Inferência Estatística II

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Programa de Pós-Graduação em Estatística Aplicada e Biometria
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Consistência

Definição 1

Seja X uma variável aleatória com função de distribuição acumulada $F(x, \theta)$, $\theta \in \mathcal{A} \subseteq \Omega$. Seja $X_1, \ldots, X_n$ uma amostra da distribuição de X e seja T_n uma estatística ($T_n = T(X_1, \ldots, X_n)$). Dizemos que T_n é um estimador consistente para θ se $T_n \xrightarrow{P} \theta$.

Exemplo

Sejam $X_1, \ldots, X_n, \ldots$ uma sequência de variáveis aleatórias iid de uma distribuição com média finita μ e variância $\sigma^2 < +\infty$, então, pela Lei Fraca dos Grandes Números, temos que, $\bar{X}_n = \frac{\displaystyle\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} \mu$. Ou seja, $\bar{X}_n$ é um estimador consistente de μ.

Exemplo

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição com média μ e variância $\sigma^2 < +\infty$. Suponha que $E[X_1^4] < +\infty$, de tal forma que $Var(S^2) < +\infty$.

$$
\begin{aligned}
S_n^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2 \right) \\
&= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \right) \\
&\xrightarrow{P} 1 \cdot [E(X_1^2) - \mu^2] = \sigma^2.
\end{aligned}
$$

Portanto, a variância da amostra é um estimador consistente de σ^2. A partir da discussão acima, temos imediatamente que $S_n \xrightarrow{P} \sigma$; ou seja, o desvio padrão da amostra é um estimador consistente do desvio padrão populacional. Vejam estes exemplos anteriores na prática, Clique aqui!.

Exemplo

Considere $X_i \overset{\text{iid}}{\sim} U(0, \theta)$, $i = 1, 2, \ldots, n$, e $Y_n = max\{X_1, \ldots, X_n\}$. Seja $\varepsilon > 0$, segue que:

$$\begin{aligned} P(|Y_n - \theta| \geq \varepsilon) &= P(\theta - Y_n \geq \varepsilon) \\ &= P(Y_n \leq \theta - \varepsilon). \end{aligned}$$

Exemplo

Considere $X_i \overset{\text{iid}}{\sim} U(0, \theta)$, $i = 1, 2, \ldots, n$, e $Y_n = max\{X_1, \ldots, X_n\}$. Seja $\varepsilon > 0$, segue que:

$$\begin{aligned} P(|Y_n - \theta| \geq \varepsilon) &= P(\theta - Y_n \geq \varepsilon) \\ &= P(Y_n \leq \theta - \varepsilon). \end{aligned}$$

Se $\theta - \varepsilon \leq 0$, então $P(Y_n \leq \theta - \varepsilon) = 0$, pois $0 \leq Y_n \leq \theta$, com $P(0 \leq Y_n \leq \theta) = 1$.

Exemplo

Considere $X_i \stackrel{\text{iid}}{\sim} U(0,\theta)$, $i = 1, 2, \dots, n$, e $Y_n = max\{X_1, \dots, X_n\}$. Seja $\varepsilon > 0$, segue que:

$$\begin{aligned} P(|Y_n - \theta| \geq \varepsilon) &= P(\theta - Y_n \geq \varepsilon) \\ &= P(Y_n \leq \theta - \varepsilon). \end{aligned}$$

Se $\theta - \varepsilon \leq 0$, então $P(Y_n \leq \theta - \varepsilon) = 0$, pois $0 \leq Y_n \leq \theta$, com $P(0 \leq Y_n \leq \theta) = 1$.

Se $0 < \varepsilon < \theta$ então,

$$P(|Y_n - \theta| \geq \varepsilon) = P(Y_n \leq \theta - \varepsilon) = F_{Y_n}(\theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \xrightarrow[n \to \infty]{} 0$$

Ou seja, $Y_n \xrightarrow{P} \theta$. Logo, Y_n é um estimador consistente para θ.

Para 🏠

Exercícios 2.8.18, 5.1.2, 5.1.3, 5.1.7 e 5.1.9

Convergência em Distribuição

Definição 2

Seja $\{X_n\}_{n\geq 1}$ *uma sequência de variáveis aleatórias com função de distribuição* F_{X_n}, $n \geq 1$. *Seja* X *uma variável aleatória com função de distribuição* F_X. *Seja* $C(F_X)$ *o conjunto de todos os pontos de continuidade de* F_X. *Dizemos que* X_n *converge em distribuição para* X *se,*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \ \forall x \in C(F_X).$$

Denotamos essa convergência por $X_n \xrightarrow{D} X$ *ou* $X_n \xrightarrow{\mathcal{L}} X$.

Convergência em Distribuição

Definição 2

Seja $\{X_n\}_{n\geq 1}$ *uma sequência de variáveis aleatórias com função de distribuição* F_{X_n}, $n \geq 1$. *Seja* X *uma variável aleatória com função de distribuição* F_X. *Seja* $C(F_X)$ *o conjunto de todos os pontos de continuidade de* F_X. *Dizemos que* X_n *converge em distribuição para* X *se,*

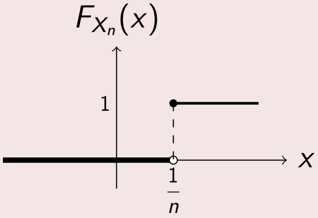
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \ \forall x \in C(F_X).$$

Denotamos essa convergência por $X_n \xrightarrow{D} X$ *ou* $X_n \xrightarrow{\mathcal{L}} X$.

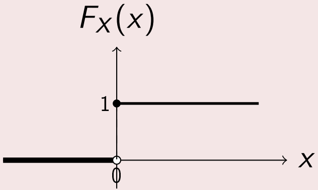
À medida que o tamanho da amostra aumenta, a distribuição das médias amostrais se aproxima da distribuição normal, veja isso acontecendo na prática, Clique aqui!.

Exemplo

$$P\left(X_n = \frac{1}{n}\right) = 1, \forall n \geq 1,\ P(X=0) = 1, \mathcal{C}(F_X(x)) = \{x \in \mathbb{R}; x \neq 0\}$$



(a) $P\left(X_n = \frac{1}{n}\right) = 1$



(b) $P(X=0) = 1$

Figura: $\lim_{n \to +\infty} F_{X_n}(x) = F_X(x),\ \forall x \neq 0,$ ou seja $X_n \xrightarrow{D} X$

Exemplo 2 (Convergência em Distribuição não Implica Convergência em Probabilidade)

Seja X uma variável aleatória contínua simétrica em torno do zero (ou seja, se f denota sua densidade, então $f(x) = f(-x), \forall x \in \mathbb{R}$. Neste caso, X e $-X$ tem a mesma distribuição (Verifiquem!). Defina a sequência de variáveis aleatórias X_n como: $X_n = \begin{cases} X, & \text{se } n \text{ é par} \\ -X, & \text{se } n \text{ é impar} \end{cases}$

Exemplo 2 (Convergência em Distribuição não Implica Convergência em Probabilidade)

Seja X uma variável aleatória contínua simétrica em torno do zero (ou seja, se f denota sua densidade, então $f(x) = f(-x), \forall x \in \mathbb{R}$. Neste caso, X e $-X$ tem a mesma distribuição (Verifiquem!). Defina a sequência de variáveis aleatórias X_n como: $X_n = \begin{cases} X, & \text{se } n \text{ é par} \\ -X, & \text{se } n \text{ é impar} \end{cases}$

É fácil ver que $F_{X_n}(x) = F_X(x)$. Logo, $X_n \xrightarrow{D} X$. Porém, $X_n \overset{P}{\not\to} X$, pois

$$P(|X_n - X| \geq \varepsilon) = \begin{cases} 0, & \text{se } n \text{ é par} \\ P(2|X| \geq \varepsilon), & \text{se } n \text{ é impar} \end{cases}$$

Exemplo 3

Seja T_n uma variável aleatória com distribuição t-Student com n graus de liberdade, ou seja, a densidade de T_n é dada por:

$$f_{T_n}(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{y^2}{n}\right)^{\frac{n+1}{2}}}, \ y \in \mathbb{R}$$

Exemplo 3

Seja T_n uma variável aleatória com distribuição t-Student com n graus de liberdade, ou seja, a densidade de T_n é dada por:

$$f_{T_n}(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{y^2}{n}\right)^{\frac{n+1}{2}}}, \quad y \in \mathbb{R}$$

Temos que,

$$\lim_{n \to +\infty} F_{T_n}(t) = \lim_{n \to +\infty} \int_{-\infty}^{t} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{y^2}{n}\right)^{\frac{n+1}{2}}} dy$$

$$\lim_{n\to+\infty} F_{T_n}(t) = \lim_{n\to+\infty}\int_{-\infty}^{t} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{y^2}{n}\right)^{\frac{n+1}{2}}}dy$$

$$\underset{\substack{\uparrow\\ \text{Teorema da}\\ \text{Convergência Dominada}}}{=} \int_{-\infty}^{t} \lim_{n\to+\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{y^2}{n}\right)^{\frac{n+1}{2}}}dy$$

$$= \star\star$$

Considere a seguinte aproximação de Stirling (Conhecida como fórmula de Stirling):

$$\Gamma(t+1) \approx \sqrt{2\pi t}\left(\frac{t}{e}\right)^t$$

Ou seja,

$$\lim_{t \to +\infty} \frac{\Gamma(t+1)}{\sqrt{2t\pi}\left(\frac{t}{e}\right)^t} = 1$$

Considere a seguinte aproximação de Stirling (Conhecida como fórmula de Stirling):

$$\Gamma(t+1) \approx \sqrt{2\pi t}\left(\frac{t}{e}\right)^{t}$$

Ou seja,

$$\lim_{t\to+\infty} \frac{\Gamma(t+1)}{\sqrt{2t\pi}\left(\frac{t}{e}\right)^{t}} = 1$$

Logo,

$$\lim_{n\to+\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{y^2}{n}\right)^{\frac{n+1}{2}}} = \lim_{n\to+\infty} \frac{\sqrt{2\pi}\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}+\frac{1}{2}} e^{-\left(\frac{n-1}{2}\right)}}{\sqrt{n}\sqrt{2\pi}\left(\frac{n-2}{2}\right)^{\frac{n-2}{2}+\frac{1}{2}} e^{-\left(\frac{n-2}{2}\right)}} \frac{1}{\left(1+\frac{y^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \star \text{ (t da fórmula de Stirling será } \frac{n-1}{2}\text{)}$$

$$
\begin{aligned}
\star &= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} \lim_{n\to\infty} \frac{(n-1)^{\frac{n}{2}}}{(n-2)^{\frac{n}{2}}(n-2)^{-\frac{1}{2}}} \frac{1}{\sqrt{n}} e^{-\frac{y^2}{2}} \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \lim_{n\to\infty} \left(1 + \frac{1}{n-2}\right)^{\frac{n}{2}} \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} e^{\frac{1}{2}} = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}
\end{aligned}
$$

$$
\begin{aligned}
\star &= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} \lim_{n\to\infty} \frac{(n-1)^{\frac{n}{2}}}{(n-2)^{\frac{n}{2}}(n-2)^{-\frac{1}{2}}} \frac{1}{\sqrt{n}} e^{-\frac{y^2}{2}} \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \lim_{n\to\infty} \left(1 + \frac{1}{n-2}\right)^{\frac{n}{2}} \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} e^{\frac{1}{2}} = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}
\end{aligned}
$$

Portanto, substituindo em $\star\star$, temos

$$
\lim_{n\to+\infty} F_{T_n}(t) = \int_{-\infty}^{t} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy
$$

Logo, $T_n \xrightarrow{D} N(0,1)$.

Teorema 1

Se $X_n \xrightarrow{P} X$, *então* $X_n \xrightarrow{D} X$.

Teorema 1

Se $X_n \xrightarrow{P} X$, *então* $X_n \xrightarrow{D} X$.

Demonstração do Teorema 1

Seja x um ponto de continuidade de $F_X(x)$, a função de distribuição acumulada (FDA) de X. Queremos mostrar que $F_{X_n}(x) \to F_X(x)$ à medida que $n \to \infty$, onde $F_{X_n}(x)$ é a FDA de X_n.
Para isso, partimos da definição de $F_{X_n}(x) = P(X_n \leq x)$. Usaremos uma técnica dividindo a probabilidade em dois pedaços.

Demonstração do Teorema 1

Dividimos o evento $\{X_n \leq x\}$ em dois subconjuntos: um onde $|X_n - X| < \varepsilon$ e outro onde $|X_n - X| \geq \varepsilon$. Assim, podemos reescrever:

$$
\begin{aligned}
F_{X_n}(x) &= P(X_n \leq x) \\
&= P(\{X_n \leq x\} \cap \{|X_n - X| < \varepsilon\}) \\
&+ P(\{X_n \leq x\} \cap \{|X_n - X| \geq \varepsilon\}) \\
&\leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon)
\end{aligned}
$$

Essa é uma decomposição da probabilidade em duas partes: uma onde X_n está "perto"de X (a diferença é menor que ε) e outra onde X_n está "longe"de X (a diferença é maior ou igual a ε).

Demonstração do Teorema 1

A probabilidade $P(\{X_n \leq x\} \cap \{|X_n - X| < \varepsilon\})$ pode ser estimada por $P(X \leq x + \varepsilon)$.

$$P(\{X_n \leq x\} \cap \{|X_n - X| < \varepsilon\}) \leq P(X \leq x + \varepsilon)$$

Isso porque, quando $|X_n - X| < \varepsilon$, sabemos que X_n está perto de X, então $X_n \leq x$ implica que $X \leq x + \varepsilon$.

Demonstração do Teorema 1

O segundo termo, $P(\{X_n \leq x\} \cap \{|X_n - X| \geq \varepsilon\})$, é menor ou igual a $P(|X_n - X| \geq \varepsilon)$, que é simplesmente a probabilidade de X_n estar longe de X. Essa probabilidade tende a 0 quando $X_n \to X$ em probabilidade, mas por enquanto, deixamos essa expressão como está:

$$P(\{X_n \leq x\} \cap \{|X_n - X| \geq \varepsilon\}) \leq P(|X_n - X| \geq \varepsilon)$$

Demonstração do Teorema 1

Juntando as duas estimativas, temos:

$$F_{X_n}(x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon)$$

Esta é a estimativa superior para $F_{X_n}(x)$.

- O primeiro termo, $P(X \leq x + \varepsilon)$, representa o evento de que X está um pouco acima de x. Isso é um "ajuste", pois estamos lidando com X_n próximo de X.

Demonstração do Teorema 1

Juntando as duas estimativas, temos:

$$F_{X_n}(x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon)$$

Esta é a estimativa superior para $F_{X_n}(x)$.

- O primeiro termo, $P(X \leq x + \varepsilon)$, representa o evento de que X está um pouco acima de x. Isso é um "ajuste", pois estamos lidando com X_n próximo de X.
- O segundo termo, $P(|X_n - X| \geq \varepsilon)$, é a probabilidade de que X_n esteja muito distante de X, ou seja, mais de ε de diferença.

Demonstração do Teorema 1

Quando $X_n \to X$ em probabilidade, sabemos que $P(|X_n - X| \geq \varepsilon) \to 0$ conforme $n \to \infty$. Portanto, com base nessa desigualdade, podemos concluir:

$$\lim_{n\to\infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

Isso nos dá a estimativa superior (upper bound) da função de distribuição acumulada de X_n.

Demonstração do Teorema 1

Agora, para obter a **estimativa inferior**, começamos reescrevendo $P(X_n \leq x)$ utilizando o complemento:

$$P(X_n \leq x) = 1 - P(X_n > x)$$

Dividimos a probabilidade $P(X_n > x)$ em dois pedaços:

$$P(X_n > x) = P(\{X_n > x\} \cap \{|X_n - X| < \varepsilon\}) + P(\{X_n > x\} \cap \{|X_n - X| \geq \varepsilon$$

Demonstração do Teorema 1

- A primeira parte $P(\{X_n > x\} \cap \{|X_n - X| < \varepsilon\})$ considera os casos em que X_n está próximo de X (a diferença é menor que ε) e, ao mesmo tempo, $X_n > x$.

Demonstração do Teorema 1

- A primeira parte $P(\{X_n > x\} \cap \{|X_n - X| < \varepsilon\})$ considera os casos em que X_n está próximo de X (a diferença é menor que ε) e, ao mesmo tempo, $X_n > x$.
- A segunda parte $P(\{X_n > x\} \cap \{|X_n - X| \geq \varepsilon\})$ considera os casos em que X_n e X estão distantes mais de ε.

Demonstração do Teorema 1

Como $P(\{X_n > x\} \cap \{|X_n - X| < \varepsilon\})$ é menor que $P(X > x - \varepsilon)$, podemos usar a seguinte desigualdade:

$$P(X_n > x) \leq P(X \geq x - \varepsilon) + P(|X_n - X| \geq \varepsilon)$$

- O primeiro termo, $P(X \geq x - \varepsilon)$, é a probabilidade de X ser maior ou igual a $x - \varepsilon$. Isso é uma aproximação para lidar com o fato de que X_n está próximo de X. - O segundo termo, $P(|X_n - X| \geq \varepsilon)$, representa a probabilidade de X_n estar distante de X (mais de ε).

Demonstração do Teorema 1

Assim, podemos expressar $P(X_n \leq x)$ como:

$$P(X_n \leq x) = 1 - P(X_n > x)$$

Substituímos o limite que encontramos para $P(X_n > x)$:

$$P(X_n \leq x) \geq 1 - P(X \geq x - \varepsilon) - P(|X_n - X| \geq \varepsilon)$$

Ou, de forma mais compacta:

$$F_{X_n}(x) \geq F_X(x - \varepsilon) - P(|X_n - X| \geq \varepsilon)$$

Demonstração do Teorema 1

Sabemos que, como $X_n \to X$ em probabilidade, temos $P(|X_n - X| \geq \varepsilon) \to 0$ conforme $n \to \infty$. Assim, no limite:

$$\lim_{n\to\infty} F_{X_n}(x) \geq F_X(x - \varepsilon)$$

Agora, combinamos as duas estimativas (superior e inferior) que obtivemos:

$$F_X(x - \varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

Finalmente, fazendo $\varepsilon \to 0$, chegamos à conclusão desejada:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Teorema 2

Se $X_n \xrightarrow{D} a,$ *então* $X_n \xrightarrow{P} a,$ *a constante.*

Teorema 3

Se $X_n \xrightarrow{D} X$ *e* $Y_n \xrightarrow{P} 0$ *então* $X_n + Y_n \xrightarrow{D} X$.

Teorema 4

Se $X_n \xrightarrow{D} X$ *e* g *é uma função contínua no suporte de* X*, então*

$$g(X_n) \xrightarrow{D} g(X).$$

Teorema de Slutsky

Teorema 5

Sejam $X_n, \; A_n$ *e* B_n, *variáveis aleatórias com* $X_n \xrightarrow{D} X, \; A_n \xrightarrow{P} a$ *e* $B_n \xrightarrow{P} b, \; a, b$ *constantes reais. Então,*

$$A_nX_n + B_n \xrightarrow{D} aX + b.$$

Para 🏠

Exercícios 5.2.2, 5.2.3, 5.2.6, 5.2.12, 5.2.15, 5.2.17, 5.2.19 e 5.2.20

Referências I

HOGG, RV; MCKEAN, J; CRAIG, AT. **Introduction to Mathematical Statistics**. Eighth Edition. [S.l.]: Pearson, 2019.