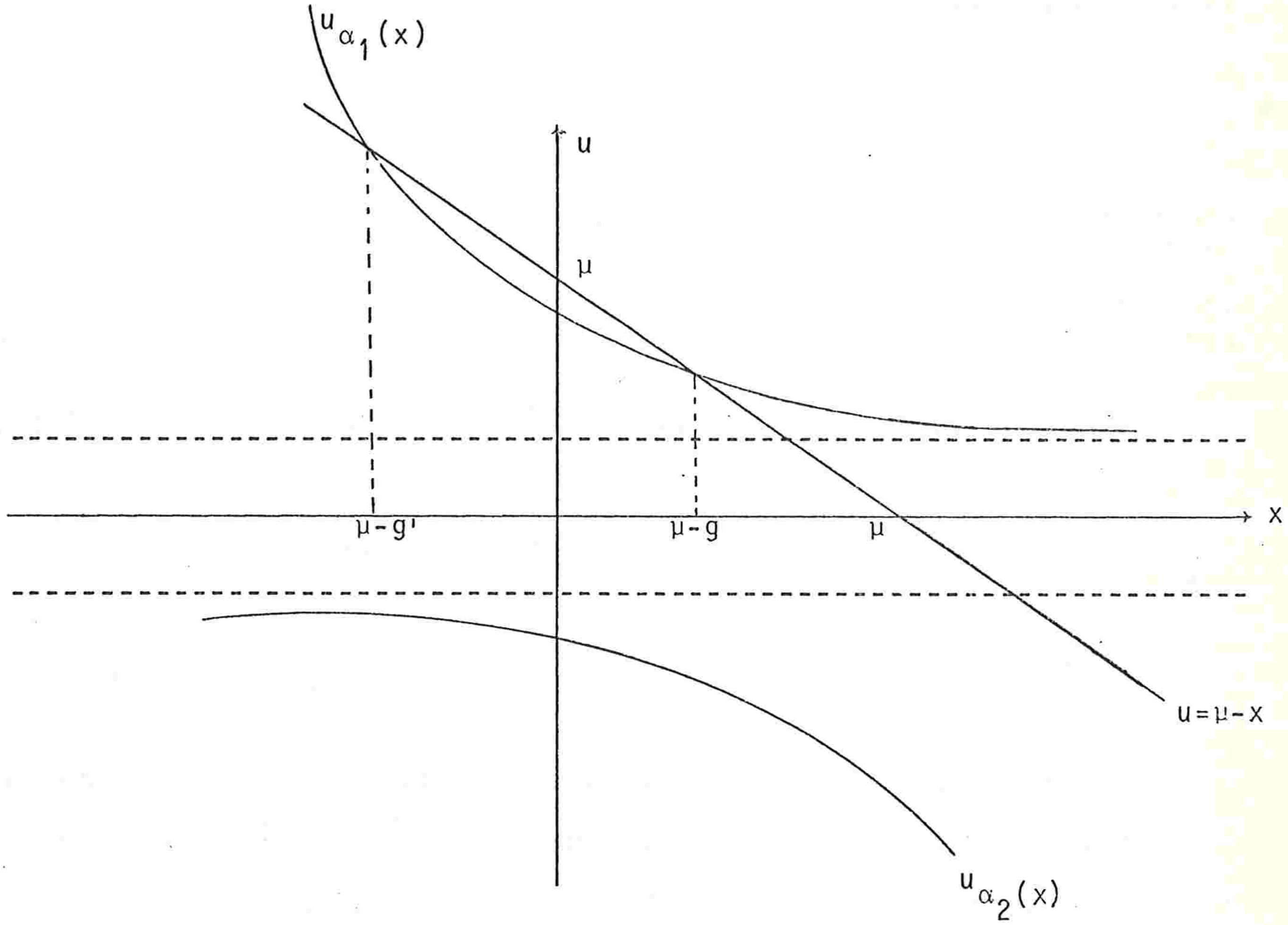


Figura 3.4: Gráfico das funções $u_{\alpha_1}(x)$, $u_{\alpha_2}(x)$ e da reta $u=\mu-x$, quando $\mu > u_{\frac{1+\alpha}{2}}$




Nestas condições quando $\mu > u_{\frac{1+\alpha}{2}}$

$$P(u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)|\mu) > P(X > \mu-g|\mu) = P(\mu-X < g|\mu)$$

Se supusermos sem perda de generalidade que $g > u_{\alpha_1}$, então

$$P(\mu-X < g|\mu) > P(\mu-X < u_\alpha|\mu) = \int_{-\infty}^{u_\alpha} \phi(\mu-x)dx = \alpha$$

consequentemente

$$P(u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)|\mu) > \alpha \text{ , } \mu > u_{\frac{u+\alpha}{2}}$$

Caso 3: Façamos $\mu < u_{\frac{1-\alpha}{2}}$ ; numa demonstração análoga ao caso 2 concluímos que

$$P(u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)|\mu) > P(\mu-X > u_{1-\alpha}|\mu) = \alpha, \quad \mu > u_{\frac{1-\alpha}{2}}$$

Com base nos três casos descritos acima temos que

$$P(-\mu_1 \leqq \mu \leqq \mu_1|\mu) = P(u_{\alpha_2}(x) \leqq \mu-X \leqq u_{\alpha_1}(x)|\mu) > \alpha, \; \forall \; \mu \in \mathbb{R}.$$

ou seja, $(-\mu_1,\mu_1)$ não é um intervalo de confiança com coeficiente de confiança $\alpha$ (note que por simplicidade usamos $\alpha$ ao invés de $1-\alpha$ para a demonstração)

c.q.d.

Em outras palavras, os intervalos $(-\infty,\mu_1^2)$ não possuem propriedade de confiança, indicando a existência de subconjuntos relevantes e portanto de valores paramétricos "reconhecíveis". Neste caso, o valor zero é "reconhecível", pois todas as probabilidades fiduciais de $\mu^2$ são obtidas da distribuição $\mu \sim N(x,1)$ e têm como ponto de referência este valor, ou seja, a probabilidade fiducial de $(-\infty,\mu_1^2)$ é a mesma de intervalos centrados no zero, $(-\mu_1,\mu_1)$, $\forall \; \mu_1 \in \mathbb{R}$. A probabilidade fiducial desses intervalos é obtida a partir de $\bar{x}$. Devemos observar que a estatística $\bar{x}$ é suficiente para $\mu$ , mas não para $\mu^2$. A estatística suficiente para $\mu^2$ é $\bar{x}^2$ e a distribuição fiducial de $\mu^2$ que é encontrada a partir da distribuição de $\bar{x}^2 \sim \chi^2_n(\mu^2)$ tem a propriedade de confiança. Portanto, esta última é a distribuição fiducial de $\mu^2$ válida.

Observamos que a propriedade de confiança vem também da relação de ordem entre o parâmetro e a estatística suficiente. Se tivermos uma estatística suficiente que possui uma relação de ordem com o parâmetro (seção 2.1.1), em geral teremos uma distribuição fiducial válida. No caso discutido aqui, a relação de ordem existente entre $\mu$ e $\bar{x}$ não existe quando confrontamos $\ell(\mu) = \mu^2$ e $\bar{x}$. No entanto, algumas função $\ell(\Theta)$ mantêm a relação de ordem inicial. Somente quando $\ell$ é inversível ou monótona é que podemos calcular a distriuição fiducial de $\ell(\Theta)$, com base na distribuição fiducial de $\Theta$.

<u>Teorema 3.2</u>: Seja t uma estatística suficiente com respeito a $\Theta$ (no caso mais geral t é condicionalmente suficiente) e que tem uma relação de ordem com $\Theta$. Então funções monótonas ou inversíveis de $\Theta, \ell(\Theta)$, mantêm esta relação de ordem.

A distribuição fiducial de $\Theta$ é definida sobre o conjunto de referência $\mathcal{D} = \{(t,\Theta); F(t/\Theta)\}$. Naturalmente quando obtemos a distribuição fiducial de $\ell(\Theta)$ com base na distribuição fiducial de $\Theta$ não podemos alterar a relação de ordem estabelecida inicialmente em $\mathcal{D}$. Assim se um intervalo $(-\infty, \Theta_\alpha(t))$ tem probabilidade fiducial $(1-\alpha)$, para todo $t \in \tau$ e todo $\alpha \in (0,1)$, a função $\ell$ deve preservar esta relação. Dessa forma se $t < t_\alpha(\Theta) \Longleftrightarrow \Theta < \Theta_\alpha(t)$, então a cada interva lo $(t_\alpha(\theta),\infty)$ no espaço amostral associamos um intervalo $(-\infty,\theta_\alpha(t))$ no espaço paramétrico. Para que a relação se mantenha é necessário que a função $\ell$ satisfaça uma das duas condições abaixo:

1) $\ell(\theta)$ é inversível

2) $\ell(\theta)$ é monótona (crescente ou decrescente).

Em (1) a cada intervalo $(-\infty,\theta_\alpha(t))$ associamos um único conjunto $\ell[(-\infty,\theta_\alpha(t))] = \{\ell(\theta);\ \theta \in (-\infty,\theta_\alpha(t))\}$. Em (2) a cada intervalo $(-\infty,\theta_\alpha(t))$ associamos ou $(-\infty,\ell(\theta_\alpha(t)))$ (se $\ell(\theta)$ é monótona crescente) ou $(\ell(\theta_\alpha(t)),\infty)$ (se $\ell(\theta)$ é monótona decrescente).

Para ilustrar vejamos dois exemplos, o primeiro envolvendo funções inversíveis e o segundo envolvendo uma função monótona.

<u>Exemplo 3.2.2.1</u>: Sejam $X_1,\ldots,X_n$ v.a.i.i.d. com função densidade de probabilidade

$$f(x_1,\ldots,x_n|\theta) \propto \exp\{-\frac{1}{2}\sum_{i=1}^{n}|x_i-\theta|\} \qquad x_i \in \mathbb{R},\ i=1,\ldots,n$$

$\theta \in \Theta = \mathbb{R}$. A função de densidade fiducial de $\theta$ é, segundo o exemplo 2.3.1,

$$g(\theta|x) \propto \exp\{-\frac{1}{2}\sum_{i=1}^{n}|x_i-\theta|\} \quad , \ \theta \in \mathbb{R}$$

Se desejamos fazer inferências sobre $\ell(\theta) = 2\theta$ ou $\ell(\theta) = \exp(\theta)$, como ambas são funções inversíveis podemos nos basear em $g(\theta|x)$ usando o cálculo de probabilidades. Assim

(1) Se $\ell(\theta) = 2\theta = \theta^*$ então $g(\theta^*|x) \propto \frac{1}{2}\exp\{-\frac{1}{2}\sum_{i=1}^{n}|x_i-\theta|\}$ , $\theta^* \in \mathbb{R}$

(2) Se $\ell(\Theta) = \exp(\Theta) = \Theta^*$ então $g(\Theta^*|x) \propto \frac{1}{\Theta^*} \exp\{-\frac{1}{2}\sum_{i=1}^{n}|x_i - \ln(\Theta^*)|\}$, $\Theta^* \in R^+$

Exemplo 3.2.2.2: Consideremos a mesma distribuição $g(\Theta|x)$ do exemplo anterior e a seguinte função monótona de $\Theta$:

$$\ell(\Theta) = \begin{cases} 0 \text{ se } \Theta < 0 \\ \\ 1 \text{ se } \Theta \geq 0 \end{cases}$$

Então a distribuição fiducial de $\ell(\Theta)$ pode ser obtida de $g(\Theta|x)$ e é dada por

$$P(\ell(\Theta) = \ell|x) = \begin{cases} G(0|x), \text{ se } \ell = 0 \\ \\ 1-G(0|x), \text{ se } \ell = 1 \end{cases}$$

com $G(0|x) = \int_{-\infty}^{0} g(\Theta|x)d\Theta$.

Os resultados contidos nesta seção indicam, mais uma vez que a probabilidade fiducial não é probabilidade no sentido comum da definição de Kolmogorov.

## 3.3. Algumas propriedades úteis para a identificação de uma estatística ancilar

Suponha que t seja uma estatística suficiente com relação ao parâmetro $\Theta$ e de mesma dimensão deste. Se esta estatística não é completa, então segundo a discussão da seção 2.3, é necessário encontrar uma estatística ancilar u de tal forma que a distribuição condicional $F(t|\Theta,u)$ possa ser utilizada para encontrar a distribuição fiducial de $\Theta$. No entanto, para determinados problemas a estatística ancilar nem sempre existe ou nem sempre é única. Neste último caso, diferentes estatísticas $u_1$ e $u_2$, que geram $F(t|\theta,u_1)$ e $F(t|\theta,u_2)$, podem produzir diferentes distribuições fiduciais.

Nesta seção discutimos algumas propriedades devidas a Buehler (1979) que auxiliam a verificar se a estatística ancilar u escolhida para um determinado problema é minimamente adequada. O modelo $F(t|\theta,u)$ deve satisfazer as condições da seção 2.1.1. para todo valor de u, então a distribuição fiducial de $\theta$ é $G(\theta|t,u) = 1-F(t|u,\theta)$ e a densidade fiducial de $\theta$, que Buehler (1979) chama de densidade induzida no espaço paramétrico é

$$g(\theta|t,u) = \left|\frac{\partial F(t|\theta,u)}{\partial\theta}\right|$$

Os quantis $(1-\alpha)$ denotados por $\theta_\alpha(t|u)$ são análogos àqueles definidos na seção 2.1.1 quando consideramos $F(t|\theta)$ e onde os quantis $(1-\alpha)$ são denotados por $\theta_\alpha(t)$.

<u>Propriedades</u>: P1) $(t,u)$ é suficiente mínima

P2) $\mathrm{Var}(t|u,\theta)$ depende de u, mas não de $\theta$

P3) para dois valores de u, $u_a$ e $u_b$ e quaisquer dois valores de $\theta$, $\theta_a$ e $\theta_b$

$$\mathrm{Var}(t|u_a,\theta_a) < \mathrm{Var}(t|u_b,\theta_a) \Longrightarrow \mathrm{Var}(t|u_a,\theta_b) < \mathrm{Var}(t|u_b,\theta_b)$$

P4) $i(\theta|u) = E[\frac{\partial}{\partial\theta}\log f(t|u,\theta)]^2$ depende de u, mas não de $\theta$

P5) para qualquer $(1-\alpha)$ e u conhecidos o sinal de $\theta_\alpha(t|u) - \theta_\alpha(t)$ é o mesmo para todo t.

P6) $P(\theta \leq \theta_\alpha(t)|u,\theta)$ depende de u, mas não de $\theta$

A propriedade P1 é a condição básica imposta por Fisher para o uso de $F(t|\theta,u)$ (vide seção 2.3). As propriedades P2, P3 e P4, formalizam a afirmação de que a estatística ancilar determina a "precisão" de t. Este fato pode ser ilustrado pelo exemplo 2.3.1.

As propriedades P5 e P6 nos mostram como evitar subconjuntos relevantes. A propriedade P4 nos mostra que se u é adequada a "informação" que foi acrescida por ela é a "mesma" para todo $\theta$. A propriedade P5, quando válida, indica que a estatística u é realmente ancilar. Em outras palavras, considerar a informação de u nos quantis $\theta_\alpha(t)$ altera o valor da probabilidade fiducial associada a $(-\infty,\theta_\alpha(t))$ em função de u, mas não de $\theta$.

Se além das propriedades acima o modelo satisfaz as condições abaixo

C1) $\theta \in \mathbb{R}$, $F(t+c|\theta+c,u) = F(t|\theta,u)$ para $\forall\ c \in \mathbb{R}$ e para todo valor de u.

C2) Existe uma priori imprópria $g^*(\theta)$ tal que qualquer que seja a estatística ancilar escolhida u

$$g(\theta|t,u) = g^*(\theta|x) \quad \text{onde } g^*(\theta|x) \propto f(x|\theta)g^*(\theta)$$

Quando o modelo satisfaz a condição C1 este modelo é de locação. Segundo Lindley (1958) temos que C1 $\Longleftrightarrow$ C2. Além disso, não é difícil mostrar que C1 $\Longrightarrow$ PI , I = 2,3,4,5,6.

<u>Prova</u>: (i) A prova da implicação C1 $\Longrightarrow$ PI, I=2,3,4 é imediata.

ii) C1 $\Longrightarrow$ P5

Temos que $P(\theta \leq \theta_\alpha(t|u)|u,\theta) = P(\theta \leq \theta_\alpha(t+c|u)|u,\theta+c)$, $\forall\ c \in \mathbb{R}$ (note que t é a v.a.). Denotando a função densidade de u por f(u) e $\mathcal{U}$ o conjunto dos possíveis valores de u, temos que

$$P(\theta \leq \theta_\alpha(t)|\theta) = \int_{\mathcal{U}} P(\theta \leq \theta_\alpha(t|u)|u,\theta)f(u)du$$

que por sua vez é igual a

$$P(\theta \leq \theta_\alpha(t+c)|\theta+c) = \int_{\mathcal{U}} P(\theta \leq \theta_\alpha(t+c|u)|u,\theta)f(u)du.$$

Então, $\theta_\alpha(t|u)-\theta_\alpha(t)$ tem o mesmo sinal de $\theta_\alpha(t+c|u)-\theta_\alpha(t+c)$ $\forall\ c \in \mathbb{R}$.

iv) C1 $\Longrightarrow$ P6

Sabemos que $P(\theta \leq \theta_\alpha(t)|\theta) = P(\theta \leq \theta_\alpha(t+c)|\theta+c)$, $\forall\ c \in \mathbb{R}$ (note que $F(t|u,\theta) = F(t+c|u,\theta+c)$ é válida para todo valor de u, portanto $F(t|\theta) = F(t+c|\theta+c)$). Então $P(\theta \leq \theta_\alpha(t)|u,\theta) = P(\theta \leq \theta_\alpha(t+c)|u,\theta+c)$ depende de u mas não de $\theta$.

Então para os modelos de locação e escala onde P1 é válida, está garantido que a estatística ancilar u determina a "precisão" de t e não forma subconjuntos relevantes. Em outros modelos não podemos garantir que exista uma estatística ancilar que satisfaça as propriedades PI, I=2,...,6.

O exemplo apresentado a seguir é de um modelo que não satisfaz a condição C1. Portanto a escolha de u deve ser cuidadosa, pois o modelo resultante pode não satisfazer as propriedades listadas nesta seção.

Exemplo 3.3.1: Considere uma observação $(x_1,x_2)$ de uma normal bivariada com médias zero, variâncias unitárias e correlação $\theta$ (desconhecida). Uma estatística suficiente com relação a $\theta$ é $t=x_1x_2$ e uma estatística ancilar é $u=x_1$ (ou $x_2$) (vide Mariotto (1982)). É claro que $f(t|u,\theta) \propto f(x_1x_2|x_1,\theta)$. Pode-se mostrar que $\mathrm{Var}\ (x_1x_2|x_1,\theta) = x_1^2(1-\theta^2)$ (De Groot, 1975 pg. 250)(1), o que contraria P2. Portanto, $u=x_1$ não é uma estatística ancilar adequada ao problema.

Além disso, se existir uma transformação $\ell(\theta)$ (a mesma para todo u) podemos definir $PI\ell$, $I=1,\ldots,6$, tal que $PI\ell$ é análoga a PI substituindo $\theta$ por $\ell(\theta)$. O mesmo pode ser feito com relação a condição C1. Dessa forma, incluímos nesta discussão os modelos de escala.

## 3.4. Método fiducial para Variáveis Discretas

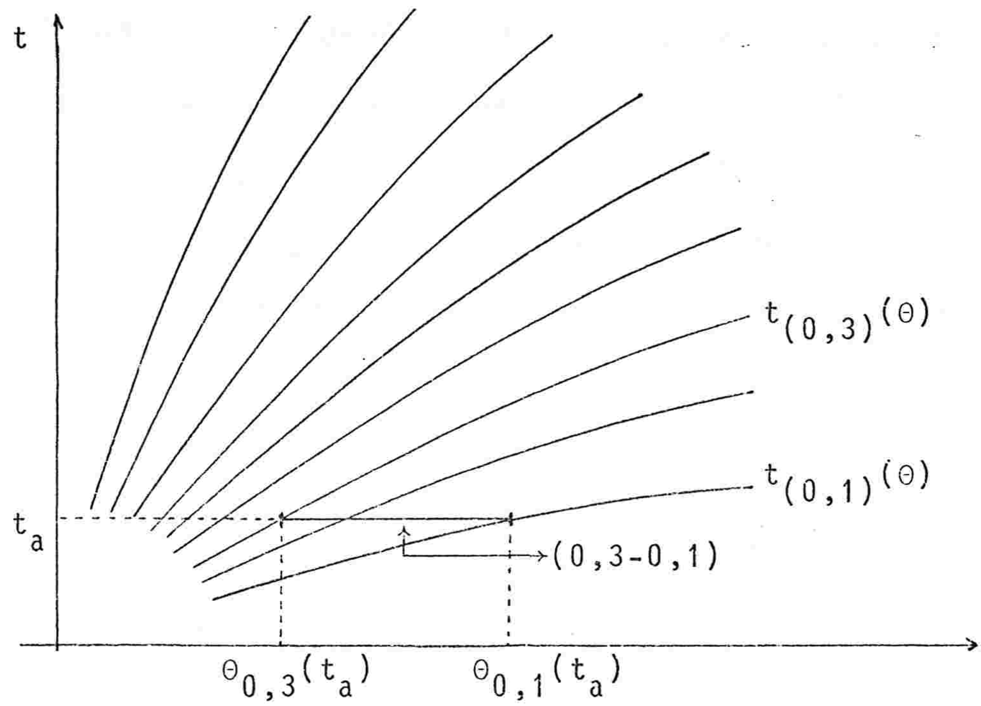
Na seção 2.1.1 vimos que para que o método fiducial possa ser utilizado é necessário que a condição $t_\alpha(\theta)$, tal que $F(t_\alpha(\theta)|\theta)=\alpha$ esteja satisfeita para todo $\theta \in \Theta$ e todo $\alpha \in (0,1)$. Em outras palavras, isto significa que t tem distribuição contínua.

---

(1) De Groot, M.H. (1975) *Probability and Statistic*. Addison - Wesley: Califórnia, 321 p.

Suponha que consideremos os valores de $\alpha$ estejam contidos em $A = \{0,1;0,2;...;0,8;0,9\}$ ; as funções $t_\alpha(\theta)$, $\alpha \in A$ dividem o plano $(t,\theta)$ em 10 regiões. Nestas condições quando $t=t_a$ o intervalo $(\theta_{\alpha_2}(t_a), \theta_{\alpha_1}(t_a))$ onde, por exemplo, $\alpha_1 = 0,3$ e $\alpha_2 = 0,1$, tem probabilidade fiducial como indicado na figura abaixo (que é $\alpha_1-\alpha_2=0,3-0,1=0,2$)

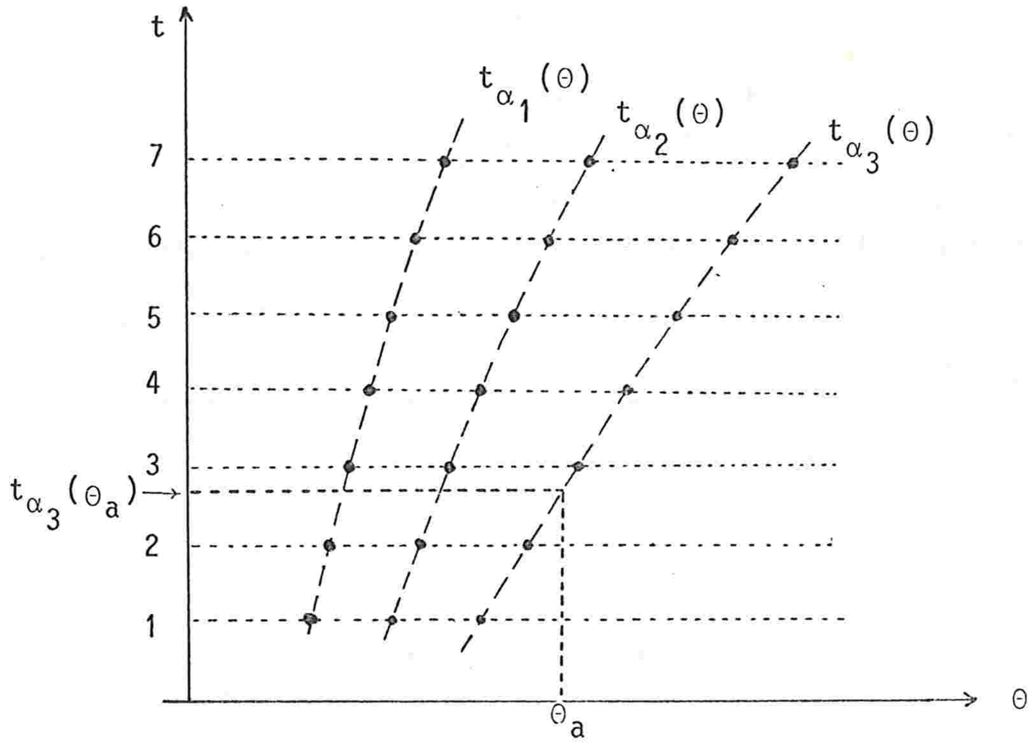
Figura 3.6: Gráfico das funções $t_\alpha(\theta)$ com $\alpha \in A$, onde $A = \{0,1;0,2;...,0,9\}$




Como $t_\alpha(\theta)$ é definida para todo $\alpha$ se incluirmos todas as funções $t_\alpha(\theta)$ para $\alpha \in (0,1)$, então quando $t=t_a$ podemos encontrar a probabilidade fiducial de $(\theta_{\alpha_i}(t_a);\theta_{\alpha_j}(t_a))$ que é $(\alpha_i-\alpha_j)$, $\forall\ \alpha_i,\alpha_j \in (0,1)$.

No caso em que t tem distribuição de probabilidades discreta $t_\alpha(\theta)$ *não* é definida para todo $\alpha \in (0,1)$ e todo $\theta \in \Theta$. Ou seja, para valores fixados de $\alpha$ e de $\theta$, *nem sempre é possível* encontrar um $t_\alpha(\theta)$ *dentre os possíveis valores de* t tal que $F(t_\alpha(\theta)|\theta) = \alpha$. A figura abaixo ilustra esta situação.

Figura 3.7: Gráfico de três funções $t_\alpha(\Theta)$ para $\alpha=\alpha_1,\alpha_2,\alpha_3$ no caso em que $t_\alpha(\Theta)$ só está definida para valores inteiros




Por este motivo, usar o método fiducial nestas condições não é aconselhável. No entanto, em alguns casos soluções aproximadas foram propostas como é o caso do exemplo abaixo, apresentado por Fraser (1968):

<u>Exemplo 3.4.1</u>: Considere que $X_1,\ldots,X_n$ v.a.i.i.d. com distribuição de Poisson com parâmetro $\Theta$ desconhecido. A distribuição da estatística suficiente com relação a $\Theta$ , $t = \sum_{i=1}^{n} X_i$, tem distribuição de Poisson com parâmetro $n\Theta$ . Então

$$P(t=K) = \frac{(n\Theta)^k}{K!} \exp(-n\Theta) \quad , \quad K=0,1,\ldots$$

Esta distribuição está fora das restrições do método fiducial. No entanto, se fizermos uma "aproximação de continuidade" para este caso, onde t é visualizada como uma variável aleatória contínua com função densidade dada pela expressão abaixo

$$(3.4.1) \quad f(t|\Theta) = \frac{(n\Theta)^t}{\Gamma(t+1)} \exp\{-n\Theta\} \;,\; t > 0$$

(segundo Fraser, 1968, pg. 260) então a função densidade fiducial de $\Theta$ pode ser encontrada tendo como base a expressão (3.4.1).

Soluções como esta são aceitáveis, segundo Fisher (1956), se o valor de n é grande.

oOo

# CAPÍTULO 4

## EXEMPLOS

O objetivo deste capítulo é apresentar as densidades fiduciais relacionadas com os modelos estatísticos mais comuns. Na maioria das vezes, o parâmetro envolvido ou é de locação ou é de escala. Mesmo quando este não é o caso, algumas vezes as técnicas de cálculo são semelhantes. Os exemplos aqui relacionados apresentam estas características e servem para ilustrar os tipos de problemas onde o método fiducial se aplica com sucesso.

Os cálculos das densidades fiduciais não estão apresentados por acreditarmos que estes são simples exercícios de aplicação dos procedimentos descritos nas seções precedentes.

Para os modelos nos quais o parâmetro $\theta$ é de locação ou de escala (mesmo certos casos onde $\theta = (\theta_1, \theta_2)$ sendo $\theta_1$ de locação e $\theta_2$ de escala), Lindley (1958) mostrou que a densidade fiducial é uma densidade a posteriori quando a densidade a priori é não informativa (imprópria). Assim, nestes casos, a densidade fiducial é proporcional ao produto da verossimilhança pela priori $g(\theta)$. No caso em que $\theta$ é de locação,

$$g(\theta) = 1 \quad , \quad \theta \in R .$$

No caso em que $\theta$ é de escala,

$$g(\theta) = \begin{cases} 1/\theta & \text{se} \quad \theta > 0 \\ 0 & \text{se} \quad \theta \leq 0 . \end{cases}$$

Recordando a notação, temos que $x = (x_1, \ldots, x_n)$ são os dados, $(Y_1, \ldots, Y_n)$ é o vetor das estatísticas de ordem e $\underline{e}(x;\theta)$ é uma quantidade pivotal cuja densidade comum a todo $\theta$ é denotada por

$p(\underline{e})$; isto é, $p(\underline{e}) = f(\underline{e}(x;\theta)|\theta)$.

## 4.1. Parâmetro de Locação

Nos modelos onde $\theta$ é um parâmetro de locação, o vetor $(x_1-\theta,\ldots,x_n-\theta)$ é uma quantidade pivotal e a densidade conjunta satisfaz a igualdade

$$f(x|\theta) = p(x_1-\theta,\ldots,x_n-\theta).$$

Assim, com a observação de x, a densidade fiducial de $\theta$ é dada por:

$$g(\theta|x) = \frac{p(x_1-\theta,\ldots,x_n-\theta)}{\int_{-\infty}^{\infty} p(x_1-\theta,\ldots,x_n-\theta)d\theta} .$$

No caso em que $x_1,\ldots,x_n$ são observações i.i.d,

$$g(\theta|x) \propto \prod_{i=1}^{n} p(x_i-\theta)$$

onde agora p é a densidade comum de $x_i-\theta$, $i=1,\ldots,n$. Uma outra representação pode ser obtida ao se considerar o vetor $(y_1,y_2-y_1,\ldots,y_n-y_1) = (y_1,u_2,\ldots,u_n)$. A estatística $(u_2,\ldots,u_n)=u$ é ancilar e a estatística $y_1$ é condicionalmente suficiente dado u. Note que $\prod_{i=1}^{n} p(x_i-\theta) = p(y_1-\theta)\prod_{i=2}^{n} p(y_1-\theta+u_i)$.

Exemplo 4.1.1: Sejam $x_1,\ldots,x_n$ observações independentes de uma uniforme em $(\theta-(1/2),\theta+(1/2))$, onde $\theta \in R$. Neste caso

$$g(\theta|x) \propto \prod_{i=1}^{n} I_{(-\frac{1}{2};\frac{1}{2})}(x_i-\theta) = \prod_{i=1}^{n} I_{(x_i-\frac{1}{2};x_i+\frac{1}{2})}(\theta) = I_A(\theta)$$

onde $\qquad A = \bigcap_{i=1}^{n} (x_i-(1/2);x_i+(1/2)) = (y_n-(1/2);y_1+(1/2)).$

Assim, $g(\theta|x) = [1-(y_n-y_1)]^{-1} I_A(\theta)$.

Consideremos agora o caso de duas amostras independentes de tamanhos r e n-r. Suponha que a densidade conjunta de x seja dada por $p(x_1-\theta,\ldots,x_n-\theta) = \prod_{i=1}^{r} p_0(x_i-\theta) \prod_{j=r+1}^{n} p_1(x_j-\theta)$ onde $p_0$ e $p_1$ são as densidade pivotais, respectivamente, das r primeiras e das n-r restante observações. Os exemplos a seguir estão relacionados com problemas de comparação de dois instrumentos de medida.

Exemplo 4.1.2: Suponha que para $1 \leq i \leq r$, $x_i \sim U(\theta-1;\theta+1)$ e que para $r < j \leq n$, $x_j \sim U(2\theta-1;2\theta+1)$. Neste caso

$$g(\theta|x) \propto I_{A\cap B}(\theta)$$

onde
$$A = \bigcap_{i=1}^{r} (x_i-1;x_i+1) = (y_r-1;y_1+1) ,$$

$$B = \bigcap_{j=r+1}^{n} \left(\frac{x_j-1}{2};\frac{x_j+1}{2}\right) = \left(\frac{y_n'-1}{2};\frac{y_{r+1}'+1}{2}\right)$$

e
$$y_{r+1}' = \min(y_{r+1},\ldots,y_n) \text{ e } y_n' = \max(y_{r+1},\ldots,y_n).$$

Exemplo 4.1.3: Suponha que para $1 \leq i \leq r$ , $x_i \sim N(\theta,1)$ e que para $r < j \leq n$ , $x_j \sim N(3\theta,1)$. Então

$$g(\theta|x) \propto \exp\{-\frac{1}{2}[\sum_{i=1}^{r}(x_i-\theta)^2+\sum_{j=r+1}^{n}(x_j-3\theta)^2]\} \propto \exp\{-\frac{9(n-r)+r}{2}(\theta-t)^2\}$$

onde
$$t = (9n-8r)^{-1}(t_0+3t_1) , \quad t_0 = \sum_{i=1}^{r} x_i \quad \text{e} \quad t_1 = \sum_{j=r+1}^{n} x_j .$$

Exemplo 4.1.4: No exemplo anterior, no lugar de $N(\theta;1)$ e $N(3\theta;1)$ considere, respectivamente, $N(\theta,\sigma_0^2)$ e $N(\theta,\sigma_1^2)$. Então, escrevendo $\bar{x}_0 = \frac{1}{r}\sum_{i=1}^{r} x_i$ e $\bar{x}_1 = \frac{1}{n-r}\sum_{j=r+1}^{n} x_j$, temos

$$g(\theta|x) \propto \exp\{-\frac{1}{2}[\frac{r}{\sigma_0^2}(\bar{x}_0-\theta)^2 + \frac{n-r}{\sigma_1^2}(\bar{x}_1-\theta)^2]\}$$

$$\propto \exp\{-\frac{1}{2}[i_0-i_1)(\theta-t)^2]\}$$

onde $i_0 = \frac{r}{\sigma_0^2}$, $i_1 = \frac{n-r}{\sigma_1^2}$ e $t = \frac{i_0\bar{x}_0 + i_1\bar{x}_1}{i_0 + i_1}$.

Isto é, $\theta|x \sim N(t, \frac{1}{i_0 + i_1})$.

Como generalização do exemplo 4.1.3, vamos considerar o modelo de regressão normal uniparamétrico.

Exemplo 4.1.5: Considere a estatística ancilar $\mu = (\mu_1,\ldots,\mu_n)$ que com $x = (x_1,\ldots,x_n)$ formam os dados $d = (x,u)$. Dessa forma temos o modelo estatístico.

$$f(x,u|\theta) = f(x|\theta,u)f(u).$$

Como apenas o primeiro fator depende de $\theta$, a inferência se restringe ao uso deste. Supondo então que

$$f(x|\theta,u) \propto \exp\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta u_i)^2\}$$

temos que $t = \frac{\Sigma u_i x_i}{\Sigma u_i^2}$ é uma estatística suficiente condicional a estatística ancilar, u. Neste caso,

$$g(\Theta|x,u) \propto \exp\{-\frac{\sum_{i=1}^{n} u_i^2}{2}(t-\Theta)^2\}$$

isto é, $\Theta|(x,u) \sim N(t,\frac{1}{\sum_{i=1}^{n} u_i^2})$.

Note que, como $u$ é ancilar, não são necessárias restrições sobre sua distribuição.

## 4.2. Parâmetros de Escala

Nos modelos onde $\Theta$ é um parâmetro de escala, o vetor $(\frac{x_1}{\Theta},\ldots,\frac{x_u}{\Theta}) = \underline{e}$ é uma quantidade pivotal. A densidade conjunta satisfaz a igualdade

$$f(x|\Theta) = (\frac{1}{\Theta})^n p(\frac{x}{\Theta}) = (\frac{1}{\Theta})^n p(\underline{e})$$

onde $(\frac{1}{\Theta})^n$ é o determinante do Jacobiano da transformação $\underline{e} = \frac{x}{\Theta}$ . Assim, após observar-se x, a densidade fiducial de $\Theta$ é dada por

$$g(\Theta|x)=\frac{(\frac{1}{\Theta})^n p(\frac{x}{\Theta})}{\int_0^\infty (\frac{1}{\Theta})^n p(\frac{x}{\Theta})d\Theta} .$$

Note que, ao considerar-se $x_i^* = \log x_i$ e $\Theta^* = \log \Theta$ , o modelo associado a $x^* = (x_1^*,\ldots,x_n^*)$ é um modelo com parâmetro de locação $\Theta^*$. A partir da distribuição de $x^*$ obteríamos a densidade fiducial de $\Theta^*$ e assim a de $\Theta$, que coincide com a densidade $g(\Theta|x)$ obtida por $f(x|\Theta)$.

Ao considerarmos observações i.i.d. temos que

$$g(\theta|x) \propto \theta^{-(n+1)} \prod_{i=1}^{n} p(\frac{x_i}{\theta}) .$$

onde p aqui é a densidade comum de $\underline{e}_i = \frac{x_i}{\theta}$ , $i=1,\ldots,n$. Uma outra representação pode ser obtida ao se considerar o vetor

$(\frac{y_1}{y_n},\ldots,(\frac{y_{n-1}}{y_n} , y_n) = (u_1,\ldots,u_{n-1},y_n)$. A estatística $y_n$ é condicionalmente suficiente dado a estatística ancilar $u = (u_1,\ldots,u_{n-1})$.

Teríamos então

$$\theta^{-n} \prod_{i=1}^{n} p(\underline{e}_i) = \theta^{-n} p(\frac{y_n}{\theta}) \prod_{i=1}^{n-1} p(\frac{y_n}{\theta} \times \frac{y_i}{y_n}) ,$$

o que mostra que o modelo

$$\frac{1}{\theta} f(y_n|\theta,\mu_1,\ldots,\mu_{n-1})$$

contém toda informação sobre $\theta$. A densidade fiducial será, proporcional ao produto da verossimilhança por $\frac{1}{\theta}$ , que é a densidade a posteriori do método de Bayes quando se considera a priori imprópria, $\frac{1}{\theta}$ .

<u>Exemplo 4.2.1</u>: Considere n observações independentes de uma distribuição uniforme em $(0,\theta)$, $0 < \theta < \infty$. A densidade fiducial será:

$$g(\theta|x) \propto (\frac{1}{\theta})^{n+1} \prod_{i=1}^{n} I_{(x_i,\infty)}(\theta) \quad ;$$

isto é, $$g(\theta|x) = \begin{cases} (n\theta)^{-1}(\theta y_n)^{-n} & \text{se} \quad \theta > y_n \\ 0 & \text{se} \quad \theta \leq y_n . \end{cases}$$

Como no caso de parâmetros de locação, vamos considerar o caso de duas amostras independentes de tamanhos r e n-r. Suponha que a densidade conjunta de x seja especificada por

$$p\left(\frac{x_1}{\Theta},\ldots,\frac{x_n}{\Theta}\right) = \prod_{i=1}^{r} p_0\left(\frac{x_i}{\Theta}\right) \prod_{j=r+1}^{n} p_1\left(\frac{x_j}{\Theta}\right)$$

onde $p_0$ e $p_1$ são as densidades pivotais comuns, respectivamente, das r primeiras observações e das n-r restantes. Este modelo é conhecido como modelo de Fisher-Pitman. Lembremos que a densidade fiducial será

$$g(\Theta|x) \propto \Theta^{-(n+1)} \prod_{i=1}^{r} p_0\left(\frac{x_i}{\Theta}\right) \prod_{j=r+1}^{n} p_1\left(\frac{x_j}{\Theta}\right)$$

Exemplo 4.2.2: Considere o caso de n observações independentes onde, $x_i \sim U(0,\Theta)$ para $1 \leqq i \leqq r$ e $x_j \sim U(0,2\Theta)$ para $r < j \leqq n$.

A densidade fiducial será

$$g(\Theta|x) \propto \Theta^{-(n+1)} I_{A\cap B}(\Theta)$$

onde, representando $y'_r = \text{máx}(x_1,\ldots,x_r)$ e $y'_n = \text{máx}(x_{r+1},\ldots,x_n)$,

$$A = \bigcap_{i=1}^{r} (x_i,\infty) = (y'_r,\infty)$$

e $$B = \bigcap_{j=r+1}^{n} \left(\frac{x_j}{2},\infty\right) = \left(\frac{y'_n}{2},\infty\right) .$$

Se $z = \text{máx}\left(y'_r, \frac{y'_n}{2}\right)$, então

$$g(\Theta|x) = \begin{cases} (n\Theta)^{-1}(z\Theta)^{-n} & \text{se} \quad \Theta > z \\ 0 & \text{se} \quad \Theta \leqq z \end{cases}$$

O exemplo seguinte é uma extensão do modelo de Fisher--Pitman.

Exemplo 4.2.3: Considere a estatística ancilar $u = (u_1,...,u_n)$ que, ao lado de $x = (x_1,...,x_n)$ formam os dados $d = (x,u)$.
Como no exemplo 4.1.5, a inferência sobre exige apenas o uso de densidade condicional de $x|u$. Suponha agora que, $x_1,...,x_n$ são condicionalmente independentes onde $x_i|u \sim N(0,u_i\theta)$. Isto é, a densidade condicional de x dado u pode ser expressa como

$$f(x|\theta,u) \propto \theta^{-\frac{n}{2}} \exp\{\frac{1}{2\theta} \sum_{i=1}^{n} \frac{x^2}{u_i}\} \quad .$$

Note que $t = \sum_{i=1}^{n} \frac{x_i^2}{u_i}$ é condicionalmente suficiente dado u e a quantidade pivotal $\frac{t}{\theta^2}$ tem distribuição qui-quadrado com n graus de liberdade. A densidade fiducial será então:

$$g(\theta|x,u) = g(\theta|t,u) \propto \theta^{-(\frac{n}{2}+1)} \exp\{-\frac{t}{2\theta}\} \ .$$

Se $u_i$ tomar apenas dois valores (1 ou 2) temos o modelo normal de Fisher-Pitman. Note que $\theta$ é escala para t e $\sqrt{\theta}$ é escala para x.

A próxima seção apresenta alguns exemplos especiais onde $\theta$ não é (apenas) parâmetro de locação nem parâmetro de escala.

## 4.3. Outros Exemplos

O seguinte exemplo é semelhante aos casos onde o modelo de Fisher-Pitman está envolvido.

Exemplo 4.3.1: Suponha que $x_1,...,x_n$ são n observações independentes exponencialmente distribuidas com parâmetro $\Theta$ nas r primeiras e parâmetro $(1/\Theta)$ nas restantes. Assim, a densidade conjunta de x é dada por

$$f(x|\Theta) = \frac{\Theta^r}{\Theta^{n-r}} \exp\{-\Theta y - \frac{z}{\Theta}\},$$

onde $$y = \sum_{i=1}^{r} x_i \quad \text{e} \quad z = \sum_{j=r+1}^{n} x_j$$

Note que $\frac{1}{\Theta}$ é o parâmetro de escala para as r primeiras observações e $\Theta$ é o parâmetro de escala para as restantes. A densidade fiducial neste caso é dada por:

$$g(\Theta|x) \propto \Theta^{2r-n-1} \exp\{ -\Theta y - \frac{z}{\Theta}\} .$$

Os dois exemplos a seguir incluem um parâmetro $\Theta$ que atua como parâmetro de locação e proporcionalmente como de escala. Isto é, $\Theta$ e $c\Theta$ são os parâmetros de locação e escala, respectivamente. Note que o valor conhecido de c informa sobre o sinal (+ ou -) de $\Theta$, visto que, como $c\Theta$ é de escala, $c\Theta > 0$ .

Considerando n observações i.i.d. a quantidade pivotal $\underline{e} = (e_1,...,e_n)$ é definida por $e_i = \frac{x_i - \Theta}{c\Theta} = \frac{x_i}{c\Theta} - \frac{1}{c}$ , $i=1,...,n$. Assim, a densidade conjunta de x se expressa como

$$f(x|\theta) = (\frac{1}{c\theta})^n \prod_{i=1}^{n} p(\frac{x_i}{c\theta} - \frac{1}{c}) .$$

A densidade fiducial neste caso será

$$g(\theta|x) \propto (\frac{1}{c\theta})^{n+1} \prod_{i=1}^{n} p(\frac{x_i}{c\theta} - \frac{1}{c})$$

Uma outra representação pode ser obtida através da utilização de $(x_1, \frac{x_2}{x_1}, \dots, \frac{x_n}{x_1})$ onde $u = (\frac{x_2}{x_1}, \dots, \frac{x_n}{x_1})$ é ancilar e $x_1$ é condicionalmente suficiente dado u.

Exemplo 4.3.2: Suponha que x é composta por n observações i.i.d. de uma distribuição uniforme $(\theta, 2\theta)$, $\theta > 0$. A densidade fiducial neste caso é uma uniforme no intervalo $(\frac{y_n}{2}, y_1)$.

Exemplo 4.3.3: Suponha, no exemplo anterior, que no lugar de $U(\theta, 2\theta)$ temos uma normal com média $\theta$ e desvio padrão $c\theta$. Evidentemente, o sinal de $\theta$ depende do coeficiente de variação, conhecido, $\theta$. Assim, se $c < 0$, $\theta < 0$ e se $c > 0$, $\theta > 0$. Neste caso, a densidade fiducial é dada por

$$g(\theta|x) \propto |\theta|^{-(n+1)} \exp\{- \frac{1}{2c^2} \sum_{i=1}^{n} (\frac{x_i}{\theta} - 1)^2\} I_A(\theta)$$

onde $A = \{\theta; c\theta > 0\}$. Ou seja, $A = (0, \infty)$ ou $A = (-\infty, 0)$.

O seguinte exemplo generaliza em parte o exemplo 4.3.1. Efrón and Hinkley (1978) denominaram o modelo envolvido de Hipérbole Gamma de Fisher, o qual foi introduzido em Fisher (1956).

Exemplo 4.3.4: Suponha que x é composto por n observações i.i.d do vetor (y,z) cuja densidade conjunta é proporcional a $\exp\{-\theta y - \frac{z}{\theta}\}$.

A densidade conjunta de x é então expressa como

$$f(x|\theta) \propto \exp\{-\theta t_0 - \frac{t_1}{\theta}\}$$

onde $t_0 = \sum_{i=1}^{n} y_i$ e $t_1 = \sum_{i=1}^{n} z_i$ .

A densidade fiducial, como antes, será

$$g(\theta|x) \propto \theta^{-1}\exp\{-\theta t_0 - \frac{t_1}{\theta}\} .$$

O exemplo a seguir foi considerado por Sprott (1961) e envolve dois diferentes tipos de distribuições.

Exemplo 4.3.5: Suponha que y e z são duas observações independentes onde z é normalmente distribuida com média $\theta$ e variância 1 e z tem distribuição gama com parâmetros m e $e^{-k\theta}$.. Note que $e_0 = y - \theta$ e $e_1 = e^{-k\theta}z$ são duas quantidades pivotais independentes onde $e_0 \sim N(0,1)$ e $e_1 \sim G(m,1)$, onde $G(m,1)$ representa a distribuição gama com parâmetros m e 1. Não é difícil ver que $\theta$ é parâmetro de locação para $(y,z')$ onde $z' = \frac{1}{k}\log z$. Para isto basta lembrar que $\log e_1 = k(\theta - z')$. Assim, se $e_1' = \frac{\log e_1}{k}$ então, a densidade ficucial de $\theta$ será

$$g(\theta|y,z) \propto p_0(e_0)p_1(e_1'),$$

onde $p_0$ e $p_1$ são as densidades de $e_0$ e $e_1'$, respectivamente.

oOo

# CAPÍTULO 5

## MÉTODO FIDUCIAL: CASO MULTIPARAMÉTRICO

A proposição deste capítulo é apresentar as condições necessárias para a utilização do método fiducial multiparamétrico e discutir alguns exemplos importantes dentro do contexto desse método.

Também aqui é necessário encontrar $t=(t_1,\ldots,t_k)$ uma estatística suficiente com relação a $\Theta$. Abordaremos apenas o caso em que a *dimensão* de t é igual à dimensão de $\Theta$ e t é suficiente completa com relação a $\Theta$. No caso em que as dimensões são diferentes e n é o tamanho da amostra é necessário encontrar uma estatística ancilar $u=(u_1,\ldots,u_{n-k})$ e proceder de forma análoga ao que foi descrito no caso uniparamétrico (seção 2.3).

Na seção 5.1 fazemos a extensão, para o caso multiparamétrico, do método de quantidades pivotais e apresentamos as respectivas implicações. Na seção 5.2. apresentamos outra forma de encontrar distribuições fiduciais multiparamétricas: é o método passo-a-passo. Esta segunda alternativa de metodologia é indicada por Fisher como sendo mais adequada. Alguns comentários à respeito dessa adequação são feitos no decorrer da exposição. Uma discussão mais detalhada é apresentada para o caso biparamétrico.

Finalmente, na seção 5.3. apresentamos um dos exemplos mais conhecidos de distribuições fiduciais: a distribuição de Behrens-Fisher e o princípio de Difusão.

## 5.1. Quantidades Pivotais

A ampliação do conceito de quantidade pivotal para o caso multiparamétrico consiste em considerarmos $\underline{e}$ como função de $t = (t_1,\ldots,t_k)$ e $\Theta = (\Theta_1,\ldots,\Theta_k)$ onde t é uma estatística suficiente com respeito ao parâmetro $\Theta$ e de mesma dimensão deste. Lembremos que a distribuição de $\underline{e}$ é independente de $\Theta$.

A quantidade pivotal $\underline{e}(\Theta;t)$ pode ser utilizada para encontrar a distribuição fiducial de $\Theta$, $g(\Theta/t)$, se satisfizer condições análogas àquelas apresentadas na seção 2.2.2, considerando na condição (ii) apenas a inversibilidade de $\underline{e}(\Theta;t)$ em $\Theta$ e substituindo (iii) por

(iii) $\underline{e}(\Theta;t)$ tem derivadas parciais contínuas com jacobiano diferente de zero.

No entanto, no caso multiparamétrico podemos ter diferentes quantidades pivotais que se encontram dentro das condições acima e qualquer uma delas pode ser utilizada para encontrar $g(\Theta/t)$. O exemplo a seguir ilustra esta situação e foi um dos primeiros apresentados para este método. (Fisher, 1941).

Exemplo 5.1.1: Sejam $X_1,\ldots,X_n$ v.a.i.i.d. com distribuição $N(\mu,\sigma^2)$

Sabemos que $e_1 = \frac{\sqrt{n}}{\sigma}(\bar{x}-\mu) \sim N(0,1)$ e $e_2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$ e que $e_1$ e $e_2$ são independentes.

Assim a quantidade pivotal $\underline{e}(x,s^2;\mu,\sigma^2) = (e_1,e_2)$ pode ser utilizada para encontrar a distribuição fiducial de $(\mu,\sigma^2)$ que tem a seguinte expressão:

$$g(\mu,\sigma^2/\bar{x},s^2) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-n}{2\sigma^2}(\bar{x}-\mu)^2\right\} \frac{(1/2)^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n}{2}} \exp\left\{\frac{-(n-1)s^2}{\sigma^2}\right\}$$

Como o método pivotal não estabelece que $\underline{e}(\bar{x},s^2,\mu,\sigma^2)$ deve ser única, podemos encontrar $g(\mu,\sigma^2/\bar{x},s^2)$ através de outra quantidade pivotal $\underline{e}'(x,s^2,\mu,\sigma^2) = (e_1',e_2')$ onde

$$e_1' = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t_{n-1} \quad e \quad e_2' = \frac{1}{\sigma^2}[(n-1)s^2 + n(\bar{x}-\mu)^2] \sim \chi^2_n$$

$\underline{e}'(\bar{x},s^2,\mu,\sigma^2)$ produz a mesma densidade fiducial encontrada com $\underline{e}(\bar{x},s^2,\mu,\sigma^2)$.

Em geral, o método pivotal multiparamétrico é pouco atrativo porque exige uma grande quantidade de operações complicadas. Além disso, dpeendendo das quantidades pivotais escolhidas, os resultados podem ser diferentes. Mauldon (1955) apresentou o seguinte exemplo onde isto ocorre:

Exemplo 5.1.2: Considere a seguinte densidade de x,y,z

$$f(x,y,z/\theta_1,\theta_2,\theta_3) = \pi^{-1}\theta_1^2\theta_2^2\cos^3\theta_3 xy \exp\{-\tfrac{1}{2}(\theta_1^2x^2+2\theta_1\theta_2 xy \operatorname{sen}\theta_3 \operatorname{sen} z + \theta_2^2y^2)\}$$

com $x > 0$ , $y > 0$ e $-\frac{\pi}{2} < z < \frac{\pi}{2}$ , $\theta_1 > 0$ , $\theta_2 > 0$ e $-\frac{\pi}{2} < \theta_3 < \frac{\pi}{2}$

Usando as quantidades pivotais $e_1 = \theta_1 x \cos\theta_3$, $e_2 = \theta_2 y \cos z$ e $e_3 = \theta_1 x \operatorname{sen}\theta_3 + \theta_2 y \operatorname{sen} z$ a densidade acima pode ser reescrita como

$$\pi^{-1}(\Theta_1 x \cos\Theta_3)(\Theta_1\Theta_2^2\cos\Theta_3 y)\exp\{-\frac{1}{2}[\Theta_1^2x^2(\cos^3\Theta_3+\text{sen}^3\Theta_3)+$$

$$+\Theta_2^2y^2(\cos^2 z+\text{sen}^2 z)+2\Theta_1\Theta_2 xy\,\text{sen}\,\Theta_3\text{sen}\,z]\}$$

donde resulta que a distribuição de $e_1,e_2,e_3$ pode ser representada através da densidade

$$(5.1.1)\qquad p(e_1,e_2,e_3)=\frac{1}{\pi}e_1\exp\{-(e_1^2+e_2^2+e_3^2)\}\text{ com } e_1>0\ ,\ e_2>0 \quad\text{e}\quad -\infty<e_3<\infty$$

Como $\underline{e}=(e_1,e_2,e_3)$ é uma quantidade pivotal com densidade (5.1.1) e satisfaz as condições indicadas no início desta seção, então a distribuição fiducial conjunta de $\Theta_1,\Theta_2,\Theta_3$ é obtida através de

$$g(\Theta_1,\Theta_2,\Theta_3/x,y,z)=p(e_1,e_2,e_3)|J|$$

onde $$|J|=\begin{vmatrix} x\cos\Theta_3 & 0 & -\Theta_1 x\,\text{sen}\,\Theta_1\\ 0 & y\cos z & 0\\ x\,\text{sen}\,\Theta_3 & y\,\text{sen}\,z & \Theta_1 x\cos\Theta_3\end{vmatrix}=x^2y\Theta_1\cos z$$

ou seja

$$(5.1.2)\qquad g(\Theta_1,\Theta_2,\Theta_3/x,y,z)=$$

$$\frac{1}{\pi}\Theta_1^2x^3y\cos\Theta_3\cos z\exp\{-\frac{1}{2}(\Theta_1^2x^2+\Theta_2^2y^2+2\Theta_1\Theta_2xy\,\text{sen}\,\Theta_3\,\text{sen}\,z)\}$$

Observe que a natureza simétrica de $f(x,y,z/\Theta_1,\Theta_2,\Theta_3)$ nos permite definir também as seguintes quantidades pivotais

$$e_1'=\Theta_1x\cos z,\ e_2'=\Theta_2y\cos\Theta_3,\ e_3'=\Theta_1x\,\text{sen}\,z+\Theta_2y\,\text{sen}\,\Theta_3$$

e procedendo de forma análoga, a distribuição conjunta de $e_1', e_2', e_3'$ pode ser escrita como:

$$p'(e_1', e_2', e_3') = \frac{1}{\pi}\, e_2' \exp\{-\frac{1}{2}(e_1'^2 + e_2'^2 + e_3'^2)\}$$

A distribuição fiducial de $\theta_1, \theta_2, \theta_3$ resultante é

$$(5.1.3) \qquad g(\theta_1, \theta_2, \theta_3 / x, y, z) =$$

$$\frac{1}{\pi}\, \theta_2^2 x y^3 \cos\theta_3 \cos z \exp\{-\frac{1}{2}(\theta_1^2 x^2 + \theta_2^2 y^2 + 2\theta_1\theta_2 xy \operatorname{sen}\theta_3 \operatorname{sen} z)\}$$

pois, neste caso $|J'| = xy^2\theta_2 \cos z$ . Note que (5.1.2) e (5.1.3) são diferentes, o que mostra que o argumento fiducial nestas condições não está bem definido. Outros exemplos como este constam de Fieller (1954) e Creasy (1954).

Devido a exemplos como este acima, Fisher (1956) definiu um método para encontrar a distribuição fiducial multiparamétrica à qual deu o nome de método "passo-a-passo" e que segundo ele é um processo "rigoroso" de construção da distribuição conjunta dos parâmetros.

## 5.2. **Método Passo-a-Passo**

Esse método é definido para situações onde temos as seguintes condições:

1) A função densidade de t pode ser fatorada como

$$f(t_1,\ldots,t_k/\theta_1,\ldots,\theta_k) = f_1(t_1/\theta_1)f_2(t_2/t_1,\theta_1,\theta_2) \times \ldots \times$$

$$\times\ f_k(t_k/t_1,\ldots,t_{k-1},\theta_1,\ldots,\theta_k)$$

2) A função distribuição de probabilidades, $F_i$, associada a $f_i$ satisfaz as condições estabelecidas no caso uniparamétrico para a obtenção da distribuição fiducial (seção 2.2.1).

Cada $f_i$ é utilizada para encontrar a distribuição fiducial de $\theta_i$ condicionada ao conhecimento de $t_1,\ldots,t_i$ e $\theta_1,\ldots,\theta_{i-1}$. As densidades condicionais serão denotadas por $g_i(\theta_i/t_1,\ldots,t_i,\theta_1,\ldots,\theta_{i-1})$. O seu produto resulta na distribuição fiducial conjunta de $\theta_1,\ldots,\theta_k$.

O seguinte exemplo devido a Fisher (1956) ilustra esta situação:

Exemplo 5.2.1: Considere o caso de uma normal bivariada com média $(\mu_1,\mu_2)$, variâncias $(\sigma_1^2,\sigma_2^2)$ e coeficiente de correlação $\rho$. Da amostra $(x_1,y_1),\ldots,(x_n,y_n))$ obtêm-se o valor da estatística suficiente $(\bar{x}_1,\bar{x}_2,s_1^2,s_2^2,r)$. Nosso interesse é obter a distribuição fiducial conjunta de $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

Passo 1: Fisher em 1915 obteve a distribuição de r, coeficiente de correlação amostral que só depende de $\rho$. Assim

$$f_1(r/\rho) = \frac{1}{\pi(n-3)!}(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\ \frac{\partial^{n-2}}{\partial(-\rho r)^{n-2}}\left(\frac{\theta}{\operatorname{sen}\theta}\right)dr$$

onde $\cos\theta = -\rho r$ e $0 \leqq \theta \leqq \pi$ .

Como a distribuição de r não depende de outro parâmetro a não ser de $\rho$, então a distribuição fiducial de $\rho$ é

$$g_1(\rho/r) = \left|\frac{F_1(r/\rho)}{\partial\rho}\right| \qquad \text{onde} \qquad F_1(r/\rho) = \int_{-\infty}^{r} f_1(r/\rho)dr$$

Passo 2: A densidade condicional de $(s_1,s_2)$ dado r será

$$f\ (s_1 s_2/r,\rho,\sigma_1,\sigma_2) =$$

$$= \frac{(n-1)^{n-1}}{(1-\rho^2)^{n-1}} \left(\frac{s_1}{\sigma_1}\frac{s_2}{\sigma_2}\right)^{n-2} \exp\left\{\frac{-(n-1)}{2(1-\rho^2)}\left[\frac{s_1^2}{\sigma_1^2} - 2r\rho\frac{s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\right]\right\} \times$$

$$\times \frac{\partial s_1 \partial s_2}{\sigma_1\sigma_2}\left[\frac{\partial^{n-2}}{(\operatorname{sen}\Theta\partial\Theta)^{n-2}}\left(\frac{\Theta}{\operatorname{sen}\Theta}\right)\right]^{-1}$$

Com base nesta densidade, $f_2$, encontramos a distribuição fiducial de $\sigma_1,\sigma_2$ condicionada a $\rho,s_1,s_2,r$. Isto é,

$$g_2(\sigma_1,\sigma_2/s_1,s_2,r,\rho) \propto f_2(s_1,s_2/r,\rho,\sigma_1,\sigma_2)\ \frac{s_1}{\sigma_1^2}\ \frac{s_2}{\sigma_2^2}$$

Passo 3: A distribuição condicional de $(\bar{x}_1,\bar{x}_2)$ dado $(s_1,s_2,r)$ é normal com média $(\mu_1,\mu_2)$, variância $((\sigma_1^2)/n,(\sigma_2^2)/n)$ e coeficiente de correlação $\rho$. A distribuição fiducial de $(\mu_1,\mu_2)$ condicionada em $\sigma_1,\sigma_2,\rho$ e $\bar{x}_1,\bar{x}_2,s_1,s_2,r$ é normal com média $(\bar{x}_1,\bar{x}_2)$ variância $((\sigma_1^2)/n,(\sigma_2^2)/n)$ e coeficiente de correlação $\rho$ e é denotada por

$$g_3(\mu_1,\mu_2/\bar{x}_1,\bar{x}_2,s_1,s_2,r,\sigma_1,\sigma_2,\rho)$$

<u>Passo 4</u>: $g(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho/\bar{x}_1,\bar{x}_2,s_1,s_2,r) =$

$$g_1(\rho/r)g_2(\sigma_1,\sigma_2/s_1,s_2,r,\rho)g_3(\mu_1,\mu_2/\bar{x}_1,\bar{x}_2,s_1,s_2,r,\rho,\sigma_1,\sigma_2)$$

Incluimos este exemplo, por ser um dos mais conhecidos da inferência fiducial multiparamétrica.

No entanto, não apresentamos todas as condições necessárias para que a distribuição fiducial conjunta seja válida. Faremos isto a seguir, onde o método passo-a-passo é apresentado em detalhes para o caso onde k=2.

<u>Caso Biparamétrico</u>: A função densidade de $(t_1,t_2)$ estatísticas suficientes para $(\Theta_1,\Theta_2)$, pode ser fatorada como

$$f(t_1,t_2/\Theta_1,\Theta_2) = f_1(t_1/\Theta_1)\ f_2(t_2/t_1,\Theta_1,\Theta_2).$$

Então a distribuição fiducial de $(\Theta_1,\Theta_2)$ será

$$g(\Theta_1,\Theta_2/t_1,t_2) = g_1(\Theta_1/t_1)g_2(\Theta_2/\Theta_1,t_1,t_2)$$

onde

$$g_1(\Theta_1/t_1) = \left|\frac{\partial F_1(t_1/\Theta_1)}{\partial\Theta_1}\right| \quad \text{com} \quad F_1(t_1/\Theta_1) = \int_{-\infty}^{t_1} f_1(v/\Theta_1)dv$$

e

$$g_2(\Theta_2/\Theta_1,t_1,t_2) = \left|\frac{\partial F(t_2/t_1,\Theta_1,\Theta_2)}{\partial\Theta_2}\right| \quad \text{com}$$

$$F(t_2/t_1,\Theta_1,\Theta_2) = \int_{-\infty}^{t_2} f_2(v/t_1,\Theta_1,\Theta_2)dv$$

Evidentemente para obtermos $g(\Theta_1,\Theta_2/t_1,t_2)$ é necessário que

I) Os domínios dos parâmetros devem ser de variação independente, ou seja, $(\theta_1,\theta_2) \in \Theta_1 \times \Theta_2$ onde $\Theta_1$ e $\Theta_2$ são os domínios de $\theta_1$ e $\theta_2$ respectivamente.

II) Definindo $\theta_1^\alpha = \theta_1^\alpha(t_1) = \{\theta_1 \; ; \; F_1(t_1/\theta_1) = \alpha\}$ e $\theta_2^\alpha = \theta_2^\alpha(t_2/t_1,\theta_1) = \{\theta_2 \; ; \; F_2(t_2/t_1,\theta_1,\theta_2) = \alpha\}$, os intervalos fiduciais $(-\infty,\theta_i^\alpha)$ de $F_i$ , $i=1,2$, com probabilidade fiducial $(1-\alpha)$ possuem a propriedade de confiança (seção 3.2)

III) A distribuição fiducial marginal de $\theta_2$ também deve possuir a propriedade de confiança, ou seja, se

$$\overline{G}_2(\theta_2/t_1,t_2) = \int_1 G_2(\theta_2/t_1,t_2,\theta_1)g_1(\theta_1/t_1)d\theta_1$$

onde $G_2(\theta_2/t_1,\theta_1,\theta_2) = \int_{-\infty}^{\theta_2} g_2(v/t_1,\theta_1,\theta_2)d\theta_2$

então, os intervalos fiduciais de $\overline{G}_2$ devem possuir uma "propriedade de confiança esperada"

Quanto as restrições acima fazemos as seguintes observações:

a) Ao encontrar $g_2(\theta_2/t_1,\theta_1,t_2)$ consideramos $\theta_1$ como um valor conhecido. Nestas condições $(t_1,\theta_1)$ nos dá a "configuração" de $\theta_2$ isto é, $(t_1,\theta_1)$ faz o "papel" da estatística ancilar u, no caso uniparamétrico (seção 2.3), contudo $\theta_1$ é um parâmetro e não uma estatística.

b) É interessante notar que se a restrição (I) não estiver satisfeita, então $\Theta_1$ contém informação relevante sobre $\Theta_2$ e assim devido a observação (a) o conjunto de referência de $\Theta_2$ contém subconjuntos relevantes (seção 2.2)

c) A restrição III indica que além de $\Theta_1$ e $\Theta_2$ terem variação independente, o quantil $(1-\alpha)$ de $\bar{G}_2$ que é $\Theta_2^{\alpha}(t_1,t_2)$ é tal que $(-\infty,\Theta_2^{\alpha}(t_1,t_2))$ possui a propriedade de confiança "*independente*" de $\Theta_1$. Em outras palavras, $\forall\ \Theta_1 \in \Theta_1$, $\exists\ \alpha(\Theta_1) \in (0,1)$ tal que

$$(5.2.1) \qquad \Theta_2^{\alpha}(t_1,t_2) = \Theta_2^{\alpha(\Theta_1)}(t_2/t_1,\Theta_1)$$

onde $\Theta_2^{\alpha(\Theta_1)}(t_2/t_1,\Theta_1)$ é o quantil $(1-\alpha(\Theta_1))$ de $G_2(\Theta_2/\Theta_1,t_1,t_2)$. Quando (5.2.1) é válida temos

$$(1-\alpha) = \bar{G}_2(\Theta_2^{\alpha}(t_1,t_2)/t_1,t_2) = \int_{\Theta_1} G_2(\Theta_2^{\alpha}(t_2,t_1)/t_1,t_2,\Theta_1)g(\Theta_1/t_1)d\Theta_1$$

$$= \int_{\Theta_1} (1-\alpha(\Theta_1))g(\Theta_1/t_1)d\Theta_1$$

o que induz uma *propriedade de confiança esperada* aos intervalos de $\bar{G}_2$, ou seja, se "eliminamos" $\Theta_1$, $\bar{G}_2$ tem toda a informação relevante sobre $\Theta_2$. Nesta condições, o conjunto de referência de $(\Theta_1,\Theta_2)$ aparentemente não possui subconjuntos relevantes.

Pedersen (1978) apresentou uma condição *suficiente* para que III esteja satisfeita: Existem duas funções m e h tais que

$$(5.2.2) \qquad f(t_1,t_2/\Theta_1,\Theta_2) = m(t_1,h(t_1,t_2;\Theta_2)/\Theta_1)\,\frac{\partial h(t_1,t_2;\Theta_2)}{\partial t_2}$$

Contudo, aparentemente não se conhece uma forma de assegurar que a restrição III é ou não válida, quando (5.2.2) não é verdadeira.

d) Se $f_2(t_2/t_1,\Theta_1,\Theta_2) = f_2(t_2/t_1,\Theta_2)$ então III é imediata. Logo as regiões $(\Theta_1 \leq \Theta_1^{\alpha_1},\ \Theta_2 \leq \Theta_2^{\alpha_2})$ têm probabilidade fiducial $(1-\alpha_1)(1-\alpha_2)$.

Um exemplo que caracteriza esta situação é apresentado a seguir

<u>Exemplo 5.2.3</u>: Considere o caso de uma distribuição $N(\mu,\sigma^2)$. Então $f(\bar{x},s/\mu,\sigma^2) = f_2(\bar{x}/s^2,\mu)f_1(s^2/\sigma^2)$. Note que $f_2(\bar{x}/s^2,\mu,\sigma^2) = f_2(\bar{x}/s^2,\mu)$ então $g_2(\mu/\bar{x},s^2)$ e $g_1(\sigma^2/s^2)$ são densidades fiduciais tais que $\mu$ e $\sigma^2$ são condicionalmente independentes (condicional a $\bar{x}$ e $s^2$). E portanto, a distribuição conjunta de $(\mu,\sigma)$ não apresenta subconjuntos relevantes.

e) Em muitas situações a densidade de $(t_1,t_2)$ pode ser fatorada de duas formas diferentes:

$$(5.2.3) \qquad f(t_1,t_2/\Theta_1,\Theta_2) = f(t_1/\Theta_1)f(t_2/t_1,\Theta_1,\Theta_2)$$

$$(5.2.4) \qquad f(t_1,t_2/\Theta_1,\Theta_2) = f(t_2/\Theta_2)f(t_1/t_2,\Theta_1,\Theta_2)$$

Nem sempre (5.2.3) e (5.2.4) produzem o mesmo resultado. Para decidir entre uma delas é importante verificar se estão satisfeitas as condições (I), (II), (III).

O seguinte exemplo (Dempster, 1963) ilustra a observação e):

Exemplo 5.2.4: Considere novamente o caso da distribuição $N(\mu,\sigma^2)$. Desejamos inferir sobre $\mu/\sigma$ e $\sigma$. A estatística suficiente neste caso é $(\frac{\bar{x}}{s},s)$, cuja função densidade pode ser fatorada como em (5.2.3) e (5.2.4), ou seja,

(5.2.5) $\quad f(\frac{\bar{x}}{s},s/\frac{\mu}{\sigma},\sigma) = f(\frac{\bar{x}}{s}/s,\frac{\mu}{\sigma},\sigma)f(s/\sigma)$

ou

(5.2.6) $\quad f(\frac{\bar{x}}{s},s/\frac{\mu}{\sigma},\sigma) = f(\frac{\bar{x}}{s}/\frac{\mu}{\sigma})f(s/\frac{\bar{x}}{s},\frac{\mu}{\sigma},\sigma)$

No primeiro caso a distribuição condicional de $\frac{\bar{x}}{s}$ dado s é $N(\frac{\sigma}{s}\frac{\mu}{\sigma},\frac{ns^2}{\sigma^2})$ e a distribuição de s é encontrada a partir da quantidade pivotal $\frac{(n-1)s^2}{\sigma^2}$ que tem distribuição $\chi^2_{n-1}$. No segundo caso a distribuição condicional de s dado $\frac{\bar{x}}{s}$ é $N\{\frac{\sigma s\mu}{\bar{x}},(\sigma^2s^2)/\bar{x}^2\}$ e $\sqrt{n}(\frac{\bar{x}}{s})$ tem distribuição t não central com (n-1) graus de liberdade e parâmetro de não centralidade $\sqrt{n}(\frac{\mu}{\sigma})$.

Dempster (1963) mostrou que a densidade fiducial marginal de $(\frac{\mu}{\sigma})$ é a mesma em ambos os casos. O mesmo não ocorre com a densidade fiducial condicional de $\sigma$ dado $\frac{\mu}{\sigma}$ encontrada através de (5.2.6) que é $(\frac{s}{\sigma})$ vezes a aquela encontrada através de (5.2.5). Assim a densidade fiducial conjunta $g(\frac{\mu}{\sigma},\sigma/\frac{\bar{x}}{s},s)$ obtida em cada um dos casos é igual a

(5.2.7) $$\frac{\sqrt{n}}{\sqrt{2\pi}}\frac{(1/2)^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})}\frac{2}{\sigma}\left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n-1}{2}}\exp\left\{\frac{-(n-1)s^2}{2\sigma^2}-\frac{n}{2}\left(\frac{s}{\sigma}\frac{\bar{x}}{s}-\frac{\mu}{\sigma}\right)^2\right\}$$

(originária de (5.2.5))

$$(5.2.8) \qquad \frac{\sqrt{n}}{\sqrt{2\pi}} \; \frac{(1/2)^{\frac{n-1}{2}} \, 2}{\Gamma(\frac{n-1}{2})\sqrt{n-1}\,\sigma} \left[\frac{(n-1)s^2}{\sigma^2}\right]^{\frac{n}{2}} \exp\left\{\frac{-(n-1)s^2}{2\sigma^2} - \frac{n}{2}\left(\frac{s}{\sigma}\frac{\bar{x}}{s} - \frac{\mu}{\sigma}\right)^2\right\}$$

(originária de (5.2.6))

Se assumirmos $t_1 = \frac{\bar{x}}{s}$, $\theta_1 = \frac{\mu}{\sigma}$, $t_2 = s$, $\theta_2 = \sigma$ a fatoração proposta no segundo caso, (5.2.6) verificamos que esta satisfaz a condição indicada em (5.2.2) quando $h(t_1,t_2;\theta_2) = \frac{s}{\sigma}$. O mesmo não ocorre no primeiro caso, (5.2.5). Com base nestes resultados a indicação é que a fatoração do primeiro caso não apresenta subconjuntos relevantes e portanto é razoável utilizá-la.

f) Nos modelos que possuem com proriedade de invariância sob um grupo de transformações G (capítulo 6) não existe problema de unicidade nas distribuições fiduciais resultantes. Além disso, essas distribuições fiduciais são idênticas às distribuições a posteriori Bayesiana com priori não informativa.

A distribuição fiducial conjunta obtida no exemplo 5.2.1, não satisfaz a restrição III e produz uma distribuição marginal de $\sigma_1$ e $\sigma_2$ incorreta(*) (Dempster, 1963). Quanto a distribuição fiducial de $\rho$ dado r, Fraser (1961 a) mostra, utilizando invariância (capítulo 6) que esta é uma distribuição fiducial válida, ou seja, possue a propriedade de confiança.

(*) Incorreta no sentido de não possuir propriedade de confiança esperada, pois depende de $\rho$ (quando $\rho=0$, essa distribuição fiducial marginal se altera).

## 5.3. **Distribuição de Behrens-Fisher**

A distribuição fiducial da diferença entre médias de duas normais independentes com variâncias não necessariamente iguais e desconhecidas, tem interesse histórico como um primeiro exemplo no qual limites fiduciais não são limites de confiança. Esta distribuição é conhecida como distribuição de Behrens-Fisher.

### 5.3.1. Duas abordagens da distribuição de Behrens-Fisher

Consideremos duas amostras de tamanhos $n_1$ e $n_2$ de duas populações normais independentes; a primeira com média $\mu_1^2$ e variância $\sigma_1^2$ desconhecidas e a segunda com média $\mu_2$ e variância $\sigma_2^2$ também desconhecidas. Desejamos inferir sobre $\mu_1-\mu_2$. A densidade fiducial de $\mu_1-\mu_2$ pode ser encontrada usando as seguintes abordagens:

(I) abordagem proposta por Fisher (1935)

$(\bar{x}_1, s_1^2)$ é uma estatística suficiente completa com relação a $(\mu_1; \sigma_1^2)$ . Definimos $\mu_1 = \bar{x}_1+s_1t_1$ onde $t_1 \sim t_{n_1-1}$.
Analogamente, temos $\mu_2 = \bar{x}_2+s_2t_2$ onde $t_2 \sim t_{n_2-1}$ e é independente de $t_1$.

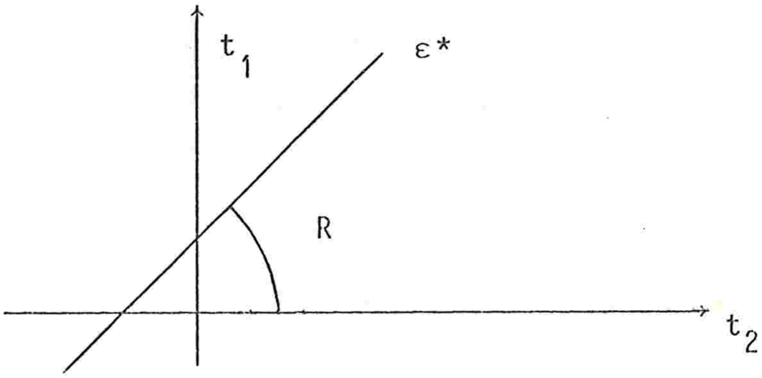
Sejam $\mu_1-\mu_2 = \delta$ e $\bar{x}_1-\bar{x}_2 = d$ então $\varepsilon = \delta-d = s_1t_1-s_2t_2$, onde $s_1$ e $s_2$ são conhecidos através da observação da amostra. Fazendo

$\operatorname{tg} R = \frac{s_1}{s_2}$ , podemos escrever

(5.3.1) $$\varepsilon^* = \frac{\varepsilon}{\sqrt{s_1^2+s_2^2}} = t_1 \cos R - t_2 \operatorname{sen} R$$

que está representada na Figura 5.1.

Figura 5.1: Gráfico de $\varepsilon^*$ como função de $t_1$ e $t_2$




Fisher (1935) encontrou a distribuição fiducial de $\varepsilon^*$ que é apresentada pela densidade abaixo:

(5.3.2) $$f(\varepsilon^*) \propto \frac{\partial t_1}{\left(1+\frac{t_1^2}{n_1-1}\right)^{n_1/2}} \cdot \frac{\partial t_2}{\left(1+\frac{t_2^2}{n_2-1}\right)^{n_2/2}}$$

Quando $n_1$ e $n_2$ crescem, a distribuição de $\varepsilon^*$ tende para uma distribuição normal, independente de R e quando R é $0^{\circ}$ ou $90^{\circ}$ a distribuição definida por (5.3.2) é t-Student . Tabelas com valores de $n_1, n_2$ e $\frac{s_1^2}{s_2^2}$ (ou o equivalente ângulo R) dão o valor de $\varepsilon^*$ correspondente à probabilidade fiducial fixada (por exemplo, Fisher (1941)).

O teste de Behrens pode ser descrito da seguinte forma: para testar $\mu_1=\mu_2$ podemos calcular $d = \frac{(\bar{x}_2-\bar{x}_1)}{[(s_1^2/n_1)+(s_2^2/n_2)]^{1/2}}$ e consultar a distribuição de $\varepsilon^*\sqrt{s_1^2+s_2^2}$ para verificar se o desvio com relação a zero é *significante*.

(II) Abordagem proposta por Fisher (1939)

Neste caso, Fisher utilizou a distribuição fiducial conjunta de $\mu_1, \mu_2, \sigma_1^2$ e $\sigma_2^2$.

Sabemos que $\dfrac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ e $\dfrac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$ e que essas quantidades pivotais são independentes. Portanto, podemos definir

(5.3.3) $$\omega = \left\{\frac{(n_1-1)s_1^2}{\sigma_1^2} + \frac{(n_2-1)s_2^2}{\sigma_2^2}\right\} \sim \chi^2_{(n_1+n_2-2)}$$

Além disso, sabemos também que

$$[(\bar{x}_1-\mu_1) - (\bar{x}_2-\mu_2)] \sim N(0,\sigma_1^2+\sigma_2^2)$$

Então, de (5.3.3) juntamente com o resultado acima, temos que

(5.3.4) $$\frac{\{(\bar{x}_1-\mu_1) - (\bar{x}_2-\mu_2)\}/\sigma_1^2+\sigma_2^2}{\left\{\dfrac{(n_1-1)s_1^2}{\sigma_1^2} \quad \dfrac{(n_2-1)s_2^2}{\sigma_2^2}\right\}} \sim t^2$$

onde $t \sim t_{(n_1+n_2-2)}$. Lembremos também que $\{(\bar{x}_1-\mu_1)-(\bar{x}_2-\mu_2)\}^2 = $ $= \varepsilon^* \; (s_1^2+s_2^2)$ ($\varepsilon^*$ é o mesmo definido em 5.3.1). A expressão em (5.3.3) pode ser reescrita como segue

$$\omega = \left\{\frac{(n_1-1)s_1^2}{\sigma_1^2} \quad \frac{(n_2-1)s_2^2}{\sigma_2^2}\right\} \Longrightarrow \frac{\sigma_1^2}{s_1^2}\,\omega = \{(n_1-1)+(n_2-1)\,\frac{p}{R'}\,\}$$

onde $p = (\sigma_1^2)/(\sigma_2^2)$ e $R' = (s_1^2)/(s_2^2)$. Com essa modificação (5.3.4) pode ser reescrita como

$$\frac{(\varepsilon^*)^2(n_1+n_2-2)\cdot(s_1^2+s_2^2)}{(1+\frac{1}{p})s_1^2[(n_1-1)+(n_2-1)\frac{p}{R'}]} = \frac{(\varepsilon^*)^2(n_1+n_2-2)(1+\frac{1}{R'})}{(1+\frac{1}{p})[n_1-1)+(n_2-1)\frac{p}{R'}]} \sim t^2$$

produz a distribuição fiducial conjunta de $\varepsilon^*$ e p. O valor de p é desconhecido, no entanto integrando a distribuição fiducial conjunta com relação a p o resultado é equivalente àquele encontrado em (I) e é chamada *distribuição de Behrens-Fisher*.

A diferença entre o método fiducial e o método dos intervalos de confiança se torna clara na distribuição de Behrens-Fisher. Olhando a distribuição fiducial conjunta de $\varepsilon^*$ e p os intervalos que esta determina para $\mu_1-\mu_2$ não são intervalos de confiança. O intervalor abaixo dá uma idéia deste fato:

$$(\bar{x}_1-\bar{x}_2) - \varepsilon_2\sqrt{s_1^2+s_2^2} \leq \mu_1-\mu_2 \leq (\bar{x}_1-\bar{x}_2) + \varepsilon_1\sqrt{s_1^2+s_2^2}$$

não é satisfeita independente de p, pois $\varepsilon_1$ e $\varepsilon_2$ são funções de $p = (\sigma_1^2)/(\sigma_2^2)$. E assim os limites para $\mu_1-\mu_2$ na distribuição de Behrens-Fisher não são limites de confiança, pois "dependem" de parâmetros que no caso são desconhecidos.

### 5.3.2. Princípio da Difusão

Na teoria dos intervalos de confiança, um intervalo (ou região) de confiança não deve depender de nenhum parâmetro desconhecido. No caso uniparamétrico, vimos (seção 3.2) que um intervalo fiducial deve possuir a propriedade de confiança, o que implica que um intervalo fiducial com probabilidade fiducial $(1-\alpha)$ deve ser um intervalo de confiança com nível de confiança $(1-\alpha)$. No entanto isto não é necessariamente verdade no caso multiparamétrico e a distribuição de Behrens-Fisher é um exemplo disto.

Alguns autores como Bartlett (1936) criticaram a distribuição de Behrens-Fisher alegando que os intervalos fiduciais desta distribuição não são intervalos de confiança, pois dependem de $\sigma_1/\sigma_2$, (como foi observado na abordagem (II). Podemos concretizar isto através da seguinte observação: considere $n_1=n_2=n$, então a estatística abaixo

$$(5.3.1) \qquad d = \frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{\sqrt{s_1^2 + s_2^2}}$$

quando $\sigma_1=\sigma_2$ tem distribuição t-Student com $(2n-2)$ g.$\ell$. Esta mesma estatística, quando consideramos a distribuição de Behrens-Fisher depende fortemente de $s_1^2/s_2^2$ quando $n$ é pequeno e assume valores que são aproximadamente iguais ao de uma t-Student com $(n-1)$ g.$\ell$. Há, portanto uma "perda" de $(n-1)$ g.$\ell$.

Observando que a distribuição de Behrens-Fisher é uma distribuição marginal, Fisher (1939) utilizou as seguintes considerações:

a) Intervalos fiduciais devem ter a propriedade de confiança mas têm interpretação diferente dos intervalos de confiança, pois a probabilidade fiducial é encarada como uma probabilidade comum.

b) se $\sigma_1$ e $\sigma_2$ são conhecidos a distribuição fiducial de $\mu_1-\mu_2$ condicionada em $\sigma_1$ e $\sigma_2$ tem a propriedade de confiança. A probabilidade fiducial associada aos intervalos fiduciais, neste caso, depende de $\sigma_1$ e $\sigma_2$.

c) Se desconhecemos $\sigma_1$ e $\sigma_2$, é razoável que nossa informação sobre $\mu_1-\mu_2$ seja mais "imprecisa" do que em b). No entanto, essa "imprecisão" não invalida, em princípio, o fato de que podemos associar probabilidades fiduciais aos intervalos paramétricos.

Com base nisto, conclui que as distribuiçoes fiduciais marginais que possuem "propriedade de confiança esperada", como mostra a restrição III da seção 5.2, são válidas como distribuições fiduciais, e justifica através da seguinte argumentação: é plausível que a estatística d perca (n-1)g.$\ell$ quando se retira a informação de que $\sigma_1=\sigma_2$, isto é interpretado como uma medida de acréscimo de "incerteza" sobre $\mu_1-\mu_2$. Portanto, a distribuição de d é menos "concentrada" quando desconhecemos o valor de $\sigma_1$ e $\sigma_2$. Esse acréscimo de incerteza se refletindo no fato da distribuição marginal ser "menos concentrada" ou com outras palavras, "mais difusa" é chamado de *Princípio de Difusão*.

Desde que uma distribuição marginal possua a "propriedade de confiança esperada" o princípio de difusão justifica a sua validade.

O caso do coeficiente de correlação $\rho$ (Fisher, 1930), aparentemente se enquadra dentro desse princípio.

Então, no caso multiparamétrico, temos que levar em conta as mesmas restrições do caso uniparamétrico acrescentando a essas considerações o princípio de Difusão. Acrescentamos que no caso multiparamétrico, ainda há muita controvérsia e não se conseguiu estabelecer uma forma de trabalhar coerentemente com essas distribuições fiduciais.

oOo

CAPÍTULO 6

# EXTENSÃO DO MÉTODO FIDUCIAL: MODELO ESTRUTURAL

Fraser (1968) desenvolveu a teoria dos modelos estruturais aparentemente como uma extensão do argumento fiducial. Esses modelos são adequados às situações onde a variável observável X, é considerada como função do parâmetro $\Theta$ , e de uma variável aleatória $\underline{e}$, cuja distribuição é conhecida e não depende de $\Theta$. Sendo assim, este tipo de modelo é composto de dois elementos:

(1) $x = h(\Theta;e), \quad x \in \chi \;, \quad \Theta \in \circledR{H} \;, \quad e \in E$

(2) a distribuição de probabilidade P, de $\underline{e}$ tal que

$$\forall \quad \Theta_j, \Theta_2 \in \circleddash{H} \quad e \quad A \subset E$$

$$P(\underline{e} \in A/\Theta_1) = P(\underline{e} \in A/\Theta_2).$$

que representaremos resumidamente por (6.1) $<x = h(\Theta;e), \; e \sim P>$. A aplicabilidade desse tipo de modelo pode ser ilustrada através do seguinte exemplo

<u>Exemplo 6.1</u>: Considere um instrumento I para medir uma certa quantidade física. Suponha que os erros de medida que I produz podem ser descritos por uma variável aleatória erro $\underline{e} \sim N(0,1)$. O "valor" da

v.a. erro é a diferença entre a leitura no instrumento x, e o valor da quantidade física $\Theta$. Então, o modelo é descrito da seguinte forma:

$$<x = \Theta + \underline{e},\ \underline{e} \sim N(0,1)>$$

Situações como esta do exemplo 6.1, são imagináveis em problemas nos quais $\underline{e}$ é visualizada como uma "fonte indefinida de variação" do tipo erro de medida (exemplo 6.1), variação na qualidade dos produtos de uma linha de fabricação, efeitos de aleatorização em experimentos e assim por diante. O termo variável erro deve ser interpretado em um sentido amplo e é esta variável que determina a estrutura probabilística do modelo.

A base da inferência estrutural é inverter a relação inicial, $x=h(\Theta;\underline{e})$ expressando $\Theta$ como função de x e $\underline{e}$ ou seja, $\Theta=h^*(x;\underline{e})$. Como a distribuição de probabilidades de $\underline{e}$ é conhecida e x é um valor observado, através da relação $\Theta = h^*(x;\underline{e})$ obtemos uma "distribuição de probabilidades" para $\Theta$ (distribuição fiducial). O exemplo abaixo ilustra este fato:

Exemplo 6.2: Considere o exemplo 6.1 onde, se observamos x=3,0 podemos escrever $\Theta = 3{,}0-\underline{e}$, donde resulta que $\Theta$ tem "distribuição"(*) normal com média 3,0 e variância 1, $\Theta \sim N(3{,}0;1)$.

Esta "distribuição de probabilidades" para $\Theta$ é chamada na *inferência estrutural* de *distribuição estrutural*.

---

(*) Colocamos "distribuição" *para* $\Theta$ entre aspas, pois esta distribuição é a distribuição fiducial, que não é uma distribuição no sentido comum (Capítulo 3).

Não basta encontrar uma relação $x=h(\Theta;\underline{e})$ e a respectiva distribuição de probabilidades de $\underline{e}$ para definir um modelo estrutural. Este tipo de modelo tem sentido apenas para a classe de modelos que respeitam o princípio de invariância que descrevemos na próxima seção. Modelos mais gerais são discutidos no capítulo 7 onde abordamos os modelos funcionais.

Nas próximas seçõs procuramos dar subsídios para que um modelo seja identificado como estrutural (seção 6.1) e de que forma nesses modelos a distribuição estrutural é encontrada (seção 6.2 e 6.3).

## 6.1. Invariância e grupos de transformação

Esta seção tem por finalidade apresentar em que condições um modelo é invariante, a formalização da invariância através de grupos de transformação e como esses grupos atuam sobre os espaços $\Theta$, $\chi$ e $E$.

Iniciamos enunciando o *Princípio de Invariância*:

Dois experimentos aleatórios, $I_1$ e $I_2$, são definidos da seguinte forma: o experimento $I_i$ consiste em observar o valor $x_i$ correspondente a uma v.a. $X_i$ com distribuição $F_i(x_i/\Theta)$ e espaço amostral $\chi_i$, $i=1,2$.

Se existe uma transformação $g: \chi_1 \to \chi_2$ tal que $F_1(x_1/\Theta) = F_2(g(x_1)/\Theta)$, $\forall\ x_1 \in \chi$ e $\forall\ \Theta \in \Theta$, então os dois experimentos contêm a mesma informação sobre $\Theta$.

O princípio de invariância nos diz que os experimentos acima são idênticos com relação à informação que possuem sobre $\Theta$. Isso implica que a mesma inferência deve resultar de $x_1 \in \chi_1$ e $g(x_1) \in \chi_2$.

O exemplo a seguir ilustra uma situação onde o princípio de invariância se verifica.

Exemplo 6.1.1: Considere que X e Y representam o tempo de decaimento de uma certa partícula atômica, medido em segundos e em minutos, respectivamente. Suponha que X e Y/60 têm distribuição exponencial com parâmetro $\Theta$ . Então, os dois experimentos, que consistem em observar x e y, respectivamente, contém a mesma informação sobre $\Theta$.

A formalização do princípio de invariância utiliza as seguintes propriedades dos grupos de transformações:

Definição 6.1.1: Um grupo de transformações G de $\chi$ é um conjunto de funções $g: \chi \to \chi$, bijetoras, tal que

(i) se $g_1, g_2 \in G$ então $g_1 o g_2 \in G$

(ii) se $g \in G$ então a inversa $g^{-1} \in G$

Note que de (i) e (ii) concluímos que a função *identidade* $g_i$ pertence a G. $(g_i(x) = x)$.

Exemplo 6.1.2: Considere $\chi = \mathbb{R}$ e $G = \{g_{a,b}: a \in \mathbb{R}$ e $b > 0\}$ onde $g_{a,b}(x) = ax + b$. Este grupo G é chamado de grupo de locação-escala ou grupo *afim*.

Se X é uma v.a. definida no espaço amostral $\chi$ e com função de probabilidade pertencente à classe $\mathcal{F} = \{f(x/\theta); \theta \in \Theta\}$ e G é um grupo de transformações de $\chi$, então:

<u>Definição 6.1.2</u>: $\mathcal{F}$ é dita *invariante* com relação a G se para todo $g \in G$ e $\theta \in \Theta$ existe um único $\theta_* \in \Theta$ tal que $Y = g(x)$ tem densidade $f(y/\theta_*)$. Em tais situações $\theta_*$ irá ser denotado por $\bar{g}(\theta)$ e

$$P(g(x) \in A/\theta) = P(X \in A/\bar{g}(\theta)) \qquad \forall\, A \subset \chi.$$

Pode-se mostrar que $\bar{G} = \{\bar{g};\ g \in G\}$ é um grupo de transformações em $\Theta$.

Esta definição formaliza e generaliza o princípio da invariância. Este fato é ilustrado pelo exemplo abaixo:

<u>Exemplo 6.1.3</u>: Considere a situação do exemplo 6.1.1: para cada $g(x) = x/60$ existe um $\bar{g}(\theta) = \theta/60$ de tal forma que se x tem distribuição exponencial com parâmetro $\theta$, g(x) tem distribuição exponencial com parâmetro $\bar{g}(\theta)$. Ambos, $f(x/\theta)$ e $f(g(x)/\bar{g}(\theta))$ pertencem à família $\mathcal{F} = \{\theta^{-1}\exp\{-x/\theta\}\, I_{(0,\infty)}(x),\ \theta \in \Theta\}$. Portanto, $\mathcal{F}$ é invariante, com relação ao grupo de transformações $G = \{g_a;\ a > 0\}$ onde $g_a(x) = x/a$.

As próximas definições mostram de que forma o grupo de transformações atua na identificação dos elementos do espaço $\Theta$, $\chi$ e $\mathcal{E}$.

<u>Definição 6.1.3</u>: Dois elementos $\theta_1, \theta_2 \in \Theta$ são ditos *equivalentes* se para algum $\bar{g} \in \bar{G}$, $\theta_1 = \bar{g}(\theta_2)$. Uma *órbita* em $\Theta$ com relação a grupo de transformações $\bar{G}$ e ao elemento $\theta_0 \in \Theta$ é o conjunto

$$\Theta(\theta_0) = \{\bar{g}(\theta_0) \; ; \; \bar{g} \in \bar{G}\}$$

Cada órbita é um conjunto de pontos equivalentes. As órbitas formam uma partição de $\Theta$.

Analogamente, podemos definir órbitas em $\chi$ com respeito ao grupo G e denotá-las por $\chi(x_0)$, $x_0 \in \chi$.

Definição 6.1.4: Um grupo de transformações $\bar{G}$ de $\Theta$ é dito *transitivo* em $\Theta$ se $\Theta(\theta_0) = \Theta$, $\forall\ \theta_0 \in \Theta$ ; equivalentemente, $\forall\ \theta_1, \theta_2 \in \Theta$ existe algum $\bar{g} \in \bar{G}$ tal que $\theta_1 = \bar{g}(\theta_2)$.

Nestas condições, todos os elementos de $\Theta$ se equivalem com relação a $\bar{G}$ e dessa forma $\Theta$ consiste numa única órbita. O mesmo pode ser definido com relação a $\chi$ e $E$.

Cada órbita é uma partição do espaço original; podemos tentar distinguir essas partições associando uma constante aos elementos de uma partição e constantes diferentes a diferentes partições. Com base nisto, podemos comparar dois elementos do espaço original e saber se são ou não equivalentes.

Definição 6.1.5: Seja G um grupo de transformações em $\chi$. A função $t(x)$ em $\chi$ é

(i) invariante com respeito a G se $t(g(x) = t(x)$ para todo $x \in \chi$ e $g \in G$

(ii) invariante maximal com respeito a G se é invariante e satisfaz

$$t(x_1) = t(x_2) \Longrightarrow x_1 = g(x_2) \text{ para algum } g \in G.$$

Assim, uma função invariante é uma função constante nas órbitas, $\chi(x_0)$, $x \in \chi$. E é invariante maximal se além de constante nas órbitas, associa valores distintos à diferentes órbitas. Note que se G é transitivo em $\chi$, $\chi$ é composto de uma única órbita, então as únicas funções invariantes maximais são as funções constantes.

O exemplo abaixo concretiza uma função invariante maximal na situação em que G é o grupo afim.

Exemplo 6.1.4: Suponha $\chi = \mathbb{R}^n$ e $G = \{g_{a,b};\ a \in \mathbb{R},\ b > 0\}$ onde

$$g_{a,b}(x_1,\ldots,x_n) = (a+bx_1,\ldots,a+bx_n).$$

Se $\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad$ e $\quad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ ,

uma função invariante maximal é

$$(6.1.1) \qquad t(x) = t(x_1,\ldots,x_n) = \begin{cases} 0 & \text{se } s_x = 0 \\ \left(\frac{x_1-\bar{x}}{s_x},\ldots,\frac{x_n-\bar{x}}{s_x}\right) & \text{se } s_x \neq 0 \end{cases}$$

Note que o valor de $t(x)$ identifica uma órbita. Por exemplo, se $t(x) = 0$, então isto significa que $x \in \{x;\ x_1=\ldots=x_n, x_i \in \mathbb{R}\}$. O conjunto dos pontos tais que $t(x) = 0$, forma uma órbita específica em $\chi$; todos os elementos de $\chi$ que não pertencem a esta órbita têm valor $t(x) \neq 0$. Valores distintos da função invariante maximal identificam órbitas distintas.

A função invariante maximal não é única. Para o caso do exemplo acima podemos definir outra função invariante maximal fazendo: $Y_1 = \min\{x_1,\ldots,x_n\}$ e $Y_n = \{\max x_1,\ldots,x_n\}$ se $y_n \neq 0$ .

$$(6.1.2) \qquad t(x_1,\dots,x_n) = \begin{cases} \left(\dfrac{x_1-y_1}{y_n},\dots,\dfrac{x_n-y_1}{y_n}\right) & \text{se } y_n \neq 0 \\ 0 & \text{se } y_n = 0 \end{cases}$$

Em qualquer uma das duas funções, (6.1.1) e (6.1.2), associamos a diferentes órbitas, valores diferentes.

Definida a função invariante maximal $t$, identificamos as órbitas. Identificando um ponto de referência ou origem $u$, de uma órbita podemos identificar os outros elementos dessa órbita como transformações dessa origem.

<u>Definição 6.1.6</u>: Seja $G$ um grupo de transformações em $\chi$. O *ponto de referência ou origem* $u$, de uma órbita de $\chi$ é tal que $\forall\, x$ pertencente à órbita, $\exists t_x \in G$ tal que $x = t_x(u)$. ~~É assim, segue a definição abaixo~~.

Podemos identificar a órbita como $\chi(u)$.

Nada mais natural que identificar a órbita pela sua origem $u$. Então, a relação entre a origem e a função invariante maximal é $t(x) = u$ ($t$ associa a cada órbita a sua origem). Por outro lado, $\forall\, x$ pertencente à órbita $x = t_x(u)$, $t_x \in G$ $\quad u = t_x^{-1}(x)$. Então $t(x) = t_x^{-1}(x)$.

Assim, chamamos $t_x^{-1}$ de transformação invariante maximal. Essa transformação, que é uma função de $x$, mostra a relação que há entre $x$ e a origem u. Dessa forma, conhecendo a origem de uma órbita, basta saber quem é $t_x^{-1}$ para conhecermos x e vice-versa.

Exemplo 6.1.5: Considere a situação do exemplo 6.1.4. A função invariante maximal t, pode ser reescrita como

$$t(x_1,\ldots,x_n) = (\frac{x_1}{s_x} - \frac{\bar{x}}{s_x},\ldots,\frac{x_n}{s_x} - \frac{\bar{x}}{s_x}) = t_{(\frac{-\bar{x}}{s_x},\frac{1}{s_x})}(x_1,\ldots,x_n)$$

onde $t_{(\frac{-\bar{x}}{s_x},\frac{1}{s_x})} \in G$, grupo afim. Não é difícil ver que

$$t_{(\frac{-\bar{x}}{s_x}\ \frac{1}{s_x})} = t^{-1}_{(\bar{x},s_x)} \in G\ .$$

Então, conhecendo a origem $u = (\frac{x_1-\bar{x}}{s_x},\ldots,\frac{x_n-\bar{x}}{s_x})$ , $t^{-1}_{(\bar{x},s_x)}$ especifica qual a relação entre u e $(x_1,\ldots,x_n)$. De forma geral, para o grupo afim, $t_x^{-1}$ pode ser escrito como $t^{-1}_{(b(x),s(x))}$ , onde $b(x)$ e $s(x)$ são funções de x.

Até o momento descrevemos alguns resultados referentes à invariância e aos grupos de transformação. Na próxima seção esses resultados são utilizados na identificação de um modelo estrutural.

### 6.1.1. Condições para a identificação de um modelo estrutural

Considere $\bar{G}$ um grupo de transformações transitivo em $\Theta$, ou seja, fixando $\theta_0 \in \Theta$ um ponto de referência, para todo outro elemento $\theta \in \Theta$ existe um $\bar{g} \in \bar{G}$ tal que $\theta = \bar{g}(\theta_0)$ e assim a cada $\theta \in \Theta$ associamos $\bar{g} \in \bar{G}$. Portanto, se existe uma transformação biunívoca entre os elementos de $\Theta$ e $\bar{G}$ então $\Theta$ e $\bar{G}$ são *isomorfos*.

Além disso, se $\bar{G}$ é um grupo de transformações definido como um conjunto de funções $\bar{g}_a$ com a mesma "forma" e "parâmetro" a, definido sobre $\Theta$ tal que $\bar{G}$ e $\Theta$ são isomorfos, então dizemos que $\Theta$ pode ser *identificado* com $\bar{G}$.

Exemplo 6.1.6: Considere a relação (1) $x = \theta_1+\theta_2\underline{e}$. O espaço paramétrico é $\Theta = \{(\theta_1,\theta_2);\ \theta_1 \in \mathbb{R}$ e $\theta_2 > 0\}$. As transformações $\bar{g}_{(a,b)}(\underline{e}) = a+b\underline{e}$ têm todas a mesma "forma" (linear) e diferem entre si pelos "parâmetros" (a,b) que estão definidos em $\Theta$. Então a relação (1) pode ser reescrita como $x = \bar{g}_{(\theta_1,\theta_2)}(\underline{e})$, onde $\bar{g}_{(\theta_1,\theta_2)} \in \bar{G}$ grupo afim. Neste grupo podemos mostrar que

(i) $$\bar{g}_{(a_1,b_1)} \circ \bar{g}_{(a_2,b_2)} = \bar{g}_{(a_1+b_1a_2,\,b_1b_2)}$$

(ii) $$\bar{g}^{-1}_{(a_1,b_1)} = \bar{g}_{\left(-\frac{a_1}{b_1},\frac{1}{b_1}\right)}$$

(iii) $$\bar{g}_i = \bar{g}_{(0,1)} = \text{identidade}$$

Fazendo $\theta_0 = (0,1)$, então $\forall\ (\theta_1,\theta_2) \in \Theta$ associamos um $\bar{g}_{(\theta_1,\theta_2)} \in \bar{G}$. Então, $\bar{G}$ e $\Theta$ são isomorfos e $\Theta$ pode ser identificado com $\bar{G}$.

O grupo afim, por ser bem simples, será utilizado mais adiante para mostrar como se processa a inferência estrutural e a partir dele generalizar para situações mais complexas.

Descrevemos a seguir as condições nas quais um modelo é identificado como estrutural. A exposição está dividida em duas partes: a primeira diz respeito à relação entre x, $\Theta$ e $\underline{e}$ chamada de equação estrutural; a segunda aborda as condições impostas à distribuição de probabilidades de $\underline{e}$.

6.1.1.1. Equação Estrutural

Se a relação da variável observável x, com o parâmetro $\Theta$, e a variável aleatória erro $\underline{e}$, puder ser expressa na forma $x=\bar{g}_\Theta(\underline{e})$, onde $\bar{g}_\Theta$ é uma transformação pertencente ao grupo de transformações $\bar{G}$ com o qual 4 pode ser *identificado*, então

(i) $\bar{G}$ é um grupo de transformações de $\chi$ e $E$

(ii) se $x = \bar{g}_\Theta(\underline{e}) \Longrightarrow$ x e $\underline{e}$ são equivalentes com relação a $\bar{G}$, ou seja, pertencem à mesma órbita

(iii) necessariamente $\chi$ e $E$ têm a mesma dimensão

O item (ii) pode ser demonstrado da seguinte forma: $\chi(u)$ é uma órbita de $\chi$ com origem u, então $\chi(u) = \{x = t_x(u);\ t_x \in \bar{G}\}$, onde $x = t_x(u)$ vem da definição 6.1.6 e $\bar{g}_\Theta \in \bar{G}$

No entanto, $x = \bar{g}_\Theta(\underline{e}) \Longrightarrow \bar{g}_\Theta(\underline{e}) = t_x(u) \Longrightarrow \underline{e} = (\bar{g}_\Theta^{-1} o t_x)(x)$ onde $\bar{g}'_\Theta = (\bar{g}_\Theta^{-1} o t_x) \in \bar{G}$ . Assim

$$\chi(u) = \{x = t_x(u);\ t_x \in \bar{G}\} = \{\underline{e} = \bar{g}_\Theta^{1}(u);\ \bar{g}'_\Theta \in \bar{G}\} = E(u)$$

O item (iii) é conclusão imediata do item (ii).

### 6.1.1.2. Medidas invariantes (ou Haar)

Para definirmos um modelo estrutural é necessário: (1) a relação $x=\bar{g}_\Theta(\underline{e})$ e (2) a distribuição de probabilidades de $\underline{e}$. As condições relativas a (1) foram definidas na seção anterior. O assunto desta seção é referente a (2). Procuramos expor as condições impostas à distribuição de probabilidades de $\underline{e}$. Estas condições associadas às condições da seção anterior configuram as situações nas quais podemos definir um modelo estrutural.

Uma das restrições necessárias para a utilização do modelo estrutural é que as medidas utilizadas sejam invariantes. Em outras palavras: temos uma medida de probabilidade definida sobre $E$; quando uma transformaçaõ de grupo $\bar{g}$, é efetuada sobre um elemento $\underline{e}$ de $E$, o valor da medida associada a $\bar{g}(\underline{e})$ é proporcional ao valor da medida associada a $\underline{e}$. A constante da proporcionalidade é dada pelas transformações de grupo.

Para que isto aconteça é necessário que tenhamos medidas de probabilidade absolutamente contínuas, ou seja, para as quais uma função densidade de probabilidade possa ser considerada.

Como a intenção é encontrar a "distribuição de probabilidades" (estrutural) de $\Theta$, deve ser válido para $\Theta$ a mesma restrição imposta a $E$ ou $\chi$ : $\Theta$ não é um conjunto enumerável e a medida associada a qualquer subconjunto é positiva, (é subconjunto de $\mathbb{R}^p$ $p \geq 1$ com medida de Lebesgue positiva)(Halmos, 1950)(*).

Representaremos a distribuição de probabilidades de $e = (e_1,\ldots,e_n)$ por $p(e_1,\ldots,e_n)de_1\ldots de_n$ , onde $p(e_1,\ldots,e_n)$ é a função densidade de probabilidade de $(e_1,\ldots,e_n)$ e $de_i$ é o diferencial de $e_i$ , $i=1,\ldots,n$. Então a distribuição de probabilidades de $(x_1,\ldots,x_n) = \bar{g}_\Theta(e_1,\ldots,e_n)$ pode ser expressa como

$$p(\bar{g}_\Theta^{-1}(x_1,\ldots,x_n))|J_{\bar{g}}^*|\,dx_1\ldots dx_n$$

onde $|J_{\bar{g}}^*|$ é o jacobiano da transformação de $(e_1,\ldots,e_n)$ em $\bar{g}_\Theta^{-1}(x_1,\ldots,x_n)$.

<u>Exemplo 6.1.1.2.1</u>: Seja $\bar{G}$ o grupo afim; então para $\bar{g}_{(\Theta_1,\Theta_2)} \in \bar{G}$ , $(x_1,\ldots,x_n) \in \chi$ e $(e_1,\ldots,e_n) \in E$ temos

(i) $$(x_1,\ldots,x_n) = g_{(\Theta_1,\Theta_2)}(e_1,\ldots,e_n) = (\Theta_1+\Theta_2 e_1,\ldots,\Theta_1+\Theta_2 e_n)$$

(ii) $$(e_1,\ldots,e_n) = \bar{g}_{(\Theta_1,\Theta_2)}^{-1}(x_1,\ldots,x_n) = \left(\frac{x_1-\Theta_1}{\Theta_2},\ldots,\frac{x_n-\Theta_1}{\Theta_2}\right)$$

O jacobiano da transformação de $E$ em $\chi$ é $|J_{\bar{g}}^*| = (1/\Theta_2)^n$. Analogamente, a transformação de $\chi$ em $E$ tem jacobiano igual a $|J_{\bar{g}}^*| = \Theta_2^n$ .

As medidas invariantes (ou Haar) definidas em um grupo de transformações $\bar{G}$ ou seja, transformações de $\bar{g}$ em $\bar{G}$ são consideradas no apêndice A. Aqui basta saber que transformações de $\bar{g}$ em $\bar{g}\circ\bar{g}_o$, onde $\bar{g}_o$ é

(*) Halmos, P.R. (1950) - Measure Theory, New York; D. Van Nostrand um elemento fixado de $\bar{G}$, tem jacobiano denotado por $|J_{\bar{g}}|$, e densidade de Haar igual a $|J_{\bar{g}}|^{-1}$ (vide apêndice A).

O modelo estrutural e a inferência resultante desse modelo são descritos na sequência.

## 6.2. **Inferência Estrutural: distribuição estrutural de $\Theta$**

Seja $\bar{G}$ um grupo de transformações em $\chi$ e $E$, com o qual $\Theta$ é identificado. Considere o seguinte modelo estrutural, utilizando a notação indicada no início deste capítulo (exemplo 6.1):

$$<x = \bar{g}_{\Theta}(\underline{e}); \quad e \sim p(\underline{e})d\underline{e}>$$

$$(x_1,\ldots,x_n) \in \chi, e=(e_1,\ldots,e_n) \in E, \ \bar{g}_{\Theta} \in \bar{G}$$

Podemos inverter a relação

(6.2.1) $$(x_1,\ldots,x_n) = \bar{g}_{\Theta}(e_1,\ldots,e_n)$$

obtendo $(e_1,\ldots,e_n) = \bar{g}_{\Theta}^{-1}(x_1,\ldots,x_n)$. Quando o valor de $(x_1,\ldots,x_n)$ é observado $(e_1,\ldots,e_n)$ é função do parâmetro $\Theta$ (através de $\bar{g}_{\Theta}^{-1}$). Num processo análogo ao da inferência fiducial devemos "transferir" a medida de probabilidade de $(e_1,\ldots,e_n)$ para $\Theta$. No entanto, as dimensões de $(e_1,\ldots,e_n)$ e $\Theta$ podem ser diferentes, o que dificulta essa "transferência". Lembremos que no método fiducial (seção 2.3) necessitavamos encontrar uma estatística suficiente mínima $(t,u)$ onde $u$ é $(n-k)$ dimensional e ancilar com relação a $\Theta$ e $t$ é suficiente condicional com relação a $\Theta$ e tem a mesma dimensão deste (K). A distribuição fiducial é encontrada tendo como base $f(t/\Theta,u)$ que é uma "redução" do modelo original $f(x_1,\dots,x_n/\Theta) \propto f(t,u/\Theta)$. Aqui a idéia é análoga:

(i) Devemos encontrar uma função invariante maximal t em $\bar{G}$ (definição 6.1.5)

(ii) A função t, define um ponto de referência na órbita à qual pertence $x \in \chi$ (que é a mesma de $\underline{e}$, seção 6.1.1.1).

Lembramos que u é definida de tal forma que $t(u) = u = (u_1,\dots,u_n)$. Segundo a definição 6.1.6, $\forall\ x \in \chi$, existe uma transformação $t_x \in \bar{G}$ tal que $x = t_x(u)$ e então $t(x) = t_x^{-1}(x)$, onde $t_x^{-1}$ é chamada transformação invariante maximal. Da mesma forma, $t_u(u) = u \Longrightarrow t_u = \bar{g}_i$ = identidade.

Podemos reescrever a relação $(x_1,\dots,x_n) = \bar{g}_\Theta(e_1,\dots,e_n)$ em função da origem u, da seguinte forma:

(6.2.2) $$t_x(u_1,\dots,u_n) = \bar{g}_\Theta(t_e(u_1,\dots,u_n))$$

(lembramos que x e $\underline{e}$ pertencem à mesma órbita, portanto faz sentido $t_e(u_1,\dots,u_n) = (e_1,\dots,e_n)$, $t_e \in \bar{G}$)

A expressão (6.2.2) pode ser reescrita como

(6.2.3) $$t_x(u_1,\dots,u_n) = (\bar{g}_\Theta o t_e)(u_1,\dots,u_n)$$

onde $(\bar{g}_\Theta o t_e) \in \bar{G}$. Portanto, conhecendo a origem $(u_1, \ldots, u_n)$ podemos reduzir (6.2.1) a uma relação entre $t_x$, $\bar{g}_\Theta$ e $t_e$.

Fazemos inicialmente essa discussão para um caso simples, no qual $\bar{G}$ é o grupo afim e em seguida generalizamos para situações mais complicadas

6.2.1. Modelo de Locação-Escala

O modelo de locação-escala é representado da seguinte forma:

$$<(x_1,\ldots,x_n) = (\Theta_1 + {}_2e_1,\ldots,\Theta_1+\Theta_2+e_n)\ ;\ (e_1,\ldots,e_n) \sim p(e_1,\ldots,e_n)de_1\ldots de_n>$$

onde $\bar{G} = \{\bar{g}_{(a,b)};\ a \in \mathbb{R} \text{ e } b > 0\}$ onde $\bar{g}_{(a,b)}(x) = a+bx$.

Podemos reescrever a relação $(x_1,\ldots,x_n) = (\Theta_1+\Theta_2 e_1,\ldots,\Theta_1+\Theta_2 e_n)$ em

$$(6.2.1.1) \qquad \begin{pmatrix} 1 \ldots 1 \\ \\ x_1 \ldots x_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \\ \Theta_1 & \Theta_2 \end{pmatrix} \begin{pmatrix} 1 \ldots 1 \\ \\ e_1 \ldots e_n \end{pmatrix}$$

onde $\bar{G}$ é agora identificado como

$$(6.2.1.2) \qquad \bar{G} = \left\{ \begin{pmatrix} 1 & 0 \\ \\ a & b \end{pmatrix} \ ;\ a \in \mathbb{R} \ \text{ é } b > 0 \right\}$$

As operações entre os elementos de $\bar{G}$ obedecem as regras de operações entre matrizes. Essa notação matricial facilita a visualização das operações entre os elementos do grupo e permite que mais adiante (seção 6.2.2) façamos a generalização dos resultados encontrados nesta seção. (observe que (a,b) são os "parâmetros" das transformações pertencentes a $\bar{G}$).

Uma transformação invariante maximal definida para o grupo $\bar{G}$ em (6.2.1.2) é

$$t_{(\cdot)} = \begin{bmatrix} 1 & 0 \\ b(\cdot) & s(\cdot) \end{bmatrix}$$

onde $b(\cdot)$ e $s(\cdot)$ são funções definidas em $\chi$ (ou $E$). Assim, o exemplo 6.1.5 pode ser representado por

$$t_x^{-1}(x) = \begin{bmatrix} 1 & 0 \\ & \\ b(x) & s(x) \end{bmatrix}^{-1} \begin{bmatrix} 1 \ldots\ldots 1 \\ \\ x_1 \ldots\ldots x_n \end{bmatrix}$$

que para $b(x) = \bar{x}$ e $s(x) = s_x$ é reescrito como

$$\begin{bmatrix} 1 & 0 \\ & \\ \bar{x} & s_x \end{bmatrix}^{-1} \begin{bmatrix} 1 \ldots\ldots\ldots 1 \\ \\ x_1 \ldots\ldots\ldots x_n \end{bmatrix}$$

A origem $u = (u_1, \ldots, u_n)$ da órbita é encontrada fazendo a transformação $t_x^{-1}(x) = u$ que na notação matricial é descrita como segue

$$(6.2.1.3) \qquad \begin{bmatrix} 1 \ldots\ldots\ 1 \\ \\ x_1 \ldots\ldots\ x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ & \\ b(x) & s(x) \end{bmatrix} \begin{bmatrix} 1 \ldots\ldots 1 \\ \\ u_1 \ldots\ldots u_n \end{bmatrix}$$

onde $(b(x), s(x))$ descrevem a "posição" de $(x_1,\dots,x_n)$ relativamente a $(u_1,\dots,u_n)$.

Analogamente, $(e_1,\dots,e_n)$ também pode ser representado em função da origem

$$(6.2.1.4)\qquad \begin{pmatrix} 1 \dots\dots 1 \\ \\ e_1 \dots\dots e_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix} \begin{pmatrix} 1 \dots\dots 1 \\ \\ u_1 \dots\dots u_n \end{pmatrix}$$

Assim a relação (6.2.1.1) pode ser reescrita em função da $(u_1,\dots,u_n)$ utilizando (6.2.1.3) e(6.2.1.4), tal que

$$(6.2.1.5)\qquad \begin{pmatrix} 1 & 0 \\ \\ b(x) & s(x) \end{pmatrix} \begin{pmatrix} 1 \dots\dots 1 \\ \\ u_1 \dots\dots u_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \\ \Theta_1 & \Theta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix} \begin{pmatrix} 1 \dots\dots 1 \\ \\ u_1 \dots\dots u_n \end{pmatrix}$$

Podemos "reduzir" (6.2.1.5) da seguinte forma:

$$(6.2.1.6)\qquad \begin{pmatrix} 1 & 0 \\ \\ b(x) & s(x) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \\ \Theta_1 & \Theta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix}$$

que dá a relação entre $(b(x), s(x))$ e $(b(\underline{e}), s(\underline{e}))$ relativamente a $u = (u_1,\dots,u_n)$.

A distribuição de probabilidades de $(e_1,\dots,e_n)$

$$(6.2.1.7)\qquad p(e_1,\dots,e_n)\,de_1\dots de_n$$

pode ser colocada em função de $(u_1,\ldots,u_n)$. Da equação (6.2.1.4) concluimos que $(e_1,\ldots,e_n) = (b(\underline{e})+s(\underline{e})u_1,\ldots,b(\underline{e})+s(\underline{e})u_n)$, então a partir de (6.2.1.7) encontramos a distribuição de probabilidades de $b(\underline{e}), s(\underline{e})$, $(u_1,\ldots,u_n)$ que é

(6.2.1.8) $\quad p((b(\underline{e})+s(\underline{e})u_1,\ldots,b(\underline{e})+s(\underline{e})u_n))|J^*_{\bar{g}}|db(\underline{e})ds(\underline{e})du_1,\ldots,du_n$

onde $p(\cdot)$ é a densidade de probabilidade de $(e_1,\ldots,e_n)$, $|J^*_{\bar{g}}|$ é o jacobiano da transformação $(e_1,\ldots,e_n)$ em $(b(\underline{e})+s(\underline{e})u_1,\ldots,b(\underline{e})+s(\underline{e})u_n)$ que é igual a $(s(e))^n$. (vide exemplo 6.1.1.2.1) e $db(\underline{e})$, $ds(\underline{e})$, $du_i$ são os diferenciais de $b(\underline{e})$, $s(\underline{e})$, $u$, respectivamente.

A partir de (6.2.1.8) encontramos a distribuição condicional de $(b(\underline{e}), s(\underline{e}))$ dado $(u_1,\ldots,u_n)$ que é igual a

(6.2.1.9) $\quad K(u)p(b(\underline{e}),s(\underline{e})/u_1,\ldots,u_n)s(\underline{e}))^n|J_{\bar{g}}|^{-1}db(\underline{e})ds(\underline{e})$

onde $K(u)$ é uma constante de proporcionalidade que faz com que a integral em $b(\underline{e})$ e $s(\underline{e})$ seja igual a 1 e $|J_{\bar{g}}|^{-1}db(\underline{e})ds(\underline{e})$ é a medida de Haar à direita associada a $(b(\underline{e}),s(\underline{e}))$ com $|J_{\bar{g}}|^{-1} = (s(\underline{e}))^{-2}$, densidade de Haar à direita (vide apêndice A).

As expressões (6.2.1.6) e (6.2.1.9) *juntas* formam o *modelo estrutural de locação-escala reduzido*. É este modelo que utilizamos para encontrar a distribuição estrutural de $\Theta$. Note que no modelo reduzido não temos o problema inicial da dimensão de $(e_1,\ldots,e_n)$ e $(\Theta_1,\Theta_2)$ serem diferentes, pois $(b(\underline{e}),s(\underline{e}))$ tem a mesma dimensão de $(\Theta_1,\Theta_2)$ (e nenhuma informação é perdida, pois a distribuição de probabilidades é condicionada à $u$ que é o ponto de referência de todas as relações).

Chamamos a atenção para o fato de que $u = (u_1,\ldots,u_n)$ pode ser identificado como uma estatística ancilar. A órbita é o conjunto de referência (seção 2.2.1) associada ao modelo e os "parâmetros" $(b(\cdot),s(\cdot))$ da transformação invariante maximal formam uma estatística condicionalmente suficiente com relação a $\Theta$ (condicional a u).

6.2.1.1. Distribuição Estrutural resultante do modelo de locação-escala

Quando $(x_1,\ldots,x_n)$ é observado, $(b(x),s(x))$ tem valor conhecido. Da mesma forma, devido a (6.2.1.3), $u = (u_1,\ldots,u_n)$ também tem valor conhecido. A "redução" (6.2.1.6) pode ser reescrita como

$$(6.2.1.10)\qquad \begin{pmatrix} 1 & 0 \\ b(e) & s(e) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \Theta_1 & \Theta_2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ b(x) & s(x) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ -\dfrac{\Theta_1}{\Theta_2} & \dfrac{1}{\Theta_2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ b(x) & s(x) \end{pmatrix}$$

Como a distribuição condicional de $(b(e),s(e))$ dado (u é conhecida e $(b(e),s(e))$ é uma função de $(b(x),s(x))$ (valor conhecido) e $(\Theta_1,\Theta_2)$ (desconhecido), utilizamos a idéia do método fiducial (seção 2.2.2) "transferindo" a medida de $(b(e),s(e))$ para $(\Theta_1,\Theta_2)$ encontrando a "distribuição de probabilidades" de $(\Theta_1,\Theta_2)$ que é a distribuição estrutural.

Como de (6.2.1.10) sabemos que $(b(e),s(e)) =$

$= (\frac{b(x)-\Theta_1}{\Theta_2}, \frac{s(x)}{\Theta_2})$ , então utilizando (6.2.1.9) encontramos a distribuição estrutural de $(\Theta_1,\Theta_2)$ que é

(6.2.1.11) $\quad g(\Theta_1,\Theta_2/b(x),s(x),u)d\Theta_1 d\Theta_2 =$

$$= K(u)p(\frac{b(x)-\Theta_1}{\Theta_2} \quad \frac{s(x)}{2}/u)(\frac{s(x)}{\Theta_2})^{n-2} |J_{\ddot{g}}^*| d\Theta_1 d\Theta_2$$

onde $|J_{\ddot{g}}^*|$ é o jacobiano da transformação de $(b(e),s(e))$ em $(\frac{b(x)-\Theta_1}{\Theta_2},\frac{s(x)}{\Theta_2})$ que é igual a $(\frac{s(x)}{\Theta_3^3})$ e $d\Theta_1$ e $d\Theta_2$ são os diferenciais de $\Theta_1$ e $\Theta_2$.

Note que a igualdade (6.2.1.11) é válida *qualquer que seja* a densidade de probabilidades $p(e_1,...,e_n)$ associada e $(e_1,..,e_n)$

O exemplo abaixo ilustra este fato:

<u>Exemplo 6.2.1</u>: Considere um modelo de locação e escala tal que

$$p(e_1,...,e_n) = \prod_{i=1}^{n} p(e_i) \text{ onde } p(e_i) = \begin{cases} 1, & 0 \leq e_i \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Uma transformação invariante maximal neste caso é

$$\begin{pmatrix} 1 & 0 \\ b(x) & s(x) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ L & R \end{pmatrix} \text{, onde } \begin{cases} L = \min\{e_1,\dots,e_n\} \\ R = \max\{e_1,\dots,e_n\} - L \end{cases}$$

sugerimos essa transformação pois ela facilita os cálculos necessá - rios para encontrar a distribuição condicional de $(b(\underline{e}),s(\underline{e}))$ dado a origem $\mu = (\mu_1,\dots,u_n)$ (estatística ancilar) que é igual a

$$\begin{pmatrix} 1 \dots\dots 1 \\ u_1 \dots\dots u_n \end{pmatrix} = \begin{pmatrix} 1 & \dots\dots\dots\dots & 1 \\ \frac{e_1-L}{R} & & \frac{e_n-L}{R} \end{pmatrix}$$

O modelo estrutural reduzido é representado por (1) e (2) abaixo

(1) $$\begin{pmatrix} 1 & 0 \\ L(x) & R(x) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \Theta_1 & \Theta_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ L & R \end{pmatrix}$$

onde $L(x) = \min\{x_1,\dots,x_n\}$ e $R(x) = \max\{x_1,\dots,x_n\} - L(x)$.

(2) $$p(L,R/\mu)dLdR = n(n-1)I_{(L,R)}R^{n-2}dLdR$$

onde $$I_{(L,R)} = \begin{cases} 1 \text{ se } & 0 \leqq L \leqq L+R \leqq 1 \\ 0 & \text{caso contrário} \end{cases}$$

A distribuição estrutural de $(\Theta_1,\Theta_2)$ é igual a

$$g(\Theta_1,\Theta_2)/L(x),R(x),\mu) = n(n-1)\left(\frac{R(x)}{\Theta_2}\right)^{n-1}\left(\frac{1}{\Theta_2}\right)^2 I_{(L(x),R(x))}d\Theta_1 d\Theta_2$$

onde $$I_{(L(x),R(x))} = \begin{cases} 1 & \text{se} \quad 0 \leq L(x) \leq L(x)+R(x) \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

Com base nos resultados desta seção podemos gerneralizar para modelos mais complexos, como é o caso da próxima seção.

### 6.2.2. Modelos Lineares

Considere o modelo linear

(6.2.2.1) $$y_i = \beta_1 x_{1i} + \ldots + \beta_r x_{ri} + \sigma e_i$$

onde $Y = (y_1,\ldots,y_n)$ é o vetor de observações $\in \mathbb{R}^n$

$$x = \begin{pmatrix} x_{11}\cdots\cdots x_{1n} \\ \vdots \qquad\qquad \vdots \\ x_{r1}\cdots\cdots x_{rn} \end{pmatrix}$$ é uma matriz de valores conhecidos (matriz de planejamento)

$(\beta_1,\ldots,\beta_r,\sigma) = (\beta\ \sigma)$ é o vetor de parâmetros $\in \mathbb{R}^n \times \mathbb{R}^+$

$\underline{e} = (e_1,\ldots,e_n)$, é o vetor dos erros $\in \mathbb{R}^n$

então, (6.2.2.1) pode ser reescrita como

$$(6.2.2.2) \qquad \begin{pmatrix} x \\ \\ y \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ \\ \beta & \sigma \end{pmatrix} \begin{pmatrix} X \\ \\ \underline{e} \end{pmatrix}$$

onde $I_r$ é a matriz identidade de dimensão r.

Associando a $(e_1,\ldots,e_n)$ uma distribuição de probabilidades $p(e_1,\ldots,e_n)de_1\ldots de_n$ e identificando um grupo de transformações em (6.2.2.2) que é

$$\bar{G} = \left\{ \begin{pmatrix} I_r & 0 \\ \\ \beta & \sigma \end{pmatrix} \; ; \; \beta \in \mathbb{R}^n \quad e \quad \sigma \in \mathbb{R}^+ \right\}$$

Fazendo a analogia com os resultados da seção 6.2.1 definimos:

(i) a transformação invariante maximal

$$\begin{pmatrix} I_r & 0 \\ \\ b(\cdot) & s(\cdot) \end{pmatrix} \quad \text{onde } b(\cdot) = (b_1(\cdot),\ldots,b_r(\cdot))$$

(ii) a origem da órbita

$$\begin{pmatrix} x \\ \\ u \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix} \begin{pmatrix} x \\ \\ \underline{e} \end{pmatrix} \quad \text{onde } u = (u_1,\ldots,u_n)$$

E assim podemos definir o modelo reduzido que é composto dos itens abaixo

$$(6.2.2.3) \quad (1) \quad \begin{pmatrix} I_r & 0 \\ b(y) & s(y) \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ \beta & \sigma \end{pmatrix} \begin{pmatrix} I_r & 0 \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix}$$

$$(6.2.2.4) \quad (2) \quad K(u)p(b(\underline{e}),s(\underline{e})/X,u)(s(\underline{e}))^{n-(r+1)}|X\ X'|^{\frac{1}{2}}db(e)ds(\underline{e})$$

onde (2) é a distribuição condicional (em u) de

$$p(b(\underline{e})X+s(\underline{e})u/X,^{u})(s(\underline{e}))^{n}|X'X|^{1/2}db(\underline{e})ds(\underline{e})du$$

com $(s(e))^n|X'X|^{1/2}$ o jacobiano de transformação de $\begin{pmatrix} X \\ u \end{pmatrix}$ em

$$\begin{pmatrix} I_r & 0 \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix} \begin{pmatrix} X \\ u \end{pmatrix}$$

Em (6.2.2.4) K(u) é a constante de proporcionaldiade (que faz com que a integral seja igual a 1 e $(s(\underline{e}))^{-(r+1)}db(\underline{e})ds(\underline{e})$ é a medida de Haar à direita associada a $(b(\underline{e}),s(\underline{e}))$.

Com base neste modelo reduzido a distribuição estrutural resultante é

$$(6.2.2.5) \qquad g(\beta,\sigma/x,b(y),s(y),u) =$$

$$K(u)p(\frac{b(y)x-\beta}{\sigma} + \frac{s(y)}{\beta}u\ /x,u)(\frac{s(y)}{\sigma})^{n-(r+1)}|XX'|^{1/2}(\frac{s(y)}{\sigma^{r+2}})d\beta d\sigma$$

onde $(\frac{s(y)}{\sigma^{r+2}})$ é o jacobiano da transformação $\begin{pmatrix} I_r & 0 \\ b(\underline{e}) & s(\underline{e}) \end{pmatrix}$ em

$$\begin{pmatrix} I_r & 0 \\ \beta & \sigma \end{pmatrix}^{-1} \begin{pmatrix} I_r & 0 \\ b(y) & s(y) \end{pmatrix}$$

Chamamos a atenção para o fato que não é necessário que $(e_1,...,e_n)$ sejam independentes e/ou identicamente distribuidas. O resultado acima é válido $\forall$ $p(e_1,...,e_n)$, como pode ser visto através do seguinte exemplo onde não existe independência entre os erros:

Exemplo 6.2.2.1: Considere o modelo linear (6.2.2.2) com

$$p(e_1,...,e_n) = \frac{|R|^{-\frac{1}{2}}}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} e\ R^{-1} e'\}$$

onde
$$R = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & \ddots & & \vdots \\ \rho^{n-2} & \cdots & \cdots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix}$$

é a matriz de correlação de $\underline{e} = (e_1,...,e_n)$.

O modelo reduzido é representado pelos dois itens abaixo:

(1) A expressão (6.2.2.3) onde

$$b(y) = YX(XX')^{-1} \quad s^2(y) = Y'[I-X'(XX')^{-1}X]Y \;;$$

analogamente para $b(\underline{e})$ e $s(\underline{e})$

(2) Com base em (6.2.2.4) encontramos a distribuição de probabilidades

$$K(u)\,\frac{|R|^{-\frac{1}{2}}}{(2\pi)^{n/2}}\,\exp\{-\tfrac{1}{2}(b(\underline{e})X+s(\underline{e})u)'R^{-1}(b(\underline{e})X+s(\underline{e})u)\}|XX'|^{1/2}(s(\underline{e}))^{n-(r+1)}db(\underline{e})ds(\underline{e})$$

a distribuição estrutural resultante é

$$g(\beta,\sigma/b(y),s(y),X,u)d\beta d\sigma =$$

$$K(u)\frac{|R|^{-\frac{1}{2}}}{(2\pi)^{n/2}}\exp\{-\frac{1}{2}(\frac{b(y)X-\beta}{\sigma}+\frac{s(y)}{\sigma}u)R^{-1}(\frac{b(y)X-\beta}{\sigma}+\frac{s(y)}{\sigma}u)\}\times$$

$$(\frac{s(y)}{\sigma})^{n-(r+1)}|XX'|^{1/2}(\frac{s(y)}{\sigma^{n+1}})d\beta d\sigma$$

6.2.3. Generalizações

Podemos estender os resultados da seção anterior no caso em que

(6.2.3.1) $$Y = \begin{bmatrix} Y_{11}\cdots Y_{1n} \\ \vdots \quad\quad \vdots \\ Y_{k1}\cdots Y_{kn}\end{bmatrix} \quad e \quad \underline{e} = \begin{bmatrix} e_{11}\cdots e_{1n} \\ \vdots \quad\quad \vdots \\ e_{k1}\cdots e_{kn}\end{bmatrix}$$

Um exemplo desta situação é o chamado modelo normal de progressão com posto de

(1) $Y = \mu 1' + \gamma \underline{e}$ onde $Y$ e $\underline{e}$ estão indicadas em (6.2.3.1)

$$\mu' = (\mu_1, \ldots, \mu_k) \quad \text{e} \quad 1' = (1 \ldots 1)$$

$$\gamma = \begin{bmatrix} \sigma_1 & & & & \\ \gamma_{21} & \gamma_2 & & & \\ \vdots & & \ddots & & \\ \gamma_{k1} & \gamma_{k2} & \cdots & \gamma_{k.k-1} & \sigma_k \end{bmatrix}$$

(2) a distribuição de probabilidades de $\underline{e}$ é

$$(2\pi)^{-\frac{Kn}{2}} \exp\{-\tfrac{1}{2}tr(e'e)\}$$ onde $tr(e'e)$ é o traço da matriz $(\underline{e}'\underline{e})$.

O grupo estrutural associado a este modelo é

$$\bar{G} = \left\{ \begin{pmatrix} I_k & 0 \\ \mu & \gamma \end{pmatrix} ; \quad \mu \in \mathbb{R}^k \quad , \gamma \in M \right\}$$

onde $M$ é o espaço das matrizes triangulares inferiores $(k \times k)$ tais que os elementos da diagonal são positivos, os elementos acima da diagonal são zero e os elementos abaixo da diagonal são reais.

Os procedimentos para a inferência são análogos ao que foi descrito no caso em que $k=1$.

Para modelos mais gerais como é o exemplo de distribuições no círculo e na esfera (Fraser, 1979, capítulo 9) a generalização dos resultados descritos aqui é possível se pudermos encontrar uma função linear que relacione os dados, o parâmetro e os erros.

## 6.3. Discussão

Os modelos que possuem a propriedade de invariância como é o caso dos modelos estruturais têm diversas vantagens: Por exemplo, o caso usual de modelos lineares pode ser analisado sem a suposição de normalidade e de independência dos erros; além disso não há necessidade da existência de estatísticas suficientes; também não há necessidade de suposição do tipo variância mínima ou poder máximo para conseguir resultados únicos. Basta a estrutura de grupo para garantir a existência e a unicidade dos resultados (Fraser, 1968). Além disso, os resultados obtidos por este modelo são idênticos àqueles correspondentes à distribuição a posteriori Bayesiana provenientes de distribuições a priori não informativas. Esse tipo de priori coincide com a medida invariante à direita definida em $\overline{G}$ (apêndice A).

As distribuições estruturais são distribuições sobre o espaço paramétrico no mesmo sentido da distribuição fiducial. No entanto Fraser (1968) faz uma interpretação dos resultados que não é indutiva (como para Fisher); Fraser coloca a inferência proposta pelo seu modelo como algo entre as escolas de Fisher e Neyman-Pearson encontrando não apenas a distribuição estrutural para $\Theta$ como propondo testes de hipóteses e regiões de confiança para a tomada de decisões.

A distribuição estrutural encarada como distribuição fiducial possue outras propriedades vantajosas: as distribuições marginais e condicionais dos problemas multiparamétricos têm propriedade de confiança (Fraser, 1961b); restrições no espaço paramétrico, $\Theta_0$ , podem ser efetuadas simplesmente multiplicando a distribuição estrutural, $g(\Theta/x)$ pelo indicador

$$I_{\Theta_0}(\Theta) = \begin{cases} 1 & \text{se } \Theta \in \Theta_0 \\ 0 & \text{caso contrário} \end{cases}$$

e normalizando o resultado através de:

$$\frac{I_{\Theta_0}(\Theta) g(\Theta/x)}{\int I_{\Theta_0}(\Theta) g(\Theta/x) d\Theta}$$

Sendo assim, apesar do modelo estrutural ser aplicável a um número restrito de problemas, nos casos em que a sua utilização faz sentido é vantajoso usá-lo.

# CAPÍTULO 7

## EXTENSÃO DO MÉTODO FIDUCIAL: MODELOS FUNCIONAIS

Os modelos funcionais introduzidos por Bunke (1975) são definidos em situações análogas a dos modelos estruturais e podem ser representados por (6.1). Enquanto o modelo estrutural é definido apenas em condições onde se verifica o princípio de invariância (seção 6.1), o modelo funcional se propõe a ser mais geral. Para concretizar este fato apresentamos o seguinte exemplo:

Exemplo 7.1: Considere $\chi = \Theta = \mathbb{R}$ e $E = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$ e o modelo funcional composto por:

(1) $$x = \frac{\Theta e_1 + e_2}{e_3} \qquad x, \Theta \in \mathbb{R}$$
$$(e_1, \ldots, e_3) \in E$$

(2) $e_1, e_2, e_3$ são v.a. independentes tal que

$$e_1^2 \sim \chi^2_{n-1} \quad , \quad e_2^2 \sim \chi^2_{n-2} \quad , \; e_3 \sim N(0,1)$$

Uma situação como esta está fora das restrições do modelo estrutural, pois aqui a dimensão de $E$ é diferente da dimensão de $\chi$ .

A inferência dos modelos funcionais é realizada da mesma forma que nos modelos estruturais: expressando $\Theta$ como função de x e $\underline{e}$ e obtendo a distribuição fiducial de $\Theta$, através da distribuição de probabilidades de $\underline{e}$ para um valor observado de x.

Bunke (1975) chama a atenção para o fato que dois modelos funcionais distintos podem gerar a mesma distribuição de probabilidades de X dado $\Theta$, como indicado no exemplo abaixo:

<u>Exemplo 7.2</u>: Considere $\chi = E = \mathbb{R}^2$ e $\Theta = \mathbb{R}$ dois modelos funcionais (I) e (II), tais que a relação entre $x, \Theta$ e <u>e</u> é expressa como

(I) $$(X_1, X_2) = (e_1, \Theta e_1 + (1-\Theta^2)^{\frac{1}{2}} e_2)$$

(II) $$(X_1, X_2) = ((1-\Theta^2)^{\frac{1}{2}} e_1 + \Theta e_2, e_2)$$

onde $e_1, e_2$ são v.a. independentes, ambas com distribuição normal padrão. Em ambos os modelos determinamos a mesma distribuição de probabilidades para $(X_1, X_2)$.

Este exemplo mostra que ignorar a relação entre $x, \Theta$ e <u>e</u> nos casos em que ela é determinável, implica em perda de informação sobre as relações entre os dados e o parâmetro que pode ter implicações na conclusão final. Este fato foi enfocado por Fisher quando definiu o método passo-a-passo em detrimento do método de quantidades pivotais, dizendo que este último é incompleto, pois mais de uma quantidade pivotal pode ser definida em um determinado problema, cada uma produzindo um resultado diferente. Portanto, é necessário saber qual delas é adequada (Fisher, 1956, pg 257 ). Encarando cada distribuição $F_i(t_i/\Theta_i, \ldots, \Theta_i, t_1, \ldots, t_{i-1})$ como uma quantidade pivotal, $i=1,\ldots,n$ determinou condições sobre elas como critério de adequação.

Neste capítulo reunimos alguns resultados importantes do modelo funcional e alguns exemplos da inferência fiducial são analisados dentro do contexto do modelo funcional, visualizando a quantidade pivotal como a variável aleatória erro $\underline{e}$ e vice-versa, fazendo a analogia das condições impostas por Fisher com as condições expostas aqui.

A seção 7.1 se ocupa de resultados preliminares que facilitam a compreensão e a expressão dos resultados futuros. A seção 7.2 aborda o modelo funcional no caso uniparamétrico e a seção 7.3 aborda o caso multiparamétrico.

## 7.1. Preliminares

Para $x \in \chi$ e $\underline{e} \in E$ diremos que x e $\underline{e}$ são compatíveis se $x = h(\Theta;e)$ para algum $\Theta \in \Theta$. Denotamos por $E_x$ o conjunto $\{e; x \text{ e } \underline{e} \text{ são compatíveis}\}$. Analogamente, definimos $\chi_{\Theta}$ e $\Theta_x$. Denotamos por $\mathcal{D}$ o conjunto $\{(x,\Theta) \in \chi \times \Theta;\ x \text{ e } \Theta \text{ são compatíveis}\}$. Esse conjunto $\mathcal{D}$ corresponde nos modelos funcionais ao conjunto de referência definido na inferência fiducial (seção 2.2.1).

Admitimos também a *invertibilidade* da relação $x=h(\Theta;e)$, ou seja, para $x \in \chi$ e $\underline{e} \in E_x$ a solução para $\Theta$ em $x=h(\Theta;\underline{e})$ é única.

Para simplificar a notação escreveremos a relação $x=h(\Theta;\underline{e})$ como $x=\Theta\underline{e}$, por exemplo $x=\Theta\underline{e} = \dfrac{\Theta+e_1}{e_2}$ (exemplo 7.1); de forma análoga o parâmetro $\Theta$ expresso como função de x e $\underline{e}$ será denotado por $\Theta = x\underline{e}^{-1}$ (no exemplo 7.1, $\Theta = x\underline{e}^{-1} = xe_2-e_1$). Quando x e $\underline{e}$ são conhecidos $\Theta = x\underline{e}^{-1}$ expressa a solução única de $\Theta$ .

A distribuição de probabilidades associada a $\underline{e}$ será denotada resumidamente por P, então $P(e \in A)$ é a probabilidade associada a $\forall A \in E$. A distribuição fiducial de $\Theta$ será denotada por $\bar{P}(\cdot/x)$, tal que $\bar{P}(\Theta \in \Delta/x)$ é a probabilidade fiducial de $\Delta \subset \Theta$ dado o valor de X=x. Analogamente, a distribuição de X será denotada por $P^*(\cdot/\Theta)$, tal que $P^*(X \in B/\Theta)$, $\forall B \subset \chi$. Ressaltamos que $P^*(X \in B/\Theta) = P(\Theta e \in B/\Theta)$ e $\bar{P}(\Theta \in A/x) = P(xe^{-1} \in A/x)$.

As definições abaixo estabelecem condições sobre um modelo funcional e serão utilizadas nas próximas seções.

<u>Definição 7.1</u>: Um modelo funcional é *monótono* se:

$$\forall\ \Theta_1, \Theta_2 \in \Theta \subset \mathbb{R}\ , \quad \Theta_1 \leq \Theta_2 \iff \Theta_1\underline{e} \leq \Theta_2\underline{e}\ .$$

<u>Definição 7.2</u>: Um modelo funcional monótono é regular se:

(i) $\chi$ e $\Theta$ são intervalos abertos

(ii) para cada $x_a \in \chi$, $P(\Theta\underline{e} \geq x_a)$ é uma função de $\Theta$, contínua estritamente crescente, assumindo todos os valores possíveis no intervalo (0,1)

(iii) para cada $\Theta_a \in \Theta$, $P(\Theta_a\underline{e} \geq x)$ é uma função de x, contínua, estritamente decrescente, assumindo todos os valores possíveis no intervalo (0,1).

Quando os dados x são observados, $E_x$ é o conjunto dos valores de $\underline{e}$ compatíveis com x. Se $E_x$ é diferente de $E$ e $x_a$ é um valor observado, para $\forall\ \Theta \in \Theta$, $P(\Theta e \geq x_a)$ está restrita a $E_{x_a}$. Então encontramos a distribuição fiducial de $\Theta$ através de $\Theta = x_a\underline{e}^{-1}$ e da distribuição de $\underline{e}$ "restrita" em $E_{x_a}$ que denotaremos por $P_a$.

O texto está dividido em partes. Inicialmente dividimos a apresentação do modelo em: caso uniparamétrico e caso multiparamétrico.

## 7.2. Caso Uniparamétrico

Os modelos funcionais uniparamétricos serão analisados em duas fases: primeiramente os modelos funcionais simples e posteriormente os modelos funcionais particionáveis. Os primeiros são definidos quando $E_x = E$ e os modelos particionáveis são definidos em situações onde $E_x \neq E$ para cada $x \in \chi$ e $\{E_x;\ x \in \chi\}$ forma uma partição de $E$.

### 7.2.1. Modelo Funcional Simples

A relação entre um modelo funcional simples e a correspondente distribuição fiducial pode ser expressa pela definição abaixo:

<u>Definição 7.2.1.1</u>: A um modelo funcional simples $< x = \Theta\underline{e}, \underline{e} \sim P >$ associamos o modelo fiducial simples que é dado por $< \Theta = x\underline{e}^{-1}, \underline{e} \sim P >$

Assim, o modelo fiducial é em si mesmo um modelo funcional, permutando x e $\Theta$. Está implícito no modelo fiducial acima que $\underline{e}$ é distribuido independentemente de x. Essa definição procura mostrar que assumindo um modelo funcional $< x=\Theta\underline{e}, \underline{e} \sim P >$ faz sentido encontrar ~~que~~ a distribuição de x em função de $\Theta$; no modelo fiducial simples $<\Theta = xe^{-1}, e \sim P>$ faz "sentido" encontrar a "distribuição" de $\Theta$ em função de x.

As condições para que um modelo funcional resulte numa distribuição fiducial válida (com propriedade de confiança, seção 3.2) serão indicadas durante a exposição que será dividida em duas etapas: (1) modelo pivotal (2) modelo funcional monótono.

7.2.1.1. Modelo Pivotal

Suponha que em um modelo funcional simples para todo $(x,\Theta) \in \mathcal{D}$ a solução para $\underline{e}$ de $x=\Theta\underline{e}$ é única; então podemos encontrar $\underline{e}=\underline{e}'(x;\Theta)$ e chamamos $\underline{e}': \mathcal{D} \to E$ de *função pivotal*. Se um modelo funcional satisfaz essas condições é chamado de *modelo pivotal simples* e pode ser expresso na forma alternativa $<\underline{e}=\underline{e}'(x;\Theta), \underline{e} \sim P>$. O exemplo a seguir ilustra uma situação como esta.

Exemplo 7.2.1.1.1: Considere a situação em que $\chi = \Theta = E = \mathbb{R}$ e o modelo funcional é $<x = \Theta+e, e \sim P>$. Este modelo funcional é pivotal e portanto pode ser expresso como $<\underline{e} = x-\Theta, \underline{e} \sim P>$. Aqui $\mathcal{D} = \chi \times \Theta$.

Um exemplo peculiar de modelo pivotal é apresentado a seguir.

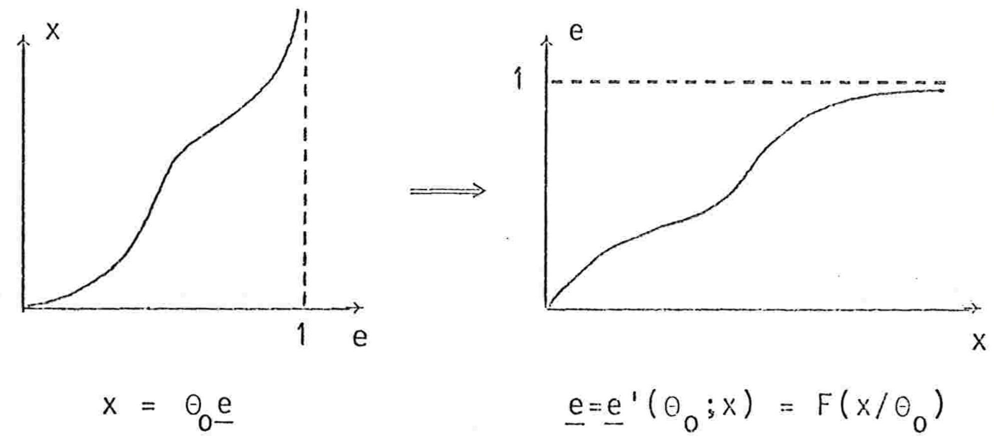
Exemplo 7.2.1.1.2: Considere a situação em que $\chi \subset \mathbb{R}$, $\Theta \subset \mathbb{R}$ e $E = (0,1)$. A distribuição de $\underline{e}$ é uniforme em $E$ e a relação $x = \Theta e$ satisfaz as condições

(i) $x = \Theta\underline{e}$ é uma função estritamente crescente de $\underline{e}$ para cada $\Theta$.

(ii) se $\Theta_1 < \Theta_2 \Rightarrow \Theta_1\underline{e} < \Theta_2\underline{e}$ , $\forall\ \underline{e} \in (0,1)$

(iii) $\forall\ (\underline{e},x) \in E \times \chi$ existe um único $\Theta$ satisfazendo $x = \Theta\underline{e}$.

Figura 7.1: Gráfico de uma relação $x=\Theta e$ e da correspondente função pivotal $\underline{e}=\underline{e}'(\Theta;x)$ para um valor de $\Theta=\Theta_0 \in \Theta$.




A figura 7.1 ilustra uma situação em que uma relação $x=\Theta\underline{e}$ que satisfaz as condições (i), (ii) e (iii) "gera" uma função pivotal $\underline{e}'=F(x/\Theta)$ que satisfaz: (1) $\underline{e}' = F(x/\Theta)$ é estritamente crescente em x para cada $\Theta \Rightarrow F(x/\Theta)$ é contínua em x; (2) $\underline{e}'=F(x/\Theta)$ é estritamente crescente em $\Theta$ para cada $x \Rightarrow F(x/\Theta)$ é contínua em $\Theta$ e portanto é derivável em $\Theta$. Uma $F(x/\Theta)$ nessas condições satisfaz as restrições de Fisher para a utilização do método fiducial (seção 2.2.1).

A finalidade deste exemplo é mostrar que modelos funcionais pivotais que satisfazem as condições (i), (ii) e (iii) produzem uma distribuição fiducial para $\Theta$ idêntica àquela obtida pela inferência fiducial que parte diretamente de $F(x/\Theta)$.

Uma distribuição fiducial é válida se tem propriedade de confiança (seção 3.2). Nos modelos pivotais, podemos mostrar que a distribuição fiducial de $\Theta$ tem propriedade de confiança da seguinte forma:

- Considere o modelo pivotal $<\underline{e}=\underline{e}'(x;\Theta),\ \underline{e} \sim P>$ tal que $\forall\ A \subset E$ com $P(\underline{e} \in A) = 1-\alpha$. Então, quando $x=x_a \in \chi$ podemos escrever

$$x_a A^{-1} = \{x_a \underline{e}^{-1};\ \underline{e} \in A\} = \{\Theta \in \Theta_{x_a};\ \underline{e}'(x_a;\Theta) \in A\}$$

Assim $\quad \overline{P}[\Theta \in x_a A^{-1}/x_a] = P[\{\underline{e}; x_a e^{-1} \in x_a A^{-1}\}/x_a] = P(\underline{e} \in A)$

que é a probabilidade fiducial de $\Theta$ pertencer ao conjunto $x_a A^{-1}$.

Por outro lado, $xA^{-1} = \{xA^{-1};\ x \in \chi\} = \{\Theta \in \Theta_x;\ e'(x;\Theta) \in A\}$ então

$$P^*(\Theta \in xA^{-1}/\Theta) = P(\underline{e}'(x;\Theta) \in A/\Theta) = P(\underline{e} \in A) = 1-\alpha$$

que é a probabilidade de $\Theta$ pertencer ao conjunto $xA^{-1}$, $\forall\ \Theta \in \Theta$. Portanto, quando $x=x_a$ o conjunto $x_a A^{-1}$ é uma região de confiança com coeficiente de confiança $(1-\alpha)$. Dessa forma a distribuição fiducial de $\Theta$ originária de um modelo pivotal tem a propriedade de confiança.

7.2.1.2. Modelo Funcional Monótono

Quando para $(x,\Theta) \in \mathcal{D}$ a solução $\underline{e}$ de $x=\Theta\underline{e}$ não é única, o modelo não é pivotal. O exemplo abaixo concretiza uma situação de modelo não pivotal.

Exemplo 7.2.1.2.1: Suponha $\chi = \Theta = \mathbb{R}$, $E = \mathbb{R} \times \mathbb{R}^+ = (e_1,e_2)$ e o modelo é expresso de tal forma que $x = \dfrac{\Theta+e_1}{e_2}$, onde $e_1$ e $e_2$ são v.a. independentes com $e_1 \sim N(0,1/n)$ e $(n-1)e_2^2 \sim \chi^2_{n-1}$ e $(e_1,e_2)$ não pode ser expressa como função de $\Theta$ e x.

Nesses casos, para que a distribuição fiducial resultante de $<\Theta = x\underline{e}^{-1}, \underline{e} \sim P>$ tenha a propriedade de confiança é necessário que o modelo funcional simples $<x = \Theta\underline{e}, \underline{e} \sim P>$ satisfaça as condições do teorema abaixo:

Teorema 7.2.1.2: Se um modelo funcional simples é *monótono regular*, então $\forall\ \alpha \in (0,1)$, existe uma função $\Theta_x: \chi \to \Theta$ tal que

$$\overline{P}(\Theta \leq \Theta_x/x) = 1-\alpha \quad , \quad \forall\ x \in \chi$$

e

$$P^*(\Theta \leq \Theta_x/\Theta) = 1-\alpha \quad , \quad \forall\ \Theta \in \Theta$$

Em outras palavras, se um modelo funcional simples é monótono regular (definição 7.2), então a distribuição fiducial de $\Theta$ tem a propriedade de confiança.

Antes de demonstrarmos o teorema acima consideremos os seguintes resultados:

<u>Lema 7.2.1.2</u>: Se o modelo é funcional monótono (definição 7.1) então $\overline{P}(\Theta \leqq \Theta_a/x_a) = P^*(x \geqq x_a/\Theta_a)$

<u>Prova</u>: $\overline{P}(\Theta \leqq \Theta_a/x_a) = P(x_a\underline{e}^{-1} \leqq \Theta_a/x_a)$ (por definição)

$= P(x_a \leqq \Theta_a\underline{e}/x_a)$ (pela monotonicidade)

$= P^*(x \geqq x_a/\Theta_a)$ (por definição)

(observe que $\Theta_a\underline{e} = x$).

Deste lema resulta o seguinte corolário

<u>Corolário 7.2.1.2</u>: Se $\Theta_x$ é uma função conhecida e estritamente crescente de x com valores em $\Theta$, então

$$\overline{P}(\Theta \leqq \Theta_{x_a}/x_a) = P^*(\Theta_{x_a} \leqq \Theta_x/\Theta_{x_a})$$

<u>Prova</u>: Observando que $\Theta_{x_a} = x_a\underline{e}$ e $\Theta_x = x\underline{e}$ então

$$P^*(\Theta_{x_a} \leqq \Theta_x/\Theta_{x_a}) = P(x_a\underline{e} \leqq x\underline{e}/x_a\underline{e}) = P^*(x \geqq x_a/\Theta_a)$$

que pelo Lema 7.2.1.2 é igual a $\overline{P}(\Theta \leqq \Theta_{x_a}/x_a)$.

Com esses resultados vamos *provar o teorema 7.2.1.2*:

Definimos $\Theta_{x_a}$ por $P(x_a \leqq \Theta_{x_a}\underline{e}/x_a) = 1-\alpha$ que é único por regularidade (definição 7.2)(*). O Lema 7.2.1.2, nos mostra que para

(*) Como $P(x \leqq \Theta_x e/x)$ é uma função de $\Theta$ contínua, estritamente crescente, então fixando o valor $(1-\alpha)$ existe um e só um $\Theta_x$ que satisfaz a igualdade $P(x \leqq \Theta_x e/x) = 1-\alpha$ .

$\forall\, x_a \in \chi$, o intervalo $\Theta \leq \Theta_{x_a}$ tem probabilidade fiducial igual a $(1-\alpha)$, ou seja, $P(\Theta \leq \Theta_{x_a} / x_a) = 1-\alpha$. Por outro lado, pela própria regularidade, $\Theta_x$ é estritamente crescente em $x$ e $\{\Theta_x : x \in \chi\} = \Theta$. Então, pelo Corolário 7.2.1.2, $P(\Theta \leq \Theta_x/\Theta) = 1-\alpha$, $\forall\, \Theta \in \Theta$ donde se conclui que $\{\Theta \leq \Theta_x ; x \in \chi\}$ define, quando $x=x_a$, uma região de confiança com coeficiente de confiança $(1-\alpha)$.

Portanto, concluímos que somente modelos pivotais ou modelos funcionais monótonos regulares produzem uma distribuição fiducial para $\Theta$ com propriedade de confiança. Cabe observar aqui, que os modelos estruturais são um caso particular de modelo pivotal, pois a relação $x = \bar{g}_\Theta(\underline{e})$ (seção 6.2.1) pode ser invertida e assim $\underline{e} = \bar{g}_\Theta^{-1}(x)$ é uma *função pivotal*.

### 7.2.2. Modelos Particionáveis

Retiremos a suposição de que $E_x = E$. Definimos um modelo funcional particionável como aquele em que $\{E_x ; x \in \chi\}$ forma uma partição de $\chi$ .

<u>Definição 7.2.2.1</u>: Um modelo é particionável se e somente se para alguma função h em $\chi$ e alguma função u em $E$ , a compatibilidade entre x e $\underline{e}$ é expressa através da igualdade $h(x) = u(e)$.

Em outras palavras, $u(\underline{e})$ e $h(x)$ são identificadores de partições, isto é, se $u(e_1) \neq u(e_2)$ isto implica que $e_1$ e $e_2$ pertencem a partições diferentes e vice-versa. Se x é compatível com $\underline{e}$ então $\underline{e} \in E_x$ uma partição de $E$ , então $h(x)$ também identifica as partições. Como

$h(x) = u(\underline{e})$ e $\underline{e} \sim P$ (independente de $\Theta$) então essas funções são chamadas funcionais ancilares. São equivalentes às estatísticas ancilares definidas no método fiducial (seção 2.3); no modelo estrutural a origem da uma órbita é um caso de funcional ancilar (capítulo 6).

No caso em que $E$ não é particionável e $E_x \neq E$ estamos na condição em que não é possível encontrar um funcional ancilar (uma estatística ancilar). O caso em que $E_x = E$ é uma situação particular de modelo particionável (equivalente às estatísticas suficientes completas seção 2.1).

A inferência nos modelos funcionais particionáveis é realizada substituindo-se o modelo original $<x = \Theta\underline{e},\ \underline{e} \sim P>$ por $<x = \Theta\underline{e},\ \underline{e} \sim P_h>$ onde $P_h$ é a distribuição condicional de $\underline{e}$ dado $u(\underline{e}) = h(x)$.

<u>Exemplo 7.2.3.1</u>: Suponha $\chi = E = \mathbb{R}^n$ $(n \geq 2)$, $\Theta = R$ e $x = \Theta\underline{e} \Longrightarrow (x_1,\ldots,x_n) = (\Theta+e_1,\ldots,\Theta+e_n)$. A distribuição de $\underline{e}$ é arbitrária. Então, fazendo $u(\underline{e}) = (e_1-\bar{e},\ldots,e_n-\bar{e}) = (x_1-\bar{x},\ldots,x_n-\bar{x}) = h(x)$ teremos o modelo funcional que pode ser expresso como $<x = \underline{e};\ e \sim P_h>$, onde $P_h$ é a distribuição condicional de $\underline{e}$ dado $h(x)$. O modelo fiducial, neste caso pode ser expresso como $<\Theta = \bar{x}-\bar{e};\ \bar{e} \sim P_h>$ (por analogia com o que foi discutido na seção 2.3 e 6.2).

Para que se verifique a propriedade de confiança a distribuição fiducial de $\Theta$ deve ser resultado de um modelo funcional *particionável* pivotal ou monótono regular.

## 7.3. Caso Multiparamétrico

No caso multiparamétrico para que a distribuição fiducial de $\Theta$ tenha propriedade de confiança é necessário que a distribuição conjunta dos componentes de $\Theta$ satisfaça algumas condições como as que serão descritas nas seções seguintes.

### 7.3.1. Marginalização Consistente

Suponha o modelo funcional $M_1 = <x=\Theta\underline{e};\ \underline{e} \sim P>$ e as seguintes considerações:

(i) nós estamos interessados somente no subparâmetro $\lambda=\lambda(\Theta)$.

(ii) Existe uma função $z = z(x)$, tal que a relação entre $z, \lambda$ e $\underline{e}$ constitue, juntamente com a distribuição de $\underline{e}$, um modelo funcional simples, expresso por $M_2=<z=\lambda\underline{e},\underline{e}\sim P>$.

Chamamos $M_2$ uma redução de $M_1$. O exemplo a seguir ilutra este fato.

<u>Exemplo 7.3.1</u>: Considere o modelo funcional abaixo

$$<(\bar{x},s) = (\mu+\sigma e_1, \sigma e_2);\ (e_1,e_2) \sim P>$$

onde $e_1$ e $e_2$ são v.a. independentes e $e_1 \sim N(0,1/n)$ e $(n-1)e_2^2 \sim \chi^2_{n-1}$.
Suponha que desejamos inferir sobre $\lambda = \frac{\mu}{\sigma}$, então existe

$z = z(x) = \frac{\bar{x}}{s}$ tal que $\frac{\bar{x}}{s} = \frac{\mu/\sigma+e_1}{e_2} \Longrightarrow z = \frac{\lambda+e_1}{e_2}$ que é uma redução do modelo original (a redução é uma função do modelo original).

Observe que a redução do modelo original implica em mudanças no espaço $\chi$ e mudanças no espaço $\Theta$, mas não necessariamente implica em mudanças no espaço $E$, como é o caso do exemplo acima.

A distribuição fiducial marginal de $\lambda$ dado $x=x_a$ em $M_1$ é a distribuição de $\lambda(x_a e^{-1})$ com $\underline{e} \sim P$. No modelo reduzido, $M_2$, a distribuição fiducial de $\lambda$ é a de $z\underline{e}^{-1}$ onde $z=z(x_a)$. A relação entre a distribuição fiducial *marginal* de $\lambda$ em $M_1$ e a distribuição fiducial de $\lambda$ em $M_2$ é dada pelo Lema abaixo:

Lema 7.3.1: $\lambda(x\underline{e}^{-1}) = z(x)\underline{e}^{-1}$ , $\forall\ x \in \chi$.

Prova: $\lambda(x\underline{e}^{-1}) = \lambda(\theta) = \lambda = z\underline{e}^{-1} = z(x)\underline{e}^{-1}$ .

Então forçosamente a distribuição fiducial marginal de $\lambda$ em $M_1$ é igual à distribuição fiducial de $\lambda$ em $M_2$ o que significa que a marginalização é consistente. É claro, então, que no exemplo 7.3.1 a distribuição fiducial de $\frac{\mu}{\sigma}$ é igual à distribuição fiducial de $\frac{\mu}{\sigma}$ obtida pela redução. A consistência é a propriedade de confiança.

O exemplo de Stein (1959) (seção 3.2) colocado sob o ponto de vista de modelos funcionais é uma situação em que não é possível marginalização consistente, como indicado abaixo:

Exemplo 7.3.2: Considere o modelo funcional

$$<(x_1,x_2) = (\theta_1+e_1,\theta_2+e_2);\ (e_1;e_2) \sim P>$$

onde $e_1$ e $e_2$ são v.a.i.i.d. com distribuição $N(0,1)$. Desejamos inferir sobre $\lambda = \theta_1^2+\theta_2^2$ , mas como é não possível encontrar $z=z(x)$ tal que

$z=\lambda\underline{e}$ então a distribuição marginal de $\lambda$ obtida da distribuição conjunta de $\Theta_1$ e $\Theta_2$ que vem do modelo $<x_1^2+x_2^2 = (\Theta_1+e_1)^2+(\Theta_2+e_2)^2;\underline{e}\sim P>$ não é consistente (não possui a propriedade de confiança).

7.3.2. Consistência Condicional

Suponha que desejamos fazer inferências sobre $\Theta$ condicionado em $x_a$ e no valor de $\lambda(\Theta) = \lambda$ conhecido. Há duas formas distintas de encontrar a distribuição fiducial de $\Theta$ condicionada em $\lambda(\Theta) = \lambda$:

(I) Encontramos a distribuição fiducial de $\Theta$ e em seguida condicionamos de forma usual em $\lambda$ encontrando $\overline{P}(\cdot/\lambda,x)$.

(II) Usamos a restrição $\Theta \in \circledR^{\lambda} = \{\Theta;\lambda(\Theta) = \lambda\}$ e consideramos $z=\lambda\underline{e}$, tal que $a(x) = z$ e $u(\underline{e}) = \lambda\underline{e}$. Isto é, $x=\Theta\underline{e}$ com $\Theta \in \circledR^{\lambda}$ é um *modelo funcional particionável* com $\underline{e}\sim P_z$, onde $P_z$ é a distribuição de $\underline{e}$ condicionada em $\lambda\underline{e} = z(x)$. Então, a distribuição condicional de $\Theta$ dado $\lambda$ é $\overline{P}_{\lambda}(\cdot/x)$.

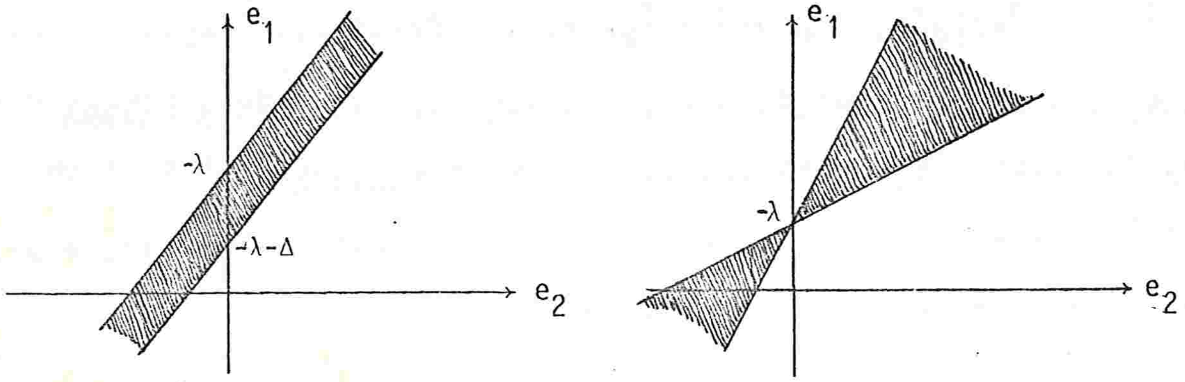
Dempster (1963) estudou o exemplo 7.3.1 onde $\lambda = \frac{\mu}{\sigma}$ e $z = \frac{\bar{x}}{s}$, nessas duas situações (I) e (II) e mostrou que produzem resultados diferentes para a distribuição fiducial condicional de $\sigma$ dado $\frac{\mu}{\sigma}$. Em cada caso a distribuição de $(e_1,e_2)$ é condicionada:

(I) nos valores de $\lambda$ da função $\lambda = z\underline{e}^{-1} = ze_2-e_1$

(II) nos valores de $z$ para a função $z = \lambda\underline{e} = \dfrac{\lambda+e_1}{e_2}$

que levam em diferentes partições do espaço Θ, como mostra a Figura 7.2.

Figura 7.2: Gráfico da partição de $(e_1,e_2)$ obtidas do condicionamento nos:




(I) Valores de $\lambda$ da função

$$\lambda = z\underline{e}^{-1} = ze_2 - e_1$$

$$\Longrightarrow \lambda \leqq ze_2 - e_1 \leqq \lambda + \Delta$$

(II) Valores de z da função

$$z = \lambda\underline{e} = \frac{\lambda + e_1}{e_2}$$

$$\Longrightarrow z \leqq \frac{\lambda + e_1}{e_2} \leqq z + \Delta$$

O exemplo 7.3.1 pode ser estudado de forma alternativa: encontrando a distribuição fiducial condicional de $\mu/\sigma$ dado $\sigma$ através de (I) e (II). Nesse caso, os dois resultados coincidem. Isto se dá segundo Dawid e Stone (1982), devido ao fato de o modelo reduzido, $\lambda = \sigma$ e $z = s$, ser pivotal. De fato, $\langle z = \lambda e,\ (n-1)e_2^2 \sim \chi^2_{n-1} \rangle$ é um modelo pivotal e $e_2$ é uma "contração" de $(e_1, e_2)$. Dawid & Stone (1982) acreditam que somente quando o modelo reduzido é pivotal é que (I) e (II) produzirão os mesmos resultados. No entanto, admitem não ter uma prova dessa afirmação. Contudo, demonstram que quando a distribuição fiducial condicional de $\Theta$ dado $\lambda$ coincide para (I) e (II) e o modelo reduzido de $\lambda$ é pivotal, essa distribuição fiducial condicional tem a propriedade de confiança: Considere o modelo reduzido pivotal $<z = \lambda f,\ f \sim P>$ onde f é uma "contração" de $\underline{e}$, e $f = f'(z,\lambda)$ é a função pivotal. Se a distribuição fiducial condicional de $\theta$ dado $\lambda$ coincide para (I) e (II) e $(x,\theta) \in A \subset \chi \times \Theta$, então

$$(7.3.1) \qquad \overline{P}[(x,\theta) \in A/x,\lambda] = \overline{P}_{\lambda}[(x,\theta) \in A/x]$$

Se $(x,\theta) \in A \subset \chi \times \Theta \longrightarrow \exists\ B \subset E \times Z$ tal que $(e,z) \in B$ quando $f=f'(z,\lambda)$ então o lado direito de (7.3.1) é igual a $P((e,z) \in B/f=f'(z,\lambda))$. Por outro lado, em modelos pivotais, a propriedade de confiança se verifica, então

$$P[(e,z) \in B/f=f'(z,\lambda)] = P^*[(x,\theta) \in A/z,\theta]$$

<u>Conclusão</u>: $\overline{P}[(x,\theta) \in A/x,\lambda] = P^*[(x,\theta) \in A/z,\theta]$. Mostra-se também que $\overline{P}[(x,\theta) \in A/x] = \int \overline{P}_{\lambda}[(x,\theta) \in A/x]dP(\lambda/z)$ o que dá a essa distribuição marginal uma propriedade de confiança esperada, ou seja, que só depende de z e é realizada como uma esperança em relação à distribuição fiducial de $\lambda$ dado z (Princípio da Difusão, seção 5.3.2).

Nas distribuições condicionais cujo elemento que condiciona $\lambda$ não provém de um modelo reduzido *pivotal* não acredita-se na validade da propriedade de confiança. Contudo, não há provas disto.

Quando mencionamos inicialmente que a relação $x = \theta\underline{e}$ é um acréscimo importante de informação, pois determina as "relações internas" de x e $\theta$ esse comentário tinha o propósito de esclarecer a diferença entre a inferência dos modelos funcionais e a inferência fiducial exposta nos primeiros capítulos deste trabalho. O exemplo 5.2.4 mostra uma situação em que duas fatorações alternativas do modelo ori ginal $f(t_1,t_2/\theta_1,\theta_2)$ produzem resultados diferentes. Isto ocorre, pois em cada uma dessas fatorações há diferentes quantidades pivotais envolvidas e é necessário escolher dentre uma delas com base em critérios de validade (seção 5.2). No modelo funcional isso não ocorre, pois estabelecendo inicialmente qual é a relação entre x, $\theta$ e $\underline{e}$, já escolhemos uma "quantidade pivotal" e toda inferência resultante é decorrente dessa escolha. Por isso, entre as duas abordagens há intersecção, mas não coincidência necessàriamente dos resultados. Por esse motivo, Fraser chama a distribuição encontrada pelo seu modelo de distribuição estrutural e não de distribuição fiducial.

oOo

# APÊNDICE A

## MEDIDAS DE HAAR (OU INVARIANTE)

Considere G um grupo de transformações. Definimos uma transformação à direita em G, tal que a cada $g \in G$ associamos $gog_0$ G onde $g_0 \in G$ é uma transformação fixa. Analogamente, a transformação à esquerda em G associa a cada $g \in G$ o elemento $g_0og \in G$.

Existe um espaço $\Theta$ que pode ser identificado com G (seção 6.1.1), então a cada $\theta \in \Theta$ associo um $g_\theta \in G$. Assim as transformações à direita e à esquerda definidas podem ser identificadas em $\Theta$ da seguinte forma: $\theta_0 \in \Theta$ é um elemento fixo então

(1) a transformação à direita associa a cada $\theta$ um $\theta_0\theta$ relativo à $g_\theta o g_{\theta_0}$.

(2) a transformação à direita associa a cada $\theta$ um $\theta\theta_0$ relativo à $g_{\theta_0} o g_\theta$.

<u>Exemplo A1</u>: Considere G um grupo afim em $\mathbb{R}^p$. As transformações $g_{a,b} \in G$ de $x \in \chi$ são $g_{a,b}(x) = ax+b$, portanto existe $\Theta = \{(a,b);\ a \quad \mathbb{R},\ b > 0\}$ que pode ser identificado com G. Então, segundo o exemplo 6.1.6 a transformação à direita em $\Theta$ é $(a,b) \to (a,b)(a_0,b_0) = (ba_0+a, b_0)$ onde $(a_0,b_0)$ é um elemento fixo de

Analogamente, a transformação à esquerda é

$$(a,b) \to (a_0,b_0)(a,b) = (b_0a-a_0, b_0b)$$

Portanto, daqui em diante quando nos referirmos a uma transformação à direita e à esquerda em G estaremos nos referindo a uma situação análoga à descrita acima, ou seja, existe um subconjunto de $\mathbb{R}^p$, $p \geq 1$, com medida de Lebesgue positiva que pode ser identificado com G de tal forma que as transformações à direita e à esquerda em G implicam em transformações à esquerda e à direita neste subconjunto.

<u>Definição A1</u>: Seja $H_{g_0}(g)$ o diferencial (isto é, a matriz das derivadas parciais) da transformação $g \to g o g_0$. Seja também $J_{g_0}(g) = |\det(H_{g_0}(g))|$ o jacobiano dessa transformação. Analogamente definimos $H^*_g(g)$ e $J^*_{g_0}(g)$ como o diferencial e o jacobiano da transformação $g \to g_0 g$.

<u>Exemplo A2</u>: Considere o exemplo B1, a transformação à direita

$$(a,b) \longrightarrow (a,b)(a_0,b_0) = (a+ba_0, bb_0)$$

pode ser representada pela função t em Ⓗ, tal que

$$t(a,b) = (t_1,t_2) = (a+ba_0, bb_0)$$

que tem diferencial

$$H_{g_0}((a,b)) = \begin{pmatrix} \dfrac{t_1(a,b)}{a} & \dfrac{t_2(a,b)}{b} \\ & \\ \dfrac{t_2(a,b)}{a} & \dfrac{t_2(a,b)}{b} \end{pmatrix} = \begin{pmatrix} 1 & a_0 \\ 0 & b_0 \end{pmatrix}$$

cujo jacobiano é $J_{g_o}((a,b)) = |\det(H_{g_o}(g))| = b_o$. Portanto, quando nos referimos ao jacobiano de uma transformação à direita estamos nos referindo a uma situação análoga a esta.

Analogamente, para o exemplo B2, a transformação à esquerda $g \to g_o o g$ tem jacobiano $J^*_{g_o}((a,b)) = b_o^2$.

Quando nos referirmos a um subconjunto $A \subset G$ estamos nos referindo ao conjunto das transformações $g_a \in G$, tal que $a \in A \subset \Theta$ Analogamente com $Ag_o$ e $g_oA$. Então, a cada $\{g_a \in G; a \in A\}$ associamos $A \subset \Theta$ e sobre A definimos diferenciais, jacobianos e a medida de Lebesgue correspondente. No entanto, trabalharemos como se G fosse o próprio $\Theta$ e cada $\{g_a \in G;\ a \in A\}$ se fosse o próprio A.

<u>Definição A2</u>: Uma densidade de Haar à direita em G é denotada por h(g) e é a densidade (com respeito à medida de Lebesgue) à qual para todo $A \subset G$ e todo $g_o \in G$.

$$\int_{Ag_o} h(y)dy = \int_A h(x)dx$$

onde $Ag_o = \{gog_o; g \in A\}$. A correspondente *medida de Haar à direita* em G é denotada $\nu(A) = \int_A h(x)dx$ e então

$$\nu(A) = \nu(Ag_o),\ \forall\ g_o \in G \qquad \text{e} \qquad d\nu(g) = h(g)dg$$

Analogamente, a medida de Haar à esquerda em G é h* e a densidade de Haar à esquerda em G são tais que:

$$\mu(g_oA) = \int_{g_oA} h^*(y)dy = \int_A h^*(x)dx = \mu(A),\ \forall g_o \in G \quad e \quad A \subset G$$

além disso $\quad d\mu(g) = h^*(g)dg$

Mostra-se que (Berger, 1980)

Resultado A1: A densidade de Haar à direita em G é $h(g) = \dfrac{1}{J_g(g_i)}$ onde $g_i$ é a transformação identidade de G e $J_g(x)$ é o jacobiano da transformação $x \to xg$.

Resultado A2: A densidade de Haar à esquerda em G é $h^*(g) = \dfrac{1}{J_g^*(g_i)}$ onde $J_g^*(x)$ é o jacobiano da transformação $x \to gx$

<u>Exemplo A3</u>: No caso do Exemplo A2 a relação entre as medidas invari antes ou Haar e os diferenciais é

$$d\mu(a,b) = \frac{dadb}{b_o^2}$$

$$d\nu((a,b) = \frac{dadb}{b}$$

<u>Propriedade</u>: As medidas de Haar à direita e à esquerda existem e são únicas com exceção de uma constante multiplicativa.

A existência é dada pelos resultados A1 e B2, quanto à unicidade Haalmos (1950) mostra que

$$\mu(Ag) = Ag\mu(A) \quad e \quad \nu(gA) = \Delta g^{-1}\nu(A) \quad ,\forall\ A \subset G \ e \ \forall g \in G-$$

onde $\Delta g$ é chamada função modular e estabelece a relação entre as medi didas de Haar à direita e à esquerda onde

$$\Delta g = \frac{J_g(i)}{J_g^*\ (i)}$$

e além disso $\mu(A) = \Delta g \nu(A)$.

<u>Observações</u>: 1) No capítulo 6, utilizamos uma notação simplificada dos jacobianos onde $(J_{\bar{g}}(i))^{-1}$ está representado por $|J_{\bar{g}}|^{-1}$ e $J_{\bar{g}}^*(i)$ está representado por $|J_{\bar{g}}^*|$.

2) As medidas de Haar coincidem com a distribuição à priori imprópria proposta por Jeffreys.

oOo

## APÊNDICE B

## MODELOS INVARIANTES E A DISTRIBUIÇÃO A POSTERIORI DE BAYES

Seja $g(\theta/x)$ a função densidade fiducial e $g^*(\theta/x)$ a função densidade a posteriori Bayesiana com priori não informativa. Vamos mostrar que quando a função densidade $f(x/\theta)$ pertence a uma família $\mathcal{F}$ invariante sob um grupo de transformações $G$ (seção 6.1), então $g(\theta/x) = g^*(\theta/x)$. Além disso, mostramos que essa igualdade no caso uniparamétrico é válida somente para os modelos de locação e escala e vice-versa (Teorema 3.1.1). Faremos as demonstrações em duas partes: primeiramente o Teorema (3.1.1) e posteriormente abordaremos os modelos invariantes.

Parte I

Teorema (3.1.1) - Sabemos que para uma $F(x/\theta) = \int_{-\infty}^{x} f(x/\theta)dx$ satisfazendo as condições da seção (2.1.1) $g(\theta/x) = \frac{-\partial F(x/\theta)}{\partial\theta}$. Por outro lado, se $n(\theta)$ é a densidade a priori não informativa de $\theta$, então $g^*(\theta/x) = \frac{\partial F(x/\theta)}{\partial x} \frac{n(\theta)}{h(x)}$ onde $h(x) = \int \frac{F(x/\theta)}{\partial x} n(\theta)d\theta$ e assim se $g(\theta/x) = g^*(\theta/x)$ isto significa que

(BI.1) $$-\frac{\partial F(x/\theta)}{\partial\theta} = \frac{\partial F(x/\theta)}{\partial x} \frac{n(\theta)}{h(x)}$$ e então

(BI.2) $$\frac{\partial F(x/\theta)}{\partial x} \frac{1}{h(x)} + \frac{\partial F(x/\theta)}{\partial\theta} \frac{1}{n(\theta)} = 0$$

Como $F(x/\theta)$ satisfaz as condições da seção (2.1.1), existe uma função que relaciona $x$ e $\theta$. Denotamos essa função por $\underline{e} = \underline{e}(x;\theta)$, e assim há uma função $P$ tal que $F(x/\theta) = P(\underline{e}(x;\theta))$.

Dessa forma, a expressão (BI.2) pode ser reescrita como

$$\frac{\partial P}{\partial \underline{e}} \frac{\partial \underline{e}}{\partial x} \frac{1}{h(x)} + \frac{\partial P}{\partial \underline{e}} \frac{\partial \underline{e}}{\partial \theta} \frac{1}{n(\theta)} = 0 \Longrightarrow \frac{\partial \underline{e}}{\partial x} \frac{1}{h(x)} = \frac{-\partial \underline{e}}{\partial \theta} \frac{1}{n(\theta)} \Longrightarrow \frac{\partial \underline{e}}{\partial x} n(\theta) = \frac{-\partial \underline{e}}{\partial \theta} h(x)$$

donde se conclui que: ou $\underline{e}(x;\theta)$ é uma constante (o que é inadimissível dado que $F(x/\theta)$ é uma função de distribuição), ou $\frac{\partial \underline{e}}{\partial x} = h(x)$ e $\frac{\partial \underline{e}}{\partial \theta} = -n(\theta)$ e assim $\underline{e}(x;\theta) = H(x) - N(\theta)$, tal que $H(x) = \int h(x)dx$ e $N(\theta) = \int n(\theta)d\theta$, onde $H$ e $N$ são funções bijetoras dado que $x$ pode ser expressa como função de $\theta$ e vice-versa (devido às condições da seção 2.1.1).

Como $n(\theta)$ é a priori não informativa (proposta por Jeffreys) então

(1) $\quad n(\theta) = 1$ , para $-\infty < \theta < \infty$

(2) $\quad n(\theta) = \frac{1}{\theta}$ , para $\theta > 0$

Isto caracteriza: (1) os modelos de locação, onde $H(x) = x$ e $N(\theta) = \theta$, $n(\theta) = 1$ e (2) os modelos escala, onde $H(x) = \ell n(x)$ e $N(\theta) = \ell n(\theta)$, $n(\theta) = \frac{1}{\theta}$. E somente nesses modelos é válida a igualdade (BI.1).

Parte II

Seja G um grupo de transformações em $\chi$ com o qual pode ser identificado (seção 6.1.1). Nessas condições as transformações de G são do tipo $g_\Theta$, por exemplo, $g_\Theta(x) = \Theta + x$. Além disso, a v.a. x tem função densidade $f(x/\Theta) \in F$, onde $F$ é invariante sob um grupo de transformações G.

Com base na invariância de $F$ e usando os resultados da seção 6.2, vamos mostrar que $g(\Theta/x) = g^*(\Theta/x)$.

a) Existe uma quantidade pivotal $\underline{e} = g_\Theta^{-1}(x)$ (seção 6.2.1) tal que

$f(x/\Theta) = f(g_\Theta^{-1}(x)/\Theta)|J^*_{g_\Theta^{-1}}(x)|$, onde $|J^*_{g_\Theta^{-1}}(x)|$ , $g_\Theta^{-1} \in G$

é o jacobiano da transformação de x em $g_\Theta^{-1}(x)$.

b) Existe uma transformação invariante maximal $t_x \in G$ associada a uma órbita u (Definição 6.1.6) tal que podemos obter a distribuição condicional de $t_x(u)$ dado u

$$K(u)f(g_\Theta^{-1}(t_x(\Theta))/\Theta,u)|J^*_{g_\Theta^{-1}}(x)||J_{t_x}|^{-1}$$

onde K(u) é uma constante de proporcionalidade e $|J_{t_x}|^{-1}$ é a densidade de Haar associada a $t_x$ (Apêndice A).

<u>Observação</u>: Lembramos que uma inferência fiducial quando $f(x/\Theta) = f(t/\Theta,u)f(u)$ devemos utilizar apenas $f(t/\Theta,u)$. No contexto do item b) a origem u faz o papel da estatística ancilar e a densidade f(u) associada a u é proporcional a $|J_{t_x}|$ que é o jacobiano da transformação de $t_x(u)$ em $t_x\, g_i(u)$, com $g_i$ = identidade de G. Portanto

$$f(t/\Theta,u) = f(x/\Theta)|J_{t_x}|K'(u).$$

c) A densidade fiducial de $\Theta$ é igual a

$$\text{(BII.1)} \quad g(\Theta/t_x,u) = K(u)f(g^{-1}(t_x(u))/\Theta,u)|J^*_{g_\Theta^{-1}}(x)||J_{t_x}|^{-1}|J^*_{g_\Theta^{-1}}(t(u))|$$

onde $|J_{g_\Theta^{-1}}(t_x(u))|$ é o jacobiano da transformação de $t_x(u)$ em $g_\Theta^{-1}(t_x(u))$.

d) Segundo Berger (1980), para qualquer que seja $t_x$ integrável

$$\text{(BII.2)} \quad \int f(g_\Theta^{-1}(t_x(u))/\Theta)|J^*_{g_\Theta^{-1}}(x)\ |J_{t_x}|^{-1}d\Theta = [K(u)|J_{g_\Theta^{-1}}(t_x(u))|]^{-1}$$

e) Segundo a metodologia Bayesiana

$$\text{(BII.3)} \quad g^*(\Theta/x) = \frac{f(g_\Theta^{-1}(x)/\Theta)|J^*_{g_\Theta^{-1}}(x)|n(\Theta)}{f(g_\Theta^{-1}(x)/\Theta)|J^*_{g_\Theta^{-1}}(x)|n(\Theta)d\Theta}$$

então, se $n(\Theta) = |J_{t_x}|^{-1}$ segue de (BII.2) que (BII.1) é igual a (BII.3) e assim a distribuição fiducial de $\Theta$ coincide com a distribuição a posteriori Bayesiana com priori imprópria igual a densidade de Haar à direita definida em G.

Para mostrar que as densidades marginais e condicionais de $\Theta$ coincidem com aquela obtida no método Bayesiano, basta considerar um subgrupo de G e trabalhar com esse subgrupo da mesma forma com que trabalhamos com G (uma referência desse tópico é Fraser (1968)). Devido a essas considerações a distribuição fiducial de $\Theta$ pode ser utilizada com priori para novas observações referentes ao modelo de distribuição F.

## Bibliografia


[1] BARNARD, G.A. (1963). Some logical aspects of the fiducial argument. *Journal Royal Statistical Society, Ser. B*, 25, 111-114

[2] BARNETT, V.D. (1973). *Comparative Statistical Inference*. New York. John Wiley. 287 p.

[3] BARTLETT, M.S. (1936). The information available in small samples. *Proceedings of the Cambridge Philosophical Society*, 32, 560-566

[4] BERGER, J.O. (1980). *Statistical decision theory: foundations, concepts and methods*. New York, Springer, 425 p.

[5] BUEHLER, R.J. (1959). Some validity criteria for statistical inferences. *Annals of Mathematical Statistics*, 30(4), 845-863

[6] ______________(1979). *Some examples of ancillary statistics and their properties*. University of Minnesota. (Technical Report, 364).

[7] BUNKE, H. (1975). Statistical inference: fiducial and structural vs. likelihood. *Mathematische Operationsforschund statistik*, 6, 667-676

[8] CREASY, M.A. (1954). Limits for the ratio of means. *Journal Royal Statistical Society. Ser. B*, 16, 186-192

[9] DAWID, A.P. (1980). A Bayesian look at nuisance parameters. In: BERNARDO, J.M.; DeGROOT, M.H.; LINDLEY, D.N.; SMITH, A.F. M., eds. *Bayesian statistics*. Valencia. Spain, University Press. p. 141-204

[10] DAWID, A.P. & STONE, M. (1982). The functional-model basis of fiducial inference. *The Annals of Statistics*, 10(4), 1054-1067

[11] DEMPSTER, A.P. (963). Further examples of inconsistencies in the fiducial argument. *Annals of Mathematical Statistics*, 34 (3), 884-891

[12] EDWARDS, A.W. F. (1976). Fiducial probability, *The Statistician* 25, 15-35

[13] ________________ (1982). Fiducial inference II. In: KOTZ, S. and JOHNSON, N.L. *Encyclopedia of Statistical Sciences*. New York, John Wiley.

[14] EFRON, B. and HINCKLEY, D.V. (1978). Assessing the occuracy of the maximum likelihood estimator: observed versus expected Fisher information, *Biometrika*, 65 (3), 457-487

[15] FIELLER, E.C. (1954). Some problems in interval estimation. *Journal Royal Statistical Society*. *Ser*. B, 16, 175-185

[16] FISHER, R.A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26, 528-535

[17] ____________ (1935a). The logic of inductive inference (with discussion). *Journal of Royal Statistical Society*. *Ser*.B, 98, 39-54

[18] _____________(1935b). The fiducial argument in statistical inference. *Annals of Eugenics*, 6, 391-398

[19] ____________ (1939). The comparison of samples with possibly inequal variances. *Annals of Eugenics*, 9, 174-180

[20] ____________ (1941). The assymptotic approach to Behrens' integral with further tables for the d test of significance. *Annals of Eugenics*, 11, 141-172

[21] ____________ (1956). *Statistical methods and scientific inference*. Edinburgh, Oliver and Boyd, 175 p.

[22] FRASER, D.A.S. (1961a) On fiducial inference. *Annals of Mathematical Statistics*, 32 (3), 661-676

[23] ______________ (1961b). The fiducial methods and invariance. *Biometrika*, 48, 261-280

[24] ______________ (1964). On the definition of fiducial probability. *Bulletin of the international Statistical Institute*, 40 (2), 842-856

[25] ______________ (1968). *The Structure of inference*. New York, John Wiley, 344 p.

[26] ______________ (1979). *Inference and linear models*. London, MacGraw-Hill. 297 p.

[27] HACKING, I. (1965). *Logic of statistical inference*. Cambridge, University Press. 232 p.

[28] HORA, R.B. & BUEHLER, R.J. (1966). Fiducial theory and invariant estimation. *Annals of Mathematical Statistics*, 37 (3), 643-656

[29] KENDALL, M.G. (1947/51). *The advanced theory of statistics*, 3.ed., London, Griffin. v. 1 e v. 2

[30] LINDLEY, D.V. (1957). Review: Statistical methods and scientific inference by R.A. Fisher, *Heredity*, 11, 280-283

[31] ______________ (1958) Fiducial distributions and Bayes' theorem. *Journal Royal Statistical Society. Ser. B.* 20, 102-107

[32] MARIOTTO, A.B. (1982). *Inferência parcial*. São Paulo, 107 p. Dissertação (Mestrado) - IME-USP

[33] MAULDON, J.G. (1955). Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory. *Journal Mathematical Statistical Society. Ser.B*, 17, 79-85

[34] NEYMAN, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal Royal Statistical Society. Ser. B*, 97, 558-625

[35] PEDERSEN, J.G. (1978). Fiducial inference. *Int. Statist. Review*, 46, 147-170

[36] ROBINSON, G.K. (1975). Some counterexamples to the theory of confidence intervals. *Biometrika*, 62 (1), 155-161

[37] SEIDENFELD, T. (1979). *Philosophical problems of statistical inference*. Boston, Reidel.

[38] SPROTT, D.A. (1961). On example an ancillary statistic and the combination of two samples by Bayes' theorem. *Annals of Mathematical Statistics*, 32 (1), 616-618

[39] ____________ (1964). A transformation model for the investigation of fiducial distributions. *Bulletin of the International Statistical Institute*, 40 (2), 856-869

[40] STEIN, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Annals of Mathematical Statistics*, 30 (4), 877-880

[41] WILKINSON, G.N. (1977). On resolving controversy in statistical inference. *Journal Royal Statistical Society. Ser. B*, 38, 119-171

[42] WILLIANS, J.S. (1966). The role of probability in fiducial inference. *Sankhyã, Ser. A*, 28, 271-296

o0o