

Testes de hipóteses paramétricos

Conteúdos das próximas aulas:

1. Estimadores pontuais e suas propriedades
2. Métodos para calcular estimadores pontuais
Estimador de Máxima Verosimilhança (EMV)
3. Testes de hipóteses paramétricos
4. Testes em famílias de razão de verosimilhança
monótona (RVM)

Estimação pontual:

- Um valor numérico único usado para fazer uma inferência sobre um parâmetro desconhecido da população
- Um **estimador** T é uma função da amostra aleatória que é usada para estimar um parâmetro desconhecido θ .
- Um estimador é uma estatística

$\mathbf{X} = \{X_1, \dots, X_n\}$ a.a $X_i \sim f(x | \theta)$, $T(\mathbf{X})$ é uma v.a.

$\mathbf{x} = \{x_1, \dots, x_n\}$ uma amostra. $T(\mathbf{x})$ é uma estimativa

Exemplos de estimadores: $\mathbf{X} = \{X_1, \dots, X_n\} X_i \sim N(\mu, \sigma^2)$

Media amostral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad E(\bar{X}) = \mu, \text{Var}(\bar{X}) = \sigma^2/n$

Variância amostral bruta:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad E[S_n^2] = \frac{n-1}{n} \sigma^2, \quad \text{Var}(S_n^2) = \frac{2\sigma^4}{n-1}$$

Variância amostral:

$$S^2 = \frac{n}{n-1} S_n^2, \quad E(S^2) = \sigma^2, \quad \text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

- Um estimador $T(\mathbf{X})$ diz-se um **estimador centrado** (ou não enviesado) de algum parâmetro θ se:

$$E(T) = \theta, \text{ para todo } \theta \in \Theta$$

Se a igualdade acima não ocorre, dizemos que T é um estimador enviesado e a diferença $E(T) - \theta$ é chamado viés de T

$$\text{vies}_{\theta}(T) := E(X) - \theta$$

Exemplos de estimadores: $\mathbf{X} = \{X_1, \dots, X_n\} X_i \sim N(\mu, \sigma^2)$

Media amostral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $E(\bar{X}) = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$

Variância amostral bruta:

$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, $E[S_n^2] = \frac{n-1}{n} \sigma^2$, $\text{Var}(S_n^2) = \frac{2\sigma^4}{n-1}$

Variância amostral:

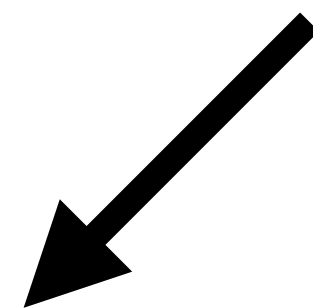
$S^2 = \frac{n}{n-1} S_n^2$, $E(S^2) = \sigma^2$, $\text{Var}(S^2) = \frac{2\sigma^4}{n-2}$

Como escolher um estimador?

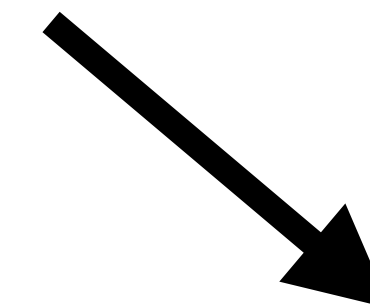
- Erro quadrático medio (EQM), Mean Squared Error (MSE).
- O EQM de um estimador W de um parâmetro θ é a função $E_{\theta}(W - \theta)^2$

$$\begin{aligned} E_{\theta}(W - \theta)^2 &= E_{\theta}(W^2 - 2W\theta + \theta^2) \\ &= E_{\theta}[W^2] - 2\theta E_{\theta}[W] + \theta^2 \\ &= E_{\theta}[W^2] - (E_{\theta}[W])^2 + (E_{\theta}[W])^2 - 2\theta E_{\theta}[W] + \theta^2 \\ &= \text{Var}_{\theta}(W) + (E_{\theta}[W] - \theta)^2 \\ &= \text{Var}_{\theta}(W) + (\text{vies}_{\theta} W)^2 \end{aligned}$$

$$E_{\theta}(W - \theta)^2 = \text{Var}_{\theta}(W) + (\text{vies}_{\theta}W)^2$$



Variabilidade



Precisão

Como escolher um estimador?

- Se o estimador é centrado, o EQM é igual a variância do estimador.
- Existe um estimador T tal que o EQM=0?
- Para resolver este problema pode-se reduzir a classe dos estimadores à classe dos estimadores centrados. T é ótimo segundo o EQM se minimizar o erro quadratico na classe dos estimadores centrados:

$$\min E_{\theta}((T - \theta)^2) \text{ na classe } T \text{ tal que } E_{\theta}(T) = \theta \Leftrightarrow \min \text{Var}_{\theta}(T)$$

Variância amostral bruta:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

$$E[S^2] = \sigma^2,$$

Centrado

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

Variância amostral:

$$S^2 = \frac{n}{n-1} S_n^2,$$

Enviesado

$$E(S^2) = \frac{n}{n-1} \sigma^2,$$

$$\text{Var}(S^2) = \frac{2(n-1)\sigma^4}{n^2}$$

$$\left(\frac{2n-1}{n^2} \right) \sigma^4 < \left(\frac{2}{n-1} \right) \sigma^4$$

O estimador enviesado têm
menos variância

Consistência de um estimador

- A consistência é uma propriedade assintótica do estimador, ou seja, o tamanho da amostra se aproxima de infinito.
- Em termos intuitivos é desejável que à medida que o tamanho da amostra aumenta o estimador esteja cada vez mais próximo do (verdadeiro valor do) parâmetro .
- Ser consistente significa que quando o tamanho da amostra é muito elevado a distribuição da estatística fica muito concentrada em torno do parâmetro da população.

Consistência

- Seja $\{T_n\}$ uma sequência de estimadores de um parâmetro de interesse θ . Dizemos que esta sequência é consistente se, dado $\varepsilon > 0$ arbitrário

$$P(|T_n - \theta| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty$$

- A sequência de estimadores $\{T_n\}$ de um parâmetro θ é consistente se

$$\lim_{n \rightarrow \infty} E_{\theta}(T_n) = \theta, \quad \text{e} \quad \lim_{n \rightarrow \infty} \text{Var}_{\theta}(T_n) = 0$$

Exemplo

Seja X_1, \dots, X_n uma a.a. i.i.d de uma população com média μ e variância σ^2 .

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad X' = \frac{1}{n+1} (2X_1 + X_2 + \dots + X_n)$$

Temos que $\mathbb{E}(\bar{X}) = \mathbb{E}(X') = \mu$, logo ambos são estimadores centrados de μ . Calculando as variâncias, temos que

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

$$\text{Var}(X') = \text{Var}\left(\frac{2X_1 + X_2 + \dots + X_n}{n+1}\right) = \frac{(n+3)\sigma^2}{(n+1)^2}.$$

Exemplo

Seja X_1, \dots, X_n uma a.a. i.i.d de uma população com média μ e variância σ^2 .

Consequentemente,

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

$$\lim_{n \rightarrow \infty} \text{Var}(X') = \lim_{n \rightarrow \infty} \frac{(n+3)\sigma^2}{(n+1)^2} = 0$$

Logo, tanto \bar{X} quanto X' são estimadores consistentes para o parâmetro μ .

Propiedades desejáveis dos Estimadores

- **Centricidade:** a média do estimador é igual ao parâmetro da população que queremos estimar. Ou seja, em média o estimador está correcto.
- **Consistência:** À medida que o tamanho da amostra é mais elevado, o estimador da-nos uma ideia cada vez mais precisa de qual é o verdadeiro valor do parâmetro.
- **Eficiência:** No conjunto de estimadores não enviesados o mais eficiente é aquele que tem menor variabilidade (que é mais preciso).

Método da Máxima Verosimilhança (MMV)

- O MMV é um dos métodos de obtenção de estimadores mais utilizados.
- Responde à pergunta: Para que valor do parâmetro os dados observados têm mais probabilidade?
- O MMV permite obter o valor mais plausível/verossímil de um parâmetro desconhecido - de entre todos os valores possíveis para esse mesmo parâmetro - tendo em conta a amostra de que dispomos.
- Ideia base: seleccionar, entre todos os valores possíveis dos parâmetros populacionais, aqueles que tornem mais verossímil a ocorrência de uma amostra idêntica àquela que efectivamente se obteve.
- Vantagens: boas propriedades assintóticas; modelo explícito; estatisticamente bem compreendido; robusto para muitas violações de condições; menor variância que outros métodos;
- Desvantagens: cálculo computacional muito intensivo e extremamente lento (cada vez menos um problema)

