Unidade II

Princípios da redução de dados

Mood
Carella
e
Bolfarine

Introdução

* Um pesquisador (experimentador) utiliza as informações em uma amostra $X_1, \ldots, X_n$ para fazer inferências (problemas da inferência: estimação e testes de hipóteses) sobre um parâmetro desconhecido θ.

* Se o tamanho da amostra n for grande, então a amostra observada $x_1, \ldots, x_n$ consiste de uma longa lista de números que podem ser difíceis de se interpretar.

* Uma das áreas da Estatística é a Análise Exploratória de Dados ou Estatística Descritiva cujo objetivo é resumir as informações em uma amostra determinando algumas características importantes (medidas resumo ou estatísticas descritivas), tais como a média amostral, a variância amostral, a menor e a maior observação.

* Observe que o resumo dos dados, em geral, é feito por meio de estatísticas, qualquer função da amostra que não depende de parâmetros desconhecidos.

* A estatística é uma variável aleatória que condensa n variáveis aleatórias $X_1, \dots, X_n$ em uma única variável aleatória. Tal condensação é atraente uma vez que trabalhar com quantidades unidimensionais é melhor (mais fácil) do que com quantidades multidimensionais. comentários extras (Mood pág 300)

* Estudaremos um princípio de redução de dados conhecido como princípio da suficiência.

* Estamos interessados em métodos de redução de dados que não desconsiderem informações importantes sobre o parâmetro desconhecido θ e em métodos que descartem, de forma bem-sucedida, informações que sejam irrelevantes para a obtenção de conhecimento sobre a que θ se refere.

suporte da distribuição

espaço amostral

A condensação pode ser visualizada também de uma outra forma. "range"

(de possibilidades)

Considere que $\mathcal{X}$ denota a gama de valores que $(X_1, \ldots, X_n)$ podem assumir. Por exemplo, se nós amostramos de uma distribuição Bernoulli, então $\mathcal{X}$ é a coleção de todos vetores n dimensionais com componentes 0 e/ou 1.

Note que uma estatística induz ou define uma partição do $\mathcal{X}$. (Lembrando a definição de partição de $\mathcal{X}$ é uma coleção de subconjuntos mutuamente exclusivos (disjuntos) de $\mathcal{X}$, tais que a união deles é o $\mathcal{X}$). Considere $t(\cdot, \ldots, \cdot)$ como sendo a função correspondente (correspondendo) à estatística $T = t(X_1, \ldots, X_n)$. A partição induzida ou definida por $t(\cdot, \ldots, \cdot)$ é provocada (dada) como segue:
Considere que t_0 qualquer valor da função $t(\cdot, \ldots, \cdot)$; que o subconjunto de $\mathcal{X}$ consistindo de todos aqueles pontos $(x_1, \ldots, x_n)$ para os quais $t(x_1, \ldots, x_n) = t_0$ é um subconjunto na coleção dos subconjuntos dos quais a partição compreende (é formada / é definida); outros subconjuntos são similarmente formados considerando outros valores de $t(\cdot, \ldots, \cdot)$. Por exemplo se uma amostra de tamanho 3 é selecionada da distribuição Bernoulli, então $\mathcal{X}$ consiste de oito pontos $(0,0,0)$, $(0,0,1)$,

Digitalizado com CamScanner

/ /

$(0,1,0)$, $(1,0,0)$, $(0,1,1)$, $(1,0,1)$, $(1,1,0)$, $(1,1,1)$. Considere $t(x_1, x_2, x_3) = x_1 + x_2 + x_3$; então $t(\cdot,\cdot,\cdot)$ toma (assume) os valores $0, 1, 2$ e 3. A partição do $\mathfrak{X}$ induzida por $t(\cdot,\cdot,\cdot)$ consiste de 4 subconjuntos

$$\Big\{ \underbrace{(0,0,0)}_{1^{\circ}}, \underbrace{\{(0,0,1), (0,1,0), (1,0,0)\}}_{2^{\circ}}, \underbrace{\{(0,1,1), (1,0,1), (1,1,0)\}}_{3^{\circ}}, \underbrace{\{(1,1,1)\}}_{4^{\circ}} \Big\}$$

correspondendo, respectivamente, aos quatro valores $0, 1, 2$ e 3 de $t(\cdot,\cdot,\cdot)$. A estatística então é realmente uma condensação do $\mathfrak{X}$. Neste exemplo acima, se nós usarmos a estatística $t(\cdot,\cdot,\cdot)$ nós temos que nos preocupar apenas com 4 diferentes valores ao invés de 8 diferentes pontos do $\mathfrak{X}$.

Observe que diferentes estatísticas podem induzir a mesma partição. De fato, se $t(\cdot,\ldots,\cdot)$ é uma estatística, então qualquer função 1 a 1 (bijetora) de t tem a mesma partição como t. No exemplo acima, $t'(x_1, x_2, x_3) = 6(x_1 + x_2 + x_3)^2$ ou $t''(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$, induzem a mesma partição de

$t(x_1, x_2, x_3) = x_1 + x_2 + x_3$.

Uma das razões para usar estatísticas é que elas fazem a condensação (condensam) X e se essa é a nossa única razão para usar estatística, então quaisquer duas estatísticas com a mesma partição são de mesma utilidade. O importante aspecto da estatística é a partição de X que ela induz, não os valores que elas assumem.

A estatística suficiente é um tipo particular de estatística. É uma estatística que condensa X de tal forma que não perde informação sobre "θ".
⇔ parâmetro de interesse.

Princípio da Suficiência

Promove um método de redução de dados que não descartam informações sobre θ ao obter algum resumo dos dados.

Uma estatística suficiente para um parâmetro θ é uma estatística que, de certa maneira, captura todas as informações sobre θ contidas na amostra. Quaisquer outras informações adicionais na amostra, além do valor da estatística suficiente, não apresentam mais nenhum detalhe sobre θ.

Princípio da Suficiência

Se $T(\underset{\sim}{X})$ é uma estatística suficiente para θ, então qualquer inferência sobre θ deverá depender da amostra $\underset{\sim}{X}$ somente pelo valor $T(\underset{\sim}{X})$. Isto é, se $\underset{\sim}{x}$ e $\underset{\sim}{y}$ são dois pontos amostrais, de modo que $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$, então a inferência sobre θ deverá ser a mesma, se $\underset{\sim}{X} = \underset{\sim}{x}$ ou $\underset{\sim}{X} = \underset{\sim}{y}$ for observado.

Uma estatística suficiente é formalmente definida da seguinte maneira.

Definição: (Estatística Suficiente)

Seja $X_1, \dots, X_n$ aa $X \sim f(\cdot|\theta)$, $\theta \in \Theta$, onde $f(\cdot|\theta)$ pode ser a fdp (caso contínuo) ou a fp (caso discreto). Dizemos que uma estatística $S = S(X_1, \dots, X_n)$ é suficiente para θ se e somente se a distribuição condicional de $X_1, \dots, X_n$ dado $S = s$ não depende de θ para qualquer valor s de S.

comentário atrás

Os exemplos a seguir ilustram a obtenção de estatísticas suficientes pela utilização da definição.

Exemplo 1: Seja X_1, X_2, X_3 uma aa de tamanho 3 proveniente da distribuição Bernoulli (p). Mostre que $S = S(X_1, X_2, X_3) = X_1 + X_2 + X_3$ é suficiente para p.

X	s (valores de S)	$f_{X_1,X_2,X_3 \mid S=s}$
$(0,0,0)$	0	1
$(0,0,1)$	1	$1/3$
$(0,1,0)$	1	$1/3$

$(1,0,0)$	1	$1/3$
$(0,1,1)$	2	$1/3$
$(1,0,1)$	2	$1/3$
$(1,1,0)$	2	$1/3$
$(1,1,1)$	3	1

Note que $f_{X_1, X_2, X_3 \mid S=1}(0,1,0 \mid 1) =$

$$P(X_1=0, X_2=1, X_3=0 \mid S=1) = \frac{P(X_1=0, X_2=1, X_3=0, S=1)}{P(S=1)} =$$

$$\frac{P(X_1=0, X_2=1, X_3=0)}{P(S=1)} = \frac{P(X_1=0)\,P(X_2=1)\,P(X_3=0)}{P(S=1)},$$

onde $S \sim \text{Binomial}(3, p)$.

$$= \frac{\cancel{(1-p)}\,\cancel{p}\,\cancel{(1-p)}}{\binom{3}{1}\cancel{p}\,\cancel{(1-p)^2}} = \frac{1}{3}.$$

Portanto, S é suficiente para p, pois a distribuição condicional da amostra dada os valores de S independe de p. ~~DUD~~ três

Exemplo 2: Seja $X_1, \dots, X_n$ aa $X \sim Bernoulli(p)$.

Mostre que $S = S(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} X_i$ é uma estatística suficiente para p.

Preciso mostrar que

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid S(X_1, \dots, X_n) = s)$$

independe de p.

Sabe-se que $f(x_1, \dots, x_n \mid p) = \prod_{i=1}^{n}$

$$p^{x}(1-p)^{1-x} \underset{\{0,1\}}{I}(x) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} I_{\{0,1\}}(x_i)$$

Assim, $P(\underset{\sim}{X} = \underset{\sim}{x} \mid S(\underset{\sim}{X}) = s) =$

$$P\left(X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^{n} X_i = s\right) =$$

$$\frac{P\left(X_1 = x_1, \dots, X_n = x_n, \sum_{i=1}^{n} X_i = s\right)}{P\left(\sum_{i=1}^{n} X_i = s\right)} =$$

$$\left\{ \frac{P(X_1 = x_1)\, P(X_2 = x_2) \dots P(X_n = x_n)}{P\left(\sum_{i=1}^{n} X_i = s\right)} \right. \quad \text{se } \sum_{i=1}^{n} x_i = s$$

$$\begin{cases} & \text{c c} \\ 0 & \end{cases}$$

$$= \begin{cases} \dfrac{\cancel{p^{\sum_{i=1}^n x_i}}\ \cancel{(1-p)^{n-\sum_{i=1}^n x_i}}\ \prod_{i=1}^n I_{\{0,1\}}(x_i)}{\binom{n}{s}\ \cancel{p^s}\ \cancel{(1-p)^{n-s}}\ I_{\{0,1,\dots,n\}}(s)} & \text{se } \sum_{i=1}^n x_i = s \\ & \text{c c} \\ 0 & \end{cases}$$

$$= \begin{cases} \dfrac{1}{\binom{n}{s}}\ \dfrac{\prod_{i=1}^n I_{\{0,1\}}(x_i)}{I_{\{0,1,\dots,n\}}(s)} & \text{se } \sum_{i=1}^n x_i = s \\ & \text{c c} \\ 0 & \end{cases}$$

$\Rightarrow$ S é suficiente para p.

Exemplo 3. Seja $X_1, \ldots, X_n$ a.a. $X \sim Poisson(\theta)$ (5)

Mostre que $S = S(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ é uma estatística suficiente para θ.

$$P(\underset{\sim}{X} = \underset{\sim}{x} \mid S(\underset{\sim}{X}) = s) = \frac{P(\underset{\sim}{X} = \underset{\sim}{x}, S(\underset{\sim}{X}) = s)}{P(S(\underset{\sim}{X}) = s)} =$$

$$\frac{P(X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^{n} X_i = s)}{P(\sum_{i=1}^{n} X_i = s)} =$$

$$\begin{cases} \dfrac{\prod_{i=1}^{n} \dfrac{e^{-\theta}\theta^{x_i}}{x_i!} I_{\{0,1,\ldots\}}(x_i)}{\dfrac{e^{-n\theta}(n\theta)^s}{s!} I_{\{0,1,\ldots\}}(s)} & \text{se } \sum_{i=1}^{n} x_i = s \\ \\ 0 & \text{c.c.} \end{cases}$$

$$
\left\{
\begin{array}{ll}
\dfrac{\dfrac{\cancel{e^{-n\theta}}\;\cancel{\theta^{\sum_{i=1}^{n} x_i}}}{\prod_{i=1}^{n} x_i!}\;\prod_{i=1}^{n} I_{\{0,1,\dots\}}(x_i)}{\dfrac{\cancel{e^{-n\theta}}\;\cancel{(n\theta)^{s}}}{s!}\;I_{\{0,1,\dots\}}(s)} & \text{se } \sum_{i=1}^{n} x_i = s \\[2ex]
0 & \text{c c}
\end{array}
\right.
$$

$$
= \left\{
\begin{array}{ll}
\dfrac{\prod_{i=1}^{n} I_{\{0,1,\dots\}}(x_i)}{\prod_{i=1}^{n} x_i!} \cdot \dfrac{s!}{n^{s}\; I_{\{0,1,\dots\}}(s)} & \text{se } \sum_{i=1}^{n} x_i = s \\[2ex]
0 & \text{c c}
\end{array}
\right.
$$

Logo, S é suficiente para θ.

Exemplo 4: Seja $X_1, \ldots, X_n$ a.a. $X \sim \exp(\lambda)$. Mostre que $S = \sum_{i=1}^{n} X_i$ é uma estatística suficiente para λ.

$$f_{X_1,\ldots,X_n \mid \sum_{i=1}^{n} X_i = s}(x_1,\ldots,x_n \mid s) = \frac{f_{X_1,X_2,\ldots,X_n,\sum_{i=1}^{n} X_i}(x_1,x_2,\ldots,x_n,s)}{f_{\sum_{i=1}^{n} X_i}(s)} =$$

$$\begin{cases} \dfrac{\prod_{i=1}^{n} \lambda e^{-\lambda x_i} I_{(0,\infty)}(x_i)}{f_{\sum_{i=1}^{n} X_i}(s)} & \text{se } \sum_{i=1}^{n} x_i = s \\ 0 & \text{c.c.} \end{cases}$$

Sabe-se que $\sum_{i=1}^{n} X_i \sim \text{Gama}(n, \lambda)$. Portanto,

$$= \begin{cases} \dfrac{\cancel{\lambda^n} e^{-\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} I_{(0,\infty)}(x_i)}{\cancel{\lambda^n}/\Gamma(n)\, s^{n-1} e^{-\lambda s} I_{(0,\infty)}(s)} & \text{se } \sum_{i=1}^{n} x_i = s \end{cases}$$

$$\begin{cases} 0 & \text{c c} \end{cases}$$

(3)

$$= \begin{cases} \dfrac{\prod_{i=1}^{n} I_{(0,\infty)}(x_i) \quad \Gamma(n)}{s^{n-1}} & \text{se} \quad \sum_{i=1}^{n} x_i = s \\ 0 & \text{c c} \end{cases}$$

Logo, $\sum_{i=1}^{n} x_i$ é suficiente para λ.

Notemos que a Definição permite, apenas, que possamos verificar se determinada estatística é ou não suficiente. Contudo não pode ser utilizada como um método para obtenção de estatísticas suficientes. Um procedimento para a obtenção de estatísticas suficientes é o critério da fatoração que apresentamos a seguir.

Teorema (critério da fatoração de Neyman) (12)

Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição da variável aleatória X com função densidade (ou de probabilidade) $f(x|\theta)$. Uma estatística $S = S(X_1, \dots, X_n)$ é a estatística suficiente para θ se, e somente se, existirem funções $g(s|\theta)$ e $h(x_1, \dots, x_n)$ de modo que, para todos os pontos amostrais $\underset{\sim}{x} = (x_1, \dots, x_n)$, e todos os pontos de parâmetro θ,

$$f(x_1, \dots, x_n|\theta) = g(S(x_1, \dots, x_n)|\theta)\, h(x_1, \dots, x_n)$$

$$= \underbrace{g(s|\theta)\, h(x_1, \dots, x_n)}_{g \text{ e } h \text{ funções não negativas!}}.$$

Comentários: Para utilizar o critério da fatoração a fim de encontrar uma estatística suficiente, fatoramos a fdp conjunta da amostra em duas partes, sendo que uma parte não depende de θ. Esta parte constitui a função $h(x_1, \dots, x_n)$. A outra, dependente de θ, geralmente depende da amostra

$x_1, \dots, x_n$ somente por meio de alguma função $S(X_1, \dots, X_n)$, e esta função é uma estatística suficiente para θ.

PROVA: Casella e Berger pág 245

Bolfarine pág 22

(idéia: utilizar a Definição de estatística suficiente).

Exemplo 1: Seja $X_1, \dots, X_n$ uma aa $X \sim$ Bernoulli (p), então

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n \mid p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} I_{\{0,1\}}(x_i)$$

$$= p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^{n} I_{\{0,1\}}(x_i)$$

$$= \underbrace{\left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} (1-p)^n}_{g\left(\sum_{i=1}^n x_i \mid p\right)} \underbrace{\prod_{i=1}^{n} I_{\{0,1\}}(x_i)}_{h(x_1, \dots, x_n)} =$$

Logo, $\sum_{i=1}^{n} X_i$ é estatística suficiente para p. (14)

Exemplo a: Seja $X_1, \ldots, X_n$ a.a. $X \sim Poisson(\lambda)$, então

$$f(x_1, \ldots, x_n \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} I_{\{0,1,\ldots\}}(x_i) =$$

$$\frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} I_{\{0,1,\ldots\}}(x_i) =$$

$$\underbrace{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}} \quad \underbrace{\frac{\prod_{i=1}^{n} I_{\{0,1,\ldots\}}(x_i)}{\prod_{i=1}^{n} x_i!}} =$$

$g\left(\sum_{i=1}^{n} x_i \mid \lambda\right) \; h(x_1, \ldots, x_n)$. Logo,

$\sum_{i=1}^{n} X_i$ é suficiente para λ.

Exemplo 3: Seja $X_1, \ldots, X_n$ aa $X \sim U(0,\theta)$,

então

$$f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{(0,\theta)}(x_i) =$$

$$\frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(x_i).$$ Note que

$$\prod_{i=1}^{n} I_{(0,\theta)}(x_i) = 1 \iff 0 \leq x_i \leq \theta \quad \forall i.$$

Então,

$$0 \leq x_{(1)} \leq x_{(n)} \leq \theta$$ e

$$f(x_1, \ldots, x_n | \theta) = \underbrace{\frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)})}_{g(x_{(n)} | \theta)} \underbrace{1}_{h(x_1, \ldots, x_n)}.$$

Logo, $X_{(n)}$ é suficiente para θ.

Família Exponencial Uniparamétrica

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

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

$x \in \mathcal{X}$ e $\theta \in \Theta$. Além disso, temos a seguinte família exponencial uniparamétrica de distribuições associada à amostra ($X_1, \dots, X_n$ aa $X \sim f(\cdot|\theta)$).

$$f(x_1,\dots,x_n|\theta) = \prod_{j=1}^{n} h(x_j)\, [c(\theta)]^n \exp\left\{w(\theta) \sum_{j=1}^{n} t(x_j)\right\}.$$

Resultado Importante !!

Note que pelo critério da fatoração $\sum_{j=1}^{n} t(X_j)$ é estatística suficiente para θ.