INFERÊNCIA PARCIAL

ANGELA BACELLAR MARIOTTO

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: ESTATÍSTICA

ORIENTADOR:

Prof. Dr. Carlos Alberto de Bragança Pereira

- SÃO PAULO, JUNHO DE 1983 -

AGRADECIMENTOS

Ao Carlinhos, pelo interesse e dedicação, pela estimulante troca de idéias, fundamental na minha formação, pelo respeito às minhas posições individuais e por compreender que a vida é bem mais que uma dissertação.

Ao professor Basu, com quem tive o privilégio do convívio intelectual por algum tempo, pelo carinho e pela força de suas idéias, inspiradoras de boa parte deste trabalho.

Aos colegas Nelson, Roberto e André, pelo gratificante e enriquecedor contato diário em que aprendi a encarar o ensino e a ciência com criatividade e inconformismo.

À Luzia do Carmo Namiki, pelo excelente trabalho de datilografia, apesar do pouco tempo que lhe dei.

E, finalmente, ao Kiko, pelo amor e apoio durante os momentos mais difíceis.

INFERÊNCIA PARCIAL

INDICE

I N F E R Ê N C I A   P A R C I A L

CAPÍTULO 1

INTRODUÇÃO

1.1 - Parâmetro Nuisance

O objetivo da Inferência Estatística é obter "Informação" a respeito de um parâmetro Θ, que representa um estado desconhecido (por nós) da natureza. O procedimento é realizar um experimento E, cujo resultado x, a ser observado, acredita-se estar relacionado ao valor desconhecido de Θ. A formalização matemática do experimento E é feita quando se considera a trinca abstrata (X,A,P), chamada de Modelo Estatístico, em que X (Espaço Amostral) é o conjunto de todos os possíveis resultados do experimento E, A a σ-álgebra gerada por X e $P = \{P_\theta;\ \theta \in \Theta\}$[1] a

(1) A notação utilizada aqui não pretende ser formal. Utilizaremos $P(x/\theta)$ para indicar tanto função densidade quanto função de probabilidade.

família de distribuições possíveis para x. A família $\mathcal{P}$ descreve o relacionamento entre x e Θ, ou seja, para alguns valores distintos de Θ, considera-se diferentes distribuições para x. Se a inferência sobre Θ é feita a partir do experimento $\mathcal{E}$ e da amostra x, a informação sobre Θ está contida nos "dados estatísticos" $(\mathcal{E}, x)$.

A maior parte dos estatísticos considera o problema da Inferência apenas para os casos em que $\mathcal{P}$ é uma família de distribuições indexada pelo próprio parâmetro de interesse Θ. Este trabalho trata de um problema mais freqüênte e mais geral que surge quando aos dados x e ao experimento $\mathcal{E}$ está associada uma família indexada por um parâmetro λ de tal modo que $\Theta = \Theta(\lambda)$ é uma "redução" de λ. Isto significa que o modelo $\mathcal{P} = \{P_\lambda, \lambda \in \Lambda\}$ depende de λ e estamos interessados em apenas uma característica de λ, denominada $\Theta(\lambda)$. Neste caso, como fazer inferência apenas sobre Θ, se existem outras características desconhecidas do modelo?

A maneira usual de enfocar o problema é introduzir um parâmetro $\phi(\lambda)$, "complementar" a $\Theta(\lambda)$ e tal que Θ e ϕ especifiquem conjuntamente a distribuição de x. Isto é, deve existir uma relação 1-1 entre λ e $(\Theta(\lambda), \phi(\lambda))$ de forma que $\phi(\lambda)$ seja a menor redução possível de λ. Desta maneira consegue-se uma relação 1-1 entre as famílias de distribuições $\mathcal{P}_{\Theta,\phi}$ e $\mathcal{P}_\lambda$ e pode-se trabalhar com uma ou com outra família indistintamente. Na reparametrização acima, ϕ é chamado de parâmetro nuisance, ou seja, a característica do modelo que não nos interessa, e Θ o parâmetro de interesse.

É importante ressaltar-se a natureza arbitrária do parâmetro $\phi(\lambda)$, uma vez que a função $\phi(\lambda)$ descrita acima não é única. Por exemplo, vamos supor que estamos interessados no comprimento (desconhecido) de certo objeto. Toma-se, então, n medidas não tendenciosas e independentes deste objeto. O modelo gerado pelo experimento pode ser a família das distribuições $N(\mu,\sigma^2)$, em que μ é o parâmetro de interesse, isto é, o comprimento do objeto e σ^2 o erro de precisão do instrumento de medida. Neste caso poderíamos tomar tanto a variância σ^2 quanto o coeficiente de variação $\frac{\sigma}{\mu}$, como parâmetro nuisance. A escolha do parâmetro nuisance dependerá muitas vezes do método de eliminação utilizado.

Por outro lado, em muitos problemas estatísticos mesmo a escolha do parâmetro de interesse é arbitrária. Em um problema de Teste de Hipótese, testar $H_o:\Theta=\mu_1-\mu_2=0$ parece ser equivalente a testar $H_o:\Theta=\frac{\mu_1}{\mu_2}=1$. Em tais casos, cabe ao estatístico encontrar a reparamenização (Θ,ϕ) mais adequada para a utilização de alguns dos métodos de eliminação. Portanto, a escolha do par "ideal" (Θ,ϕ) é um problema de ordem prática e para o qual não existe um procedimento esquemático. A seguir tentaremos exemplificar como uma "boa" reparametrização pode levar a uma solução simples.

<u>EXEMPLO 1.1</u> (Poisson) - Deseja-se testar a Igualdade das médias de duas v.as. independentes com distribuição de Poisson. Sejam

x_1 ~ Poisson (λ_1)

x_2 ~ Poisson (λ_2)

O modelo é dado por,

$$p(x_1,x_2/\lambda_1,\lambda_2) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\,\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\;;\qquad x_1\geq 0,\; x_2\geq 0$$

e o espaço paramétrico é,

$$\Lambda=\{(\lambda_1,\lambda_2)\text{ t.q. }\quad \lambda_1\geq 0,\quad \lambda_2\geq 0\}.$$

Estamos interessados em verificar apenas uma característica do modelo; a igualdade ou não dos parâmetros, ou seja, se $\lambda_1 = \lambda_2$ ou se $\lambda_1 \neq \lambda_2$. O problema é encontrar θ que especifique esta característica.

Considere a seguinte reparametrização do modelo: $\theta = \dfrac{\lambda_1}{\lambda_1+\lambda_2}$, $\phi = \lambda_1+\lambda_2$. Não é difícil ver que entre (λ_1,λ_2) e (θ,ϕ) existe uma relação 1-1. Nestas condições a hipótese nula a ser testada é $H_o:\theta = \frac{1}{2}$, de modo que θ é o parâmetro de interesse e ϕ o parâmetro nuisance. A escolha de (θ,ϕ) definida acima não é um problema trivial e adiante veremos porque ela foi considerada. Observe que outras reparametrizações, como por exemplo, $\theta = \dfrac{\lambda_1}{\lambda_2}$ ou $\theta = \lambda_1-\lambda_2$, poderiam testar da mesma forma a hipótese nula.

A questão que se coloca é encontrar um método de Inferência sobre θ quando a distribuição dos dados x envolve também um valor desconhecido ϕ, que não nos interessa. Da maneira como foi colocado o problema conclui-se que, na verdade, estamos procurando métodos de eliminação do parâmetro nuisance ϕ. Existem vários métodos com esse objetivo, como por exemplo substituir ϕ pelo seu estimador $\hat{\phi}$. Uma breve descrição de alguns métodos pode ser encontrada em Basu [1977]. O objetivo do presente trabalho é estudar e analisar em detalhes uma pequena classe destes métodos.

1.2 - Redução de Experimento

Os métodos de eliminação que aqui serão analisados consistem basicamente na simplificação do modelo original (X, A, P) para um modelo, redução deste, cuja família de distribuições depende apenas de θ. Esta simplificação do modelo original para um modelo dependendo apenas de θ é o objetivo principal, uma vez que, deste modo a Inferência sobre θ segue das técnicas usuais.

A dificuldade está em justificar a particular simplificação, que muitas vezes não é evidente.

Vimos que quando se realiza um experimento E, objetivando-se Inferências sobre θ, o resultado x e o modelo estatístico (X, A, P) (ou E), contém toda a informação disponível sobre θ. Logo, a substituição do modelo original deve ter como justificativa a não-perda de Informação sobre θ. Conseqüentemente, para as inferências sobre θ a partir de uma "redução" do modelo original, dois objetivos estão identificados:

(A) <u>Simplicidade</u>: O experimento (modelo) reduzido deve possuir uma distribuição que dependa apenas de Θ.

(B) Informatividade: A redução do experimento original não deve perder informação relevante para as Inferências sobre Θ.

O próximo passo é traduzir em termos matemáticos os dois objetivos acima. É necessário definir o que seja redução do experimento (ou modelo).

Seja $t:X \longrightarrow \tau$ uma estatística. É sempre verdade que $p(x/\lambda)=p(x/t,\lambda).p(t/\lambda)$. Associada a esta fatoração podemos considerar a seguinte fatoração do modelo:

- Modelo Marginal: (t,P_t) onde $P_t=\{p(t/\lambda),\ \lambda\in\Lambda\}$ é a família de distribuições marginais de t.

- Modelo Condicional: $(x,P(./t))$ onde $P(./t)=\{p(x/t,\lambda),\ \lambda\in\Lambda\}$ é a família das distribuições condicionais de x dado $t(x)=t$.

Estes dois modelos são chamados de reduções do modelo (X,A,P).

Note que associados aos modelos reduzidos pode-se considerar reduções do experimento E. Respectivamente temos:

- Experimento Marginal E^t, gerando os dados t.

- Experimento Condicional E_t, que gera os dados x, para cada valor fixado de t.

Desta maneira considera-se os dados (E,x) composto por (E^t,t) e (E_t,x). O primeiro experimento pode ser operacionalmen te definido como "realiza-se E e observa-se apenas t". O segundo é apenas um experimento conceitual pois não é realizável.

Os objetivos (A) e (B) podem ser agora formulados em termos das duas possíveis reduções do experimento. Esta formulação originará duas versões da Inferência Parcial (ou separável) de Θ em termos de dois Princípios.

S: Princípio da Marginalização (ou Suficiência). Se o modelo $P = \{p(x/\theta,\phi),\ \theta\in\Theta\ ,\ \phi\in\Phi\}$ satisfizer:

a) O modelo marginal P_t depende apenas de θ, isto é, $p(t/\lambda_1)=p(t/\lambda_2)$ se $\theta(\lambda_1)=\theta(\lambda_2)$.

b) O modelo condicional $(x,P(./t))$ é não informativo com respeito a θ.

Então Inferências sobre θ podem ser feitas a partir do Modelo Marginal (P_t,t).

C: Princípio do Condicionamento (ou Ancilaridade): Se o modelo P satisfizer:

a) O modelo condicional $P(./t)$ depende apenas de θ.

b) O modelo marginal (P_t,t) é não informativo com respeito a θ.

Então Inferências sobre θ podem ser feitas a partir do Modelo Condicional $(P(./t),x)$.

O objetivo (A), de simplicidade está claramente expli citado nos dois Princípios acima pelas condições (a). Assim, em S exige-se que o Modelo Marginal dependa apenas de θ e em C que isto aconteça com o modelo condicional. O mesmo não acontece com o objetivo (B) uma vez que as condições (b) dos princípios expressam apenas um conceito vago de não-informação. É esta ausência de uma definição precisa de não-informação que origina os vários métodos estudados aqui. Cada um destes métodos tenta capturar, em termos de uma frase matemática, o significado de Informação e Não-Informação.

1.3 - Suficiência Parcial e Ancilaridade Parcial

Os conceitos de informação e não-informação estão estreitamente ligados aos conceitos de suficiência parcial[1] e ancilaridade parcial[1].

Esta relação tem como motivação generalizações dos conceitos de suficiência e ancilaridade usuais que foram introduzidos por Fisher, para os casos em que o parâmetro de interesse coincide com o parâmetro do modelo. Analisaremos para esta situação o relacionamento entre informação e suficiência (não-informação e ancilaridade).

Seja $P=\{P_\theta,\ \theta\in\Theta\}$ o modelo estatístico e $t:X \longrightarrow \tau$ uma estatística.

(1) A denominação parcial vem do fato de se tratar de conceitos de suficiência e ancilaridade para um subparâmetro do modêlo.

DEFINIÇÃO 1.1 - A estatística t é suficiente(1) com respeito a Θ se $p(x/\Theta)=p(x/t).p(t/\Theta)$, em palavras, se a distribuição condicional de x dado t independe de Θ.

DEFINIÇÃO 1.2 - A estatística t é ancilar com respeito a Θ se $p(x/\Theta)=p(x/t,\Theta).p(t)$, em palavras, se a distribuição marginal de t independe de Θ.

Como já foi dito no início do trabalho, o único ente que relaciona o resultado x ao parâmetro Θ, é a verossimilhança $L(\Theta/x)=p(x/\Theta)$. Assim, espera-se que ao variar-se o valor de Θ, o valor de $L(\Theta/x)$ se altere. Caso contrário não deve haver relacionamento entre x e Θ, tornando-se impossível retirar de x alguma informação relevante sobre Θ. Como consequência intuitiva, acreditamos que ao fatorarmos $p(x/\Theta)$, todo o fator que não se altera ao variarmos Θ, pode ser abandonado e deve-se considerar , apenas, aqueles fatores que são funções de Θ.

A consequência dessa discussão é a eliminação dos fatores $p(x/t)$ e $p(t)$ nas definições 1.1 e 1.2, respectivamente . Assim se t é uma estatística suficiente com respeito a Θ, qualquer inferência sobre Θ pode ser baseada apenas na função $p(t/\Theta)$. Analogamente, se t é uma estatística ancilar, as inferências sobre Θ podem ser baseadas apenas na função $p(x/t,\Theta)$.

---

(1) A expressão "t é suficiente c.r. a Θ", não é muito adequada Expressões mais propícias seriam, por exemplo, "t é suficiente para a amostra x" ou "t isola e exaure toda a informação sobre Θ, contida em x".

A conclusão a que se chega com a discussão acima é que se t é suficiente com respeito a Θ, então $t=t(x)$ contém toda a informação relevante sobre Θ que x carrega. Deste modo, é suficiente que se observe t no lugar de x. A situação em que t é ancilar é um pouco mais delicada. Não há dúvida que toda a informação relevante sobre Θ está descrita pelo modelo condicional $p(x/t,\Theta)$. Contudo, a afirmação que t é não informativa com respeito a Θ traz alguns problemas sérios. Se encararmos o modelo $p(x/\Theta)$ como uma afirmação condicional sobre x dado Θ, dizer que t é ancilar é dizer que t é independente de Θ ($t \perp\!\!\!\perp \Theta$). Poderá existir uma outra estatística ancilar t', ou seja $t' \perp\!\!\!\perp \Theta$ e tal que (t,t') não é independente de Θ; isto é, t é ancilar, t' é ancilar porém (t,t') contém informação sobre Θ. Um exemplo extremo deste fato, devido a Basu [1964], é apresentado a seguir. Aqui, (t,t') na verdade é suficiente com respeito a Θ.

<u>EXEMPLO 1.2</u>: Considere o vetor aleatório (x_1,x_2) com distribuição $N_2(0,0\ ,\ 1,1\ ,\ \rho)$. Observe que (x_1,x_2) é uma estatística suficiente com respeito ao coeficiente de correlação ρ. Entretanto como $x_1 \sim N(0,1)$ e $x_2 \sim N(0,1)$ temos que x_1 é ancilar com respeito a ρ, idem para x_2.

No lugar de afirmarmos que se t é ancilar então t é não informativa com respeito a Θ, preferimos estabelecer que se t é ancilar então nenhuma informação com respeito a Θ pode ser extraída de t.

Na discussão acima a condição de não-informação (B) está claramente determinada; um modelo é não informativo com respeito a Θ se este modelo não depender de Θ. Entretanto para o caso em que os modelos envolvem parâmetros nuisance teremos não apenas um conceito de não-informação mas alguns, que foram desenvolvidos por autores diferentes. A proposta dos próximos capítulos é exatamente discutir estes conceitos.

A generalização dos métodos de redução, relacionados com as definições de suficiência e ancilaridade de Fisher (1.1 e 1.2), para o caso em que não estamos livres de parâmetros nuisance pode ser feita como segue:

Se t é uma estatística que satisfaz a condição (b) de não informação do Princípio S[Princípio C] segundo um autor P, então t é dita parcialmente suficiente [ancilar] com respeito a Θ, segundo o autor P, ou, t é dita P-suficiente [P-ancilar] com respeito a Θ. Esta notação é a mesma utilizada por Barndorff - Nielsen [1980].

CAPÍTULO 2

S-SUFICIÊNCIA E S-ANCILARIDADE

Antes de se introduzir as definições de S-suficiência e S-ancilaridade serão necessárias algumas considerações sobre o parâmetro nuisance ϕ.

No capítulo 1 discutiu-se a natureza arbitrária do parâmetro nuisance. A consequência deste fato é a não existência, em toda a literatura, de uma definição precisa para parâmetro nuisance. Entretanto, exige-se do parâmetro ϕ que conjuntamente com o parâmetro de interesse Θ especifiquem λ, e além disso que nenhuma informação sobre ϕ, é de interesse nas inferências sobre Θ (complementar). Essas "qualidades" do parâmetro nuisance são de certa forma explicitadas na definição abaixo, que pretende ser apenas uma definição funcional.

DEFINIÇÃO 2.1 - Os subparâmetros $\Theta:\Lambda \longrightarrow \Theta$ e $\phi:\Lambda \longrightarrow \Phi$ são de variação independente se o campo de variação de Θ é o mesmo para cada valor de ϕ e vice-versa, ou seja, o campo de variação de (θ,ϕ) é $\Theta\times\Phi$.

Na sequência desta seção exigiremos que θ e ϕ sejam de variação independente.

<u>DEFINIÇÃO 2.2</u> - A estatística t é S-suficiente com respeito a θ se $p(x/\theta,\phi)=p(t/\theta)\ p(x/t,\phi)$.

<u>DEFINIÇÃO 2.3</u> - A estatística t é S-ancilar com respeito a θ se $p(x/\theta,\phi)=p(t/\phi).p(x/t,\theta)$.

A definição de S-suficiência acima, é geralmente atribuída a Fraser [1956], e pode ser considerada uma extensão das definições de suficiência e ancilaridade de Fisher. A condição de não-informação nelas contida é decorrência da discussão feita na seção 1.3. Exigia-se lá, que o modelo não-informativo para θ independesse deste parâmetro. O mesmo acontece com as definições acima.

Portanto se t é S-suficiente com respeito a θ, estão satisfeitas as condições do Princípio S, da seguinte maneira:

(a) A distribuição marginal de t depende apenas do parâmetro de interesse θ.

(b) A distribuição condicional de x dado t independe de θ e consequentemente não contém informação sobre θ.

Assim, inferências sobre θ são feitas considerando-se apenas o modelo marginal (P_t, t).

Raciocínio análogo pode ser feito quando considera-se t S-ancilar com respeito a θ.

É interessante notar que as definições acima são equivalentes no sentido em que elas fatoram a função de verossimilhança em duas partes uma que depende apenas de θ e outra que depende apenas de ϕ. Nestes casos t define um corte de Barndorff-Nielsen.

DEFINIÇÃO 2.4 - A estatística t define um corte de Barndorff-Nielsen no modelo P_λ se existem dois subparâmetros de variação independente $\theta=\theta(\lambda)$ e $\phi=\phi(\lambda)$ tal que o modelo marginal P_t dependa apenas de θ e que para cada t fixado o modelo condicional $P(./t)$ dependa apenas de ϕ.

Portanto se t define um corte de Barndorff-Nielsen t é S-suficiente com respeito a θ e S-ancilar com respeito a ϕ.

Apresentamos a seguir alguns exemplos clássicos da literatura, com o objetivo de facilitar o entendimento das definições acima e realçar o fato de que nem sempre é possível aplicá-las.

CAPÍTULO 3

EXEMPLOS

3.1 - Teste de Igualdade das Médias de Duas Poissons Independentes

Considere novamente o exemplo 1.1. Suponha que estamos interessados em inferir apenas sobre uma característica do modelo; a igualdade ou não dos parâmetros. Vimos que essa característica pode ser expressa em termos do parâmetro $\Theta = \dfrac{\lambda_1}{\lambda_1+\lambda_2}$, quando considera-se a seguinte hipótese $H_o:\Theta = \frac{1}{2}$.

A seguinte reparametrização do modelo, foi proposta: $\Theta = \dfrac{\lambda_1}{\lambda_1+\lambda_2}$, $\phi=\lambda_1+\lambda_2$. O espaço paramétrico $\Lambda=\{(\lambda_1,\lambda_2)\text{ t.q } \lambda_1>0 , \lambda_2>0\}$ passa a ser $\Lambda' = \{(\Theta,\phi) \text{ t.q } 0<\Theta<1, \phi>0\}$. É evidente então, que Θ e ϕ são parâmetros de variação independente (definição 2.1).

Com esta reparametrização, o modelo é equivalente a:

$$p(x_1,x_2/\Theta,\phi) = \frac{e^{-\phi}\,\phi^{x_1+x_2}}{(x_1+x_2)!}\;\frac{(x_1+x_2)!}{x_1!\,x_2!}\;\Theta^{x_1}(1-\Theta)^{x_2};\quad x_1\geq 0, x_2\geq 0$$

Considere a estatística $t=x_1+x_2$. Observe que esta estatística determina um corte de Barndorff-Nielsen no modelo, pois:

$$p(x_1,x_2/\theta,\phi) = p(t/\phi).p(x/t,\theta)$$

onde a distribuição marginal de t é Poisson (ϕ) e a distribuição condicional de x dado t é Binomial (t,θ).

Portanto, t é uma estatística S-ancilar com respeito a θ e o teste da Igualdade das médias é realizado utilizando-se apenas o modelo condicional de x dado t, ou seja, o modelo Binomial (t,θ).

Neste exemplo a inferência parcial (separável) é elegantemente aplicável. Se o parâmetro de interesse fosse ϕ o modelo considerado seria o modelo marginal de t, Poisson (ϕ). Nos exemplos que seguem essa técnica não parece ser aplicável, pois não conseguimos encontrar uma reparametrização que permita um corte de Barndorff-Nielsen.

3.2 - Teste Exato de Fisher para Tabela de Contingência 2×2.

Considere uma população de indivíduos, que pode ser classificada segundo duas características A e B (A^c e B^c são os complementares de A e B), gerando a tabela,

$$\begin{array}{cccc} & B & B^c & \\ A & p_{11} & p_{12} & p_{1.} \\ A^c & p_{21} & p_{22} & p_{2.} \\ & p_{.1} & p_{.2} & p_{2.} \end{array} \qquad (3.1)$$

onde p_{11}, p_{12} e p_{21} são os parâmetros populacionais desconhecidos correspondentes às proporções das categorias AB, AB^c e A^cB respectivamente.

O interesse do estatístico é testar a independência entre as características A e B, ou seja, se o fato do indivíduo ter característica B (ou B^c) não influencia a probabilidade de ele ter a característica A e vice-versa. A hipótese a ser testada é $H_o: p(A/B)=p(A/B^c)$ (ou $P(A^c/B)=P(A^c/B^c)$). Segundo a estrutura paramétrica da tabela (3.1) temos

$$H_o: \frac{p_{11}}{p_{.1}} = \frac{p_{12}}{p_{.2}} \qquad (\text{ou } \frac{p_{21}}{p_{.1}} = \frac{p_{22}}{p_{.2}})$$

O resultado de um experimento, conduzido para testar a hipótese acima, pode ser disposto em uma tabela 2×2 como segue:

	B	B^c	
A	a	b	n_1
A^c	c	d	$n-n_1$
	t	n-t	n

onde a,b,c e d são as frequências obtidas pelo experimentador em cada uma das categorias AB, AB^c, A^cB e A^cB^c respectivamente.

Como a Inferência Clássica exige o conhecimento do espaço amostral, vamos considerar dois experimentos que podem gerar a tabela acima.

(i) Multinomial; retira-se uma amostra de tamanho n da população, com reposição, classificando-os nas categorias. Os dados para este experimento são (a,b,c) e seguem o modelo multinomial (n,p_{11},p_{12},p_{21}).

(ii) Binomial; toma-se n_1 indivíduos com características A e $n-n_1$ com característica A^c. Dentre estes, observa-se o número de indivíduos com característica B. Os dados são (a,c) e seguem a distribuição a - Binomial $(n_1 , \frac{p_{11}}{p_{1.}})$ e c - Binomial $(n_2 , \frac{p_{21}}{p_{2.}})$, independentes.

O Modelo Binomial pode ser considerado como uma redução do modelo multinomial, basta ver que a distribuição multinomial pode ser escrita como

$$p(a,b,c/p_{11},p_{12},p_{21})=p(a,c/n_1,p_{11},p_{12},p_{21}) \times$$
$$\times\ P(n_1/p_{11},p_{12},p_{21})$$

onde $p(a,c/n_1,p_{11},p_{12},p_{21})$ é justamente a distribuição do experimento ii.

O "teste exato de Fisher" é idêntico para as duas situações. Vamos considerar primeiramente o caso mais geral ou seja o modelo multinomial.

Trabalharemos com a estatística (a,n_1,t) em lugar de (a,b,c). O espaço paramétrico do modelo é o conjunto $\Lambda=\{(p_{11},p_{12},p_{21})/p_{11}+p_{12}+p_{21} \leq 1 \ , \ p_{11}\geq 0,\ p_{12}\geq 0,\ p_{21}\geq 0\}$. Fisher considerou a seguinte reparametrização do modelo para construção do seu teste: $\Theta = \frac{p_{11}p_{22}}{p_{12}p_{21}}$, $\phi = \frac{p_{21}}{p_{2.}}$ e $\Psi=p_{1.}$ Não é difícil ver que a condição de bijeção entre (Θ,ϕ,Ψ) e (p_{11},p_{12},p_{21}) está satisfeita. O novo espaço paramétrico é $\Lambda'=\{(\Theta,\phi,\Psi)$ t.q. $\Theta>0,\ 0<\phi<1,\ 0<\Psi<1\}$ e neste caso temos que Θ,ϕ e Ψ são parâmetros de variação independente. A tabela (3.1) pode ser escrita como:

	B	B^c	
A	$\dfrac{\phi\Theta}{1-\phi+\Theta\phi}\Psi$	$\dfrac{(1-\phi)\Psi}{1-\phi+\Theta\phi}$	Ψ
A^c	$\phi(1-\Psi)$	$(1-\Psi)(1-\phi)$	$(1-\Psi)$

Das relações $\dfrac{p_{11}}{p_{.1}}=\dfrac{p_{12}}{p_{.2}}$ e $\dfrac{p_{21}}{p_{.1}}=\dfrac{p_{22}}{p_{.2}}$ segue que a hipótese nula pode ser equivalentemente escrita como $H_o:\Theta=1$, assim Θ é o parâmetro de interesse e (ϕ,Ψ) são os parâmetros nuisance.

O próximo passo é encontrar algum modelo reduzido que dependa apenas do parâmetro de interesse Θ, portanto que elimine ϕ e Ψ.

A distribuição dos dados (a,n_1,t) de acordo com a reparametrização é dada por:

$$p(a,n_1,t/\Theta,\phi,\Psi)=\binom{n}{n_1}\binom{n_1}{a}\binom{n-n_1}{t-a}\frac{\Theta^a\phi^t(1-\phi)^{n-t}\Psi^{n_1}(1-\Psi)^{n-n_1}}{(1-\phi+\Theta\phi)^{n_1}}.$$

Considerando a seguinte fatoração da distribuição acima

$$p(a,n_1,t/\Theta,\phi,\Psi)=p(n_1/\Theta,\phi,\Psi).p(t/n_1,\Theta,\phi,\Psi).p(a/t,n_1,\Theta,\phi,\Psi)$$

verificamos que a distribuição condicional de a dado (t,n_1) depende apenas do parâmetro de interesse θ. Vamos supor que $t \leq n_1$ sem perda de generalidade.

$$p(a/t,n_1,\theta) = \frac{\binom{n_1}{a}\binom{n-n_1}{t-a}\theta^a}{\sum_{k=0}^{t}\binom{n_1}{t}\binom{n-n_1}{t-k}\theta^k}, \quad t \leq n_1$$

$$p(t/n_1,\phi,\theta) = \frac{\phi^t(1-\phi)^{n-t}}{(1-\phi+\theta\phi)^{n_1}} \sum_{k=0}^{t}\binom{n_1}{t}\binom{n-n_1}{t-k}\theta^k, \quad t \leq n_1$$

$$P(n_1/\Psi) = \binom{n}{n_1}\Psi^{n_1}(1-\Psi)^{n-n_1}$$

Observe também que se considerarmos o Modelo Binomial $p(a,t/n_1,\theta,\phi)$ a mesma redução pode ser conseguida, uma vez que na fatoração $p(a,t/n_1,\theta,\phi)=p(a/t,n_1,\theta).p(t/n_1,\theta,\phi)$ o modelo condicional de a dado (t,n_1) depende apenas de θ.

Para os dois experimentos considerados, conseguiu-se um modelo reduzido $p(a/t,n_1,\theta)$ dependente apenas do parâmetro de interesse θ, e é esta distribuição que é utilizada no "Teste exato de Fisher".

O argumento de Fisher na construção do teste, parece ter sido que as estatísticas marginais (n_1,t) não conteriam informação sobre θ. Consequentemente a distribuição marginal de (n_1,t) não é levada em consideração neste teste.

Contudo, a distribuição marginal de (n_1,t) envolve o parâmetro θ e segundo a discussão levada na seção 1.3 deve conter informação sobre θ. Além disso (n_1,t) não é S-ancilar com respeito a θ e não define um corte de Barndorff-Nielsen para o modelo.

Para alguns estatísticos que tentaram justificar o "teste exato de Fisher" a condição de S-não informação, isto é, a condição de que o modelo não informativo independesse de θ é muito restritiva. Deste modo, êles tentaram, em seus trabalhos, responder a seguinte pergunta: Quando um modelo reduzido, embora dependendo de θ, pode ser considerado não informativo com respeito a θ e assim ser eliminado da análise?

Será esta a questão que tentaremos analisar nos próximos capítulos.

3.3 - Normal Univariada

Considere uma população normal com média μ e variância σ^2. Suponha que $\lambda=(\mu,\sigma^2)$ é desconhecido. Estaremos interessados em inferências separadamente para μ e σ^2.

O objetivo aqui é procurar uma estatística cuja distribuição dependa apenas do sub-parâmetro de interesse e sumarize toda a informação sobre este, contida na amostra.

A relevância deste exemplo está em proporcionar uma discussão dos conceitos de suficiência e ancilaridade parcial que estão estreitamente ligados a idéia de uma estatística conter ou não informação sobre um parâmetro. Além disso, do ponto de vista histórico este exemplo é importante, pois o conceito de suficiência foi introduzido por Fisher quando discutia a precisão de dois estimadores da variância da normal quando a média era desconhecida; Fisher [1920], [1922].

Seja então uma a.a.s $x=(x_1,\dots,x_n)$ de uma população $N(\mu,\sigma^2)$, μ e σ^2 desconhecidos. Sabe-se que a estatística suficiente com respeito a (μ,σ^2) é $(\bar{x},S^2)$ e que $\bar{x}$ e S^2 são independentes. Seria, então, intuitivo esperar que $\bar{x}$ contenha toda a informação da amostra sobre μ idem para S^2 sobre σ^2. Analisaremos em que sentido $\bar{x}(S^2)$ pode ser considerada parcialmente suficiente com respeito a $\mu(\sigma^2)$.

Suponha que estamos interessados apenas na média da população normal. Será que $\bar{x}$ resume toda a informação relevante na amostra sobre μ?

A resposta à pergunta acima é negativa. O problema é que a distribuição de $\bar{x}$ depende de σ^2; $\bar{x} \sim N(\mu,\sigma^2/n)$, de maneira que não se elimina o parâmetro nuisance σ^2. Além disso vimos que $\bar{x}$ "depende" de σ^2, por sua vez σ^2 "depende" de S^2 logo $\bar{x}$ deve "depender" de S^2 e S^2 deve conter informação sobre a média μ. Isto pode ser facilmente compreendido quando percebe-se que uma vez abandonada a estatística S^2 torna-se impossível analisar $\bar{x}$ como estimador de μ em termos de sua precisão. Portanto perde-se informação sobre μ quando se considera apenas $\bar{x}$.

Se a variância é o parâmetro de interesse o problema torna-se menos trivial que no caso anterior, pois consegue - se eliminar o parâmetro nuisance μ quando se considera a distribuição marginal de S^2. Resta discutir se o problema da informação está resolvido, isto é, se S^2 é totalmente informativa com respeito a σ^2.

A função de verossimilhança dos dados x pode ser escrita como função de $(\bar{x},S^2)$ como segue

$$p(s/\mu,\sigma^2) = A(\sigma)\exp\left[-\frac{nS^2}{2\sigma^2}\right]\exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right]$$

onde $A(\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n$

Note que a distribuição condicional dos dados x dado S^2 depende de σ^2, isto é,

$$p(x/\mu,\sigma^2)\,\alpha^{(1)}\,P(\bar{x},S^2/\mu,\sigma^2) = P(S^2/\sigma^2)P(\bar{x}/S^2,\mu,\sigma^2) =$$

$$= P(S^2/\sigma^2)P(\bar{x}/\mu,\sigma^2).$$

Como no "Teste exato de Fisher" a parte do modelo que estamos abandonando depende do parâmetro de interesse σ^2, e S^2 não é S-suficiente com respeito a σ^2.

A pergunta que fica é a mesma que anteriormente. Será que a condição de não informação, que exige do modelo a não dependência do parâmetro é muito restritiva?

(1) O símbolo "α" indica proporcionalidade.

CAPÍTULO 4

G-ANCILARIDADE (GODAMBE)

4.1 - Introdução

A definição de ancilaridade parcial, G-ancilaridade, apresentada neste capítulo foi desenvolvida por Godambe [1980] a partir de uma generalização do número de Informação de Fisher.

Como no capítulo 2, exigiremos que θ e ϕ sejam de variação independente, além disso que Θ, o campo de variação de θ, seja um intervalo real.

<u>DEFINIÇÃO 4.1</u> - A estatística t é G-ancilar com respeito a θ se satisfizer as condições

a) A distribuição condicional de x dado t depende somente de θ, $p(x/\theta,\phi)=p(t/\theta,\phi).p(x/t,\theta)$.

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

$$P_{t,\theta} = \{p(t/\theta,\phi),\ \phi \in \Phi\} \text{ é completa.}$$

A condição (a) se refere claramente à condição de simplicidade do Princípio C, garantindo que inferências sobre θ podem ser feitas a partir do modêlo condicional $p(x/t,\theta)$, sem problemas. Resta saber se as condições (a) e (b) garantem que o modelo marginal $p(t/\theta,\phi)$ é não informativo com respeito a θ, e portanto pode ser descartado da análise.

Para descrever melhor esta questão precisaremos, primeiramente, trabalhar com a situação em que o modelo está livre de parâmetros nuisance.

4.2 - Estatística Ancilar para o Modelo Livre de Parâmetros Nuisance

Considere o modelo (X,A,P) onde P é a família de distribuições $\{p(x/\theta),\ \theta \in \Theta\}$ e θ é o parâmetro de interesse.

O principal resultado desta seção mostra que a distribuição condicional $p(x/t,\theta)$ e a distribuição $p(x/\theta)$ contém a mesma "informação sobre o parâmetro θ" se e somente se a estatística t for ancilar com respeito a θ. O sentido de "informação sobre o parâmetro θ" aqui adotado é uma generalização do proposto por Fisher, o qual é definido como:

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \ell n\ p(x/\theta)\right]^2 = - E\left[\frac{\partial^2}{\partial\theta^2} \ell n\ p(x/\theta)\right]$$

A ressalva que pode ser feita às conclusões desse capítulo diz respeito evidentemente a se tomar $I(\theta)$ como a informação sobre θ contida na distribuição $p(x/\theta)$. Entretanto nem a justificativa nem a crítica deste procedimento serão aqui tratados.

O enfoque utilizado por Godambe [1960] no desenvolvimento de sua teoria é mais geral que o enfoque usual, pois estuda funções de estimação em lugar de estimadores.

As funções de estimação são funções reais, $g(x,\theta)$, dos dados x e do parâmetro θ, tal que o estimador de θ é obtido resolvendo-se a equação $g(x,\hat{\theta})=0$ em $\hat{\theta}$. Uma função g é não tendenciosa se sua média é zero para todo valor θ, isto é, $E(g/\theta)=0$. Note que esta definição tem a vantagem de ser invariante sob transformações 1-1 de θ.

Na estimação de θ, considere as funções de estimação $g(x,\theta)$, $g: X \times \Theta \longrightarrow R$ que satisfaçam às condições:

(i) $E\left[g(x,\theta)\,|\,\theta\right] = 0, \quad \forall\theta \in \Theta$,

(ii) $\frac{\partial g}{\partial\theta}$ existe para todo $\theta \in \Theta$, q.t.p ,

(iii) $\int g(x,\theta)p(x/\theta)dx$ é diferenciável sob o sinal de integração e

(iv) $E\left[\left(\frac{\partial g}{\partial\theta}\right)\Big|\,\theta\right] > 0$ para todo $\theta \in \Theta$.

Qualquer função satisfazendo as condições (i)-(iv) é chamada de função regular de estimação. Denotaremos por G a classe das funções regulares de estimação.

O procedimento para estimar θ é encontrar $g^* \in G$, "ótima" segundo algum critério e resolver a equação $g^*(x,\hat{\theta})=0$ em $\hat{\theta}$.

Suponha que a família $\mathcal{P}$ das distribuições satisfaça as condições usuais de regularidade (Godambe [1960]).

DEFINIÇÃO 4.2 (Godambe [1980]) - A informação $I(\theta)$ da distribuição $p(x/\theta)$ é dada por

$$I(\theta) = \max_{g \in G} \left\{ 1/E \left[g/E \left(\frac{\partial g}{\partial \theta} \middle| \theta \right) \middle| \theta \right]^2 \right\}$$

O teorema abaixo mostra que $I(\theta)$ definida acima é exatamente a Informação de Fisher.

TEOREMA 4.1 - (Godambe [1960]) A função $g^* = \frac{\partial}{\partial \theta} \ln p(x/\theta)$ satisfaz a desigualdade

$$E \left[g^*/E\left(\frac{\partial g^*}{\partial \theta} \middle| \theta \right) \middle| \theta \right]^2 \leq E \left[g/E \left(\frac{\partial g}{\partial \theta} \middle| \theta \right) \middle| \theta \right]^2, \quad \forall g \in G.$$

DEMONSTRAÇÃO - Não é difícil ver que $g^* \in G$. Substituindo-se g^* por $\frac{\partial}{\partial \theta} \ln p(x/\theta)$ e utilizando-se a igualdade

$$E\left[\frac{\partial}{\partial \theta} \ln p(x/\theta) \right]^2 = -E\left[\frac{\partial^2}{\partial \theta^2} \ln p(x/\theta) \right]$$ tem-se que

$$E\left[g^*/E(\partial g^*|\theta)|\theta\right]^2 = 1/E\left[\left(\frac{\partial \ln P}{\partial \theta} \right)^2 |\theta\right] = \frac{1}{I(\theta)}$$

Portanto basta demonstrar que:

$$\frac{E(g^2|\theta)}{\left[E(\frac{\partial g}{\partial \theta}|\theta)\right]^2} \geq 1/E\left[\left(\frac{\partial \ln P}{\partial \theta} \right)^2 |\theta\right], \ \forall g \in G$$

Mas

$$0 = \int \frac{\partial}{\partial \theta} (g(x,\theta).p(x/\theta))dx = \qquad \text{(i e iii)}$$

$$= E\left[\frac{\partial g}{\partial \theta} |\theta\right] + E\left[\left(\frac{\partial \ln p}{\partial \theta} \right) g|\theta\right]$$

Portanto $E^2\left[\frac{\partial g}{\partial \theta} |\theta\right] = E^2\left[\left(\frac{\partial \ln p}{\partial \theta} \right) g|\theta\right]$ e

$$E\left[\left(\frac{\partial \ln p}{\partial \theta} \right).g|\theta\right]^2 \leq E(g^2|\theta)E\left[\left(\frac{\partial \ln p}{\partial \theta} \right)^2 |\theta\right] \qquad \text{(Desig.Cauchy-Schwarz)}$$

Das desigualdades acima vem que

$$\left[E\left(\frac{\partial g}{\partial \theta} |\theta\right)\right]^2 \leq E(g^2|\theta).\ E\left[\left(\frac{\partial \ln p}{\partial \theta} \right)^2 |\theta\right]$$

c.q.d.

Algumas considerações sob a teorema acima se fazem necessárias. Primeiramente, a definição 4.2 é exatamente a definição do número de Informação de Fisher, pois

$$\max_{g \in G} \left\{ 1/E\left[g/E\left(\frac{\partial g}{\partial \theta} \,|\theta\right) |\theta \right]^2 \right\} = 1/E\left[g^*/E\left(\frac{\partial g^*}{\partial \theta} |\theta\right) |\theta \right]^2$$

$$= E\left[\left(\frac{\partial \ell np}{\partial \theta} \right)^2 |\theta \right] = I(\theta)$$

Em segundo lugar, a função de estimação "ótima" é a função de estimação da verossimilhança, ou seja, $\frac{\partial \ell nP(x/\theta)}{\partial \theta}$. E em terceiro lugar, o teorema acima nada mais é, que uma generalização da Desigualdade de Cramer-Rao.

Vamos supor que os modelos marginal P_t e condicional $P(./t)$ satisfaçam as condições de regularidade. Neste caso denota-se $I_t(\theta)$ a informação sobre θ contida na distribuição condicional ~~margi-~~ de x dado ~~nal~~ de t por,

$$I_t(\theta) = \max_{g \in G_t} \left\{ 1/E \left[g/E \left(\frac{\partial g}{\partial \theta} \Big|\, t,\theta \right) \Big| t,\theta \right] \right\}$$

onde G_t é a classe das funções de estimação que satisfazem as condições (i) - (iv) para a família $P(./t,\theta)$. Observe que $I_t(\theta)$ depende de t.

De maneira análoga pode-se definir $I^t(\theta)$ como a informação contida na distribuição marginal de $t, p(t/\theta)$.

A Informação $I(\theta)$ pode ser decomposta em duas partes: uma que depende da informação contida na distribuição condicional $p(x/t,\theta)$ e outra dependente da distribuição marginal $p(t/\theta)$. Isto é traduzido pelo resultado abaixo.

TEOREMA 4.2 - $I(\theta) = E[I_t(\theta)|\theta] + I^t(\theta)$

DEMONSTRAÇÃO -

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

$$= -E\left\{E\left[\left(\frac{\partial^2}{\partial\theta^2} \ln p(x/t,\theta).p(t/\theta)\right)|t\right]\right\}$$

$$= -E\left\{E\left[\frac{\partial^2}{\partial\theta^2} \ln p(x/t,\theta)|t\right]\right\} - E\left\{E\left[\frac{\partial^2}{\partial\theta^2} \ln p(t/\theta)|t\right]\right\}$$

$$= E\left[I_t(\theta)|\theta\right] + I^t(\theta)$$

DEFINIÇÃO 4.3 - A distribuição condicional $p(x/t,\theta)$ contém a mesma informação sobre θ que a distribuição $p(x/\theta)$ se

$$I(\theta) = E[I_t(\theta)|\theta].$$

TEOREMA 4.3 - $E[I_t(\theta)|\theta] = I(\theta)$ se e somente se t for uma estatística ancilar c.r.[1] a θ.

DEMONSTRAÇÃO:- Se t é ancilar c.r. a θ a demonstração é imediata uma vez que a distribuição marginal de t não depende de θ e portanto

$$I^t(\theta) = 0$$

Vamos mostrar então que se $I^t(\theta) = 0$ então t é ancilar c.r. a θ.

$$I^t(\theta) = \int\left(\frac{\partial}{\partial\theta} \ln p(t/\theta)\right)^2.p(t/\theta)dt = 0$$

(1) c.r. = com respeito

portanto como o integrando é positivo

$$\left(\frac{\partial}{\partial \Theta} \ell n \, p(t/\Theta)\right)^2 p(t/\Theta) = 0 \quad , \text{ q.t.p.}$$

$$\Rightarrow \frac{\partial}{\partial \Theta} \ell n \, p(t/\Theta) = \frac{1}{p(t/\Theta)} \frac{\partial}{\partial \Theta} p(t/\Theta) = 0 \quad \text{q.t.p.}$$

$$\Rightarrow \frac{\partial}{\partial \Theta} p(t/\Theta) = 0 \quad \text{q.t.p.}$$

$\therefore$ $p(t/\Theta)$ independe de Θ, q.t.p. e t é uma estatística ancilar c.r. a Θ.

c.q.d.

Aceitamos no capítulo 1 que se t é uma estatística ancilar com respeito a Θ então as distribuições $p(x/\Theta)$ e $p(x/t,\Theta)$ contém a mesma informação sobre Θ, uma vez que elas se diferenciam por uma constante que não depende de Θ. Portanto, pelo Teorema 4.3 a definição 4.3 é coerente, segundo nosso ponto de vista. Assim se $I(\Theta) = E(I_t(\Theta)|\Theta)$, $p(x/\Theta)$ e $p(x/t,\Theta)$ contém a mesma informação sobre Θ.

A partir dessas considerações será feito a generalização para o modelo envolvendo parâmetros nuisance.

4.3 - Estatística G-Ancilar

Considere a família $\mathcal{P} = \{P_\lambda, \lambda \in \Lambda\}$ de distribuições e $\Theta(\lambda)$, $\phi(\lambda)$ uma reparametrização variação independente. Estamos interessados apenas na Informação sobre Θ.

O método que vamos descrever abaixo, supõe que a família de distribuições P e que a classe das funções de estimação $g(x,\Theta)$ satisfaçam generalizações das condições de regularidade (Godambe [1976]) apresentadas na seção anterior. Para que a explanação não se interrompa pom a enumeração destas condições, elas poderão ser encontradas, juntamente com as demonstrações de alguns teoremas, no apêndice.

DEFINIÇÃO 4.4 - A informação $i(\Theta)^{(1)}$ sobre θ (ignorando ϕ) na distribuição $p(x/\lambda)$ é dada por

$$i(\Theta) = \max_{g \in G_1} \left\{ 1/E \left[g/E \left(\frac{\partial g}{\partial \Theta} \middle| \lambda \right) \middle| \lambda \right]^2 \right\}$$

onde G_1 é a classe das funções de estimação de Θ que satisfazem as condições de regularidade (apêndice).

Da mesma maneira que na seção anterior, definiremos aqui a informação $i_t(\Theta)$, sobre Θ (ignorando ϕ) contida na distribuição condicional de x dado t. Portanto

$$i_t(\Theta) = \max_{G_{1t}} \left\{ 1/E \left[g/E \left(\frac{\partial g}{\partial \Theta} \middle| \lambda , t \right) \middle| t, \lambda \right]^2 \right\}$$

onde G_{1t} é o conjunto das funções de estimação de Θ que satisfazem as condições de regularidade (apêndice) para $p(x/t,\lambda)$.

---

(1) O símbolo i (minúscula) indica que estamos tratando de Informação sobre um subparâmetro do modelo.

A definição abaixo tem como inspiração a definição 4.3 apresentada na seção anterior.

DEFINIÇÃO 4.5 - A distribuição condicional $p(x/t,\lambda)$ é dita conter a mesma informação sobre Θ que a distribuição dos dados $p(x/\lambda)$ se $i(\Theta) = E\left[i_t(\Theta)|\lambda\right]$.

Uma vez aceita a definição acima como medida de igualdade de informação com respeito a Θ, dos modelos $p(x/\lambda)$ e $p(x/t,\lambda)$, a estatística t ancilar (segundo Godambe) deve ser tal que a definição seja satisfeita. O teorema abaixo expressa exatamente esta idéia.

TEOREMA 4.4 - Se t é uma estatística G-ancilar com respeito a Θ (definição 4.1) então

$$i(\Theta) = E[i_t(\Theta)|\lambda]$$

Para a demonstração deste teorema, faremos uso do Lema a seguir. As demonstrações do Teorema e do Lema serão apresentadas no apêndice.

LEMA - Se t é uma estatística G-ancilar com respeito a Θ(definição 4.1) então:

$$\max_{g \in G_1}\left\{ 1/E\left[g/E\left(\frac{\partial g}{\partial \Theta}|\lambda\right)|\lambda\right]^2\right\} \text{ é atingido}$$

para $g^* = \dfrac{\partial \ell n p(x/t,\Theta)}{\partial \Theta}$, e é a única função tal que isto aconteça.

Os resultados acima exigem certas considerações. A definição de G-ancilaridade não exige que o modelo marginal $p(t/\theta,\phi)$, que é abandonado quando se considera inferências sobre θ, independa de θ. Entretanto, se t é G-ancilar com respeito a θ a função de estimação ótima é dada pela função de verossimilhança do modelo condicional $p(x/t,\theta)$. Portanto, segundo Godambe, o modelo $p(t/\theta,\phi)$, apesar de depender de θ deve conter informação zero sobre θ.

Algumas objeções podem ser feitas a este resultado. Uma delas, citada anteriormente, diz respeito a se tomar Informação de Fisher como conceito de informação contida em uma distribuição sobre o parâmetro θ. Outra diz respeito a definição 4.4. Esta definição se inspira na definição 4.3 da seção anterior, a qual foi "provada" ser coerente pelo Teorema 4.3. O problema com a definição 4.4 parece ser o fato de que primeiramente aceita-se como coerente esta definição e depois encontra-se a estatística t que a satisfaça. O caminho utilizado aqui foi o inverso da seção anterior. Portanto, deve-se aceitar com precaução os resultados acima.

Uma generalização do Teorema de Lehman-Scheffé pode ser considerada para a definição de G-Ancilaridade.

Note que para $\theta_0 \in \Theta$ fixado, mas arbitrário, a condição (a) da definição 4.1;

$$p(x/\theta_0,\phi) = p(t/\theta_0,\phi)\cdot p(x/t,\theta_0),$$

pode ser considerada como a estatística t ser suficiente na família $P_{\theta_0} = \{p(x/\theta_0,\phi)\ ;\phi \in \Phi\}$, $\forall\theta_0 \in \Theta$. A condição (b) impõe que a família $P_{t,\Theta_o} = \{p(t/\Theta_o,\phi)\ ,\phi \in \Phi\}$ é completa. Logo estamos nas condições do Teorema de Lehman-Scheffé e t é uma estatística suficiente minimal para a família

$$P_{\Theta_o} = \{p(x/\Theta_o,\phi),\phi \in \Phi\},\quad \Theta_o \in \Theta \text{ e é única}$$

<u>TEOREMA 4.5</u> (Lehmann-Scheffé generalizado) - Se a estatística t é G-ancilar com respeito a Θ então para $\Theta_o \in \Theta$, fixado e arbitrário t é suficiente minimal para a família

$$P_{\Theta_o} = \{p(x/\Theta_o,\phi)\ ,\phi \in \Phi\} \quad \text{e é única.}$$

O Teorema acima tem uma interpretação interessante;a estatística G-suficiente é função de toda estatística t' para a qual vale a condição $p(x/\Theta,\phi)=p(x/t',\Theta).p(t'/\Theta,\phi)$. Observe que a condição acima explicita o objetivo maior do estatístico que é encontrar uma parte do modelo que dependa apenas de Θ. Portanto, se o estatístico concorda em usar o modelo reduzido $p(x/t',\Theta)$ nas inferências sobre Θ, está descartando a outra parte do modelo $p(t'/\Theta,\phi)$ a qual também depende de Θ. Neste caso é melhor que "minta o menos possível" e considere a estatística G-ancilar $t=t(t')$ como conhecida (em lugar de t'), e consequentemente utilize o modelo $p(x/t,\Theta)$, (em lugar de $p(x/t',\Theta)$).

<u>EXEMPLO 3.2</u> -(continuação). Teste Exato de Fisher - Mostrou-se na seção 3.2 que a distribuição dos dados pode ser decomposta em $p(a,t,n_1/\Theta,\phi,\Psi) = p(a/t,n_1,\Theta).p(t,n_1/\phi,\Psi,\Theta)$.

O objetivo aqui é justificar a construção do Teste E xato de Fisher que utiliza apenas o modelo condicional $p(a/t,n_1,\theta)$ e abandona $p(t,n_1/\phi,\Psi,\theta)$, que depende de θ. Isto é, como garantir que não estamos perdendo informação sobre θ?

Observe que a condição (a) da definição de G-ancilaridade é satisfeita pelas estatísticas marginais (t,n_1). Resta saber se para $\theta\in\Theta$ fixado a família de distribuições $\mathcal{P}_\theta = \{p(t,n_1/\phi,\Psi,\theta),\ 0<\phi<1,\ 0<\Psi<1\}$ é completa.

$$p(t,n_1/\theta,\phi,\Psi) = \frac{\phi^t(1-\phi)^{n-t}}{(1-\phi+\theta\phi)^{n_1}}\ \Psi^{n_1}(1-\Psi)^{n-n_1}\binom{n}{n_1}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k,$$

$$t\leq n_1$$

É necessário mostrar que se

$$E[u(t,n_1)] = 0 \quad \Rightarrow \quad u(t,n_1)\equiv 0$$

$$E[u(t,n_1)] = \sum_{n_1=0}^{n}\sum_{t=0}^{n_1} u(t,n_1)\frac{\phi^t(1-\phi)^{n-t}}{(1-\phi+\theta\phi)^{n_1}}\ \Psi^{n_1}(1-\Psi)^{n-n_1}\binom{n}{n_1}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k =$$

$$= \sum_{n_1=0}^{n}\left\{\frac{\Psi^{n_1}(1-\Psi)^{n-n_1}\binom{n}{n_1}}{(1-\phi+\theta\phi)^{n_1}}\left[\sum_{t=0}^{n_1} u(t,n_1)\phi^t(1-\phi)^{n-t}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k\right]\right\}=0$$

Para θ e ϕ fixados, a equação acima pode ser encarada como a esperança da função

$$h(n_1) = \frac{1}{(1-\phi+\theta\phi)^{n_1}}\left[\sum_{t=0}^{n_1} u(t,n_1)\phi^t(1-\phi)^{n-t}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k\right]$$

Pela completividade da distribuição Binomial tem-se que $h(n_1)=0$, $n_1=0,\dots,n$.

Seja então n_1 fixado, arbitrario tem-se então que

$$\frac{1}{(1-\phi+\theta\phi)^{n_1}}\left[\sum_{t=0}^{n_1} u(t,n_1)\phi^t(1-\phi)^{n-t}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k\right]=0$$

Note que $1+\phi(\theta-1)>0$, portanto podemos analisar apenas o somatório.

$$\therefore \sum_{t=0}^{n_1} u(t,n_1)\phi^t(1-\phi)^{n-t}\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k=0$$

Novamente pela comptividade da distribuição Binomial,

$$u(t,n_1)\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k=0 \qquad t\leq n_1 .$$

Como $\sum_{k=0}^{t}\binom{n_1}{k}\binom{n-n_1}{t-k}\theta^k>0$ tem-se que

$$u(t,n_1)=0 \text{ para } n_1=0,\dots,n \text{ e } t=0,\dots,n_1$$

Logo a família $P_\theta=\{p(t,n_1/\phi,\Psi,\theta);\ 0<\phi<1,\ 0<\Psi<1\}$ é completa e a estatística (t,n_1) é G-ancilar com respeito a θ.

Parece então, que com base na G-ancilaridade, o teste exato de Fisher tem um suporte lógico.

CAPÍTULO 5

H-SUFICIÊNCIA

As definições de suficiência e ancilaridade se baseiam, ambas, no Princípio de verossimilhança. No entanto, a definição de suficiência é mais aceita e conhecida pela comunidade estatística que a definição de ancilaridade. Esta maior credibilidade se justifica em parte pelo sentido estatístico de se trabalhar com estatísticas, mas também devido a uma crença de que o sentido da suficiência é provado matematicamente pelo Teorema de Rao-Blackwell. Este Teorema mostra basicamente que o melhor estimador de θ deve ser função da estatística suficiente com respeito a θ.

<u>TEOREMA 5.1</u> (Rao-Blackwell) - Seja t uma estatística suficiente com respeito a θ e s um estimador para $a(\theta)$ então $t^*=E[s/t]$ é tal que

$$E\,[\ell(t^*,\theta)/\theta] \leq E[\ell(s,\theta)/\theta]$$

onde $\ell(y,\theta)$ é uma função de perda convexa em $a(\theta)$.

Este resultado será generalizado para a situação em que o modelo envolve parâmetros nuisance.

A quantidade que se deseja estimar é $a(\theta)$ e o modelo é dado por

$$\mathcal{P} = \{p(x/\theta,\phi) \quad , \quad \theta \in \Theta \quad , \quad \phi \in \Phi \ \} \ .$$

A função de perda ao se estimar $a(\theta)$ por u é $\ell(\theta,y)$ que é uma função convexa em $a(\theta)$, $\theta \in \Theta$.

Seja U a classe dos estimadores u de $a(\theta)$ tal que a função de risco,

$$r_u(\theta,\phi) = E\ [\ell(\theta,u)/\theta,\phi] = r_u(\theta),$$

depende de (θ,ϕ) somente através de θ.

Portanto, $U = \{u \text{ t.q. } r_u(\theta,\phi) = r_u(\theta)\}$.

Neste caso temos o seguinte teorema.

<u>TEOREMA 5.2</u> (Fraser) - Se t é uma estatística S-suficiente com respeito a θ, então para qualquer $u \in U$ existe uma função de t, $u_o = u_o(t)$ tal que

$$r_{u_o}(\theta) \leq r_u(\theta).$$

<u>DEMONSTRAÇÃO</u> - Seja $\phi_o \in \phi$ fixado.

Define-se $u_o(t) = E(u/t,\theta,\phi_o)$, que está bem definido como estimador, uma vez que a distribuição condicional de x dado t independe de Θ.

$$r_u(\Theta) = E[\ell(u,\Theta)/\Theta,\phi_o] =$$

$$=E\{E[\ell(u,\Theta)/t,\Theta,\phi_o]\} \geq E\{\ell[E(u/t,\Theta,\phi_o),\Theta]\} = \text{[Desigualdade de Jensens]}$$

$$= E[\ell(u_o(t),\Theta)] \qquad \forall \Theta \in \Theta$$

$$u_o(t) \in U \text{ portanto } E[\ell(u_o(t),\Theta)] = r_{u_o}(\Theta)$$

c.q.d.

Este teorema pode ser generalizado expandindo-se a classe U dos estimadores da seguinte maneira. Seja a classe U' dos estimadores u tal que a função de risco $r_u(\Theta,\phi)$ está bem definida mas não é necessariamente livre de ϕ;

$$U' = \{ u \text{ tal que } r_u(\Theta,\phi) < \infty \}$$

Utilizando-se o princípio minimax para eliminação do parâmetro ϕ, define-se:

$R_u(\Theta) = \sup\limits_{\phi \in \Phi} r_u(\Theta,\phi)$ como a função de risco "eliminado ϕ" associada a u.

<u>TEOREMA 5.3</u> (Hájek) - se t é S-suficiente para Θ então $\forall$ $u \in U'$ existe $u_o = u_o(t)$ tal que $R_{u_o}(\Theta) \leq R_u(\Theta)$ para todo $\Theta \in \Theta$.

DEMONSTRAÇÃO - Seja $\phi_o \in \phi$ fixado.

$$u_o(t) = E[u/t,\theta,\phi_o] \quad \text{independe de } \theta$$

$$R_u(\theta) \geq r_u(\theta,\phi) = E\,[\ell(u,\theta)/\theta,\phi_o]$$

$$= E\{E\,[\ell(u,\theta)/t,\theta,\phi_o]\} \geq \text{[Desigualdade de Jensen]}$$

$$\geq E\,\{\ell[E(u/t,\theta,\phi_o),\theta]\}.$$

$$= r_{u_o}(\theta,\phi_o) = R_{u_o}(\theta)$$ uma vez que u_o é função de t e a distribuição marginal de t depende apenas de θ.

s.q.d.

O Teorema é uma generalização do Teorema de Fraser . Note que se $u \in U$, $r_u(\theta) = R_u(\theta)$.

A caracterização de S-suficiência dada pelos teoremas acima é semelhante a caracterização de suficiência dada pelo Teorema de Rao-Blackwell. Dado qualquer estimador de θ é possível se encontrar um novo estimador, função de estatística S-suficiente com respeito a θ, com menor risco médio.

O caminho utilizado aqui é semelhante ao que foi utilizado por Godambe. Sabe-se que se t é suficiente [FISHER] com respeito a θ então t satisfaz o Teorema de Rao-Blackwell. No caso em que o modelo envolve parâmetros nuisance, procura-se a estatística que satisfaça aos teoremas 5.2 e 5.3 e esta estatística será denominada parcialmente suficiente com respeito a θ no sentido de Hájek.

Pelas demonstrações dos teoremas pode-se observar que a condição de S-suficiência é mais forte que o necessário. As demonstrações exigem que a distribuição marginal de t, $p(t/\theta)$, dependa apenas de θ. Quanto à distribuição condicional, $p(x/t, \theta,\phi)$, exige-se apenas que para $\phi_o \in \Phi$ independa de θ. Esta observação sugere uma generalização da noção de S-suficiência, uma vez que é possível encontrar uma estatística t que satisfaça condições mais fracas e para a qual os teoremas são satisfeitos.

Para cada $\theta \in \Theta$, seja o conjunto $\Omega_\theta = \{\lambda : \theta(\lambda) = \theta\}$ dos pontos de λ que geram o mesmo θ.

Seja $\bar{P}_\theta$ o "envoltório convexo" da família $P_\theta = \{p_\lambda, \lambda \in \Omega_\theta\}$. Isto é, $\bar{P}_\theta$ é o conjunto das medidas Q_θ da forma:

$$Q_\theta(A) = \int_{\Omega_\theta} p_\lambda(A)\, dq_\theta(\lambda), \quad \text{para todo } A \in \mathcal{A},$$

onde q_θ é alguma medida de probabilidade em Ω_θ.

Segue a definição de H-suficiência.

<u>DEFINIÇÃO 5.1</u> - A estatística t é H-suficiente com respeito a θ, se:

(a) A distribuição marginal de t depende apenas de θ, $p(x/\lambda) = p(t/\theta) . p(x/t,\lambda)$.

(b) Se para todo $\theta \in \Theta$, existe uma medida de probabilidade $q_\theta(\lambda)$ em Ω_θ tal que: t é suficiente com respeito a θ no modelo

$$Q = \{Q_\Theta,\ \Theta\in\Theta\} \quad \text{onde} \quad Q_\Theta = \int_{\Omega_\Theta} p(x/\lambda)\ dq_\Theta(\lambda)$$

A definição de H-suficiência é mais geral que S-suficiência e não exige, como na segunda, que Θ e ϕ sejam de variação independente.

Suponha t S-suficiente com respeito a Θ. Para cada $\Theta\in\Theta$ fixado, considere a medida $q_\Theta(\phi)$ que dá valor 1 para $\phi_o\in\Phi$ fixado e arbitrário. Neste caso $Q_\Theta(x) = p(x/\Theta,\phi_o)=p(x/t,\phi_o).p(t/\Theta)$, e t é suficiente com respeito a Θ para a família Q.

Os teoremas 5.2 e 5.3 continuam válidos para a definição de H-suficiência. Consideraremos apenas o teorema mais geral.

Nas mesmas condições que o Teorema 5.3, sendo que $R_u(\Theta)$ pode ser definido como

$R_u(\Theta) = \sup\limits_{\lambda\in\Omega_\Theta} r_u(\lambda)$, dado que Θ e ϕ não necessitam ser de variação independente.

<u>TEOREMA 5.4</u> - Se t é H-suficiente com respeito a Θ, então para qualquer $u\in U'$ existe $u_o=u_o(t)$ tal que $R_{u_o}(\Theta) \leq R_u(\Theta)$.

<u>DEMONSTRAÇÃO</u> - Define-se $u_o(t) = E_{Q_\Theta}[u/t]$ a esperança condicional de u dado t com respeito a família Q. A estatística t é suficiente (Fisher) com respeito a Θ para a família Q, logo $u_o(t)$ está bem definido.

Tomando todas as esperanças com relação a $\{Q_\Theta,\ \Theta\in\Theta\}$, vamos demonstrar primeiramente que

$$\int_X \ell(u,\Theta)\,dQ_\Theta \geq \int_X \ell(u_o,\Theta)\,dQ_\Theta$$

$$\int_X \ell(u,\Theta)\,dQ_\Theta = E_{Q_\Theta}\,[\ell(u,\Theta)] =$$

$$= E_{Q_\Theta}\,\{E_{Q_\Theta}\,[\ell(u,\Theta)/t]\} \geq \text{[Desigualdade de Jensens]}$$

$$\geq E_{Q_\Theta}\,\{\ell[E_{Q_\Theta}(u/t),\Theta]\} = E_{Q_\Theta}[\ell(u_o,\Theta)] = \int_X \ell(u_o,\Theta)\,dQ_\Theta$$

Observe que:

$$\int_X \ell(u_o,\Theta)\,dQ_\Theta = \int_X \ell(u_o,\Theta)\,Q_\Theta(dx) =$$

$$= \int_X \ell(u_o,\Theta) \int_{\Omega_\Theta} p_\lambda(dx)\,dq_\Theta(\lambda) =$$

$$= \int_{\Omega_\Theta} \left\{\int_X \ell(u_o,\Theta)\,p_\lambda(dx)\right\} dq_\Theta(\lambda) =$$

Mas $\int_X \ell(u_o,\Theta)\cdot p_\lambda(dx) = r_{u_o}(\lambda)$

Como u_o é uma função de t e a distribuição de t depende apenas de Θ então;

$$r_{u_o}(\lambda) = \bar{r}_{u_o}(\Theta) \quad \text{para todo } \lambda \in \Omega_\Theta.$$

Então

$$\int_X \ell(u_o,\Theta)\,dQ_\Theta = \int_{\Omega_\Theta} \bar{r}_{u_o}(\Theta)\,dq_\Theta(\lambda) = \bar{r}_{u_o}(\Theta) = R_{u_o}(\Theta)$$

Pelos mesmos argumentos vem que

$$\int_X \ell(u,\Theta)\,dQ_\Theta = \int_{\Omega_\Theta} r_u(\lambda)\,dq_\Theta(\lambda) \leq$$

$$\int_{\Omega_\Theta} \sup_{\lambda\in\Omega_\Theta} r_u(\lambda) dq_\Theta(\lambda) = \sup_{\lambda\in\Omega_\Theta} r_u(\lambda) = R_u(\Theta)$$

Portanto,

$$R_{u_o}(\Theta) = \int \ell(u_o,\Theta) dQ_\Theta \leq \int \ell(u,\Theta) dQ_\Theta \leq R_u(\Theta)$$

c.q.d.

A caracterização de suficiência parcial dada por Hájek pode ser colocada nos seguintes termos: se u é um estimador para Θ consegue-se um estimador função da estatística t, H-suficiente com respeito a Θ, com menor risco que o anterior. Portanto t deve conter toda a informação relevante sobre Θ, e as inferências sobre Θ podem ser feitas utilizando-se apenas o modelo marginal $p(t/\Theta)$.

Novamente o modelo $p(x/t,\Theta,\phi)$ depende de Θ e apesar disso considera-se que ele seja não-informativo com respeito a Θ.

EXEMPLO 3.3 (continuação) Normal Univariada - Sabe-se que a estatística S não é S-suficiente com respeito a σ. Vamos analisar se S é H-suficiente com respeito a σ.

Como já foi dito a condição (a) da definição de H-suficiencia (5.1) está satisfeita uma vez que a distribuição de S depende apenas de σ.

A função de verossimilhança dos dados se fatora como

$$p(x/\mu,\sigma^2) = A(\sigma) \exp\left(-\frac{nS^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)$$

onde $A(\sigma) = (\sqrt{2\pi}\sigma)^{-n}$

Mostrar a condição (b) da definição 5.1 é equivalente a mostrar que para todo $\sigma>0$, existe alguma medida q_σ em $\Omega_\sigma = \{(\mu,\sigma)\ ;\ \mu\in \mathbb{R}\}$ tal que S seja suficiente em Q_σ, com respeito a σ;

$$Q_\sigma(x) = \int_{\Omega_\sigma} p(x/\mu,\sigma^2)dq_\sigma(\mu)$$

$$= A(\sigma)\exp\left[-\frac{nS^2}{2\sigma^2}\right]\int_{-\infty}^{\infty}\exp\left[-\frac{n(\bar{x}-\mu)^2}{\partial\sigma^2}\right]dq_\sigma(\mu)$$

Para que S seja suficiente com respeito a σ, na família Q devemos ter a seguinte fatoração:

$$\int_{-\infty}^{\infty}\exp\left[-\frac{n(\bar{x}-\mu)^2}{\partial\sigma^2}\right]dq_\sigma(\mu) = A(\bar{x}).B(\sigma^2)$$

Pode-se mostrar que se o campo de variação de σ é $[0,\infty)$ não se consegue a fatoração acima para nenhuma medida própria $q_\sigma(\mu)$ em $\mathbb{R}$. Entretanto, quando o campo de variação de σ é limitado superiormente por uma constante K, para a família $q_\sigma = N(0,\sqrt{(K^2-\sigma^2)/n})$, $0<\sigma<K$, a fatoração acima é verdadeira.

Devemos ressaltar que na prática, é natural considerarmos um conjunto limitado para σ^2 e na verdade esta condição não é uma restrição.

<u>TEOREMAS 5.5</u> - Seja $(x_1,\dots,x_n)$ uma a.c.s. da $N(\mu,\sigma^2)$. Considere a estatística $S^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$. Então:

(i) S^2 é H-suficiente com respeito a σ^2, no caso de σ^2 ser limitado, $0\leq\sigma^2<k$.

(ii) S^2 não é H-suficiente com respeito a σ^2, no caso de $0 \leq \sigma^2 < \infty$ (σ^2 ilimitado).

DEMONSTRAÇÃO -

(i) Para demonstrar a condição (i) basta considerar a seguinte medida para μ, $q_\sigma = N(0, \frac{\sqrt{k^2-\sigma^2}}{n})$. Obtém-se que:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right] dq_\sigma(\mu) = \exp\left(-\frac{n}{2}\frac{\bar{x}^2}{k^2}\right).\frac{\sigma}{k} = A(\bar{x})B(\sigma).$$

(ii) Quer-se mostrar que:

$$\int_{\infty} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n(\bar{x}-\mu)^2}{\partial\sigma^2}\right) dq_\sigma(\mu) \neq A(\bar{x})B(\sigma),$$ no caso de σ ser ilimitado.

Observe que isto equivale a mostrar que

$$\int_{-\infty}^{\infty} f(\bar{x}/\mu,\sigma).f(\mu/\sigma)d\mu \neq A(\bar{x}).B(\sigma)$$

onde $f(\bar{x}/\mu,\sigma)$ é a $N(\mu,\sigma^2/n)$.

Isto é, que $\bar{x}$ independe de σ para qualquer $f(\mu/\sigma)$.

A demonstração desta condição pode ser obtida como um caso particular do lema abaixo.

LEMA[1]: Sejam x, y, z variáveis aleatórias tal que

---

(1) A demonstração deste lema é devida à Professora Maria Eulália Vares.

1) A distribuição condicional de x dado y,z é $N(z,y^2)$; $x/(y,z) \sim N(z,y^2)$.

2) $P(y>N)>0, \ \forall N$.

Então x e y não são independentes.

<u>DEMONSTRAÇÃO</u>: Suponhamos por absurdo que x e y sejam independentes.

Para $\forall N$, temos que

$P[-N<x<N] = P[-N<x<N/y>N] =$ (hipótese de absurdo)

$$= \frac{P[-N<x<N,y>N]}{P[y>N]} = \frac{1}{P[y>N]}\int_N^\infty P[-N<x<N/y]\,df(y) =$$

$$= \frac{1}{P[y>N]}\int_N^\infty P[-N<x<N,-\infty<z<\infty/y]\ f(y)dy =$$

$$= \frac{1}{P[y>N]}\int_N^\infty\int_{-\infty}^\infty P(-N<x<N/y,z]\ f(z/y)dz\ f(y)dy =$$

$$= \frac{1}{p[y>N]}\int_N^\infty\int_{-\infty}^\infty P[\frac{-N-z}{y} <\Psi(x)< \frac{N-z}{y}]\ f(z/y)dz\ f(y)dy \leq$$

onde $\Psi(x) \sim N(0,1)$.

$$\leq \frac{1}{P[y>N]}\int_N^\infty\int_{-\infty}^\infty P[-1<\Psi(x)<1]\ f(z/y)dz\ f(y)dy \leq$$

$$\leq P\ [-1<\Psi(x)< 1]< 1$$

$\therefore \ \forall N \ \ P[-N<x<N]<1$ o que é absurdo.

c.q.d.

Aplicando o lema obtém-se que $\bar{x}$ e σ não são independentes e portanto é impossível conseguir-se a fatoração $A(\bar{x})B(\sigma)$.

A seguir analisaremos dois exemplos de H-suficiência (um deles também de G-ancilaridade) para os quais a intuição estatística vai de certa forma contra os resultados obtidos.

<u>EXEMPLO 5.1</u> (Neyman and Scott) - A amostra x consiste de n-pares (x_{i1},x_{i2}), $i=1,2,\ldots,n$ em que $x_{ij} \sim N(\mu_i,\sigma^2)$, $i=1,\ldots,n$ e $j=1,2$ independentes.

O parâmetro do modelo é $(\mu_1,\ldots,\mu_n,\sigma)$ sendo que σ é o parâmetro de interesse e $\mu = (\mu_1,\ldots,\mu_n)$ o parâmetro nuisance (desconhecido).

Considere as estatísticas:

$$S^2 = \sum_{i=1}^{n} (x_{i1}-x_{i2})^2$$

$$\bar{x}_i = \frac{x_{i1}+x_{i2}}{2}, \quad i=1,\ldots,n$$

A função de verossimilhança de x, pode ser exprimida em termos das estatísticas acima,como abaixo;

$$p(x/\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2n} \exp\left(-\frac{S^2}{4\sigma^2}\right)\exp\left(-\sum_{i=1}^{n}\frac{(\bar{x}_i-\mu_i)^2}{\sigma^2}\right).$$

Pela decomposição acima fica evidente que a distribuição marginal de S^2 depende apenas de σ, logo seria interessante marginalizar em S^2 quando σ é o parâmetro de interesse.

Pode-se mostrar que se tomamos $(\mu_1,\ldots,\mu_n)$ como variá veis normais i.i.d. com média zero e variância $(\frac{\mu^2-\sigma^2}{2})$ (σ^2 limitado por n) a estatística S^2 é H-suficiente com respeito a σ.

Resultado similar pode ser obtido utilizando-se G-suficiência.

As estatísticas aqui consideradas são equivalentes as anteriores:

$$S^2 = \sum_{i=1}^{n} (x_{i1}-x_{i2})^2 \quad e$$

$$t = (x_{11}+x_{12}, x_{21}+x_{22}, \ldots, x_{n1}+x_{n2}) =$$

$$= (2\bar{x}_1, 2\bar{x}_2, \ldots, 2\bar{x}_n)$$

Note que t e S^2 são suficientes para o modelo abaixo e são independentes

$$p(x/\mu,\sigma) = (\frac{1}{\sqrt{2\pi}\sigma})^{2n} \exp(-\frac{S^2}{4\sigma^2}) \exp\left[-\frac{(\bar{x}_i-\mu_i)^2}{\sigma^2}\right]$$

A distribuição marginal de t é

$$p(t/\mu,\sigma) = (\frac{1}{2\sigma\sqrt{\pi}})^n \exp\left[-\sum_{i=1}^{n}\frac{(\bar{x}_i-\mu_i)^2}{\sigma^2}\right]$$

Portanto a distribuição condicional $p(x/t,\sigma) \alpha\, p(S^2/\sigma)$ depende apenas de σ e a condição (a) da definição de G-ancilaridade está satisfeita. Não é difícil ver também que para σ fixado a família das distribuições marginais de t é completa. Logo, t é uma estatística G-ancilar com respeito a σ e a inferência sobre σ é feita utilizando-se apenas o modelo condicional p(x/t,

σ) ou seja o modelo de S^2, $p(S^2/\sigma)$.

Quando se marginaliza em S^2 abandona-se $(\bar{x}_1,\ldots,\bar{x}_n)$ e de acordo com H-suficiência e G-ancilaridade não se perde informação sobre σ. Contudo a distribuição de $(\bar{x}_1,\ldots,\bar{x}_n)$ depende de σ e espera-se que contenha informação sobre σ. Este fato é reforçado quando se considera que para $(\bar{x}_1,\ldots,\bar{x}_n)$ próximas espera-se σ pequeno e para $(\bar{x}_1,\ldots,\bar{x}_n)$ afastadas espera-se σ grande. A pergunta que fica é de quão seguro se está em marginalizar para S^2 sem antes considerar-se $(\bar{x}_1,\ldots,\bar{x}_n)$?

O exemplo a seguir de Basu[1977] realça com maior clareza esse problema com H-suficiência.

<u>EXEMPLO 5.2</u> - Seja $\underset{\sim}{x} = (x_1,\ldots,x_m,y_1,\ldots,y_n)$ m+n variáveis aleatórias independentes com distribuição normal com variância unitária e média,

$$E(x_i) = \theta \quad i=1,\ldots,n$$

$$E(y_i) = \theta\phi \quad j=1,\ldots,n, \text{ onde } -\infty<\theta<\infty \text{ e } \phi=0 \text{ ou } 1.$$

O parâmetro de interesse é θ.

A função de verossimilhança pode ser escrita como:

$$p(x/\theta,\phi) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2}\left(\Sigma x_i^2 - m\bar{x}^2 + \Sigma y_i^2 - n\bar{y}^2\right) - \frac{m(\bar{x}-\theta)^2}{2} - \frac{n(\bar{y}-\theta\phi)^2}{2}\right]$$

$$= A(\underset{\sim}{x}) \cdot \exp\left[-m\,\frac{(\bar{x}-\theta)^2}{2}\right] \exp\left[-n\,\frac{(\bar{y}-\theta\phi)^2}{2}\right]$$

A estatística $(\bar{x},\bar{y})$ é suficiente com respeito a (θ,ϕ), portanto faremos uma primeira redução, isto é, trabalharemos apenas com o modelo de $(\bar{x},\bar{y})$. Além disso, a distribuição de $\bar{x}$ depende apenas de θ e se $\phi=0$ $\bar{x}$ é suficiente com respeito a θ. Basta considerar, para θ fixado, $q_\theta(\phi) = \begin{cases} 1 & \text{para } \phi=0 \\ 0 & \text{c.c.} \end{cases}$

Logo $\bar{x}$ é H-suficiente com respeito a θ.

Observe que a marginalização em $\bar{x}$ é perigosa, uma vez que a estatística $(\bar{x},\bar{y})$ é muito informativa com respeito a ϕ e se percebe que se ϕ fosse 1 perder-se-ia muito informação sobre θ quando $\bar{y}$ é abandonada.

CAPÍTULO 6

I-SUFICIÊNCIA[1]

A definição de I-suficiência tem seu principal fundamento no Princípio da Invariância. A fim de se introduzir esta definição analisaremos o problema da suficiência parcial de S com respeito a σ, segundo o ponto de vista de Barnard [1963].

A definição de suficiência (Fisher) exige que a distribuição de qualquer função das observações, condicionada a um valor fixado da estatística suficiente, seja independente do parâmetro em questão, e não há dúvidas de que, para esta definição, S não é suficiente com respeito a σ.

---

(1) A notação utilizada aqui não é usual. O símbolo I vem de Invariância. Hall, Wijsman & Ghosh [1965] denotam Suficiência Invariante ou estatística invariantemente suficiente. Já Dawid [1981] denota G-suficiência, por se trabalhar com Grupo de transformações.

Entretanto, segundo Barnard, a definição de suficiência (Fisher) destinava-se a incorporar uma noção lógica, a de prover o conjunto de toda a informação relevante, disponível, para um dado parâmetro. Como a disponibilidade ou não de infornação depende criticamente do conhecimento ou falta de conhecimento, a definição acima não é bem sucedida com respeito a seu objetivo. Pois, se σ já é conhecido, S não nos fornece nenhuma informação. Por outro lado, lembra Barnard, o fracasso de S em satisfazer a definição de suficiência, vem do fato de que a distribuição de $\bar{x}-\mu$ depende também de σ ; isso implica que se μ fosse conhecido, $\bar{x}-\mu$ poderia ser usado para nos dar informação sobre σ. Entretanto, μ é dado como desconhecido de modo que a informação em $\bar{x}-\mu$ não é disponível.

Em seguida, Barnard propõe que se expresse, sob forma matemática precisa, as noções correspondentes a "conhecido" e "desconhecido". E conclui que, no caso de parâmetros de locação e escala, as noções acima poderiam ser expressas em termos de propriedades de invariância para grupos, em relação aos quais os problemas considerados são invariantes.

Concretamente, em termos do exemplo da $N(\mu,\sigma^2)$, as idéias de Barnard se traduzem em considerar transformação dos dados correspondente a translação da origem, isto é, $g_a(x_1,\ldots,x_n) = (x_1+a,\ldots,x_n+a)$. Com esta transformação a estatística suficiente $(\bar{x},s)$ se torna $(\bar{x}+a,s)$ enquanto que (μ,σ) se transforma em $(\mu+a,\sigma)$. Observe também que σ, o parâmetro de interesse, permanece inalterado (invariante) e se μ é desconhecido no primeiro modelo $\mu+a$ também o é no modêlo transformado. Além disso o modêlo é o mesmo nas duas situações. Portanto, o problema de estimação de σ permanece invariante sob estas transformações e neste caso deve-se considerar apenas estimadores invariantes para σ, ou seja, que satisfaçam $t(x_1,\ldots,x_n)=t(x_1+a,\ldots,x_n+a)$.

O ponto principal do argumento de Barnard, parece ser o de restringir a classe dos estimadores de σ para uma classe de estimadores invariantes com relação às transformações de locação. Esta restrição se baseia na idéia de que se não há mudança relevante com relação ao problema de estimação de σ, dois estatísticos, um trabalhando com a amostra original e outro com a amostra transformada, devam produzir a mesma inferência sobre σ.

Antes de se formalizar as idéias acima, é interessante notar que o que está por trás da argumentação é o Princípio de Invariância.

Princípio de Invariância: Se $E_1=(\mathcal{X}_1,\Lambda,P_1)$ e $E_2=(\mathcal{X}_2,\Lambda,P_2)$ são dois experimentos tais que existe uma transformação 1-1 $g:\mathcal{X}_1\to\mathcal{X}_2$ tal que $P_1(x_1/\lambda)=P_2(g(x_1)/\lambda)$, $\forall\ x_1\in\mathcal{X}_1$, $\forall\lambda\in\Lambda$. Então (E_1,x_1) e $(E_2,g(x_1))$ contém a mesma informação.

Fica evidente que os experimentos E_1 e E_2 são idênticos em todos os aspectos com exceção da maneira de se rotular os pontos. Isso implica que a mesma inferência deve resultar de (E_1,x_1) e $(E_2,g(x_1))$.

A teoria que descrevemos a seguir exige a consideração de um grupo de transformações.

DEFINIÇÃO 6.1 - Um grupo G é um conjunto de transformações $g: X \to X$, 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 $g^{-1} \in G$

(iii) A transformação identidade I é um elemento de G, isto é, $I \in G$

Note que a condição (iii) é decorrente de (i) e (ii).

DEFINIÇÃO 6.2 - O modelo P é invariante com respeito ao grupo de transformações G se para todo $g \in G$ e $\lambda \in \Lambda$ existe um único $\lambda^* = \bar{g}(\lambda)$, $\lambda^* \in \Lambda$ tal que se $x \sim P_\lambda$ então $g(x) \sim P_{\lambda^*}$.

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

A definição acima fornece uma condição próxima da condição do princípio de invariância. As duas condições seriam exatamente as mesmas se se exigisse na definição 6.2, que G fosse composto apenas pela função identidade. Desta maneira se $x \sim P_\lambda$ então $g(x) \sim P_\lambda$.

Observe que o modelo P é invariante se o modelo dos dados transformados continua sendo P.

DEFINIÇÃO 6.3 - Uma órbita de X com respeito ao grupo G e ao ponto $x_o \in X$ é o conjunto

$\mathcal{X}(x_o) = \{g(x_o),\ \forall g \in G\}$

<u>DEFINIÇÃO 6.4</u> - Uma função $t(x)$ é

(i) invariante com respeito a G se $t(g(x))=t(x)$, $\forall\ x \in \mathcal{X}$, $\forall\ g \in G$.

(ii) maximal invariante com respeito a G se é invariante e se $t(x_1)=t(x_2) \rightarrow x_1=g(x_2)$ para algum $g \in G$.

A definição 6.4 pode ser expressa em termos de órbitas. A estatística t é invariante com respeito a G se for constante nas órbitas de $\mathcal{X}$; é maximal se além de ser constante, assume diferentes valores para diferentes órbitas.

No caso do modêlo P ser invariante com respeito a G , define-se um grupo de transformações $\bar{G}$ em Λ . Portanto é possível aplicar as definições acima para $\theta(\lambda)$, o parâmetro de interesse. Assim $\theta(\lambda)$ é uma função de λ invariante com respeito a G se $\theta(\lambda)=\theta(\bar{g}(\lambda))$, $\forall\ \bar{g} \in \bar{G}$, $\forall\ \lambda \in \Lambda$. Observe que neste caso, quando os dados sofrem uma transformação g, causando uma transformação $\bar{g}$ em λ, o parâmetro de interesse θ permanece o mesmo (invariante) e segundo o princípio de invariância os dois modêlos , original e transformado, devem fornecer as mesmas informações sobre θ . Logo, a regra de decisão deve ser também invariante com respeito a G já que se deva chegar à mesma inferência com qualquer dos dois modêlos.

<u>DEFINIÇÃO 6.5</u> - Seja P invariante com respeito a G . A estatística t é <u>I-suficiente</u> com respeito a θ se

(a) $\theta(\lambda)$ é uma função maximal invariante com respeito a $\bar{G}$.

(b) t é uma estatística maximal invariante com respeito a G .

Veremos como as condições (a) e (b) satisfazem as suposições de simplicidade e informação, expostos no Capítulo 1.

<u>TEOREMA 6.1</u> - A estatística $t(x)$ é maximal invariante com respeito a G, então a função $h(x)$ é invariante com respeito a G se e somente se $h(x)$ é uma função de $t(x)$.

<u>Demonstração</u>:

($\Longrightarrow$) Sejam x_1, $x_2 \in X$ t.q $t(x_1)=t(x_2)$.

Então $\exists\, g \in G$ t.q $x_1=g(x_2)$ (t maximal invariante). Mas h é invariante, portanto

$$h(x_1) = h(g(x_1)) = h(g(x_2)) = h(x_2)$$

Logo h é uma função de t.

($\Longleftarrow$) se $h(x)=s(t(x))$ então

$$h(g(x))=s(t(g(x)))=s(t(x))=h(x)$$

$\therefore$ h é invariante com respeito a G.

c.q.d.

Note então que toda estatística invariante é função da estatística maximal invariante. Isto fica evidente quando percebe-se que a estatística maximal invariante é a única estatística invariante que consegue distinguir duas órbitas específicas. Portanto se o modêlo é invariante com respeito a G e aceita-se trabalhar apenas com estatísticas invariantes, a estatística maximal invariante é a que contém mais informação com respeito a θ, dentro dessa classe. A estatística maximal invariante parece ser uma boa candidata à definição de I-suficiência. Resta saber se sua distribuição depende apenas de θ. Este resultado é obtido através do teorema abaixo.

<u>TEOREMA 6.2</u> - Seja P invariante com respeito a G e $\theta(\lambda)$ uma função invariante maximal com respeito a $\bar{G}$. Neste caso, se $h(x)$ é invariante com respeito a G, a distribuição de $h(x)$ depende apenas de $\theta(\lambda)$.

<u>Demonstração</u> - Para qualquer $A \subset X$, fixado tem-se

$$P[h(x) \in A/\lambda] = p[h(g(x)) \in A/\lambda] = \quad \text{(invariância de h)}$$

$$= p[h(x) \in A/\bar{g}(\lambda)] \quad \text{(invariância de } P\text{)}$$

Portanto $P[h(x) \in A / \lambda]$ é uma função de λ invariante com respeito a $\bar{G}$.

Aplicando o Teorema 6.1:

$\theta(\lambda)$ invariante maximal com respeito a $\bar{G}$,

$P[h(x) \in A/\lambda]$ invariante com respeito a $\bar{G}$, então $P[h(x) \in A / \lambda]$ é função apenas de $\theta(\lambda)$.

c.q.d.

Em resumo, a condição de que $\theta(\lambda)$ seja invariante maximal com respeito a $\bar{G}$ juntamente com invariância de t garantem que a distribuição de t dependa apenas de θ e consequentemente que as inferências sobre θ podem ser feitas utilizando-se o modêlo marginal de t. A exigência de que t seja maximal está relacionada ao fato de t ser a estatística invariante com mais informação sobre θ .

Alguns exemplos de I-insuficiência são apresentados a seguir.

<u>EXEMPLO 2.3</u> - (continuação)

Seja $(x_1,\ldots,x_n)$ a.c.s da $N(\mu,\sigma^2)$.

Considere o grupo G de transformações de $\mathbb{R}^n$ em $\mathbb{R}^n$ em que $g_a(x_1,\ldots,x_n)=(x_1+a,\ldots,x_n+a)$, $a \in \mathbb{R}$.

A este grupo está associado o grupo $\bar{G}$ de transformações de $(\mathbb{R},\mathbb{R}_+)$ em $(\mathbb{R},\mathbb{R}_+)$ em que $\bar{g}_a(\mu,\sigma)=(\mu+a,\sigma)$.

Alguns resultados podem ser observados:

1. P é invariante com respeito a G .
2. O espaço paramétrico de (μ,σ) é o produto cartesiano dos espaços de μ e de σ. Observe que $\bar{G}$ deixa σ invariante e age transitivamente em μ. Portanto, uma órbita de (μ,σ) é o conjunto

$\{(\mu,\sigma)\ ;\ \mu \in \mathbb{R}\}$, logo σ distingue todas as órbitas e é invariante maximal com respeito a $\bar{G}$.

3. A estatística $t=(x_1-\bar{x},\ldots,x_n-\bar{x})$ é invariante maximal com respeito a G. Note que, $t(g(x))=t(x), \forall\, g\in G$ e que se $t(x)=t(x')$ existe $a=\bar{x}-\bar{x}'$ tal que $g_a(x')=x$.

Dos três resultados segue que $t=(x_1-\bar{x},\ldots,x_n-\bar{x})$ é I-suficiente com respeito a σ.

Uma vez que S^2 é suficiente com respeito a σ, no modêlo marginal de t, é possível reduzir-se mais uma vez para o modêlo de S^2.

O mesmo resultado é obtido se a redução por suficiência (Fisher) antecede a redução por I-suficiência, ou seja, se trabalhamos com o modêlo de $(\bar{x},S^2)$. Em qualquer um dos procedimentos obtém-se a estatística S^2 como aquela que contem toda a informação relevante sobre σ .

<u>EXEMPLO 6.1</u> - Seja uma amostra de tamanho n da normal bivariada de parâmetro $\lambda=(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ desconhecido. O parâmetro de interesse é o coeficiente de correlação ρ. Veremos que a correlação amostral r é I-suficiente com respeito a ρ.

Uma primeira redução do modêlo é considerada; trabalharemos com a estatística suficiente $t = (\bar{x}_1,\bar{x}_2,\Sigma(x_{1i}-\bar{x}_1)^2,\Sigma(x_{2i}-\bar{x}_2)^2,$

$$\frac{\Sigma(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sqrt{\Sigma(x_{1i}-\bar{x}_1)^2\ \Sigma(x_{2i}-\bar{x}_2)^2}}\ ) \ .$$

Considere o grupo de transformações de locação e escala agindo separadamente em cada componente da amostra, porém identicamente, isto é,

$$G = \{g_{a,b}; a=(a_1,a_2) \ \text{ e } b=(b_1,b_2);\ a \in \mathbb{R}^2_+ \ \text{ e } b \in \mathbb{R}^2\}$$

onde

$$g_{a,b}(x_{1i},x_{2i})=(a_1x_{1i}+b_1, a_2x_{2i}+b_2)\ ,\ i=1,\ldots,n.$$

Ao grupo G está associado o grupo $\bar{G} = \{g_{ab};\ a=(a_1,a_2),$ $b=(b_1,b_2), a \in \mathbb{R}^2_+\ , b \in \mathbb{R}^2\}$, onde

$$\bar{g}_{a,b}(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)=(a_1\mu_1+b_1, a_2\mu_2+b_2, a_1\sigma_1, a_2\sigma_2, \rho).$$

Tem-se então que o modêlo P é invariante com respeito a G. O parâmetro ρ é invariante maximal com respeito a G, pelas mesmas razões que no exemplo anterior, ou seja, ρ e $(\mu_1,\mu_2,\sigma_1,\sigma_2)$ são de variação independente e G age transitívamente em $(\mu_1,\mu_2,\sigma_1,\sigma_2)$ e deixa ρ invariante. A estatística invariante maximal com respeito a G é

$$r = \frac{\Sigma(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sqrt{\Sigma(x_{1i}-\bar{x}_1)^2\ \Sigma(x_{2i}-\bar{x}_2)^2}}$$

pelas mesmas razões anteriores.

Logo, o coeficiente de correlação amostral r é I-sufuciente com respeito a ρ.

No próximo capítulo (seção 7.6), uma interpretação Bayesiana será analisada para o método descrito acima.

CAPÍTULO 7

MÉTODO BAYESIANO

7.1 - Suficiência Marginal e Ancilaridade Marginal

Um Bayesiano B não tem dificuldades conceituais em fazer inferências sobre um subparâmetro $\theta(\lambda)$.

Da mesma forma que o estatístico clássico seus dados x dependem de uma quantidade desconhecida λ e esta relação é dada em termos da função de verossimilhança $p(x/\lambda)$.

O Bayesiano B utiliza a função de verossimilhança para transformar a opinião a priori sobre λ, especificada por B em termos da priori $\Pi(\lambda)$, na opinião a posteriori dada em termos da distribuição a posteriori de λ, isto é,

$$\Pi(\lambda/x) = \frac{\Pi(\lambda)\ p(x/\lambda)}{\int_\lambda \Pi(\lambda) p(x/\lambda) d\lambda} \propto \Pi(\lambda) p(x/\lambda).$$

Se $\Theta(\lambda)$ é seu parâmetro de interesse, B simplesmente calcula a distribuição marginal posteriori de $\Theta(\lambda)$, $\Pi(\Theta/x)$, induzida por $\Pi(\lambda/x)$.

No desenvolvimento do método acima não foi necessário, em nenhum momento, especificar o parâmetro nuisance ϕ. No entanto, para simplificar o trabalho de B em especificar a priori $\Pi(\lambda)$, pode-se proceder a uma escolha do parâmetro ϕ. Assim, para uma dada escolha de ϕ, a priori de B para λ pode ser dada como:

$$\Pi(\lambda) = \Pi(\Theta).\Pi(\phi/\Theta) \qquad \text{onde}$$

$\Pi(\Theta)$ é a distribuição a priori marginal para Θ, a qual é facilmente especificada e $\Pi(\phi/\Theta)$ a distribuição condicional a priori de ϕ dado Θ, a qual não é facilmente especificada e depende da escolha de ϕ.

Neste caso, a distribuição a posteriori de Θ, $\Pi(\Theta/x)$ é dada por:

$$\Pi(\Theta,\phi/x) \;\alpha\; \Pi(\Theta,\phi).p(x/\Theta,\phi) \quad [1]$$

$$\therefore\; \Pi(\Theta/x) \;\alpha \int \Pi(\Theta).\Pi(\phi/\Theta).p(x/\Theta,\phi)\, d\phi =$$

[1] A proporcionalidade é usada com base no Princípio da Verossimilhança (ou Suficiência) uma vez que o fator de proporcionalidade não depende dos parâmetros.

$$= \Pi(\Theta) \int p(x/\Theta,\phi)\Pi(\phi/\Theta)\, d\phi =$$

$$= \Pi(\Theta) \cdot \bar{p}(x/\Theta)^{[2]}$$

onde $\bar{p}(x/\Theta)$ define a distribuição dos dados x, dependendo apenas de Θ, para B, pois envolve sua opinião subjetiva sobre $\phi|\Theta$, $\Pi(\phi/\Theta)$. Essa distribuição é chamada de modelo marginal para o Bayesiano B e exprime (para B) a relação entre os dados x e o parâmetro de interesse Θ. Tendo em mãos seu modelo marginal, B define facilmente a posteriori de Θ utilizando a priori $\Pi(\Theta)$.

Do ponto de vista do Bayesiano B, a distribuição $\bar{p}(x/\Theta)$ é tão boa quanto qualquer outra função de verossimilhança. No entanto, quando toda comunidade estatística é envolvida não há concordância com $\bar{p}(x/\Theta)$ em vista da subjetividade introduzida por $\Pi(\phi/\Theta)$, ou seja, Bayesianos com diferentes $\Pi(\phi/\Theta)$ definem diferentes modelos marginais.

O próximo passo é tentar encontrar uma definição de suficiência parcial com respeito a Θ, que esteja de acordo com o ponto de vista Bayesiano. Este problema não tem interesse imediato para o Bayesiano B, pois ele não tem dificuldades conceituais em trabalhar com o modelo indexado por λ. O objetivo de se encontrar tal definição é principalmente a comparação com os métodos clássicos e uma simplificação do trabalho do Bayesiano.

---

[2] A notação $\bar{p}(x/\Theta)$ significa que eliminamos por integração o parâmetro nuisance ϕ do modelo $p(x/\Theta,\phi)$.

A suficiência de uma estatística t com respeito a θ garante que não haja perda de informação ao se trabalhar apenas com o modelo de t. As definições de suficiência e ancilaridade, segundo a visão Bayesiana têm como base a idéia acima, ou seja, de se ter a mesma "informação" (distribuição a posteriori) sobre θ a partir do modelo original de x ou do modelo de t.

Elas são definidas a partir do modelo marginal $\bar{p}(x/\theta)$ e portanto dependem da escolha de $\Pi(\phi/\theta)$ (ou de B), mas não dependem de $\Pi(\theta)$.

<u>DEFINIÇÃO 7.1</u> - A estatística t é marginalmente suficiente com respeito a θ(para B), se t for suficiente com respeito a θ no modelo marginal $\bar{p}(x/\theta)$.

<u>DEFINIÇÃO 7.2</u> - A estatística t é marginalmente ancilar com respeito a θ(para B), se t for ancilar com respeito a θ no modelo marginal $\bar{p}(x/\theta)$.

Trabalhar-se-á apenas com a definição de suficiência marginal pois os resultados para ancilaridade marginal seguem por analogia.

Seja t marginalmente suficiente (para B) com respeito a θ. Neste caso a distribuição posteriori de θ depende apenas de t pois:

$$\Pi(\theta/x) \alpha \Pi(\theta).\bar{p}(x/\theta) = \Pi(\theta) a(x) \bar{p}(t/\theta) \alpha \Pi(\theta) \bar{p}(t/\theta).$$

e coincide com a posteriori de θ, calculada a partir do modelo da estatística t.

Basta ver que

$$\Pi(\theta/t)\alpha\Pi\ (\theta) . \int_{\phi} p(t/\theta,\phi)d\phi = \Pi(\theta)\ \bar{p}(t/\theta)$$

Concluindo, se t é marginalmente suficiente (para B), pode-se trabalhar com o modelo reduzido de t sem que haja perda de informação sobre θ, isto porque consegue-se a mesma distribuição a posteriori que no caso de se trabalhar com o modelo original dos dados.

O Bayesiano B desfruta de vantagens das definições acima, uma vez que os cálculos a partir do modelo reduzido, $p(t/\lambda)$, ficam simplificados. Contudo, quando B descobre que t é marginalmente suficiente com respeito a θ, esta informação não tem mais utilidade: ou ele já calculou a posteriori de θ, a partir de $p(x/\lambda)$ e observou que depende apenas de t, ou calculou o modelo $\bar{p}(x/\theta)$ e observou que t é suficiente com respeito a θ neste modelo.

Da comparação com os métodos clássicos veremos que o modelo $p(x/\lambda)$ e a priori $\Pi(\lambda)$ devem satisfazer certas condições para que as definições 7.1 e 7.2 sejam satisfeitas. Conhecendo estas propriedades o Bayesiano B pode trabalhar com o modelo reduzido sem que antes tenha que fazer cálculos que tornariam inúteis as definições acima citadas.

7.2 - Suficiência de Fisher e Suficiência Marginal

Se t é suficiente com respeito a λ então para qualquer priori $\pi(\lambda)$ (para qualquer Bayesiano B), t é marginalmente suficiente com respeito a θ. Basta ver que:

$$\begin{aligned}
\bar{p}(x/\theta) &= \int_{\phi} p(x/\theta,\phi)\pi(\phi/\theta)\, d\phi = \\
&= \int_{\phi} p(x/t)p(t/\theta,\phi)\pi(\phi/\theta)d\phi = \\
&= p(x/t) \int_{\phi} p(t/\theta,\phi)\pi(\phi/\theta)\, d\phi \,.
\end{aligned}$$

Portanto t é suficiente com respeito a θ no modelo marginal $\bar{p}(x/\theta)$ para qualquer priori $\pi(\phi/\theta)$ considerada.

Este resultado é lógico pois se t é suficiente com respeito a λ o modelo reduzido contém toda informação sobre λ e consequentemente sobre θ. Assim, tanto os estatísticos clássicos quanto os Bayesianos concordam em que não estão perdendo informação sobre θ ao se trabalhar com o modelo reduzido $p(t/\lambda)$. Para os clássicos ainda fica o problema da eliminação do parâmetro nuisance.

7.3 - S-Suficiência e Suficiência Marginal

A definição de S-suficiência de t com respeito a θ exige que θ e ϕ sejam de variação independente. Este conceito é uma maneira que o estatístico clássico encontra para expressar a idéia de que ϕ não contém informação sobre θ. Para um Bayesiano esta idéia é facilmente explicitada em termos da independência entre Θ e ϕ, pois para ele Θ e ϕ são (técnicamente) variáveis aleatórias.

Seja t uma estatística S-suficiente com respeito a Θ e Θ e ϕ a priori independentes. Neste caso t é marginalmente suficiente com respeito a Θ, pois:

$$\bar{p}(x/\Theta) = \int_\phi p(x/\lambda).\Pi(\phi/\Theta)\, d\phi$$

$$= p(t/\Theta). \int_\phi p(x/t,\phi).\Pi(\phi)\, d\phi$$

A conclusão que se pode tirar, é que para Bayesianos com prioris independentes para Θ e ϕ, as inferências sobre Θ podem ser feitas utilizando-se apenas a parte do modelo que depende de Θ, ou seja, no caso de S-suficiência $p(t/\Theta)$. Os resultados para S-ancilaridade são análogos.

A próxima seção continua esta discussão generalizando o conceito de S-suficiência, pois muitas vezes a função de verossimilhança pode ser fatorada em duas funções uma dependendo apenas de Θ e outra dependendo apenas de ϕ, sem que essas funções sejam modelos probabilísticos com relação aos dados.

7.4 - Parâmetros não Relacionados

No início, foi colocado que como o interesse eram apenas as inferências sobre Θ, deveríamos isolar a parte da informação sobre θ. Para isto é importante que θ e ϕ sejam não "relacionados", de alguma maneira, de modo que ϕ não contenha informação sobre θ. Em Basu[1977], temos uma definição de "não relacionamento".

Seja a família de distribuições à priori em $\Lambda = \Theta \times \Phi$, independentes isto é,

$$\Pi_0 = \{\pi(\lambda) = \pi(\theta).\pi(\phi) \ ; \ \theta \in \Theta, \ \phi \in \Phi \}$$

DEFINIÇÃO 7.3 - Os parâmetros θ e ϕ são não-relacionados relativamente ao modelo P se para toda priori $\pi \in \Pi_0$ e para todo $x \in X$ a dist. à posteriori $\pi(\lambda/x) \in \Pi_0$.

Em outras palavras temos que θ e ϕ são não relacionados relativamente ao modelo P se distribuições à priori independentes para θ e ϕ gerarem distribuições à posteriori independentes. Vamos analisar agora para que tipo de modelo esta definição é satisfeita.

TEOREMA 7.1 - Os parâmetros θ e ϕ são não relacionados relativamente ao modelo P se e somente se a função de verossimilhança se fatora como

$$p(x/\theta,\phi) = A(x,\theta) \ B(x,\phi) \quad [1]$$

[1] Note que as funções $A(x,\theta)$ e $B(x,\phi)$ não são necessariamente modelos de probabilidade.

<u>DEMONSTRAÇÃO</u> - Se $p(x/\theta,\phi) = A(x,\theta)\ B(x,\phi)$ então

$$\Pi(\theta,\phi/x) = \frac{A(x,\theta)\Pi(\theta)}{\int_{\Theta} A(x,\theta)\Pi(\theta)d\theta} \cdot \frac{B(x,\phi)\Pi(\phi)}{\int_{\Phi} B(x,\phi)\Pi(\phi)d\phi}$$

$$= \Pi(\theta/x)\ \Pi(\phi/x)$$

Da mesma maneira, se

$$\Pi(\theta,\phi/x) = \Pi(\theta/x)\Pi(\phi/x)\ \alpha\ p(x/\theta,\phi)\Pi(\theta)\Pi(\phi)$$

então $p(x/\theta,\phi)$ deve se fatorar como $A(x,\theta)\ B(x,\phi)$.

c.q.d.

Fica fácil reconhecer quando o parâmetro de interesse θ é não relacionado com o parâmetro nuisance ϕ. Basta verificar se a função de verossimilhança se fatora como acima e se a priori pertence a Π_o. Se isto acontece o trabalho do Bayesiano é simplificado pois não é necessário especificar a priori para ϕ e além disso a posteriori de θ é calculada a partir do fator da verossimilhança que depende apenas de θ, ou seja,

$$\Pi(\theta/x) = \frac{A(x,\theta)\Pi(\theta)}{\int_{\Theta} A(x,\theta)\Pi(\theta)d\theta}$$

É claro que se t é S-suficiente com respeito a θ e $\Pi\in\Pi_o$, os parâmetros θ e ϕ são não relacionados. O que o teorema 7.1 coloca é que pode <u>não</u> existir tal estatística t e mesmo assim a verossimilhança se fatorar em duas partes, uma dependendo de θ, $A(x,\theta)$, e outra de ϕ, $B(x,\phi)$. Neste caso $A(x,\theta)$ e $B(x,\phi)$ não são modelos probabilísticos e é impossível para o estatístico clássico trabalhar apenas com $A(x,\theta)$ ou $B(x,\phi)$.

Antes de se analisar alguns exemplos deste resultado, são feitas algumas considerações assintóticas.

Uma objeção ao resultado acima, é o fato de considerar-se independência a priori. Muitas vezes é difícil julgar que as prioris para θ e ϕ sejam independentes. Entretanto para amostras grandes sabe-se que a força da verossimilhança é grande e tende a anular as informações da priori. Se neste caso a função de verossimilhança se fatorar como no teorema 7.1 as posterioris de θ e ϕ tendem a ser independentes.

EXEMPLO 7.1 - Tábua de vida com retirada.

Este exemplo pode ser encontrado em Lindley[1979].

Um número n de indivíduos é observado durante um período de tempo. Sem perda de generalidade podemos supor este período $[0,1]$. A cada indivíduo estão associadas duas variáveis supostas independentes; t_1 o tempo de vida do indivíduo no intervalo $[0,1]$ e t_2 o tempo até a retirada em $[0,1]$. As variáveis para diferentes indivíduos são independentes e identicamente distribuídas.[1]

[1] Outros enfoques podem ser considerados neste problema. Por exemplo considerar t_1, o tempo de morte devido a uma doença e t_2, o tempo de morte devido a outra doença. Ou ainda, para uma máquina composta de duas componentes em série independentes, t_1 seria o tempo de falha da primeira componente e t_2 o tempo de falha da segunda componentes.

Estaremos interessados nas inferências sobre a taxa de mortalidade durante o período, ou seja, em $p(t_1 \leq 1)$.

A dificuldade deste problema está no fato de que o número observado de mortes é menor que o número real de mortes pois a morte de um indivíduo pode ocorrer depois de ele ter sido retirado da população, porém ainda no intervalo $[0,1]$.

Note, entretanto, que este é um problema de eliminação de parâmetros nuisance pois estamos interessados em apenas uma característica do modelo, $p(t_1 \leq 1)$. Para resolvê-lo é necessário calcular a função de verossimilhança.

Se n indivíduos são observados no início do período, a amostra consiste em observar durante o período: n_1 o número de indivíduos que morreram, n_2 o número de indivíduos que se retiraram do estudo e n_3 o número de indivíduos que sobreviveram. Estes dados seguem uma distribuição multinomial nas seguintes categorias.

$$[t_1 < t_2 \ , \ t_1 < 1] \ , \ [t_2 < t_1 \ , \ t_2 < 1] \quad e$$

$$[t_1 > 1 \ , \ t_2 > 1].$$

Como o parâmetro de interesse é $P(t_1 \leq 1)$, faz-se a seguinte reparametrização das probabilidades das categorias acima;

$$\theta = p(t_1 \leq 1) \ ,$$

$$\phi = p(t_2 \leq 1) \ \text{e}$$

$$\rho = p(t_1 < t_2 \mid t_1 \leq 1, \ t_2 \leq 1) \ ,$$

onde θ é a probabilidade do indivíduo morrer em $[0,1]$, ϕ a probabilidade do indivíduo se retirar do estudo em $[0,1]$ e ρ a probabilidade de que o indivíduo morra dado que os dois eventos ocorram em $[0,1]$: o indivíduo morre e o indivíduo se retira.

É necessário calcular $p(t_1<t_2,\ t_1<1)$, $p(t_2<t_1,\ t_2<1)$ e $p(t_1>1,\ t_2>1)$ em termos de θ, ϕ e ρ, para encontrar a função de verossimilhança segundo a reparametrização.

$$p(t_1<t_2,\ t_1<1) = p(t_1<1)p(t_1<t_2/t_1<1)$$

mas,

$$p(t_1<t_2/t_1<1) = 1- \frac{p(t_2<t_1,\ t_1<1)}{p(t_1<1)} =$$

$$= 1 - \frac{p(t_2<t_1,\ t_1<1)}{p(t_1<1)} \cdot \frac{p(t_2<1)}{p(t_2<1)} = \text{(independência de } t_1 \text{ e } t_2\text{)}$$

$$= 1 - \frac{p(t_2<t_1,\ t_1<1, t_2<1)p(t_2<1)}{p(t_1<1,\ t_2<1)} =$$

$$= 1 - p(t_2<t_1/t_1<1, t_2<1)p(t_2<1).$$

Portanto

$$p(t_1<t_2, t_1<1) = p(t_1<1)[1-p(t_2<t_1/t_1<1, t_2<1)p(t_2<1)]$$

e substituindo θ, ϕ e ρ na expressão acima tem-se que

$$p(t_1<t_2,\ t_1<1) = \theta[1-(1-\rho)\phi] \tag{1}$$

É imediato, pela independência de t_1 e t_2, que

$$p(t_1>1,\ t_2>1) = (1-\theta).(1-\phi) \tag{2}$$

Da diferença das duas probabilidades acima de 1, obtém-se

$$p(t_2<t_1,\ t_2<1) = \phi(1-\rho\theta) \tag{3}$$

Portanto, substituindo-se as probabilidades (1), (2) e (3) no modelo multinomial, obtém-se a seguinte função de verossimilhança:

$$p(x/\theta,\phi,\rho) = \theta^{n_1}\left[1-(1-\rho)\phi\right]^{n_1}(1-\theta)^{n_2}(1-\phi)^{n_2}\phi^{n_3}(1-\rho\theta)^{n_3}.$$

Alguns comentários podem ser feitos a propósito da expressão acima. Primeiramente, que esta verossimilhança contém informação sobre três aspectos, apenas, das distribuições de t_1 e t_2, a saber θ,ϕ e ρ. Os parâmetros (θ,ϕ,ρ) são de variação independente pois cada um deles varia no intervalo $[0,1]$, independentemente dos valores dos outros. Em segundo lugar se as probabilidades da distribuição multinomial nas expressões (1), (2) e (3) são conhecidas $(n\to\infty)$, os valores de θ, ϕ e ρ não ficam especificados e sabe-se unicamente que eles pertencem a uma curva no cubo $[0,1]^3$. O fato de existir uma curva na qual θ, ϕ e ρ variem e as probabilidades (1), (2) e (3) permaneçam inalteradas, é denominado de "problema de não-identificação" pelos economistas: Por exemplo, para $\rho = \frac{1}{2}$ e $\theta=\phi$ as probabilidades (1), (2) e (3) permanecem constantes. Este fato gera dificuldades pois embora se encontre estimadores para (1), (2) e (3) é impossível encontrar os estimadores para θ, ϕ e ρ pelos métodos clássicos conhecidos. Em terceiro lugar, se ρ é conhecido a função de verossimilhança se fatora no produto de dois termos, um dependendo de θ apenas e outro dependendo de ϕ. Consequentemente se θ e ϕ são à priori independentes ou se a amostra é grande, as inferências sobre θ ou ϕ são feitas separadamente. Entretanto, se ρ é desconhecido, a situação é bem mais complicada. Não existe nenhuma razão evidente para se eliminar o fator que depende de ϕ, quando θ é o parâmetro de interesse, pois esta parte parece conter informação sobre ρ o qual é importante na interpretação do fator de θ. Portanto, ρ tem um papel fundamental nesta análise.

Analisaremos aqui apenas o primeiro ponto de vista em que ρ é suposto conhecido[1]. O procedimento é fazer considerações sobre as funções distribuição de probabilidade de t_1 e t_2, F e G respectivamente, de modo que ρ fique especificado. Por exemplo se $F(x) = cH(x)$ não é difícil provar que $\rho = \frac{1}{2}$, e neste caso a função de verossimilhança é dado por:

$$p(x/\theta,\phi)=[\theta^{n_1}(1-\theta)^{n_2}(1-\tfrac{\theta}{2})^{n_3}]\,[\phi^{n_3}(1-\phi)^{n_2}(1-\tfrac{\phi}{2})^{n_1}]$$

Esta verossimilhança se fatora em duas partes, uma dependente de θ e outra de ϕ, sendo que as duas partes dependem de toda a amostra (n_1,n_2,n_3). Portanto, não existe estatística t que defina um corte de Barndorff-Nielsen para este problema.

---

[1] O segundo ponto de vista é analisado em Lindley [1979] .

Se as prioris para θ e ϕ podem ser consideradas independentes, o que não é de forma alguma absurdo, θ e ϕ são "não-relacionados" e as inferências sobre θ são feitas utilizando-se

$$A(x,\theta) = \theta^{n_1}(1-\theta)^{n_2}(1-\frac{\theta}{2})^{n_3} \quad \text{e } \Pi(\theta).$$

A resolução deste problema, para um estatístico clássico é impossível, uma vez que $A(x,\theta) = \theta^{n_1}(1-\theta)^{n_2}(1-\frac{\theta}{2})^{n_3}$ não é um modelo de probabilidade em (n_1,n_2,n_3).

7.5 - H-Suficiência e Suficiência Marginal

A definição de H-suficiência tem uma interpretação Bayesiana, que será dada a seguir.

A estatística t é H-suficiente com respeito a θ se existe alguma medida de probabilidade em ϕ tal que t é suficiente para as distribuições.

$$Q_\theta = \int p(x/\theta,\phi)\, dq_\theta(\phi).$$

Comparando esta definição com a definição de suficiência marginal fica evidente que t é H-suficiente com respeito a θ se existir algum Bayesiano B (alguma priori $\Pi(\phi/\theta)$) tal que t é marginalmente suficiente com respeito a θ para B, no modelo $\bar{p}(x/\theta)=\int p(x/\theta,\phi).\Pi(\phi/\theta)d\phi$.

Portanto, quando toda a comunidade estatística é envol vida, a definição de H-suficiência não tem muito sentido.

Além disso, no exemplo 5.2, a estatística $\bar{x}$ era H-suficiente com respeito a Θ quando considerava-se a priori

$$\Pi(\phi/\Theta) = \begin{cases} 1 & \text{se} \quad \phi=0 \\ 0 & \text{se} \quad \phi=1 \end{cases}$$

Como confiar na definição de H-suficiência se ela é válida para uma priori tão extremática quanto esta, ou seja, que não considera a possibilidade de $\phi=1$?

7.6 - I-Suficiência e Suficiência Marginal.

Suponha $P = \{p(x/\lambda),\ \lambda \in \Lambda\}$ uma família de distribuições invariantes com respeito a G e t uma estatística I-suficiente com respeito a Θ.

Considere também a família de prioris para λ, F, que são invariantes em relação a $\bar{G}$, portanto se $\Pi \in F$, $g \in G$ então

$$\lambda \sim \Pi \Rightarrow \bar{g}(\lambda) \sim \Pi.$$

Nesta situação a família de distribuições posteriori para Θ dado x é equivariante sob a ação de $\bar{G}$ em Θ e G em x (estamos oonsiderando Θ como "variável aleatória" cuja distribuição é governada pelo "parâmetro" x). Basta ver que:

$$\Pi(\bar{g}(\lambda)/x) = \frac{P(x/\bar{g}(\lambda)).\Pi(\bar{g}(\lambda))}{\int_{\Lambda} p(x/\bar{g}(\lambda))\ \Pi(\bar{g}(\lambda))} =$$ [Invariância de P com respeito a G e de F com respeito a $\bar{G}$]

$$= \frac{P(g(x)/\lambda)\ \pi(\lambda)}{\int_\Lambda P(g(x)/\lambda)\pi(\lambda)} = \pi(\lambda/g(x))$$

Estamos então na condição do teorema 6.2 pois:

1. A família de distribuições posteriori para θ dado x é invariante com respeito a $\bar{G}$.

2. $\theta(\lambda)$ é invariante maximal sob $\bar{G}$.

3. $t(x)$ é invariante maximal sob G.

Portanto a distribuição posteriori de θ depende apenas de t. Desta maneira se t é I-suficiente com respeito a G e a distribuição à priori para λ é invariante com respeito a $\bar{G}$, $\pi \in F$, então t é marginalmente suficiente com respeito a θ. Para Bayesianos com prioris em F as duas definições são equivalentes.

A teoria desenvolvida acima tem aplicação limitada. Uma distribuição a priori própria $\bar{G}$-invariante existe apenas quando $\bar{G}$ é um grupo topológico compacto. Geralmente esta condição não é satisfeita como por exemplo nos casos da variância da população normal[1] e do coeficiente de correlação da normal bivariada. Nestes casos as prioris $\bar{G}$-invariante que existem são medidas impróprias. Este fato pode gerar dificuldades uma vez que a distribuição posteriori de θ pode depender apenas de t mas pode não ser derivável do modelo marginal de t. Segundo Dawid[1981],

[1] Mostramos no capítulo 5 que se σ era ilimitado não existia uma medida própria para μ tal que S fosse suficiente com respeito a σ no modelo marginal $p(\bar{x},s/\sigma)$.

um problema tecnicamente difícil é descobrir se uma estatística I-suficiente pode ser marginalmente suficiente para prioris não-invariantes, no caso de grupos topológicos não compactos. Estudos de caso sugerem que isto não é normalmente possível.

Portanto, a redução para uma estatística I-suficiente, quando o grupo $\bar{G}$ é não-compacto é incoerente do ponto de vista Bayesiano, em vista de que as únicas distribuições à priori, para as quais tal redução é possível são impróprias e possivelmente paradoxais.

CAPÍTULO 8

VEROSSIMILHANÇA PARCIAL

8.1 - Modelo com Taxa de Falha Proporcional

Todos os métodos clássicos de inferência parcial estudados aqui consideravam apenas duas reduções do modelo original de x; o modelo condicional de x dado t e o modelo marginal de t. O método desenvolvido por Cox [1975] considera um modelo reduzido mais geral o qual é denominado de verossimilhança parcial. Para sua apresentação consideraremos o exemplo do "Modelo com Taxa de Falha Proporcional", Cox [1972] , abaixo.

O Modelo com Taxa de Falha Proporcional é um caso particular de um modelo mais geral que estuda os tempos de falha (ou tempo de vida) de uma população.

Primeiramente vamos estudar a distribuição de variáveis que representam o tempo de falha.

Seja T uma v.a. não negativa representando o tempo de falha de um elemento. Vamos considerar T uma v.a. contínua pois é este o caso que nos interessará.

A distribuição da v.a. T pode ser especificada de várias maneiras dadas a seguir:

$f(t)$ = função densidade de probabilidade de T;

$F(t) = \int_0^t f(\mu)d\mu$, função distribuição de probabilidade de T;

$\bar{F}(t) = 1 - F(t)$, função de sobrevivência de T;

$\lambda(t) = \lim_{\Delta t \to 0} \dfrac{P[t<T<t+\Delta t \mid T \geq t]}{\Delta t}$, a taxa de falha no instante t e

$\Lambda(t) = \int_0^t \lambda(\mu)d\mu$, o funcional de falha.

Uma relação entre as representações acima pode ser:

$$f(t) = \frac{d}{dt} F(t) = - \frac{d}{dt} \bar{F}(t) = \underbrace{\lambda(t)e^{-\Lambda(t)} = \lambda(t)\,\bar{F}(t)}_{\bar{F}(t) = e^{-\Lambda(t)}} \qquad (8.1)$$

Observe que f(t) e F(t) são representações comuns da distribuição de uma variável aleatória, enquanto que a taxa de falha $\lambda(t)$ é uma caracterização especial no estudo de variáveis de tempo de falha. Idem para $\Lambda(t)$ e $\bar{F}(t)$.

A taxa de falha $\lambda(t)$ representa a probabilidade do indivíduo morrer em $(t,t+dt)$ dado que viveu até t e é também co nhecida como força de mortalidade. Sua importância se deve ao fato de que a informação disponível sobre o tempo de falha é dada, geralmente, em termos da variação da taxa de falha com relação a quantidade de tempo já vivido. Conhecido $\lambda(t)$ pode-se obter facilmente $f(t)$ e $F(t)$, e portanto a taxa de falha especifica totalmente o modelo.

Uma descrição de vários Modelos de Tábuas de Vida pode ser encontrada em Kalbfleisch & Prentice [1980] .Estudaremos aqui apenas o Modelo com Taxa de Falha Proporcional.

De uma população $S=\{1,\ldots,n\}$ de indivíduos observa-se as variáveis aleatórias:

t_i = tempo de vida do i-ésimo indivíduo; $i=1,\ldots n$.

A amostra obtida x é o conjunto dos pares (i,t_i),que representam o i-ésimo indivíduo e seu tempo de falha, assim,

$$x =\{(1,t_1),(2,t_2),\ldots,(n,t_n)\}$$ $T_i \sim Exp(\lambda)$

Se a população considerada não for homogênea a função de verossimilhança para o modelo geral é dada por

$$L(f_n) = f_1(t_1)\ f_2(t_2)\ \ldots\ f_n(t_n)$$ em que

$(f_1,\ldots,f_n)$ são desconhecidas e são os parâmetros.

O modelo com Taxa de Falha Proporcional impõe hipóteses simplificadoras:

$$\lambda_i(t) = \lambda_o(t)\ .\ \delta_i \quad , \; i = 1,\ldots,n$$

Neste modelo, $\lambda_o(t)$ é uma função comum a todos indivíduos, que descreve o comportamento comum da taxa de falha e δ_i é a característica de cada indivíduo, que influencia na taxa de falha multiplicativamente.

No modelo introduzido por Cox [1972], também conhecido por Modelo de Regressão de Cox ou Modelo Log-Linear, impõe-se as seguintes restrições:

$$\delta_i = e^{<\beta, z_i>} \quad \text{em que}$$

$\beta = (\beta_1, \ldots, \beta_p)$ são parâmetros de regressão e

$z_i = (z_{i1}, \ldots, z_{ip})$, $i = 1, \ldots, n$; as características conhecidas de cada indivíduo.

Estamos interessados apenas na relação entre o tempo de falha e as características z_i, ou seja, estamos interessados em inferências sobre β . Portanto β é o parâmetro de interesse e $\lambda_o(t)$ o parâmetro nuisance.

O procedimento é encontrar a função de verossimilhança de modo a fazer inferências sobre β.

A amostra x pode ser equivalentemente descrita como:

$$x = \{(t_{(1)}, (1)); (t_{(2)}, (2)); \ldots; (t_{(n)}, (n))\}$$

em que $t_{(i)}$ é o i-ésimo tempo de falha e (i) o indivíduo correspondente. Isto é, $t_{(1)}, \ldots, t_{(n)}$ são as estatísticas de ordem de $t_1, \ldots, t_n$ e (1), ..., (n) os postos dos indivíduos. Não suporemos empates nos tempos de falha. Observar (i), no Modelo de Regressão de Cox, é observar as características $z_{(i)}$

Com os dados x representados desta forma a função de verossimilhança pode ser fatorada da seguinte maneira:

Aconjunto

$$L(\lambda_o(\cdot),\beta)=p[t_{(1)}/\lambda_o(\cdot),\beta]\,p[(1)/t_{(1)},\lambda_o(\cdot),\beta]\,p[t_{(2)}/(1),t_{(1)},\lambda_o(\cdot),\beta]\,\ldots$$

$$\ldots p[(n)/t_{(1)},(1),\ldots,(n-1),t_{(n)},\ \lambda_o(\cdot),\beta]\ .$$

Cox coloca que a função de verossimilhança relevante sobre β ,deve ser considerada condicionalmente no conjunto de instantes $\{t(i)\}$ em que as falhas ocorrem. Explicitamente, ele considera apenas os fatores da verossimilhança acima do tipo: $p[(i)/t_{(1)},(1),\ldots,(i-1),\ t_{(i)},\ \lambda_o,\ \beta]$.

O argumento utilizado foi o de que sendo a função $\lambda_o(\cdot)$ arbitrária, nos intervalos entre falhas consecutivas esta função deve ser conceitualmente nula, uma vez que não há falha nestes intervalos. Portanto nenhuma informação pode ser obtida sobre β nestes intervalos. Assim condiciona-se no conjunto de instantes $\{t_{(i)}\}$ em que as falhas ocorrem. Além disso, os fatores $p[(i)/t_{(1)},(1),\ldots,(i-1),t_{(i)},\ \lambda_o(\cdot),\ \beta]$ independem do parâmetro nuisance $\lambda_o(\cdot)$. É isto que mostraremos a seguir.

A distribuição de $t_{(1)}$, o primeiro tempo de falha, pode ser dada equivalentemente por $f_{(1)}$, $F_{(1)}$, $\bar{F}_{(1)}$, $\lambda_{(1)}$ e $\Lambda_{(1)}$. Segue facilmente que:

$F_{(1)}(t) = P[(1) \leq t] = 1 - P[(1) > t]$

$$\bar{F}_{(1)}(t) = \prod_{i=1}^{n} \bar{F}_i(t) = e^{-\sum \Lambda_i(t)}$$

$\bar{F}_{(1)}(t) = 1 - F_{(1)}(t)$

$= 1 - \prod S_i(t)$

$\bar{F}_{(1)}(t) =$

$= 1 - \prod \bar{F}_i(t)$ $S_{(1)}(t) = \prod S_i(t)$

$F_{(1)}(t) = 1 - \prod S_i(t) \Rightarrow 1 - F_{(1)}(t) = \prod S_i(t)$

Da relação (8.1) vem que:

$$\lambda_{(1)}(t) = \sum_{i=1}^{n} \lambda_i(t) \qquad \lambda(t) = \frac{f(t)}{\bar{F}(t)} \Rightarrow f_{(1)}(t) = \lambda_{(1)}(t).\ \bar{F}_{(1)}(t)$$

e

$$f_{(1)}(t) = \lambda_{(1)}(t) \prod_{i=1}^{n} \bar{F}_i(t) = \lambda_{(1)}(t)\ \bar{F}_{(1)}(t)$$

Considere agora a probabilidade condicional

$$P[(1) = i \,|\, t_{(1)} \in (t,t+dt)] = \frac{P[(1)=i \cap t_{(1)} \in (t,t+dt)]}{\underbrace{P[t_{(1)} \in (t,t+dt)]}_{P[T_{(1)} \approx t] = f_1(t)dt}} =$$

$$= \frac{P[t_i \in (t,t+dt),\ t_j > t,\ j \neq i]}{f_1(t)dt} = \qquad \text{(Independência)}$$

$$= \frac{dt\ \lambda_i(t)\ \bar{F}_i(t) \prod_{j\neq i} \bar{F}_j(t)}{dt\ \lambda_{(1)}(t)\ \bar{F}_{(1)}(t)} = \frac{\lambda_i(t)}{\lambda_{(1)}(t)} = \frac{\lambda_i(t)}{\sum_{j=1}^{n} \lambda_j(t)}$$

Segue então que:

<u>TEOREMA 8.1</u>: $P[(1)=i \,|\, t_{(1)}] = \dfrac{\lambda_i(t_{(1)})}{\sum_{j=1}^{n} \lambda_j(t_{(1)})}$; $i=1,\ldots,n$.

Se o modelo considerado é o Modelo com Taxa de Falha Proporcional, e neste caso $\lambda_i(t) = \lambda_o(t)\ \delta_i$, $i=1,\ldots,n$, $\forall t$ então temos que $(1) \perp\!\!\!\perp t_{(1)}$ pois

$$P[(1)=i/t_{(1)}] = \frac{\delta_i}{\sum_{j=1}^{n} \delta_j}\ , \quad i=1,\ldots,n.$$

A interpretação deste resultado é de que neste modelo existe a independência entre o posto do indivíduo que falha e seu tempo de falha. O teorema abaixo mostra que esta é uma característica específica deste modelo.

TEOREMA 8.2: $(1) \perp\!\!\!\perp t_{(1)}$ se e somente se $\lambda_i(t)=\lambda_o(t)\ \delta_i$ para $i=1,\ldots,n$ e $\forall t$.

Demonstração:

Mostrou-se que se $\lambda_i(t)=\lambda_o(t).\delta_i$ então $(1) \perp\!\!\!\perp t_{(1)}$.

Se $pi = P[(1) = i] = P[(1) = i \mid t_{(1)}]$.

$$pi = \frac{\lambda_i(t_{(1)})}{\sum_{j=1}^{n} \lambda_j(t_{(1)})}, \text{ para todos valores de } t_{(1)}.$$

$$\text{Portanto } \lambda_i(t) = pi \sum_{j=1}^{n} \lambda_j(t) = pi\ \lambda_o(t).$$

c.q.d.

Antes de se encontrar as outras distribuições de probabilidade condicionais, duas observações importantes sobre a distribuição da variável tempo de falha, no modelo com taxa de falha proporcional, devem ser feitas.

A distribuição condicional do tempo de vida T condicionado em $T > t_o$ é dada por:

$$P[T > t \mid T > t_o] = \frac{P[T > t]}{P[T > t_o]} = \frac{\bar{F}(t)}{\bar{F}(t_o)} =$$

$$= e^{-(\Lambda(t)-\Lambda(t_o))}, \quad t \geq t_o.$$

Portanto

$$\bar{F}(t/t \geq t_o) = e^{-(\Lambda(t)-\Lambda(t_o))}, \quad t \geq t_o$$

$$f(t/t \geq t_o) = -\frac{d\,\bar{F}(t|t \geq t_o)}{dt} = \lambda(t)\, e^{-[\Lambda(t)-\Lambda(t_o)]}, t \geq t_o$$

A primeira observação, relacionada apenas com variáveis que representam tempos de falha, é que dada a informação que $T > t_o$ o funcional de falha se transforma em $\Lambda(t)-\Lambda(t_o)$, enquanto que, a taxa de falha $\lambda(t)$ permanece inalterada. Isto significa que apesar de ser dada a informação de que $T > t_o$ o tempo de falha continua tendo a mesma taxa de falha.

A segunda consideração está relacionada com o modelo de Taxa de Falha Proporcional. Se $t_1,\ldots,t_n$ são os tempos de falha mutuamente independentes com taxa de falha $\lambda_i(t)=\lambda_o(t)\delta_i$, $i=1,\ldots,n$, pela observação feita anteriormente tem-se que, dado $\{t_i > t_o,\ i=1,\ldots,n\}$, a distribuição condicional conjunta de $t_1,\ldots,t_n$ muda mas eles continuam sendo mutuamente independentes com a mesma taxa de falha proporcional $\lambda_i(t)=\lambda_o(t)\delta_i; i=1,\ldots,n$, $t > t_o$. Em outras palavras, continua-se trabalhando com um Modelo de Taxa de Falha Proporcional e a caracterização dada pelo Teorema 8.2, permanece válida.

A distribuição condicional de $t_{(2)}$ e (2) dado $t_{(1)}$ e (1), será considerada.

Seja o conjunto de sobreviventes em $t_{(j)}$ $R_{(j)} = \{i: t_i \geq t_{(j)}\}$

conjunto de indivíduos em Risco

$$R_{(j)} = \{(1,\ldots,n)-(1),\ldots,(j)\} \; ; \; j=1,\ldots,n.$$

Dar a informação $t_{(1)}$ e (1) significa dizer que (2) é um valor no conjunto $R_{(1)}$ e que $\{t_i > t_{(1)} : i \in R_{(1)}\}$. Desta maneira, pelas observações anteriores temos ainda um Modelo com Taxa de Falha Proporcional em $R_{(1)}$, com as mesmas taxas de falha e portanto pelo Teorema 8.1,

$$P[(2)/t_{(1)},(1),\, t_{(2)}] = \frac{\delta_{(2)}}{\sum_{i \in R_{(2)}} \delta_i}$$

Neste caso $(2) \perp\!\!\!\perp t_{(2)} \mid t_{(1)}, (1)$ [1] e além disso $(2) \perp\!\!\!\perp (t_{(1)}, t_{(2)}) \mid (1)$. Se continuarmos este argumento passo a passo:

TEOREMA 8.3: No caso do Modelo com Taxa de Falha Proporcional, com δ_i como fator de proporcionalidade e $\lambda_o(t)$ como a base da taxa de falha, o processo estocástico $t_{(1)},(1),t_{(2)},(2),\ldots,t_{(n)},(n)$ possui a seguinte propriedade de independência condicional

$$(1) \perp\!\!\!\perp t_{(1)}; \quad (2) \perp\!\!\!\perp (t_{(1)},t_{(2)}) \mid (1); (3) \perp\!\!\!\perp (t_{(1)},t_{(2)},t_{(3)}) \mid (1),(2),\ldots$$

assim por diante e as probabilidades condicionais

[1] . A notação $x \perp\!\!\!\perp y/z$ indica que os elementos aleatórios x e y são condicionalmente independentes dado z, ou seja, $p(x/y,z) = p(x/z)$.

$$P[(i)/t_{(1)},(1),\dots,(i-1),t(i)]=P[(i)/(1),(2),\dots,(i-1)] =$$

$$= \frac{\delta_i}{\sum_{j\in R_{(i)}} \delta_j}$$

No Modelo de Regressão de Cox, as probabilidades condicionais acima são dadas por:

$$P((i)/t_{(1)},(1),t_{(2)},\dots,(i-1),t(i))= \frac{e^{<\beta,z_{(i)}>}}{\sum_{j\in R_{(i)}} e^{<\beta,z_{(j)}>}}$$

A função de "verossimilhança parcial", que considera apenas os fatores $P((i)/(1),\dots,(i-1))$ é

$$L(\beta) = \prod_{i\in S} \frac{e^{<\beta,z_{(i)}>}}{\sum_{j\in R_{(i)}} e^{<\beta,z_{(j)}>}}$$

e independe dos parâmetros nuisance $\lambda_o(t)$.

8.2 - Modelo com Dados Censurados

Considere uma população $S=\{1,\ldots,n\}$ de indivíduos; para cada indivíduo observa-se ou o tempo de falha ou o tempo de retirada do estudo. Assim, para indivíduos que se retiram, sabe-se apenas que o tempo de falha é maior que o tempo de saída da observação.

Sejam as variáveis aleatórias t_i o tempo de falha do i-ésimo indivíduo e w_i a tempo de retirada do i-ésimo indivíduo, $i=1,\ldots,n$. O modelo aqui considerado supõe (t_i,w_i), $i=1,\ldots,n$, variáveis aleatórias mutuamente independentes e que para cada indivíduo $i=1,\ldots,n$ as variáveis aleatórias w_i e t_i são independentes. Esta última suposição embora não seja muito realista, tem sido utilizada com frequência e será também aqui mantida.

Pode-se caracterizar a distribuição de w_i por g_i, G_i, $\bar{G}_i$, γ e Γ.

Considere os conjuntos aleatórios $I=\{i:t_i<w_i\}$ e $I=\{j:t_j>w_j\}$ onde I consiste dos indivíduos que se observou falharem e J daqueles que se observou se retirarem.

A amostra pode ser dada por

$$x = \left\{[(i,t_i);\ i\in I] \cup [(j,w_j),\ j\in J]\right\}$$ e a função de verossimilhança como:

$$L(f,g/x) = \left[\prod_{i\in I} f_i(t_i)\ \bar{G}_i(t_i)\right]\left[\prod_{j\in J} g_i(w_j)\ \bar{F}_j(w_j)\right]$$

$$= \left[\prod_{i \in I} f_i(t_i) \prod_{j \in J} \bar{F}_j(w_j)\right]\left[\prod_{i \in I} \bar{G}_i(t_i) \prod_{j \in J} g_i(w_j)\right] =$$

$$= A(f,x)\ B(g,x).$$

Observe que os parâmetros são os vetores das funções densidades de probabilidade dos tempos de falha e tempos de retirada, f e g respectivamente. Como estamos interessados apenas em estudar os tempos de falha, o vetor g é o parâmetro nuisance e f é o parâmetro de interesse.

Por outro lado, a função de verossimilhança se fatora em uma parte dependente apenas de f e outra apenas de g. Como considerou-se t_i e w_i independentes para $i=1,\ldots,n$, o Bayesiano considera f e g não relacionados e se seu parâmetro de interesse é f ele utilizará como função de verossimilhança A(f,x). Portanto

$$L(f|x) = \prod_{i \in I} f_i(t_i) \prod_{j \in J} \bar{F}_j(w_j).$$

É interessante notar que a função de verossimilhança acima é equivalente a função de verossimilhança para o caso em que se supõe w_j, $j=1,\ldots,n$, constantes desconhecidas. Basta ver que:

$$\prod_{i \in I} P[T_i \in (t_i, t_i+dt), T_i < w_i] \prod_{j \in J} P[T_j > w_j] =$$

$$= \prod_{i \in I} P[T_i < w_i / T_i \in (t_i, t_i+dt]. P[T_i \in (t_i, t_i+dt)] \prod_{j \in J} P[T_j > w_j]$$

$$= \prod_{i \in I} f_i(t_i) \prod_{j \in J} \bar{F}_j(w_j).$$

Portanto, vamos supor, no decorrer desta seção, que w_i, $i=1,\ldots,n$ não são mais variáveis aleatórias.

Para o modelo com Taxa de Falha Proporcional as suposições são como antes:

$$\lambda_i(t) = \lambda_o(t)\, e^{\langle \beta, z_i \rangle} \quad , \; i=1,\ldots,n$$

Estamos interessados em encontrar uma parte da verossimilhança que depende apenas de β. Para isto é necessário considerar:

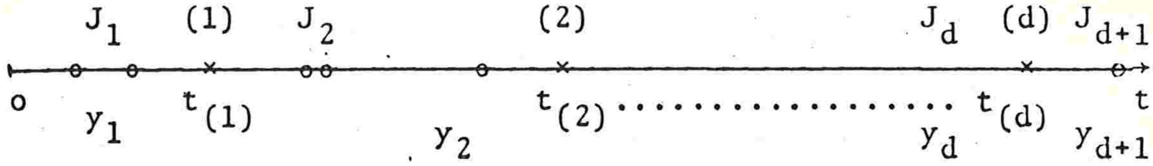
- $(t_{(1)},\ldots,t_{(d)})$, a estatística de ordem dos tempos de falha, onde $I=\{1,\ldots,d\}$;

- $((1),\ldots,(d))$, a estatística dos postos dos indivíduos que falharam;

- $J_i=\{j:\ j\in J \ \text{ e } \ t_{(i-1)}<w_j<t_{(i)}\}$, o conjunto dos indivíduos que se retiraram entre as $(i-1)$-ésima e a i-ésima falhas, $i=1,\ldots,d+1$ e

- $R_i = \left\{ S - \bigcup_{j=1}^{i} J_i - \bigcup_{j=1}^{i-1} \{(j)\} \right\}$, o conjunto dos indivíduos que estão em risco de falhar em $t_{(i)}$, $i=1,\ldots d$.

Uma possível amostra pode ser esquematizada como abaixo:



onde x: falha

o: retirada

Sejam as variáveis aleatórias:

y_1: $\{w_j: j \in J_1\}$, $\{t_{(1)}\}$, história até a 1ª falha;

(1): posto do indivíduo da 1ª falha;

y_2: $\{w_j: j \in J_2\}, \{t_{(2)}\}$, história depois da 1ª falha até a e2ª falha;

.
.
.
.
.
.

y_d: $\{w_j: j \in Jd\}$, $\{t_{(d)}\}$, história depois (d-1)-ésima falha até a d-ésima falha;

(d): posto do indivíduo da d-ésima falha.

y_{d+1}: $\{w_j: j \in J_{d+1}\}$, história depois da d-ésima falha.

A função de verossimilhança pode ser fatorada como segue:

$$L(\beta,\lambda_o(.)/x)=p(y_1)p(_{(1)}/y_1)p(y_2/y_1,(1))\ p(_{(2)}/y_1,(1),y_2)\ldots.$$

$$\ldots..\ p((d)/y_1,(1),\ldots.,(d-1),y_d).p(y_{d+1}/y_1,(1),y_2,\ldots,y_d,(d)).$$

O argumento anterior pode ser extendido a este modelo, uma vez que $\lambda_o(.)$ continua sendo desconhecida (arbitrária), logo, se β é o parâmetro de interesse considera-se apenas as probababilidades condicionais do tipo; $p((i)/y_1,(1),y_2,\dots,y_{i-1})$ e a função de verossimilhança parcial é dada por:

$$L(\beta/x) = p((1)/y_1)p((2)/y_1,(1),y_2)\dots p((d)/y_1,(1),y_2,\dots y_{d-1}).$$

Note que dar a informação y_1 significa dizer que (1) será um valor no conjunto R_1 e que $\{t_j \geq t_{(1)}: j \in R_1\}$. Assim, pelas duas observações feitas em 8.1, temos que dado y_1 os tempos de falha dos indivíduos de R_1 seguem um modelo com taxa de falha proporcional e com as mesmas taxas de falha. Portanto, pelo Teorema 8.1

$$p((1)/y_1) = \frac{\delta_{(1)}}{\sum_{j\in R_1} \delta_j}$$

Repetindo o argumento teremos:

$$p((i)/y_1,(1),y_2,\dots y_{i-1}) = \frac{\delta_{(i)}}{\sum_{j\in R_1} \delta_j} = \frac{e^{<\beta,z(i)>}}{\sum_{j\in R_i} e^{<\beta,z_j>}}$$

$$i=1,\dots,d$$

A função de verossimilhança parcial é dada por

$$L(\beta/x) = \prod_{i=1}^{d} \frac{e^{<\beta.z(i)>}}{\sum_{j\in R_i} e^{<\beta,z_j>}}$$

8.3 - Verossimilhança Parcial

Os métodos de eliminação de parâmetros nuisance aqui estudados (com exceção dos métodos do capítulo 7) se baseiam exclusivamente nas funções de verossimilhança marginal e condicional. Cox[1975] propôs uma generalização das verossimilhanças acima e a denominou de "Verossimilhança Parcial".

Suponha que a experiência total do cientista possa ser representada por

$$x = (y_1, z_1, y_2, z_2, \ldots, y_m, z_m, y_{m+1})^{[1]}$$

A função de verossimilhança da sequência pode ser dada por:

$$p(x/\theta,\phi) = p(y_1/\theta,\phi).p(z_1/y_1,\theta,\phi).p(y_2/y_1,z_1,\theta,\phi)\ldots$$

$$\ldots p(y_{m+1}/y_1,z_1,\ldots,z_m,\theta,\phi).$$

Suponha ainda que:

(a) Todos os fatores da forma $p(z_i/y_1,z_1,\ldots,y_i,\theta)$ são livres do parâmetro nuisance ϕ, o mesmo não acontecendo com nenhum dos fatores restantes.

---

[1] O número m pode depender da amostra x.

(b) As variáveis $(y_1,\dots,y_m)$ não devem "conter informação relevante sobre o parâmetro de interesse Θ, tal que a distribuição dos y_i's devem depender de maneira essencial do parâmetro nuisance".

Neste caso Cox[1975] chamou de verossimilhança parcial para Θ o produto:

$$L(\Theta/x) = p(z_1/y_1,\Theta)p(z_2/y_1,z_1,y_2,\Theta)\dots p(z_m/y_1,z_1,\dots,y_m,\Theta)$$

Algumas observações podem ser feitas. Em primeiro lugar, que dificilmente podemos chamar a isto de definição uma vez que não existe um procedimento que forneça a decomposição de x em $(y_1,z_1,\dots,y_m,z_{m+1})$. Uma amostra pode ser decomposta em mais de uma maneira e para outra pode não existir tal decomposição . Em segundo lugar é evidente que as verossimilhanças marginal e condicional são casos particulares de verossimilhança parcial . Note, entretanto, que em geral a verossimilhança parcial não pode ser interpretada como probabilidade. E finalmente que a condição (b) expressa uma noção vaga e intuitiva de não-informação, uma vez, que não se exige que a parte descartada independa de Θ.

Para finalizar nossa discussão, gostaríamos de chamar atenção para o fato de que quando a verossimilhança se fatora em duas componentes, uma dependendo apenas do parâmetro de interesse, o fato de basearmos nosso estudo apenas nesse fator pode, muitas vezes, simplificar o problema de tal forma que a perda de informação em não se considerar o segundo fator fica compensada. É bom lembrar que, na verdade, como não conseguimos controlar to dos os fatores que envolvem um experimento, na maioria das vêzes os nossos estudos são baseados em verossimilhanças parciais cujo fator complementar não é por nós conhecido.

APÊNDICE

Vamos supor as seguintes condições de regularidade para a família de distribuições $\mathcal{P} = \{p(x/\theta,\phi);\ \theta \in \Theta,\ \ \phi \in \Phi\}$.

(1) Θ é um intervalo real,

(2) $\dfrac{\partial \ell n p}{\partial \theta}$ e $\dfrac{\partial^2 \ell n p}{\partial \theta^2}$ existem para todo $\theta \in \Theta$, q.t.p,

(3) $\int p\, d\mu$ é diferenciável sob o sinal de integração com respeito a θ,

(4) quando a condição (a) da Definição 4.1 valer então as condições (2) e (3) devem valer também para os modêlos: condicional $p(x/t,\theta)$ e marginal $p(t/\theta,\phi)$.

As funções regulares de estimação $g(x,\theta) \in G_1$ devem satisfazer às seguintes condições: para todo $\lambda \in \Lambda$

(i) $E[g/\lambda] = 0$,

(ii) $\dfrac{\partial g}{\partial \theta}$ existe,

(iii) $\int g\, p\, d\mu$ é diferenciável com respeito a θ sob o sinal de integração,

(iv) $\left| E\left[\dfrac{\partial g}{\partial \theta} / \lambda \right] \right|^2 > 0$

Quando a condição (a) da Definição 4.1 valer então devem valer também as seguintes condições:

$$\text{(v)} \int g\left(\frac{\partial \ell n\ p(x/t,\theta)}{\partial \theta}\right) p\ d\mu \quad \text{e}$$

$$\int g\left(\frac{\partial \ell n\ h}{\partial \theta}\right) p\ d\mu \quad \text{existem, para todo } \lambda \in \Lambda\ .$$

<u>LEMA</u>: Se t é uma estatística G-ancilar (Definição 4.1) com respeito a θ então

$$\max_{g \in G_1}\{1/E\left[g/E\left(\frac{\partial g}{\partial \theta}|\lambda\right)|\ \lambda\right]^2\} \quad \text{é atingido para}$$

$$g^* = \frac{\partial \ell n\ p(x/t,\theta)}{\partial \theta}$$ e esta é a única função tal que isto acontece.

<u>Demonstração</u>:

Para simplificar a notação denota-se a "forma padronizada" da função de estimação por gs, ou seja,

$$gs = g/E\left(\frac{\partial g}{\partial \theta}\ |\lambda\ \right)$$

Não é difícil ver que $g^* = \dfrac{\partial\ \ell n\ P(x/t,\theta)}{\partial \theta} \in G_1$.

Quer-se demonstrar que

$$E(gs^{*2}|\lambda) \leq E(gs^2|\lambda)^2 \quad , \forall\ g \in G_1,$$

onde $gs^* = g^*/E\left(\dfrac{\partial g^*}{\partial \theta}\ |\lambda\right)$.

Tem-se que

$$E[(gs-gs^*)^2|\lambda]=E[gs^2|\lambda]-2E[gsgs^*|\lambda]+E[gs^{*2}|\lambda] \geq 0$$

Portanto, basta demonstrar que

$$E[gsgs^*|\lambda] = E[gs^{*2}|\lambda] \quad .$$

Desenvolvendo o primeiro membro da igualdade acima tem-se que:

$$E[gsgs^*/\lambda] = E\left[gg^*/E(\frac{\partial g}{\partial\theta}|\lambda)\ E(\frac{\partial g^*}{\partial\theta}|\lambda)|\lambda\right]$$

$$= E(gg^*|\lambda)/E(\frac{\partial g}{\partial\theta}|\lambda)\ E(\frac{\partial g^*}{\partial\theta}|\lambda) \qquad [1]$$

Mas

$$E(gg^*|\lambda) = E(g \cdot \frac{\partial}{\partial\theta} \ln p(x/t,\theta)|\lambda) =$$

$$= E(g \cdot \frac{\partial}{\partial\theta} p(x/\lambda)|\lambda) - E(g \cdot \frac{\partial}{\partial\theta} \ln p(t/\lambda)|\lambda) \qquad [2]$$

Desenvolvendo as duas esperanças acima

$$E(g \cdot \frac{\partial}{\partial\theta} \ln p(x/\lambda)) = \int g \frac{\partial \ln p(x/\lambda)}{\partial\theta} \cdot p(x/\lambda)\, d\mu =$$

$$= \int g \frac{\partial}{\partial\theta} \cdot p(x/\lambda) d\mu = \int \frac{\partial}{\partial\theta}[g \cdot p(x/\lambda)]\, d\mu - \int (\frac{\partial}{\partial\Theta} g)\ p(x/\lambda) d\mu$$

$$= - E(\frac{\partial}{\partial\theta} g) \quad [3] \quad [\text{por (i) e (iii)}] \; .$$

$$E\left[g \frac{\partial}{\partial\theta} \ell n\, p(t/\lambda) | \lambda\right] = E\left\{E\left[g.\frac{\partial}{\partial\theta} \ell n\, p(t/\lambda) | t,\lambda\right] \lambda\right\}$$

$$= E\left[\frac{\partial}{\partial\theta} \ell n\, P(t/\lambda)\; E(g/t,\lambda) |\; \lambda\right] = [\text{Propriedade de Esperança Condicional}]$$

Pela condição (b) da DEFINIÇÃO 4.1, para θ fixado , tem-se que

$$P_{\theta,t} = \{P(t/\theta,\phi);\; \phi \in \Phi\} \text{ é completa.}$$

Logo, se

$$E(g|\lambda\,) = E[E(g|t,\lambda)|\lambda] = 0 \, , \quad [\text{por (i)}]$$

pela condição de completividade acima,

$$E[g|t,\lambda\,] = 0.$$

Portanto

$$E[\frac{\partial}{\partial\theta} \ell n\, P(t/\lambda)\; E(g/t,\lambda)|\lambda\,] = 0$$

Substituindo as esperanças em [2], vem que

$$E\,[gg^*] = - E[\frac{\partial}{\partial\theta} g|\; \lambda\,]$$

Logo

$$E[gsgs^* \;\; \lambda] = \frac{- E[\frac{\partial}{\partial\theta} g|_\lambda\,]}{E[\frac{\partial}{\partial\theta} g|_\lambda]\cdot E[\frac{\partial}{\partial\theta} g^*|\lambda\,]} =$$

$$= - \frac{1}{E\left[\frac{\partial^2}{\partial\theta^2} \ell n\ p(x/t,\theta)\,|\,\lambda\right]} = E[g_s^{*2}|\lambda]$$

c.q.d.

A demonstração da unicidade da função $\frac{\partial \ell n\ p(x|t,\theta)}{\partial\theta}$ é omitida aqui por se tratar de uma demonstração técnica apenas. Pode-se encontrá-la em Godambe [1976].

TEOREMA 4.4 - Se a estatística t é G-ancilar com respeito a θ então $i(\theta)=E[i_t(\theta)|\lambda]$

Demonstração:

Pelo Lema tem-se que a informação $i(\theta)$, da distribuição $p(x|\lambda)$, sobre θ é dada por:

$$i(\theta) = \left\{1/E\left[\frac{\partial \ell n\ p(x/t,\theta)}{\partial\theta} \Big/ E\left(\frac{\partial^2 \ell n\ p(x/t,\theta)}{\partial\theta^2}/\lambda\right)\right]^2 |\lambda\right\}$$

$$= E\left[-\frac{\partial^2 \ell n p(x/t,\theta)}{\partial\theta^2}\,\Big|\,\lambda\right]$$

Como t é G-ancilar com respeito a θ, a distribuição condicional $p(x/t,\theta)$ não depende de ϕ e portanto a informação $i_t(\theta)$ coincide com $I_t(\theta)$.

$$i_t(\theta) = I_t(\theta) = E\left[-\frac{\partial^2 \ell n\ p(x/t,\theta)}{\partial\theta^2}\,\Big|\,t,\theta\right]$$

$$\therefore\ E\left[i_t(\theta)\right] = E\left\{ E\left[- \frac{\partial^2 \ell n\ p(x/t,\theta)}{\partial\theta} \,\middle|\, t,\theta \right] \middle|\, \lambda \right\}$$

$$= E\left[- \frac{\partial\ \ell n\ p(x/t,\theta)}{\partial\theta^2} \,\middle|\, \lambda \right] = i(\theta)$$

c.q.d.

REFERÊNCIAS BIBLIOGRÁFICAS

- Barnard, G.A. (1963): Some logical aspects of the Fiducial Argument. J.R.S.S., B25, 111-4.

- Barndorff - Nielsen, O. (1978): Information and Exponential Families in Statistical Theory. John Wiley, New York.

- Basu, D. (1977): On the Elimination of Nuisance Parameters. JASA 72, 355-66.

- ——— (1978): On Partial Sufficiency: A review. J.Statist. Plan. Inf., 2, 1-13.

- ——— (1979): Discussion of Joseph Berkson's Paper "In dispraise of the exact test". J. Statis. Plan. Inf., 3, 189-197

- ——— (1982): A Note on likelihood. Atas do 5º Simpósio Nacional de Probabilidade e Estatística (a ser publicado).

- Berger, J.O. (1980): Statistical Decision Theory.Springer--Verlag, New York.

- Cox, D.R. (1972): Regression models and life tables (with discussion) J.R.S.S., B 34, 187-220.

- Cox, D.R. (1975): Partial likehood. Biometrika, 62, 269--276.

- Dawid, A.P. (1981): A Bayesian Look at Nuisance Parameters. Bayesian Statistics Proceeding. Valência, University Press.

- Ferreira, P.E. (1980): Comments on Berkson's Paper "In Dispraise of the exact test". (unpuplished report).

- Fisher, R.A. (1920): A Mathematical Examination of the Methods of determining the accuracy of an observation by the mean error, and the mean square error. Monthy Notices of the Royal Astronomical Soc., 80, 758-770. (também reproduzido em Fisher (1950)).

- ——— (1922): On the Mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. London A 222, 309-368 (também reproduzido em Fisher (1950)).

- ——— (1950): Contributions to Mathematical Statistics, John-Wiley & Sons.

- Fraser D.A.S. (1956): Sufficient Statistics with nuisance parameters. Ann. Math. Statist. 27, 838-842.

- Godambe, V.P. (1960): An optimum property of regular maximum likelehood estimation. Ann Math. Statist., 31, 1205-11.

- Godambe, V.R. (1976): Conditional likelihood and unconditional optimun estimating equations. Biometrika, 63, 2, 277-284.

- ——— (1980): On Sufficiency and Ancilearity in presence of a nuisance parameter. Biometrika, 67, 155-162.

- Hall, W.J.; Wijsman, R.A. & Ghosh, J.K. (1965): The relationship between sufficiency and Invariance with applications in sequential analysis. Ann. Math. Statist.,36, 575-614.

- Hajek, S. (1965): On Basic Concepts of Statistics. Proceedings of the fifth Berkeley Simposium, 1, 139-62.

- Kalbfleisch, J.D. and Prentice, R.L. (1980): The Statistical Analysis of failure time data. John Wiley & Sons.

- Lindley, D.V. (1979): Analysis of the Tables with groufing and with drawals. Biometrics, 35, 605-612.