Estatísticas Suficientes, Ancilares e Completas

Uma estatística suficiente mínima é aquela que atingiu a máxima redução de dados possível, ao mesmo tempo mantendo todas as informações sobre o parâmetro θ.

Uma vez que a distribuição de uma estatística ancilar não depende de θ, é possível suspeitar que uma estatística suficiente mínima não esteja relacionada (ou, matematicamente falando, não é funcionalmente dependente de) a uma ancilar. Contudo isso não é necessariamente verdade.

Por exemplo, considere $X_1, \dots, X_n$ a.a $X \sim U(\theta, \theta+1)$, então pelo teorema de Lehmann e Scheffé pode-se mostrar que $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ é uma estatística suficiente mínima.
Além disso, uma estatística suficiente mínima não é única, pois qualquer função um a um de uma estatística suficiente mínima, é também uma estatística suficiente mínima. Portanto, $T_1(\underset{\sim}{X}) = \left(X_{(n)} - X_{(1)}, \frac{X_{(n)} + X_{(1)}}{2}\right)$ também é uma estatística suficiente mínima. Note que a família de distribuições

$U(\theta, \theta+1)$ pertence à família de locação, 1

então foi demonstrado que $R = X_{(n)} - X_{(1)}$ é

uma estatística ancilar.

$X_1, \ldots, X_n$ a a $X \sim U(\theta, \theta+1)$, $\theta \in \mathbb{R}$.

→ FAMÍLIA LOCAÇÃO

→ $R = X_{(n)} - X_{(1)}$ é ancilar.

→ Pelo Teorema de LS, temos que $(X_{(1)}, X_{(n)})$ é estatística suficiente minimal e portanto

$\left(R, \dfrac{X_{(n)} + X_{(1)}}{2}\right)$ também é suficiente minimal, pois função 1-1, ou seja, são estatísticas equivalentes.

→ $T = (X_{(1)}, X_{(n)})$ é estatística completa para θ?

<u>NÃO!</u> Note que <u>R é função da estatística suficiente mínima</u> e R sendo ancilar sua distribuição não depende de θ e consequentemente $E(R) = a$ (constante que não depende de θ).

Portanto, $g(T) = X_{(n)} - X_{(1)} - a = R - a$ e

$E[g(T)] = E[R - a] = E(R) - a = 0$ $\not\Rightarrow$

$P(g(T) = 0) = 1$.

(ver exercício 6.10 do Casella pág 269)

Digitalizado com CamScanner

No entanto, nossa intuição de que uma estatística suficiente mínima é independente de qualquer estatística ancilar não está totalmente equivocada.

Segue abaixo uma condição na qual uma estatística suficiente mínima é independente de toda estatística ancilar.

Teorema (Teorema de Basu)

Se $T(\underset{\sim}{X})$ é uma estatística suficiente mínima e completa, então $T(\underset{\sim}{X})$ é independente de toda estatística ancilar. $(\Rightarrow)$

vs atrás

O inverso do Teorema de Basu

Um interessante fato estatístico é que o inverso do teorema de Basu é falso. Isto é, se $T(\underset{\sim}{X})$ é independente de toda estatística ancilar, não segue, necessariamente, que $T(\underset{\sim}{X})$ é uma estatística suficiente mínima completa.

Comentários: * O teorema de Basu apresenta uma relação entre estatísticas suficientes e estatísticas ancilares utilizando o conceito de estatísticas completas.

* Mesmo que a palavra "mínima" seja redundante na declaração do Teorema de Basu, ele foi estabelecido desta maneira como um lembrete de que a estatística $T(\underset{\sim}{X})$ no teorema é uma estatística suficiente mínima.

* O teorema de Basu é útil no sentido que permite deduzir a independência de duas estatísticas sem jamais encontrar a distribuição conjunta das duas estatísticas.

* Para utilizar este teorema precisamos mostrar que uma estatística é completa, o que, algumas vezes, é um problema de análise realmente difícil.

Exemplo 1: Sejam $X_1, \dots, X_n$ observações exponenciais iid com parâmetro θ. Considere calcular o valor esperado de $g(\underset{\sim}{X}) = \dfrac{X_n}{X_1 + \dots + X_n}$.

Note que a família de distribuições exponencial(θ) pertence à família de escala e desta maneira, foi demonstrado que qualquer estatística que dependa da amostra somente por meio dos $n-1$ valores $\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}$ é uma estatística ancilar. Assim, $g(\underset{\sim}{X}) = \frac{X_n}{X_1 + \cdots + X_n}$ é uma estatística ancilar, pois $g(\underset{\sim}{X}) = \frac{1}{\frac{X_1 + \cdots + X_n}{X_n}} =$

$$\frac{1}{\boxed{\frac{X_1}{X_n} + \cdots + \frac{X_{n-1}}{X_n}} + 1}.$$

Além disso, sabendo que a família de distribuição $exp(\theta)$ pertence à família exponencial uniparamétrica, tal que associada à amostra, temos que $T(\underset{\sim}{X}) = \sum_{j=1}^{n} X_j$ é estatística suficiente mínima e completa.

Portanto, pelo teorema de Basu, $T(\underset{\sim}{X})$ e $g(\underset{\sim}{X})$ são independentes. Logo,

$\Rightarrow$ $X_1, \ldots, X_n$ a a $X \sim exp(\theta)$.

1.ª parametrização $f(x|\theta) = \theta e^{-\theta x}$, $x > 0$ e $\theta > 0$.

função padrão $f(x) = e^{-x}$, $x > 0$ e

o parâmetro de escala $\rho = \frac{1}{\theta} > 0$.

2.ª parametrização $f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x > 0$ e $\theta > 0$.

função padrão $f(x) = e^{-x}$, $x > 0$ e

o parâmetro de escala $\rho = \theta > 0$.

1ª parametrização $f(x|\theta) = \theta e^{-\theta x}, x > 0$ e $\theta > 0$.

Note que

$$f(x|\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$$, onde

$h(x) = I_{(0,\infty)}(x)$, $c(\theta) = \theta$, $w(\theta) = -\theta$ e

$t(x) = x$. Portanto, associada à amostra, temos

$$T(\underset{\sim}{x}) = \sum_{j=1}^{n} x_j .$$

2ª parametrização $c(\theta) = \frac{1}{\theta}$ e $w(\theta) = \frac{-1}{\theta}$,

as demais são iguais.

2.a parametrização

$E(X_n) = \theta = E[T(\underset{\sim}{X})g(\underset{\sim}{X})] =$

$E[T(\underset{\sim}{X})]E[g(\underset{\sim}{X})] = n\theta E[g(\underset{\sim}{X})]$, pois

$T(\underset{\sim}{X}) = \sum_{j=1}^{n} X_j \sim \text{gama}(n, \theta)$.

Portanto, $E[g(\underset{\sim}{X})] = n^{-1}$ //

1.a parametrização

$E(X_n) = \frac{1}{\theta} = E[T(\underset{\sim}{X})g(\underset{\sim}{X})] = E[T(\underset{\sim}{X})]E[g(\underset{\sim}{X}$

$= \frac{n}{\theta} E[g(\underset{\sim}{X})]$, pois $T(\underset{\sim}{X}) = \sum_{j=1}^{n} X_j \sim$

Gama(n, θ). Portanto,

$E[g(\underset{\sim}{X})] = n^{-1}$. //

Exemplo 2 ou Exercício 1 :

Lembrete :

Teorema : Seja $X_1, \ldots, X_n$ a.a $X \sim N(\mu, \sigma^2)$ e considere $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ e $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$.

Então,

a) $\bar{X}$ e S^2 são variáveis aleatórias independentes. Casella, Teorema 5.3.1

b) $\bar{X} \sim N(\mu, \sigma^2/n)$.

c) $\frac{(n-1) S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Parte I :

Seja $X_1, \ldots, X_n$ a.a $X \sim N(\mu, \sigma^2)$, onde σ^2 é conhecida.

Sabendo que a família de distribuições $N(\mu, \sigma^2)$, para σ^2 conhecida, pertence à família exponencial e o $\Theta = \mathbb{R}$ contém um retângulo no $\mathbb{R}$ (intervalo), então $\bar{X}$ é uma estatística suficiente mínima e completa p/ μ. Observe que S^2 é uma estatística ancilar. A ancilaridade de S^2 pode ser verificada

Utilizando o resultado que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ ou utilizando a família de locação. Portanto, $\bar{X}$ e S^2 são independentes pelo Teorema de Basu.

Além disso, considere $R = X_{(n)} - X_{(1)}$. Então, $\bar{X}$ e R são independentes pelo Teorema de Basu.

Note que a ancilaridade de R pode ser verificada utilizando a família de locação.

Pergunta: $\bar{X}$ e $\frac{S}{R}$ são independentes?

Sim!

Demonstramos que para quaisquer μ e σ^2 fixa, $\bar{X}$ e S^2 são independentes. Note que a nova considera σ^2 arbitrário!!

Parte II Seja $X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$, onde μ é conhecida.

Sabemos que a família de distribuições $N(\mu, \sigma^2)$, para μ conhecida, pertence à família exponencial e o $\Theta = \mathbb{R}_+$ contém um retângulo no $\mathbb{R}$ (intervalo), então $U^2 = \sum_{i=1}^{n} (X_i - \mu)^2$ ("conhecida") é uma estatística suficiente mínima e completa para σ^2. Observe que

$W_1 = \frac{(X_1 - \bar{X})}{U^2}$ é uma estatística ancilar e

então, U e W_1 são independentes pelo Teorema de Basu. Além disso, $U^2 = \sum_{i=1}^{n} (X_i - \mu)^2$ e $W_2 = \frac{(X_{(n)} - \bar{X})}{S}$ são independentes pelo Teorema de Basu. Note que a ancilaridade de W_1 e W_2 podem ser verificadas utilizando a família de escala.

Parte III Seja $X_1,\dots,X_n$ a.a. $X \sim N(\mu,\sigma^2)$, onde $\theta=(\mu,\sigma^2)$ e $\Theta=\{(\mu,\sigma^2)$, tal que $\mu \in \mathbb{R}$ e $\sigma^2>0\}$. Neste caso, também podemos aplicar o Teorema de Basu para mostrar que $\bar{X}$ e S^2 são independentes.

Agora considere σ^2 fixa, mas assume qualquer valor arbitrário $\sigma_0^2>0$, e que μ varia, tal que $\mu \in \mathbb{R}$. Neste caso, temos a situação da parte I e assim, $\bar{X}$ e S^2 são independentes pelo teorema de Basu, quando $\sigma^2=\sigma_0^2$ fixo em um valor arbitrário. Entretanto, como o σ^2 era arbitrária, temos que a $\bar{X}$ e S^2 são independentes para qualquer escolha de μ e σ^2, tal que $\mu \in \mathbb{R}$ e $\sigma^2>0$. (Mostra que essa independência prossegue mesmo se σ^2 for desconhecido, uma vez que o conhecimento de σ^2 não tem nenhuma relação com as

Comentário. Proveniente da parte III, podemos compreender que usamos a suficiência de $\bar{X}$ e a ancilaridade de S^2, mas, se μ e σ^2 são ambos desconhecidos, então $\bar{X}$ não é suficiente e S^2 não é ancilar.

Porém, note que usamos os seguintes dois fatos: quando $\sigma^2=\sigma_0^2$ é fixo, mas arbitrário, $\bar{X}$ é suficiente e S^2 é ancilar.

Parte IV Seja $X_1,\dots,X_n$ a.a. $X \sim N(\mu,\sigma^2)$, onde μ e σ^2 são desconhecidos. Então, sabemos que a família de distribuições $N(\mu,\sigma^2)$ pertence à família exponencial

→ Comentários

Note que a independência de $\bar{X}$ e S^2 é determinada pela distribuição conjunta de $(\bar{X}, S^2)$ para cada valor de (μ, σ^2). Pela parte I, para cada valor de (μ, σ^2), $\bar{X}$ e S^2 são independentes.

// cqd.

k-paramétrica para $k=2$ e $\Theta = \mathbb{R} \times \mathbb{R}_+$ contém um retângulo no $\mathbb{R}^2$ e $U = (\bar{X}, S^2)$ é estatística suficiente mínima e completa para $\underset{\sim}{\theta} = (\mu, \sigma^2)$. Note que $W = \frac{(X_1 - X_2)}{S}$ é uma estatística ancilar, tal que a ancilaridade pode ser demonstrada utilizando a família de locação-escala. Portanto,

U e W são independentes pelo Teorema de Basu.