Modelos Estatísticos
Inferência - EST187

Profa. Denise Duarte

2020/01

- Seja Y uma v.a.com função de distribuição desconhecida $F(\cdot)$ e considere $y = (y_1, ..., y_n)$ n realizações da v.a. Y.
- Objetivo: tirar informações sobre F com base nos valores observados de y.
- Obviamente, as informações sobre F estarão sujeitas a incerteza.
- Assim, devemos assegurar que:
    - o grau de incerteza é o menor possível, considerando a aleatoriedade de Y.
    - seja possível acessarmos o grau de incerteza em nossas decisões.
- A natureza física do problema que gera y, o esquema amostral, bem como outras possíveis informações adicionais, irão impor limites sobre as possíveis escolhas de F.

Definição

Modelo Estatístico: *é o conjunto $\mathscr{F}$ formado por todas as possíveis candidatas a $F(\cdot)$.*

- A classe $\mathscr{F}$ é muito grande, pois, em princípio, $\mathscr{F}$ pode ser qualquer conjunto de funções de distribuições.
- Intuitivamente, nossas inferências serão mais acuradas se formos capazes de selecionar o conjunto $\mathscr{F}$ de tal forma que ele tenha o menor tamanho possível.

Definição

Suporte: *o suporte de uma v.a. Y corresponde ao conjunto*

$$\mathscr{Y}_\theta = \{y : f(y;\theta) > 0\} \tag{1}$$

Definição

Modelo Paramétrico: *é o conjunto $\mathscr{F}$ formado por todas as funções da forma*

$$\mathscr{F} = \{F(\cdot;\theta) : \theta \in \Theta \subset \mathbb{R}^k\}, \tag{2}$$

em que, para todo θ fixado, $F(\cdot;\theta)$ corresponde a uma função de distribuição cujo suporte é um subconjunto de $\mathbb{R}$, com k e n inteiros positivos. Alternativamente, poderiamos definir $\mathscr{F}$ em termos de f.d.p.'s ou f.m.p.'s, isto é,

$$\mathscr{F} = \{f(\cdot;\theta) : \theta \in \Theta \subset \mathbb{R}^k\}. \tag{3}$$

Obervações

- *A quantidade θ é o parâmetro, que pode ser um vetor;*
- *Θ é o espaço paramétrico associado a θ;*
- *$\mathscr{F}$ definido em (2) ou (4) é chamada classe paramétrica ou modelo (estatístico) paramétrico;*
- *Note que os elementos de $\mathscr{F}$ estão associados aos elementos de Θ.*
- *Nossas inferências serão sobre $\theta_* \in \Theta$, associado a F, chamado valor verdadeiro do parâmetro, a partir de uma amostra.*

Definição

Espaço Amostral: *é o conjunto* $\mathscr{Y}$ *formado por todos os possíveis resultados* y *(realizações da v.a.* Y*) compatíveis com um dado modelo paramétrico, ou seja,*

$$\mathscr{Y} = \cup_{\theta \in \Theta} \mathscr{Y}_\theta. \tag{4}$$

Em geral, $\mathscr{Y}_\theta$ *é o mesmo para todo* $\theta \in \Theta$ *e* $\mathscr{Y}_\theta = \mathscr{Y}$.

Exemplo

Assuma que $Y \sim U(0, \theta)$. *Então*

$$f(y|\theta) = \begin{cases} \frac{1}{\theta}, se\, 0 < y < \theta, \\ 0, caso\ contrário. \end{cases}$$

Podemos reescrever a densidade como

$$f(y|\theta) = \frac{1}{\theta} I_{(0,\theta)}(y),$$

onde

$$\begin{cases} 1, se\ a < y < b; \\ 0, caso\ contrário. \end{cases}$$

é a função indicadora do intervalo aberto (a, b).

Logo, $\Theta = \{\theta : \theta > 0\}$. *Note que, neste caso, o suporte da distribuição,* $\mathscr{Y}_\theta$, *depende do espaço paramétrico* Θ *e* $\mathscr{Y} = \cup_{\theta \in \Theta} \mathscr{Y}_\theta$.

Exemplo

Assuma que $Y \sim Bin(n, \theta)$. *Então*

$$f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \tag{5}$$

com $\Theta = \{\theta : 0 < \theta < 1\}$ *e* $\mathscr{Y} = \{y : y = 0, 1, ..., n\}$, *pois o suporte da distribuição* $\mathscr{Y}_\theta$ *não depende do espaço paramétrico* Θ.

Exemplo

Se dois valores são amostrados independentemente de $N(\theta, 1)$, *então* $y = (y_1, y_2)'$, *com* $y_i \in \mathbb{R}$ $(i = 1, 2)$. *Logo,* $Y \sim N_2\left(\left(\begin{array}{c}\theta\\ \theta\end{array}\right); I_2\right)$, *onde* I_2 *matriz identidade de ordem 2, e*

$$\begin{aligned} f(y|\theta) &= \left[\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(y_1-\theta)^2\right\}\right] \times \left[\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(y_2-\theta)^2\right\}\right] \\ &= \frac{1}{2\pi}\exp\left\{-\sum_{i=1}^{2}\frac{1}{2}(y_i-\theta)^2\right\} \qquad (6) \end{aligned}$$

Neste caso, $\Theta = \{\theta : \theta \in \mathbb{R}\}$ *e* $\mathscr{Y} = \mathbb{R} \times \mathbb{R}$.

- Para evitar ambiguidade, é essencial que valores distintos de θ correspondam a distindas distribuições de probabilidade.
- Sem tal propriedade, poderiam existir 2 valores de θ, digamos θ' e θ", associados à mesma distribuição verdadeira F, o que tornaria impossível dizer qual valor (θ' ou θ") é o verdadeiro.
- Neste sentido, dizemos que um modelo é identificável se existir pelo menos um conjunto B tal que, para $\theta' \in \Theta$ e $\theta" \in \Theta$, com $\theta' \neq \theta"$, valer

$$P(Y \in B; \theta') \neq P(Y \in B; \theta"). \tag{7}$$

- A especificação de $\mathscr{F}$ pode ser feita de diversas maneiras, de acordo com a parametrização adotada.
- Seja h uma função um a um de Θ para Ψ. Então podemos redefinir $\mathscr{F}$ em (4) da seguinte forma

$$\begin{aligned}\mathscr{F} &= \{f(\cdot;\psi) : \psi = h(\theta) \in \Theta\} \\ &= \{f(\cdot;\psi) : \psi \in \Psi\},\end{aligned}$$

em que $\Psi = \{\psi : \psi = h(\theta) \in \Theta\}$.

- A parametrização a ser adotada depende do problema específico em estudo.
- Em geral, a parametrização é escolhida de tal forma a fornecer uma melhor interpretação física, embora em alguns casos a escolha da parametrização seja inteiramente arbitrária.

Exemplo

Seja Y a v.a. que representa o tempo entre ocorrências de um determinado evento de interesse. Assuma que $Y \sim Exp(\lambda)$, em que λ corresponde ao número esperado de eventos por unidade de tempo. Então

$$f(y|\lambda) = \lambda e^{-\lambda y}, \quad \lambda > 0, \quad y > 0. \tag{8}$$

Alternativamente, se definirmos $\theta = 1/\lambda$, então

$$f(y|\theta) = \frac{1}{\theta} e^{-\frac{1}{\theta} y}, \quad \theta > 0, \quad y > 0. \tag{9}$$

Neste caso, θ corresponde ao tempo médio entre duas ocorrências do evento de interesse.