# Inferência Estatística

Professor: Pedro M. Almeida-Junior

22 de dezembro de 2021

Departamento de Estatística (UEPB)

# Consistência

- Um conceito importante é o da consistência do estimador, ou seja, à medida que o tamanho da amostra aumenta, os estimadores ficam tão próximos do parâmetro que está sendo estimado quanto desejado

- Seja $X_1, \ldots, X_n$ uma AAS de uma distribuição com parâmetro desconhecido $\theta$. Dizemos que o estimador $\widehat{\theta}$ é consistente para o parâmetro $\theta$, se

$$\lim_{n \to \infty} P\left(\mid \widehat{\theta} - \theta \mid > \epsilon\right) = 0 \quad \text{ou} \quad \widehat{\theta} \xrightarrow{P} \theta$$

Em geral, usamos a desigualdade de Markov para verificação dessa propriedade. Pois, segundo tal desigualdade, seja $Z$ uma v.a. assumindo valores reais positivos. Então, temos que para qualquer $\epsilon > 0$:

$$P(Z > \epsilon) \leq \frac{\mathbb{E}(Z)}{\epsilon}$$

Assim, podemos observar que, seja $Z = |\widehat{\theta} - \theta|$ e $\epsilon > 0$, temos que

$$\begin{aligned} P(|\hat{\theta} - \theta| > \epsilon) &= P(|\hat{\theta} - \theta|^2 > \epsilon^2) \\ &\leq \frac{\mathbb{E}\left((\hat{\theta} - \theta)^2\right)}{\epsilon^2} \\ &= \frac{Var(\hat{\theta}) + b^2(\hat{\theta})}{\epsilon^2} \end{aligned}$$

ou seja,

$$P(|\hat{\theta} - \theta| > \epsilon) \leq \frac{Var(\hat{\theta}) + b^2(\hat{\theta})}{\epsilon^2}$$

Se o estimador $\widehat{\theta}$ é um **não viesado** para $\theta$, então podemos usar a desigualdade de Chebyschev, e neste caso, se $Var(\widehat{\theta}) < \infty$, então,

$$P(|\hat{\theta} - \theta| > \epsilon) \leq \frac{Var(\hat{\theta})}{\epsilon^2}$$

**Exemplo:** Suponha que $X_1, \ldots, X_n$ uma AAS da v.a. $X$ tal que $\mathbb{E}(X) = \mu$ e $V(X) = \sigma^2 < \infty$. Então, $\hat{\mu} = \bar{X}$ é um estimador não viesado e consistente para o parâmetro $\mu$.

# Suficiência

- Até agora vimos formas de verificar se um estimador possui boas propriedades;
- Porém, precisamos estabelecer métodos de construções de estimadores que tenham alguma propriedade interessante, ou que levem a estimadores com 'boas' propriedades
- Antes, vamos considerar estatísticas que reduzam os dados sem que haja perda de informação; Daí, vem o conceito de estatísticas suficientes.

- Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X$ com função de densidade ou de probabilidade $f(x \mid \theta)$. Quando resumimos a informação que os dados contêm sobre $\theta$, utilizando uma estatística, é importante que não haja perda de informação sobre $\theta$.

- Ou seja, a estatística a ser considerada deve, dentro do possível, conter toda a informação sobre $\theta$ presente na amostra.

# Princípio da Suficiência

- Em outras palavras, se pudermos usar uma estatística $T = T(X_1, \ldots, X_n)$ para extrairmos toda informação que a amostra $X_1, \ldots, X_n$ contém sobre $\theta$, então dizemos que $T$ (que pode ser um vetor) é suficiente para $\theta$.

- Desse modo, o conhecimento apenas de $T$ (e não necessariamente da amostra completa $X_1, \ldots, X_n$ ) é suficiente para que sejam feitas inferências sobre $\theta$.

**Definição:** Sejam $X_!, \ldots, X_n$ uma amostra aleatória simples de uma variável aleatória com distribuição envolvendo o parâmetro desconhecido $\theta$. Dizemos que a estatística $T = T(X_1, \ldots, X_n)$ é suficiente para $\theta$, quando a distribuição condicional de $X_1, \ldots, X_n$ dado $T$ for independente de $\theta$.

**Exemplo 1:** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição Binomial$(1, \theta)$, ou seja, de Bernoulli$(\theta)$. Vamos verificar se a estatística $T = \sum_{i=1}^{n} X_i$ é suficiente para $\theta$.

**Exemplo 2:** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição Poisson com parâmetro $\theta$. Vamos verificar se a estatística $T = \sum_{i=1}^{n} X_i$ é suficiente para $\theta$.

**Teorema: (Critério da Fatoração de Neyman)** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória $X$ com função de densidade (ou de probabilidade) $f(x \mid \theta)$ e função de verossimilhança $L(\theta; \mathrm{x})$. Temos, então, que a estatística $T = T(X_1, \ldots, X_n)$ é suficiente para $\theta$, se e somente se pudermos escrever

$$L(\theta; \mathrm{x}) = h(x_1, \ldots, x_n)\, g_\theta\left(T(x_1, \ldots, x_n)\right),$$

onde $h(x_1, \ldots, x_n)$ é uma função que depende apenas de $x_1, \ldots, x_n$ (não depende de $\theta$) e $g_\theta\left(T(x_1, \ldots, x_n)\right)$ depende de $\theta$ e de $x_1, \ldots, x_n$ somente através de $T$.

**Exemplo 3:** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição Binomial$(1, \theta)$, ou seja, de Bernoulli$(\theta)$. Vamos verificar se a estatística $T = \sum_{i=1}^{n} X_i$ é suficiente para $\theta$ usando o critério da Fatoração.

**Exemplo 4:** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição Poisson com parâmetro $\theta$. Vamos verificar se a estatística $T = \sum_{i=1}^{n} X_i$ é suficiente para $\theta$ usando o critério da Fatoração.

# Estatísticas conjuntamente Suficientes

- Vamos considerar agora o caso multiparamétrico em que $\boldsymbol{\theta}$ é um vetor de parâmetros.

- No caso do modelo normal $\mathcal{N}(\mu, \sigma^2)$, com $\boldsymbol{\theta} = (\mu, \sigma^2)$, sendo $\mu$ e $\sigma^2$ desconhecidos é um exemplo de distribuições com mais de um parâmetros.

**Teorema (Critério da fatoração. Caso Multiparamétrico):** Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória $X$, com função de densidade (ou de probabilidade) $f(x \mid \theta)$. Temos, então, que a estatística $r$ dimensional $\mathrm{T} = (T_1, \ldots, T_r)$, $T_i = T_i(\mathrm{X})$ é conjuntamente suficiente para $\theta$ se

$$
\begin{aligned}
L(\theta; \mathrm{x}) &= f(x_1, \ldots, x_n \mid \theta) \\
&= \prod_{i=1}^{n} f(x_i \mid \theta) \\
&= h(x_1, \ldots, x_n)\, g_\theta(T_1(\mathrm{x}), \ldots, T_r(\mathrm{x}))
\end{aligned}
$$

onde $h(x_1, \ldots, x_n)$ é uma função que não depende de $\theta$ e $g_\theta(T_1(\mathrm{x}), \ldots, T_r(\mathrm{x}))$ depende de $\theta$ e de $\mathrm{x} = (x_1, \ldots, x_n)$ somente por meio de $(T_1(\mathrm{x}), \ldots, T_r(\mathrm{x}))$.

**Exemplo:** Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho $n$ da variável aleatória $X \sim N\left(\mu, \sigma^2\right)$, onde $\mu$ e $\sigma^2$ são desconhecidos. Temos, então, que $\theta = \left(\mu, \sigma^2\right)$.