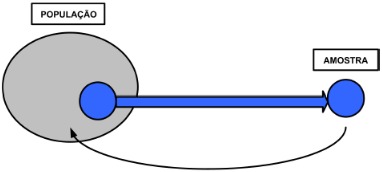
Métodos de Estimação Pontual

Ben Dêivide de Oliveira Batista

6 de março de 2016

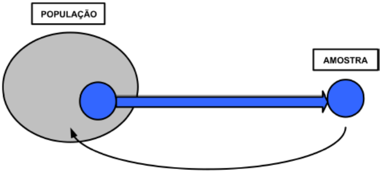
Sumário

Estimação



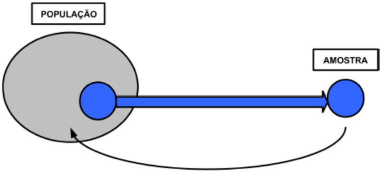
Estimação

- $X_1, X_2, \ldots, X_n$



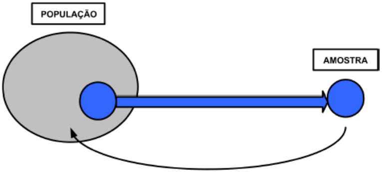
Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$



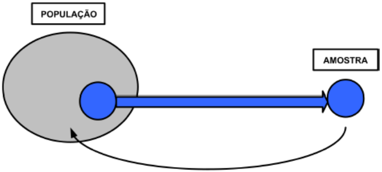
Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$
- T_n (Estimador Pontual) $\rightarrow \theta$ (Parâmetro)



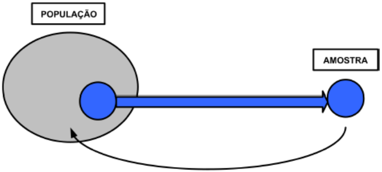
Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$
- T_n (Estimador Pontual) $\rightarrow \theta$ (Parâmetro)
- Métodos de estimação pontual:



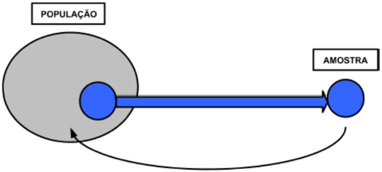
Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$
- T_n (Estimador Pontual) $\rightarrow \theta$ (Parâmetro)
- Métodos de estimação pontual:
    - Método dos momentos



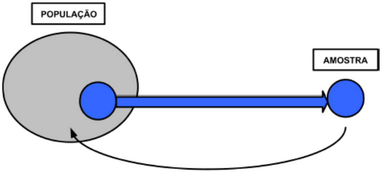
Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$
- $T_n\,(\text{Estimador Pontual}) \rightarrow \theta$ (Parâmetro)
- Métodos de estimação pontual:
  - Método dos momentos
  - Método da Máxima Verossimilhança



Estimação

- $X_1, X_2, \ldots, X_n \Rightarrow T_n = t(X_1, X_2, \ldots, X_n)$
- T_n (Estimador Pontual) $\to \theta$ (Parâmetro)
- Métodos de estimação pontual:
    - Método dos momentos
    - Método da Máxima Verossimilhança
    - Método dos Mínimos Quadrados



Definição 1 (Momentos amostrais)

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *com fdp ou fp* $f_X(x; \boldsymbol{\theta})$, *com* $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]' \in \Theta$ *em que* Θ *é o espaço paramétrico. O* k*-ésimo momento amostral, denotado por* M_k, *é definido por*

$$M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k, \tag{1}$$

e o k*-ésimo momento amostral em torno da média amostral* $\bar{X}$, *denotado por* M_k', *é definido por*

$$M_k' = \frac{1}{n} \sum_{i=1}^{n} (X - \bar{X})^k. \tag{2}$$

Definição 2 (Momentos populacionais)

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *com fdp ou fp* $f_X(x; \boldsymbol{\theta})$, *com* $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]' \in \Theta$ *em que* Θ *é o espaço paramétrico. O* k*-ésimo momento populacional, denotado por* μ_k, *é definido por*

$$\mu_k = E[X^k], \tag{3}$$

e o k*-ésimo momento populaciona em torno da média populacional* $\mu = E[X]$, *denotado por* μ'_k, *é definido por*

$$\mu'_k = E[(X - \mu)^k]. \tag{4}$$

□

Definição 2 (Momentos populacionais)

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *com fdp ou fp* $f_X(x; \boldsymbol{\theta})$, *com* $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]' \in \Theta$ *em que* Θ *é o espaço paramétrico. O* k*-ésimo momento populacional, denotado por* μ_k, *é definido por*

$$\mu_k = E[X^k], \tag{3}$$

e o k*-ésimo momento populaciona em torno da média populacional* $\mu = E[X]$, *denotado por* μ_k', *é definido por*

$$\mu_k' = E[(X - \mu)^k]. \tag{4}$$

□

Geralmente μ_k ou μ_k' é função dos k parâmetros $\theta_1, \theta_2, \ldots, \theta_k$

Definição 3 (Método dos momentos)

O método dos momentos consiste em igualar (1) *e* (3) *ou* (2) *e* (4)*, formando as k equações*

$$M_j = \mu_j(\theta_1, \theta_2, \ldots, \theta_k), \ \ para \ j = 1, 2, \ldots, k, \tag{5}$$

ou

$$M'_j = \mu'_j(\theta_1, \theta_2, \ldots, \theta_k), \ \ para \ j = 1, 2, \ldots, k, \tag{6}$$

sendo $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ suas soluções. Diremos que $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ são os estimadores de $\theta_1, \theta_2, \ldots, \theta_k$ pelo método dos momentos.

Exemplo 1

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *de uma normal com média* μ *e variância* σ^2 *desconhecidos. Denote* $(\theta_1, \theta_2) = \mu, \sigma^2$*. Vamos estimar os parâmetros* μ *e* σ^2 *pelo método dos momentos.*

Definição 4 (Função de verossimilhança)

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *(iid) com fdp ou fp conjunta* $f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$, *com* $\theta \in \Theta$ *em que* Θ *é o espaço paramétrico. Considere ainda* $x_1, x_2, \ldots, x_n$ *a realização da amostra aleatória* $X_1, X_2, \ldots, X_n$, *então a função de verossimilhança é definida por*

$$L(\theta; x_1, x_2, \ldots, x_n) = L(\theta; \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta). \quad (7)$$

Definição 5 (Método da máxima verossimilhança)

Seja uma função de verossimilhança $L(\theta; x_1, x_2, \ldots, x_n)$ *para uma amostra aleatória* $X_1, X_2, \ldots, X_n$. *Então o método da máxima verossimilhança é a forma de encontrar um* $\hat{\theta} = \vartheta(x_1, x_2, \ldots, x_n)$, *que é o valor estimado de* $\theta \in \Theta$ *que maximiza* $L(\theta; x_1, x_2, \ldots, x_n)$. *Dizemos que* $\hat{\theta} = \vartheta(X_1, X_2, \ldots, X_n)$ *é o estimador de máxima verossimilhança de* θ.

- Para maximizar $L(\theta; x_1, x_2, \ldots, x_n)$, tomamos a sua derivada em relação a θ, igualamos a zero e resolvemos o sistema para obtenção de $\hat{\theta} = \vartheta(X_1, X_2, \ldots, X_n)$;

- Para maximizar $L(\theta; x_1, x_2, \ldots, x_n)$, tomamos a sua derivada em relação a θ, igualamos a zero e resolvemos o sistema para obtenção de $\hat{\theta} = \vartheta(X_1, X_2, \ldots, X_n)$;
- Posteriormente, devemos identificar se a segunda derivada de $L(\theta; x_1, x_2, \ldots, x_n)$ é negativa para saber se $\hat{\theta}$ é um ponto de máximo.

- Para maximizar $L(\theta; x_1, x_2, \ldots, x_n)$, tomamos a sua derivada em relação a θ, igualamos a zero e resolvemos o sistema para obtenção de $\hat{\theta} = \vartheta(X_1, X_2, \ldots, X_n)$;
- Posteriormente, devemos identificar se a segunda derivada de $L(\theta; x_1, x_2, \ldots, x_n)$ é negativa para saber se $\hat{\theta}$ é um ponto de máximo.
- Muitas vezes esse processo torna-se complicado.

Definição 6 (Função de Log-verossimilhança)

Se $L(\theta; x_1, x_2, \ldots, x_n)$, expressão (7), é a função de verossimilhança, então

$$l(\theta; \boldsymbol{x}) = l(\theta; \boldsymbol{x}) = \log L(\theta; \boldsymbol{x}), \tag{8}$$

é a função de log-verossimilhança, para $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]'$.

Exemplo 2

Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ *de uma normal com média* μ *e variância* $\sigma^2 = 1$. *Vamos obter o estimador de* μ *pelo método da máxima verossimilhança.*

- Modelo de regressão:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{9}$$

$$\boldsymbol{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{X}_{n\times p'} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

$$\boldsymbol{\theta}_{p'\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ e } \boldsymbol{\varepsilon}_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, p' = p + 1$$

- Modelo de regressão:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{10}$$

As pressuposições para esse modelo são:

- $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, sendo $\mathbf{0}$ um vetor de zeros de dimensão $n \times 1$, ou equivalentemente $E[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\theta}$;

- Modelo de regressão:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{10}$$

As pressuposições para esse modelo são:

- $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, sendo $\mathbf{0}$ um vetor de zeros de dimensão $n \times 1$, ou equivalentemente $E[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\theta}$;
- $Var[\boldsymbol{\varepsilon}] = \boldsymbol{I}\sigma^2$, sendo $\boldsymbol{I}$ uma matriz identidade de dimensão $n \times n$, ou equivalentemente $Var[\boldsymbol{Y}] = \boldsymbol{I}\sigma^2$;

- Modelo de regressão:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{10}$$

As pressuposições para esse modelo são:

- $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, sendo $\mathbf{0}$ um vetor de zeros de dimensão $n \times 1$, ou equivalentemente $E[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\theta}$;
- $Var[\boldsymbol{\varepsilon}] = \boldsymbol{I}\sigma^2$, sendo $\boldsymbol{I}$ uma matriz identidade de dimensão $n \times n$, ou equivalentemente $Var[\boldsymbol{Y}] = \boldsymbol{I}\sigma^2$;
- $cov[\varepsilon_i, \varepsilon_j] = 0$ para todo $i \neq j$, ou equivalentemente, $cov[Y_i, Y_j] = 0$.

Teorema 1 (Método de mínimos quadrados)

Se $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \varepsilon$, em que $\boldsymbol{X}$ é $n \times p'$ de posto $p' < n$, então o valor de $\hat{\boldsymbol{\theta}} = [\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p]'$ que minimiza $\varepsilon'\varepsilon$ é igual a

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{11}$$

Assim, $\hat{\boldsymbol{\theta}}$ é conhecido como estimador de mínimos quadrados de $\boldsymbol{\theta}$.

Exemplo 3

Seja um modelo de regressão linear simples do tipo $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ *para* $i = 1, 2, \ldots, n$. *De modo matricial, temos*

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

em que está desejando estudar a relação entre a distância (pés) que um carro percorre até sua parada em função da velocidade limite (milhas por hora). (Dados: `cars` *do programa **R**)*