ACH2053 – Introdução à Estatística

Aula 09b: Estimadores

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2025

Exemplo

Considere uma caixa com 10 dados seguindo a seguinte distribuição: 5 dados com faces (111223), 3 dados com faces (112233), e 2 dados com faces (122333). Considere o seguinte experimento:

1. um dado foi retirado aleatoriamente da caixa;
2. o dado foi jogado 6 vezes e os seguintes valores foram obtidos: 3, 2, 1, 2, 3, 2.

Responda:

1. Como são as faces do dado retirado?
2. Repita o exercício considerando que na caixa exista apenas um dado de cada um dos 3 tipos.
3. Repita o exercício considerando que você não tem nenhuma informação sobre os dados na caixa.

Estimadores Bayesianos

Considere que o parâmetro $\theta_0 \in \Omega$ é uma variável aleatória e é distribuida de acordo com a p.d.f. (p.m.f.) $f(\theta)$ sobre o espaço de parâmetros $\Omega \in \mathbb{R}^d$.

Considere que as n variáveis aleatórias $X_1, \ldots, X_n$ observadas são independentes e identicamente distribuidas de acordo com a p.d.f (p.m.f.) condicional $f(x|\theta)$.

Então, seguindo o teorema de Bayes, temos que:

$$
\begin{aligned}
\Pr(\theta_0 = \theta | X_1 = x_1, \ldots, X_n = x_n) &= \frac{f(x_1|\theta) \cdots f(x_n|\theta) f(\theta)}{\Pr(X_1 = x_1, \ldots, X_n = x_n)} \\
&= \frac{\prod_{i=1}^{n} f(x_i|\theta) f(\theta)}{\int_{\theta'} \prod_{i=1}^{n} f(x_i|\theta') f(\theta') d\theta'}
\end{aligned}
$$

Estimadores Bayesianos

Considere a p.d.f. (p.m.f.) condicional dada por:

$$f_n(\theta|x_1, \ldots, x_n) = \Pr(\theta_0 = \theta|X_1 = x_1, \ldots, X_n = x_n).$$

O estimador bayesiano:

$$\hat{\theta} = \arg\max_{\theta \in \Omega} f_n(\theta|x_1, \ldots, x_n)$$

é chamado de estimador bayesiano Maximum a Posteriori (MAP).

O estimador bayesiano:

$$\hat{\theta} = \int_\theta \theta f_n(\theta|x_1, \ldots, x_n) d\theta = \mathsf{E}_{\theta_0|X_1,\ldots,X_n}(\theta_0)$$

é chamado de estimador bayesiano Expectation a Posteriori (EAP).

Estimador de Máxima Verossimilhança

Considere a p.d.f (p.m.f) conjunta $f_n(\mathbf{x}|\theta)$. Se essa função é interpretada como uma função de θ com parâmetros $\mathbf{x} = (x_1, \ldots, x_n)$, então ela é chamada de função de Verossimilhança (likelihood) e é denotada por $L(\theta; \mathbf{x})$.

Suponha que as n variáveis aleatórias $X_1, \ldots, X_n$ formam uma amostra aleatória de uma distribuição para qual a p.d.f. (p.m.f.) condicional é $f(X|\theta)$. Então:

$$L(\theta; \mathbf{x}) = f(x_1|\theta) \cdots f(x_n|\theta).$$

Para cada possível vetor de observação $\mathbf{x} = (x_1, \ldots, x_n)$, defina $\hat{\theta} = \arg\max_{\theta \in \Omega} L(\theta; \mathbf{x})$. A estimativa $\hat{\theta}$ é a estimativa de máxima verossimilhança (M.L.E. - maximum likelihood estimator).

Estimador de Máxima Verossimilhança

Definition (Divergência de Kullback-Leibler)

Seja $p(x)$ e $q(x)$ duas p.d.f. sobre $\mathbb{R}$. A divergência (distância) de Kullback-Leibler, de $q(x)$ com respeito a $p(x)$ é definida como:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx.$$

Teorema

Uma estimativa $\hat{\theta}$ baseada nas amostras $\mathbf{x} = (x_1, \ldots, x_n)$ é um M.L.E. se e somente se para todo $\theta \in \Omega$:

$$KL[\widehat{f_n}(x)||f(x;\hat{\theta})] \leq KL[\widehat{f_n}(x))||f(x;\theta)],$$

onde $\widehat{f_n}(x)$ é a distribuição discreta com base na amostra.

Função Log-likelihood

Seja $\hat{\theta}$ o M.L.E. de θ, se $g : \mathbb{R} \to \mathbb{R}$ é uma função estritamente crescente, então $\hat{\theta} = \arg\max_{\theta \in \Omega} g[L(\theta; \mathbf{x})]$.

Para encontrar o M.L.E. usualmente considera-se a transformação $\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x})$ e resolve-se a seguinte equação:

$$\nabla_\theta \log L(\theta) = 0.$$

O estimador M.L.E. não necessariamente é único e também pode não existir dependendo da classe de distribuição.

Exercício: encontre o estimador M.L.E. para uma amostrada obtida de uma distribuição de Bernoulli.

Estimador de Máxima Verossimilhança

Considere um exame com 3 questões. Cada questão i é modelada por uma dificuldade $b_i \in \mathbb{R}$ e cada aluno é modelado por uma habilidade $\theta \in (-\infty, +\infty)$. Considere então o seguinte experimento: sorteie uma habilidade θ da distribuição $f(\theta) = \frac{e^\theta}{(1+e^\theta)^2}$ e aplique o exame.
Considere as variáveis aleatórias binárias independentes X_i que indica se o aluno acertou ou não a questão i com distribuição condicional:

$$f(x_i = 1|\theta) = \begin{cases} 0 & \text{, se } \theta < b_i - 2 \\ 0.5 + 0.25(\theta - b_i) & \text{, se } \theta \in [b_i - 2, b_i + 2] \\ 1 & \text{, se } \theta > b_i + 2 \end{cases} .$$

Considere que o exame é formado por questões com as seguintes dificuldades: $b_1 = -0.7$, $b_2 = 0.1$ e $b_3 = 0.5$ e que T é a variável aleatória representando a habilidade do aluno.

Seja $s(\theta) = \Pr(X_1 = 1, X_2 = 1, X_3 = 0|T = \theta)$, encontre θ que maximize $s(\theta)$

Lei dos Números Grandes

Inequação de Markov
Suponha que X é uma variável aleatória tal que $\Pr(X \geq 0) = 1$. Então para todo número real $t > 0$,

$$\Pr(X \geq t) \leq \frac{\mathsf{E}(X)}{t}.$$

Inequação de Chebyshev
Suponha que X é uma variável aleatória tal que exista $\mathsf{Var}(X)$. Então para todo número real $t > 0$,

$$\Pr(|X - \mathsf{E}(X)| \geq t) \leq \frac{\mathsf{Var}(X)}{t^2}.$$

Teorema

Seja $X_1, \dots, X_n$ amostras aleatórias de uma distribuição com média μ e desvio padrão σ. Seja $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média das amostras. Então:

$$\mathsf{E}(\overline{X}_n) = \mu \text{ e } \mathsf{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Teorema (Lei dos Números Grandes)

Suponha que $X_1, \dots, X_n$ forme uma amostra aleatória de uma distribuição com média μ e variância finita. Seja $\overline{X}_n$ a média das amostras e $g(z)$ uma função contínua em $z = \mu$. Então:

$$\overline{X}_n \xrightarrow{p} \mu \text{ e } g(\overline{X}_n) \xrightarrow{p} g(\mu).$$

Método dos Momentos

Seja $\mathcal{F}_\theta$ um espaço de c.d.f., encontre funções $U : \mathbb{R} \to \mathbb{R}^d$ e $V : \Omega \to \mathbb{R}^d$ inversível, tal que $\mathsf{E}[U(X)] = V(\theta)$. Então construa o estimador:

$$\hat{\theta} = V^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} U(x_i)\right).$$

Teorema

O Método dos Momentos produz um estimador consistente.

> Exercício: considerando o Métodos dos Momentos encontre um estimador para a distribuição uniforme contínua entre a e b.