Para alguns problemas, não existe única estatística suficiente. No entanto, existirão estatísticas conjuntamente suficientes. (19)

Problema 1: Até o momento, vimos o caso uniparamétrico, ou seja, a distribuição dos dados depende de um único parâmetro θ. Vamos considerar o caso multiparamétrico em que θ é um vetor de parâmetros, que denotamos por $\underset{\sim}{\theta}$. Em muitas situações, o modelo estatístico depende de mais de um parâmetro.

* Modelo $N(\mu, \sigma^2)$, em que $\underset{\sim}{\theta} = (\mu, \sigma^2)$, sendo μ e σ^2 desconhecidos.

* Modelo Gama(α, β), em que α e β são desconhecidos e, portanto, $\underset{\sim}{\theta} = (\alpha, \beta)$.

Problema 2: Nos exemplos a seguir, observamos que não é possível uma redução substancial da amostra. Suponha que f seja a fdp de Cauchy $f(x|\theta) = \frac{1}{\pi(x-\theta)^2}$ ou a fdp logística $f(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$. Então, segue que a densidade da amostra é dada por

$$f(\underset{\sim}{x}|\theta) = f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} f(x_{(i)}),$$

onde $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ são as estatísticas de ordem. De acordo com o critério da fatoração, podemos mostrar que as estatísticas de ordem são uma estatística suficiente. Assim, a redução para as estatísticas de ordem é o máximo que conseguimos obter nestas famílias.

Casella e Berger pág 245

Comentário: (Lehmann e Casella, 1998, Seção 1.6)

Acontece que, fora da família de distribuições exponenciais, é raro ter uma estatística suficiente de dimensão menor que o tamanho da amostra, portanto, em muitos casos as estatísticas de ordem são o melhor que podemos fazer.

Definição: (Estatísticas conjuntamente suficientes)

Seja $X_1, \dots, X_n$ aa $X \sim f(\cdot \mid \underset{\sim}{\theta})$, $\underset{\sim}{\theta} \in \Theta$. Dizemos que as estatísticas $S_1, \dots, S_k$ são conjuntamente suficientes para $\underset{\sim}{\theta}$ se e somente se a distribuição condicional de $X_1, \dots, X_n$ dado $S_1 = s_1, \dots, S_k = s_k$ não depende de $\underset{\sim}{\theta}$.

Comentários: * Pode ser complicado utilizar a Definição de estatísticas suficientes com a finalidade de encontrar estatísticas suficientes para um modelo particular. Em geral, a Definição é útil para mostrar que uma particular estatística não é suficiente.

* O critério da fatoração permite encontrar estatísticas suficientes pela simples inspeção da fdp ou fp da amostra.

* O critério da fatoração é útil para

mostrar que uma estatística ou um conjunto de estatísticas é suficiente. Em geral, não é útil para mostrar que uma estatística ou um conjunto de estatísticas não é suficiente. O fato de não conseguir fatorar a densidade conjunta não significa que não existem estatísticas suficientes (pode ser que não seja capaz de encontrar a fatoração correta).

Teorema : (critério da fatoração - estatísticas conjuntamente suficientes). (21)

Sejam $X_1, \ldots, X_n$ uma amostra aleatória proveniente da densidade (ou função de probabilidade) $f(\cdot \mid \underset{\sim}{\theta})$. O conjunto de estatísticas $S_1 = S_1(X_1, \ldots, X_n), \ldots, S_r = S_r(X_1, \ldots, X_n)$ é conjuntamente suficiente se, e somente se, a densidade conjunta de $X_1, \ldots, X_n$ pode ser fatorada como

$$f(x_1, \ldots, x_n \mid \underset{\sim}{\theta}) = g(s_1, \ldots, s_r \mid \underset{\sim}{\theta})\, h(x_1, \ldots, x_n),$$

onde a função $h(x_1, \ldots, x_n)$ é não negativa e não envolve o parâmetro $\underset{\sim}{\theta}$ e a função $g(s_1, \ldots, s_r \mid \underset{\sim}{\theta})$ é não negativa e depende de $x_1, \ldots, x_n$ apenas através das funções $s_1 = S_1(x_1, \ldots, x_n), \ldots, s_r = S_r(x_1, \ldots, x_n)$.

Comentário: No caso do Teorema, dizemos que a estatística suficiente é de dimensão r, ou seja, $\underset{\sim}{S} = (S_1, \ldots, S_r)$, $S_i = S_i(\underset{\sim}{X})$, que em muitos casos é também a dimensão do espaço paramétrico Θ. Porém, existem situações em que tal fato não ocorre, ou seja, a dimensão de Θ é menor que r, por exemplo. 2k

<u>Exemplo 1</u>: Seja $X_1, \dots, X_n$ a.a. $X \sim U(\theta_1, \theta_2)$

Encontre a estatística suficiente para $\underset{\sim}{\theta} = (\theta_1, \theta_2)$.

$$f(x_1, \dots, x_n \mid \underset{\sim}{\theta}) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \theta_2)}(x_i) =$$

$$\frac{1}{(\theta_2 - \theta_1)^n} \prod_{i=1}^{n} I_{(\theta_1, \theta_2)}(x_i).$$ Note que

$$\prod_{i=1}^{n} I_{(\theta_1, \theta_2)}(x_i) = 1 \iff \theta_1 < x_i \leq \theta_2 \quad \forall i.$$

Então,

$$\theta_1 \leq x_{(1)} \leq x_{(n)} \leq \theta_2 \quad \text{e}$$

$$f(x_1, \dots, x_n \mid \underset{\sim}{\theta}) = \frac{1}{(\theta_2 - \theta_1)^n} I_{(\theta_1, x_{(n)})}(x_{(1)}) \, I_{(x_{(1)}, \theta_2)}(x_{(n)}).$$

Portanto, pelo critério da fatoração, temos que $h(x_1, \dots, x_n) = 1$ e $g(x_{(1)}, x_{(n)} \mid \underset{\sim}{\theta}) =$

$$\frac{1}{(\theta_2-\theta_1)^n} \; I_{(\theta_1,\, x_{(n)})}(x_{(1)}) \; I_{(x_{(1)},\, \theta_2)}(x_{(n)})$$ e as

estatísticas $X_{(1)}$ e $X_{(n)}$ são conjuntamente suficientes ou $(X_{(1)}, X_{(n)})$ é uma estatística suficiente p/ $\underset{\sim}{\theta}$.

<u>Exemplo 2</u>: Seja $X_1, \dots, X_n$ a.a $X \sim N(\mu, \sigma^2)$.

Encontre a estatística suficiente para $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

Note que $$f(x_1,\dots,x_n \mid \underset{\sim}{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-1}{2\sigma^2}(x_i-\mu)^2 \right\} I_{(0,\infty)}(x_i) =$$

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 \right\} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i^2 - 2x_i\mu + \mu^2) \right\} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)$$

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n}{2\sigma^2}\mu^2 \right\}}_{g\left(\sum_{i=1}^{n} x_i,\, \sum_{i=1}^{n} x_i^2 \,\mid\, \mu, \sigma^2\right)} \underbrace{\prod_{i=1}^{n} I_{(0,\infty)}(x_i)}_{h(x_1,\dots,x_n)}$$

Assim, as estatísticas $\sum_{i=1}^{n} X_i$ e $\sum_{i=1}^{n} X_i^2$ são conjuntamente suficientes ou $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ é uma estatística suficiente p/ $\underset{\sim}{\theta}$. 24

<u>Observação</u>: Note que $\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$.

Assim,

$$f(x_1, \ldots, x_n \mid \underset{\sim}{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right\} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)$$

e pelo critério da fatoração, teríamos que $\left(\bar{X}, \sum_{i=1}^{n} (X_i - \bar{X})^2\right)$ é ~~...~~ uma estatística suficiente p/ $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

Além disso, observe que podemos reescrever $f(x_1, \ldots, x_n \mid \underset{\sim}{\theta})$ como

$$f(x_1, \ldots, x_n \mid \underset{\sim}{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2}(n-1)s^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2 \right\} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)$$

, onde

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

e pelo critério da fatoração, teríamos que $(\bar{X}, S^2)$ é uma estatística suficiente para $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

Vd comentário folha 26 (a)

<u>Comentário</u>: * Este exemplo demonstra que, para o modelo normal, fica justificada a prática comum de resumir um conjunto de dados relatando somente a média e a variância amostrais. A estatística suficiente $(\bar{X}, S^2)$ contém todas as informações sobre (μ, σ^2) que estão disponíveis na amostra. Porém, devemos ~~nos~~ lembrar de que a definição de uma estatística suficiente depende do modelo. Para outro modelo, isto é, outra família

de densidades, a média e a variância (26)
amostrais podem não ser uma estatística
suficiente para a média e a variância da
população. O pesquisador que calcula
somente $\bar{X}$ e S^2, e ignora totalmente
o restante dos dados está acreditando muito
na suposição do modelo normal.

* **Definição**: Dizemos que duas estatísticas
T_1 e T_2 são equivalentes se existir uma
relação 1:1 entre elas.

Em outras palavras, T_1 e T_2 são equivalentes
se T_1 puder ser obtida a partir de T_2 e
vice-versa. Neste caso, temos que, se
T_1 é suficiente para θ, então T_2 também
é suficiente para θ. Esse resultado
vale também para o caso multidimensional.

<u>Teorema</u>: Se $S_1, \dots, S_r$ é um conjunto de estatísticas suficientes, então qualquer conjunto de funções 1:1, ou transformações, de $S_1, \dots, S_r$ é também conjuntamente suficiente.

<u>Exemplo 1</u>: Seja $X_1, \dots, X_n$ a.a. $X \sim N(\mu, 1)$ e pelo critério da fatoração, temos que $\sum_{i=1}^{n} X_i$ é suficiente para μ / (exercício 1, pág 16).

Como $S_1 = \sum_{i=1}^{n} X_i$ é equivalente a $S_2 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, temos que S_2 também é suficiente para μ.

$$g: \mathbb{R} \to \mathbb{R}$$
$$(x) \to \left(\frac{1}{n} x\right)$$

<u>Exemplo 2</u>: Seja $X_1, \dots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$ e pelo critério da fatoração, temos que $\sum_{i=1}^{n} X_i$ e $\sum_{i=1}^{n} X_i^2$ são conjuntamente suficientes para $\underset{\sim}{\theta} = (\mu, \sigma^2)$ / (exemplo 2, pág 23).

Como $\underset{\sim}{S_1} = \left(\sum_{i=1}^{n} X_i , \sum_{i=1}^{n} X_i^2 \right)$ é equi- (28)

valente a $\underset{\sim}{S_2} = (\bar{X} , S^2)$, temos

que $\bar{X}$ e S^2 são também conjun-

tamente suficientes para $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

Bickel pág 85

exemplos

25 ②

Comentário

Considere $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ sendo um parâmetro bivariado. Suponha que $T_1(\underset{\sim}{X})$ é suficiente para θ_1 se θ_2 é fixo e conhecido, enquanto $T_2(X)$ é suficiente para θ_2 se θ_1 é fixo e conhecido. Assuma que θ_1 e θ_2 variem independentemente, $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ e que o conjunto $S = \{x \mid p(x \mid \underset{\sim}{\theta}) > 0\}$ não depende de θ.

suporte da distribuição

1 → Se T_1 e T_2 não dependem de θ_2 e θ_1, respectivamente, então $(T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X}))$ é suficiente para $\underset{\sim}{\theta}$. ⇒ EXEMPLOS: Distribuição gama / Distribuição Beta, ...

exercícios dos folhas 3 e 4 da parte de família exponencial

$X_1, \ldots, X_n$ a.a. $N(\mu, \sigma^2)$, μ conhecido

$$f(x \mid \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\left(\prod_{i=1}^{n} I_{(-\infty,\infty)}(x_i)\right)$$

$$\boxed{t(\underset{\sim}{X}) = \sum_{i=1}^{n}(x_i-\mu)^2.}$$ suficiente para σ^2 quando μ fixo e conhecido.

$X_1, \ldots, X_n$ a.a. $N(\mu, \sigma^2)$, σ^2 conhecido.

Neste caso, $\boxed{t(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i}$ suficiente para μ quando σ^2 fixo e conhecido.

$X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$, μ e σ^2 desconhecidos

$$t(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right) \text{ ou}$$

$t(\underset{\sim}{X})$ $(\bar{X}, S^2)$ é suficiente para $\underset{\sim}{\theta} = (\mu, \sigma^2)$.

Exemplo no qual $(T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X}))$ é suficiente para $\underset{\sim}{\theta}$, $T_1(\underset{\sim}{X})$ é suficiente para θ_1 se θ_2 é fixo e conhecido, mas $T_2(\underset{\sim}{X})$ não é suficiente para θ_2, quando θ_1 é fixo e conhecido.

É fácil encontrar uma estatística suficiente para uma família de distribuições exponenciais utilizando o Teorema da Fatoração.

Família Exponencial Multiparamétrica

Dizemos que a distribuição da variável aleatória X pertence à família exponencial multiparamétrica de distribuições, se pudermos escrever sua fp ou fdp como

$$f(x|\underset{\sim}{\theta}) = h(x)\, c(\underset{\sim}{\theta}) \exp\left\{ \sum_{i=1}^{k} w_i(\underset{\sim}{\theta})\, t_i(x) \right\},$$

$x \in \mathcal{X}$ e $\underset{\sim}{\theta} \in \Theta$, onde $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_d)$, $d \leq k$. Além disso, temos a seguinte família exponencial multiparamétrica de distribuições associada à amostra $(X_1, \dots, X_n$ a a $X \sim f(\cdot|\underset{\sim}{\theta}))$.

$$f(x_1, \dots, x_n|\underset{\sim}{\theta}) = \prod_{j=1}^{n} h(x_j) \left[c(\underset{\sim}{\theta})\right]^n \exp\left\{ \sum_{i=1}^{k} w_i(\underset{\sim}{\theta}) \sum_{j=1}^{n} t_i(x_j) \right\}.$$

Resultado Importante !!

Note que pelo critério da fatoração, as estatísticas $\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)$

são conjuntamente suficientes ou

$\left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j) \right)$ é uma estatística suficiente para $\underset{\sim}{\theta}$.

Comentários Finais: (Resultados úteis) !!

Comentário I:

É possível verificar que os estimadores de máxima verossimilhança dependem da amostra $X_1, \ldots, X_n$ através de estatísticas suficientes.

Teorema: Um estimador de máxima verossimilhança ou um conjunto de estimadores de máxima verossimilhança dependem da amostra através de algum conjunto de estatísticas suficientes (ou estatísticas conjuntamente suficientes).

PROVA: Se $S_1, \dots, S_k$ são estatísticas conjuntamente suficientes, então a função de verossimilhança pode ser escrita como $L(\underset{\sim}{\theta} \mid x_1, \dots, x_n) =$

$$L(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(x_i \mid \underset{\sim}{\theta}) = f(x_1, \dots, x_n \mid \underset{\sim}{\theta}) =$$

$g(s_1, \dots, s_k \mid \underset{\sim}{\theta})\, h(x_1, \dots, x_n)$. Como função de $\underset{\sim}{\theta}$, $L(\underset{\sim}{\theta})$ terá seu máximo no mesmo lugar que $g(s_1, \dots, s_k \mid \underset{\sim}{\theta})$ (no mesmo local) tem seu máximo, mas o local (o ponto) onde g atinge seu máximo pode depender de $x_1, \dots, x_n$ apenas através de $s_1, \dots, s_k$.

É importante notar que os estimadores de momentos podem não ser funções de estatísticas suficientes.

Comentário II :

A própria amostra $X_1, \dots, X_n$ é sempre conjuntamente suficiente uma vez que a distribuição condicional da amostra dado a amostra não depende de θ (pode ser vetor, $\underset{\sim}{\theta}$).

Além disso, as ~~es~~ estatísticas de ordem $X_{(1)}, \dots, X_{(n)}$ são também conjuntamente suficientes ~~xxxxx~~ para θ (pode ser vetor, $\underset{\sim}{\theta}$).

Mood pág 254

Teorema (caso contínuo)

Seja $X_1, \dots, X_n$ a a $X \sim f(\cdot)$ com fda $F(\cdot)$. Considere $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ as estatísticas de ordem correspondentes, então

$$f_{X_{(1)}, \dots, X_{(n)}}(y_1, \dots, y_n) = \begin{cases} n!\, f(y_1) \cdots f(y_n) & \text{se } y_1 < y_2 < \dots < y_n \\ 0 & \text{cc} \end{cases}$$

Pela Definição ou

Pelo critério da fatoração, pode-se mostrar

que $X_{(1)}, \dots, X_{(n)}$ são conjuntamente

suficientes para θ (pode ser vetor, $\underset{\sim}{\theta}$) !!