Estatísticas Suficientes Minimais ou Estatísticas Suficientes Mínimos

Casella e Mood

Introdução

* Quando nós introduzimos o conceito de suficiência, dissemos que nosso objetivo era a redução dos dados sem perda de qualquer informação sobre o parâmetro. * Vimos que há um conjunto de estatísticas suficientes ou em outras palavras, todos os modelos têm estatística suficiente (própria amostra). * Por exemplo, seja $X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$, vimos que $X_1, \ldots, X_n$ (própria amostra), $X_{(1)}, \ldots, X_{(n)}$ (estatísticas de ordem) e $\bar{X}$ e S^2 são três conjuntos de estatísticas conjuntamente suficientes. * Naturalmente, preferimos as estatísticas conjuntamente suficientes $\bar{X}$ e S^2 uma vez que elas condensam os dados mais do que os outros dois conjuntos de estatísticas. * Questão: Existe um conjunto de estatísticas suficientes que condensa os dados mais que $\bar{X}$ e S^2? Até o momento, não temos ferramentas para responder tal

questionamento. A questão de que estamos aludindo é a de um conjunto mínimo de estatísticas suficientes, que chamaremos de estatísticas suficientes mínimas ou mínimais.

Este questionamento nos leva ao conceito de estatística suficiente minimal ou estatística suficiente mínima.

Comentários ① ⑤

Objetivo: Procurar a maior redução dos dados sem perda de informação sobre o parâmetro θ.

Lembre-se de que o propósito de uma estatística suficiente é obter a redução de dados sem perder informações sobre o parâmetro θ. Deste modo, uma estatística que consegue a maior redução de dados enquanto ainda mantém todas as informações sobre θ pode ser considerada preferível.

A definição de tal estatística é formalizada a seguir.

Notemos anteriormente que uma estatística induz a uma partição do espaço amostral ($\mathcal{X}$). O mesmo é verdadeiro para um conjunto de estatísticas. Um conjunto de estatísticas induz uma partição do $\mathcal{X}$.

De uma forma geral, a condensação dos dados que uma estatística ou um conjunto de estatísticas exibe pode ser medida (mensurada) pelo número de subconjuntos na partição induzida pela estatística ou pelo conjunto de estatísticas. Se um conjunto de estatísticas tem menos subconjuntos na sua partição induzida do que na partição induzida por outro conjunto de estatísticas, então dizemos que a 1ª estatística condensa mais os dados do que a última.

Desta forma, um conjunto de estatísticas suficientes minimais é então um conjunto de estatísticas suficientes que tem menos subconjuntos na sua partição do que a partição induzida por qualquer outro conjunto de estatísticas suficientes. Então um conjunto de estatísticas suficientes é minimal se não há outro conjunto de estatísticas suficientes que condensa mais os dados.

<u>Definição</u> : (<u>Estatística Suficiente Minimal</u>)

Uma estatística suficiente é minimal se for função de qualquer estatística suficiente. Isto é, $T = T(\underset{\sim}{X})$ é suficiente minimal se $\forall\ U = U(\underset{\sim}{X})$ suficiente $\exists$ uma função ρ, tal que $T = \rho(U)$.

<u>Objetivo</u> : Procurar a maior redução de informação.

<u>Comentários</u> : 1 - T produz uma maior redução de dados do que U a menos que ρ seja bijetora (função 1-1), em cujo caso T e U são ditas equivalentes.

2 - As estatísticas suficientes mínimas são únicas?

(Infinitas)! Qualquer função um a um pode ser considerada, mas são ditas equivalentes, e neste sentido poderíamos dizer que são únicas.

3 - Nem sempre $\exists$ estatística suficiente minimal.

1 ① - ⑤

Notemos anteriormente que uma estatística induz a uma partição do espaço amostral (X). O mesmo é verdadeiro para um conjunto de estatísticas, um conjunto de estatísticas induz uma partição do $\mathcal{X}$.

De uma forma geral, a condensação dos dados que uma estatística ou um conjunto de estatísticas exibe pode ser medida (mensurada) pelo número de subconjuntos na partição induzida pela estatística ou pelo conjunto de estatísticas. Se um conjunto de estatísticas tem menos subconjuntos na sua partição induzida do que na partição induzida por outro conjunto de estatísticas, então dizemos que a 1ª estatística condensa mais os dados do que a última.

Desta forma, um conjunto de estatísticas suficientes minimais é então um conjunto de estatísticas suficientes que tem menos subconjuntos na sua partição do que a partição induzida por qualquer outro conjunto de estatísticas suficientes. Então um conjunto de estatísticas suficientes é minimal se não há outro conjunto de estatísticas suficientes que condensa mais os dados.

Casella

Definição: Uma estatística suficiente $T(\underset{\sim}{X})$ é chamada estatística suficiente mínima se, para qualquer outra estatística suficiente $T'(\underset{\sim}{X})$, $T(\underset{\sim}{X})$ é uma função de $T'(\underset{\sim}{X})$.

Dizer que $T(X)$ é uma função de $T'(X)$ simplesmente significa que se $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$, então $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$. Em termos de partição, se $\{B_{t'} : t' \in \mathcal{T}'\}$ é a partição para $T'(\underset{\sim}{x})$

↳ subconjuntos da partição induzida por T

e $\{A_t : t \in \mathcal{T}\}$ é a partição induzida por $T(\underset{\sim}{x})$. Então, pela definição temos que cada $B_{t'}$ é um subconjunto de algum A_t.

Exemplo 1: Seja $X_1, \ldots, X_n$ a a $X \sim \text{Bernoulli}(\theta)$

Mostre que $T = \sum_{i=1}^{n} X_i$ é suficiente minimal para θ.

Já demonstramos que $T = \sum_{i=1}^{n} X_i$ é suficiente para θ.

(qualquer outra estatística suficiente)

Seja $S(\underset{\sim}{X})$ uma estatística suficiente para θ.

Como S é suficiente para θ, então pelo teorema da fatoração $\exists$'n g e h tais que

$$f(\underset{\sim}{x} \mid \theta) = \theta^{t} (1-\theta)^{n-t} \prod_{i=1}^{n} I_{\{0,1\}}(x_i) = g(S(\underset{\sim}{x}) \mid \theta)\, h(\underset{\sim}{x}),$$

onde $t = \sum_{i=1}^{n} x_i$, $\forall \theta \in \Theta$.

para quaisquer dois valores fixos de θ

Considere θ_1 e θ_2 (fixos) dois possíveis valores de θ, tal que

$$\frac{f(\underset{\sim}{x} \mid \theta_1)}{f(\underset{\sim}{x} \mid \theta_2)} = \frac{\theta_1^{t} (1-\theta_1)^{n-t}}{\theta_2^{t} (1-\theta_2)^{n-t}} \frac{\prod_{i=1}^{n} I_{\{0,1\}}(x_i)}{\prod_{i=1}^{n} I_{\{0,1\}}(x_i)} = \frac{g(s \mid \theta_1)\, \cancel{h(\underset{\sim}{x})}}{g(s \mid \theta_2)\, \cancel{h(\underset{\sim}{x})}}$$

$$\doteq H(s \mid \theta_1, \theta_2).$$

$$\Rightarrow \quad \left(\frac{\theta_1}{\theta_\alpha}\right)^t \left(\frac{1-\theta_1}{1-\theta_\alpha}\right)^{n-t} = H.$$

Vamos aplicar o logaritmo para isolar t.

Assim,

$$t \ln\left(\frac{\theta_1}{\theta_\alpha}\right) + (n-t)\ln\left(\frac{1-\theta_1}{1-\theta_\alpha}\right) = H \quad e$$

$$t\left[\ln\left(\frac{\theta_1}{\theta_\alpha}\right) - \ln\left(\frac{1-\theta_1}{1-\theta_\alpha}\right)\right] = H - n\ln\left(\frac{1-\theta_1}{1-\theta_\alpha}\right)$$

$$\Leftrightarrow \quad t = \frac{H - n\ln\left(\frac{1-\theta_1}{1-\theta_\alpha}\right)}{\ln\left(\frac{\theta_1}{\theta_\alpha}\right) - \ln\left(\frac{1-\theta_1}{1-\theta_\alpha}\right)} \quad e$$

$\rightsquigarrow$ sfrás

portanto, t é uma função de s, ou seja, $T = g(S(\underset{\sim}{X}))$ e então,

T é suficiente minimal.

Utilizar a <u>Definição</u> para encontrar uma estatística suficiente minimal não é prático. Talvez seja preciso adivinhar que $T(\underset{\sim}{X})$ era uma estatística suficiente mínima e, então verificar a condição na <u>Definição</u>. Felizmente, o resultado a seguir, de Lehmann e Scheffé, apresenta uma maneira mais fácil para encontrar uma estatística suficiente mínima.

<u>Teorema</u>: Sejam $X_1, \ldots, X_n$ aa $X \sim f(\cdot|\theta)$ e $S(\underset{\sim}{X})$ uma estatística. Se para cada dois pontos amostrais $\underset{\sim}{x}$ e $\underset{\sim}{y}$, a razão

$$\frac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)} = c(\underset{\sim}{x}, \underset{\sim}{y}) \iff S(\underset{\sim}{x}) = S(\underset{\sim}{y}),$$

onde $c(\underset{\sim}{x}, \underset{\sim}{y})$ é uma função somente de $\underset{\sim}{x}$ e $\underset{\sim}{y}$ e $\mathcal{X}$. Então, $S(\underset{\sim}{X})$ é uma estatística suficiente minimal.

($\Rightarrow$ só vale a ida: condição necessária mas não suficiente).

Exemplo 2: Seja $X_1, \dots, X_n$ aa $X \sim \exp(\lambda)$, $\lambda > 0$. (6)

Encontre uma estatística suficiente minimal para λ.

Considere $\underset{\sim}{x}$ e $\underset{\sim}{y} \in \mathcal{X}$, a razão

$$\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)} = \frac{\cancel{\lambda^n} \, e^{-\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)}{\cancel{\lambda^n} \, e^{-\lambda \sum_{i=1}^{n} y_i} \prod_{i=1}^{n} I_{(0,\infty)}(y_i)} =$$

$$e^{-\lambda \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right)} \frac{\prod_{i=1}^{n} I_{(0,\infty)}(x_i)}{\prod_{i=1}^{n} I_{(0,\infty)}(y_i)} = c(\underset{\sim}{x}, \underset{\sim}{y})$$

$\Leftrightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. Então, $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$

é uma estatística suficiente minimal para λ.

Exemplo 3: Seja $X_1, \ldots, X_n$ a.a. $X \sim U(0, \theta)$. Encontre uma estatística suficiente minimal para θ. 7

Considere $\underset{\sim}{x}$ e $\underset{\sim}{y} \in X$, a razão

$$\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)} = \frac{\cancel{\left(\frac{1}{\theta}\right)^n} I_{(0,\theta)}(x_{(n)})}{\cancel{\left(\frac{1}{\theta}\right)^n} I_{(0,\theta)}(y_{(n)})} = c(\underset{\sim}{x}, \underset{\sim}{y}) \iff$$

$x_{(n)} = y_{(n)}$. Então, $S(\underset{\sim}{x}) = X_{(n)}$ é uma estatística suficiente minimal para θ.

Exemplo 4: Seja $X_1, \ldots, X_n$ a.a. $X \sim U(\theta, \theta+1)$, ~~[illegible]~~ $\theta > 0$.

Encontre a estatística suficiente minimal para θ.

Considere $\underset{\sim}{x}$ e $\underset{\sim}{y} \in X$, a razão

$$\frac{f(\underset{\sim}{x} \mid \theta)}{f(\underset{\sim}{y} \mid \theta)} = \frac{1 \prod_{i=1}^{n} I_{(\theta,\theta+1)}(x_i)}{1 \prod_{i=1}^{n} I_{(\theta,\theta+1)}(y_i)} = \frac{I_{(\theta, x_{(n)})}(x_{(1)}) \, I_{(x_{(1)}, \theta+1)}(x_{(n)})}{I_{(\theta, y_{(n)})}(y_{(1)}) \, I_{(y_{(1)}, \theta+1)}(y_{(n)})} =$$

$$= I_{(x_{(n)}-1,\ x_{(1)})}(\theta) \Big/ I_{(y_{(n)}-1,\ y_{(1)})}(\theta) \quad , \text{ pois}$$

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

Então,

$$\theta \leq x_{(1)} \leq x_{(n)} \leq \theta+1 .$$

Note que a razão $\dfrac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)} = c(\underset{\sim}{x}, \underset{\sim}{y}) \iff$

$X_{(1)} = Y_{(1)}$ e $X_{(n)} = Y_{(n)}$. Então,

$T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ é uma estatística suficiente mínima para θ ou

$X_{(1)}$ e $X_{(n)}$ são estatísticas conjuntamente suficientes minimais para θ.

* Este é um exemplo no qual a dimensão de uma estatística suficiente mínima não corresponde à

dimensão do parâmetro.

Comentário: Uma estatística suficiente minimal não é única. Qualquer função um a um, de uma estatística suficiente mínima, é também uma estatística suficiente mínima.

Assim, por exemplo, temos que $T'(\underset{\sim}{X}) = \left(X_{(n)} - X_{(1)} , \dfrac{X_{(n)} + X_{(1)}}{2} \right)$ também é uma estatística suficiente mínima, considerando o exemplo anterior (EXEMPLO 4).

$$\left\{ \begin{array}{l} T' : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \\ (x, y) \longrightarrow \left(y - x , \dfrac{y + x}{2} \right) \end{array} \right\}$$

Família Exponencial Uniparamétrica

dos dados originais

Seja X uma variável aleatória tq a distribuição pertence à família exponencial uniparamétrica, isto é,

$$f(x|\theta) = h(x)\, c(\theta) \exp\{w(\theta)\, t(x)\},$$

$x \in \mathcal{X}$ e $\theta \in \Theta$. Então, para uma amostra de tamanho n, $S(\underset{\sim}{X}) = \sum_{j=1}^{n} t(X_j)$ é uma estatística suficiente mínima para θ. Fácil de verificar utilizando o resultado proposto por Lehmann e Scheffé (Teorema).

Família Exponencial multiparamétrica

Seja X uma variável aleatória tal que a distribuição pertence à família exponencial multiparamétrica, isto é,

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

$x \in \mathcal{X}$ e $\underset{\sim}{\theta} \in \Theta$. Então, para uma amostra de tamanho n, $S_1(\underset{\sim}{X}), \ldots, S_k(\underset{\sim}{X})$ são estatísticas conjuntamente suficientes mínimas para $\underset{\sim}{\theta}$, onde $S_i(\underset{\sim}{X}) = \sum_{j=1}^{n} t_i(X_j)$, $i = 1, \ldots, k$.