Estatística II

Estimação Paramétrica

Matemática e Física ∘ 1º semestre 2006 ∘ Turma 4M

Wagner de Souza Borges

FCBEE, Universidade Presbiteriana Mackenzie

wborges@mackenzie.com.br

Introdução

No cálculo de probabilidades vimos que as probabilidades de eventos definidos através de uma característica observável no resultado de um *processo aleatório* Ψ são calculadas a partir da *distribuição de probabilidades* da *variável aleatória* que representa essa característica. Veremos agora, de que maneira os valores de uma *variável aleatória*, observada em *ensaios repetidos* de um *processo aleatório* Ψ pode ser utilizada para fazer um prognóstico acerca de um objeto desconhecido ξ, de natureza numérica, relacionado com a *distribuição de probabilidades* dessa *variável aleatória*, ao qual daremos o nome de *parâmetro*. Esse problema será tratado aqui sob o nome de *estimação paramétrica*

Para introduzir de maneira simples alguns conceitos importantes da estimação paramétrica, considere o seguinte exemplo.

Exemplo 1. Suponha que você está interessado em determinar a probabilidade de cara ξ de uma moeda. A coisa mais intuitiva que nos ocorre numa situação como esta é arremessar a moeda um certo número de vêzes, e observar o que acontece. Digamos que você tenha arremessadoa moeda, de forma independente e sob as mesmas condições, 13 vezes e tenha observado o seguinte resultado:

$$cara,\ coroa,\ coroa,\ cara,\ cara,\ coroa,\ cara,$$

$$cara,\ coroa,\ cara,\ coroa,\ coroa,\ cara.$$

Com base neste resultado, um prognóstico para o *parâmetro* ξ é a

$$\textit{proporção observada de caras nos 13 arremessos} = \tfrac{7}{13}.$$

Representando *cara* por 1 e *coroa* por 0, o resultado obtido, nos 13 arremessos,

$$(1,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1),$$

nada mais é do que o conjunto das observações das *variáveis aleatórias*

$$X_1, X_2, X_3, \ldots, X_{13},$$

em que X_i é o *número de caras* observado no *i-ésimo* arremêsso da moeda, $i = 1, 2, \ldots, 13$.

Nas condições em que os arremessos foram executados, podemos admitir que essas *variáveis aleatórias* são independentes e identicamente distribuídas (IID) com função de probabilidade

$$f(x;\xi) \;=\; \xi^x(1-\xi)^{1-x} \;,\; x \;=\; 0,\; 1 \text{ e } 0 \leq \xi \leq 1.$$

Observação 1.1.1. Lembre que a expressão acima admite a seguinte forma tabular:

x	$P\{X = x\}$
0	$1-\xi$
1	ξ

Nessas condições, diz-se que

$$X_1, X_2, X_3, \ldots, X_{13},$$

é uma *amostra casual simples* (ACS) da distribuição $f(x;\xi)$. Diz-se ainda que o conjunto de dados

$$(1,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1),$$

é a ACS *observada* da distribuição $f(x;\xi)$ e que o prognóstico $\frac{7}{13}$ para a *probabilidade de cara*, ξ , da moeda, é uma *estimativa* desse *parâmetro*.

Observação 2. Note que a *estimativa* de ξ é obtida substituindo-se na expressão

$$\frac{X_1 + X_2 + \ldots + X_{13}}{13}$$

os valores observados

$$x_1 = 1,\ x_2 = 0,\ x_3 = 0,\ x_4 = 1,\ \ldots,\ x_{12} = 0,\ x_{13} = 1$$

de $X_1, X_2, \ldots, X_{13}$. Por esse motivo, diz-se que

$$T(X_1, X_2, \ldots, X_{13}) \;=\; \frac{X_1 + X_2 + \ldots + X_{13}}{13}$$

é o estimador de ξ.

$\triangle$

No **Exemplo 1**, três conceitos básicos foram estabelecidos. São eles: o de *amostra casual simples*, o de *estimativa* e o de *estimador*. Mais precisamente, tem-se:

Amostra Casual Simples

Diz-se que um conjunto de dados

$$x_1,\ x_2,\ \ldots,\ x_n$$

é uma amostra casual simples (ACS) observada de uma distribuição de probabilidades $f(x;\theta)$ quando ele é constituido pelos valore observados das *variáveis aleatórias*

$$X_1\ ,\ X_2\ ,\ X_3, \ldots,\ X_n\ ,$$

independentes e identicamente distribuídas (IID), com distribuição de probabilidades $f(x;\theta)$.

Observação 1.2.1. O têrmo *amostra casual simples* será aplicado também em referência a $\mathbf{X}$.

Observação 1.2.2. Serão tratados apenas conjuntos de dados reais univariados.

Observação 1.2.3. A distribuição de probabilidades $f(x;\theta)$ será sempre expressa por uma função de probabilidades (componentes de $\mathbf{X}$ discretas)ou por uma função densidade de probabilidades (componentes de $\mathbf{X}$ contínuas). O símbolo θ representa um escalar ou um vetor real que, se conhecido, especifica completamente tanto a distribuição de probabilidades de X quanto o parâmetro de interêsse ξ.

△

Definição 1.3: Estimativa.

Uma estimativa para um parâmetro de interêsse ξ é um prognóstico do seu valor desconhecido obtido através de uma função,

$$(x_1,\ x_2,\ \ldots,\ x_n) \mapsto T(x_1,\ x_2,\ \ldots,\ x_n),$$

dos dados disponíveis.

△

Definição 1.3: Estimador.

Se a estimativa para um parâmetro de interêsse ξ é obtida através da função

$$(x_1,\ x_2,\ \ldots,\ x_n) \mapsto T(x_1,\ x_2,\ \ldots,\ x_n),$$

dos dados disponíveis, diz-se que a variável ou vetor aleatório

$$T(\mathbf{X}) \ = \ T(X_1,\ X_2,\ \ldots,\ X_n)$$

é um estimador de ξ.

△

1 Estimativas de Máxima Verossimilhança

Exemplo 2.1. Nas condições do **Exemplo 1.1**, não é difícil ver que a *estimativa*

$$T(x_1, x_2, \ldots, x_{13}) = \frac{x_1 + x_2 + \ldots + x_{13}}{13} = \frac{7}{13}$$

de ξ maximiza a

$$\begin{aligned} P\{X_1 = x_1, \ldots, X_{13} = x_{13}\} &= \prod_{i=1}^{13} f(x_i; \xi) \\ &= \prod_{i=1}^{13} \xi^{x_i}(1-\xi)^{1-x_i} \\ &= \xi^{x_1+x_2+\ldots+x_{13}}(1-\xi)^{13-(x_1+x_2+\ldots+x_{13})}, \end{aligned}$$

isto é,

$$\frac{7}{13} = \frac{x_1 + x_2 + \ldots + x_{13}}{13} \in argmax\{L(\xi) :\ 0 \leq \xi \leq 1\},$$

em que

$$\begin{aligned} L(\xi) &= \xi^{x_1+x_2+\ldots+x_{13}}(1-\xi)^{13-(x_1+x_2+\ldots+x_{13})} \\ &= \xi^7(1-\xi)^6, \quad 0 \leq \xi \leq 1 \ . \end{aligned}$$

Por esse motivo diz-se que essa *estimativa* é de *máxima verossimilhança.* A função $L(\xi)$, $0 \leq \xi \leq 1$, é denominada função de verosimilhança da amostra (ou dos dados).

$\triangle$

Definição 2.2: Estimativas de Máxima Verossimilhança.

Se $(x_1, x_2, \ldots, x_n)$ é uma ACS de uma distribuição $f(x; \theta)$, diz-se que uma estimativa $T(x_1, x_2, \ldots, x_n)$ de θ, é de *máxima verossimilhança* se

$$T(x_1, x_2, \ldots, x_n) \in argmax\{L(\theta);\ \theta \in \Theta\}$$

em que

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Para distingui-la, representa-se essa *estimativa* por

$$\hat{\theta}(x_1, x_2, \ldots, x_n)$$

ao invés de $T(x_1, x_2, \ldots, x_n)$. A variável ou vetor aleatório definido por

$$\hat{\theta}(X_1, X_2, \ldots, X_n),$$

é denominado *estimador de máxima verossimilhança* de θ.

Observação 2.2.1. A função $L(\theta)$ definida acima é denominada *função de verossimilhança* da amostra (ou dos dados).

Observação 2.2.2. $argmax\{L(\theta);\ \theta \in \Theta\}$ é conjunto dos valores de $\theta \in \Theta$ que maximizam $L(\theta)$.

Observação 2.2.3. A função

$$\begin{aligned} l(\theta) &= \ln L(\theta), \\ &= \sum_{i=1}^{n} \ln f(x_i;\theta), \quad \theta \in \Theta, \end{aligned}$$

é denominada *função escore* da amostra (ou dos dados).

Observação 2.2.4. $argmax\{L(\theta);\ \theta \in \Theta\} = argmax\{l(\theta);\ \theta \in \Theta\}$.

Observação 2.2.5. Se o parâmetro ξ a ser estimado é uma função de θ, isto é,

$$\xi = g(\theta),\ \theta \in \Theta,$$

a *estimativa de máxima verossimilhança* de ξ é dada por

$$\hat{\xi}(x_1, x_2, \ldots, x_n) = g(\hat{\theta}(x_1, x_2, \ldots, x_n)).$$

O *estimador de máxima verossimilhança* de ξ, neste caso, é

$$\hat{\xi}(X_1, X_2, \ldots, X_n) = g(\hat{\theta}(X_1, X_2, \ldots, X_n)).$$

$\triangle$

Exemplo 2.3. Suponha que o tempo de duração de um componente eletrônico é uma variável aleatória com distribuição exponencial com parâmetro $\lambda > 0$ desconhecido. Qual a *estimativa de máxima verossimilhança* do tempo médio de duração ξ desse componente se a soma das durações observadas de 10 componentes colocadas em teste foi de 7350 horas?

Sejam $x_1, x_2, \ldots, x_{10}$ as durações observadas dessas componentes. Admitindo que elas foram testadas independentemente sob as mesmas condições, essas observações constituem uma ACS de tamanho 10 da distribuição de probabilidades

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0 \ \text{ e } \ \lambda > 0.$$

Como tempo médio de duração ξ desse componente é igual a $\frac{1}{\lambda}$, sua *estimativa de máxima verossimilhança* é dada por

$$\hat{\xi}(x_1, x_2, \ldots, x_{10}) = \frac{1}{\hat{\lambda}(x_1, x_2, \ldots, x_{10})},$$

em que

$$\hat{\lambda}(x_1, x_2, \ldots, x_{10}) \in argmax\{l(\lambda) :\ \lambda > 0\}.$$

e

$$
\begin{aligned}
l(\lambda) &= \ln(\prod_{i=1}^{10} \lambda e^{-\lambda x_i}) \\
&= \ln(\lambda^{10} e^{-\lambda \sum_{i=1}^{10} x_i}) \\
&= 10 \ln \lambda - \lambda \sum_{i=1}^{10} x_i \ , \quad \lambda > 0.
\end{aligned}
$$

Portanto,

$$
argmax\{l(\lambda) : \ \lambda > 0\} = \{\lambda > 0 \ : \ \frac{10}{\lambda} = \sum_{i=1}^{10} x_i\}
$$

e

$$
\begin{aligned}
\hat{\xi}(x_1, x_2, \ldots, x_{10}) &= \frac{1}{\hat{\lambda}(x_1, x_2, \ldots, x_{10})} \\
&= \frac{1}{\frac{10}{x_1 + x_2 + \ldots + x_{10}}} \\
&= \frac{x_1 + x_2 + \ldots + x_{10}}{10} \ .
\end{aligned}
$$

Portanto, a *estimativa de máxima verossimilhança* do tempo médio de duração desse componente, ξ , é

$$
\frac{x_1 + x_2 + \ldots + x_{10}}{10} = \frac{7350}{10} = 735 horas.
$$

$\triangle$

2 Vício, Consistência e Eficiência

Definição 3.1: Vício.

Sejam: $(x_1, x_2, \ldots, x_n)$ uma ACS de uma distribuição $f(x; \theta)$, $\theta \in \Theta$; $T(X_1, X_2, \ldots, X_n)$ um *estimador* de $\xi = g(\theta) \in \Xi \subset \Re$; e

$$
\mu_T(\theta) = E(T(X_1, X_2, \ldots, X_n)).
$$

Diz-se que T é *não viciado* se

$$
\mu_T(\theta) = g(\theta).
$$

para todo $\theta \in \Theta$.

$\triangle$

Exemplo 3.2. No **Exemplo 1.1** O *estimador* de ξ é *não viciado*.

$\triangle$

Definição 3.3: Consistência.

Sejam: $(x_1, x_2, \ldots, x_n)$ uma ACS de uma distribuição $f(x;\theta)$, $\theta \in \Theta$; $T(X_1, X_2, \ldots, X_n)$ um *estimador* de $\xi = g(\theta) \in \Xi \subset \Re$; e

$$\sigma_T^2(\theta) = Var(T(X_1, X_2, \ldots, X_n))/.$$

Diz-se que T é *consistente* se

$$\lim_{n\to\infty} \sigma_T^2(\theta) = 0$$

para todo $\theta \in \Theta$.

Exemplo 3.4. No **Exemplo 1.1** O *estimador* de ξ é *consistente*.

$\triangle$

Definição 3.5: Eficiência.

Sejam: $(x_1, x_2, \ldots, x_n)$ uma ACS de uma distribuição $f(x;\theta)$, $\theta \in \Theta$; e $T_1(X_1, X_2, \ldots, X_n)$ e $T_1(X_1, X_2, \ldots, X_n)$ dois *estimadores não viciados* de $\xi = g(\theta) \in \Xi \subset \Re$. Diz-se que T_1 é *mais eficiente* que T_2 se

$$\sigma_{T_1}^2(\theta) \leq \sigma_{T_2}^2(\theta) \ ,$$

para todo $\theta \in \Theta$, e

$$\sigma_{T_1}^2(\theta_0) < \sigma_{T_2}^2(\theta_0)$$

para algum $\theta_0 \in \Theta$.