Tópico 4: Teoremas Limite

Ben Dêivide

6 de outubro de 2021

Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, em que Ω é o espaço amostral, $\mathcal{F}$ é a σ-álgebra que contém uma coleção de subconjuntos de Ω que pode ser atribuído uma probabilidade P. Trataremos *iid* uma amostra aleatória $X_1, X_2, \ldots, X_n$ independente e identicamente distribuída. Denote $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias $X_1, X_2, \ldots$.

Em se tratando de sequência de variáveis aleatórias, às vezes estamos interessados no seu limite e a sua convergência. Muitas das teorias de probabilidades clássicas se baseiam nos teoremas limites, tais como o estudo de estimadores pontuais, intervalos de confiança, testes de hipóteses. Assim, apresentaremos diversas formas de convergência de interesse na Teoria das Probabilidades.

1 Revisão sobre convergência de funções

No material escrito existe a distinção entre os dois tipos de convergência de funções: convergência pontual e convergência uniforme.

2 Tipos de convergência de variáveis aleatórias

Usaremos a notação $(X_n)_{n\in\mathbb{N}}$ para definir uma sequência de variáveis aleatórias.

Definição 1 (Convergência em probabilidade)**.** *Uma sequência de variáveis aleatórias X_n converge em probabilidade para X, denotado por $X_n \xrightarrow{p} X$, se para cada $\epsilon > 0$*

$$\lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1, \tag{1}$$

ou equivalentemente,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0. \tag{2}$$

Uma condição suficiente[1] para a convergência em probabilidade é a desigualdade de Chebychev.

[1] Raciocínio Lógico: Considere uma condicional p $->$ q, dizemos que p é condição SUFICIENTE para q, e também dizemos que q é condição NECESSÁRIA para p. Em uma bicondicional p $<->$ q, dizemos que p é condição NECESSÁRIA e SUFICIENTE para q, e vice-versa.

Teorema 1 (Desigualdade de Chebychev)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, com $f_X(x;.)$. Considere ainda $g(x)$ uma função não negativa de X. Então para qualquer $k > 0$,*

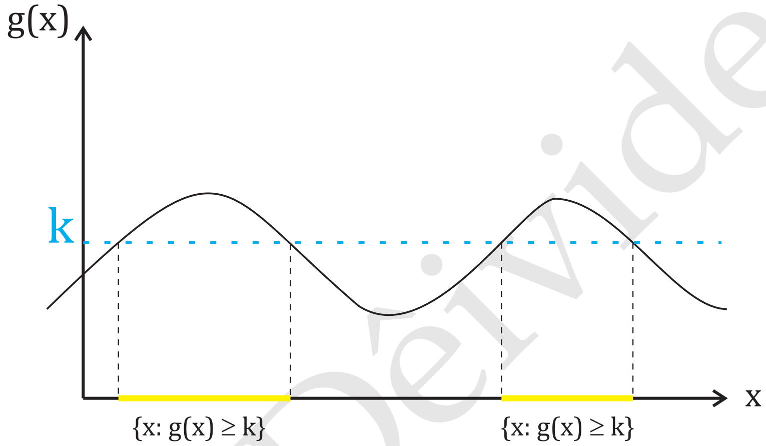
$$P(g(x) \geq k) \leq \frac{E[g(x)]}{k}. \tag{3}$$

□

Demonstração. A prova será realizada considerando X uma variável aleatória absolutamente contínua. Então poderemos expressar esperança de $g(x)$ da seguinte forma:

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x;.) dx. \tag{4}$$

Observe pelo gráfico:



que $E[g(x)]$ pode ser particionada em

$$\begin{aligned} E[g(x)] &= \int_{\{x:g(X) \leq k\}} g(x) f_X(x;.) dx + \int_{\{x:g(X) \geq k\}} g(x) f_X(x;.) dx \\ &\geq \int_{\{x:g(X) \geq k\}} g(x) f_X(x;.) dx \\ &\geq \int_{\{x:g(X) \geq k\}} k f_X(x;.) dx = k \int_{\{x:g(X) \geq k\}} f_X(x;.) dx = kP(g(x) \geq k). \end{aligned} \tag{5}$$

Logo,

$$P(g(x) \geq k) \leq \frac{E[g(x)]}{k}, \tag{6}$$

como queríamos provar. □

Se a distribuição de X for discreta, basta substituir as integrais por somatórios.

Teorema 2. *Uma condição suficiente para $X_n \rightarrow X$ é que*

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0, \tag{7}$$

isto é, X_n converge em média quadrática para X, em notação $X_n \xrightarrow{r} X$.

Demonstração. Se

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0, \tag{8}$$

então

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - X| > \epsilon) &= \lim_{n\to\infty} P[(X_n - X)^2 > \epsilon^2] \\
&\leq \lim_{n\to\infty} \frac{E[(X_n - X)^2]}{\epsilon^2} \qquad \text{(Desigualdade de Chebychev)} \\
&\leq \frac{1}{\epsilon^2} \lim_{n\to\infty} E[(X_n - X)^2] \\
&= 0.
\end{aligned}$$

□

Exemplo 1 (Binomial). *Seja uma amostra aleatória $Y_1, Y_2, \ldots, Y_n$ com distribuição de Bernoulli(p). Considere ainda $S_n = \sum_{i=1}^{n} Y_i$ o número de sucessos em n experimentos de sucesso p. Se denotarmos $X_n = S_n/n$ e $X = p$, então*

$$\lim_{n\to\infty} E[(X_n - X)^2] = \lim_{n\to\infty} Var[X_n] = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0. \tag{9}$$

De (19) *segue que* $\frac{S_n}{n} \xrightarrow{p} p$.

Uma importante aplicação da convergência em probabilidade é a consistência de estimadores.

Definição 2 (Consistência de estimadores). *Uma sequência de estimadores W_n de uma função paramétrica $g(\theta)$ é consistente se*

$$W_n \xrightarrow{p} g(\theta). \tag{10}$$

Uma convergência mais forte é a convergência quase certa ou convergência com probabilidade 1.

Definição 3 (Convergência quase certa). *Uma sequência de variáveis aleatórias X_n converge quase certamente para X, denotado $X_n \xrightarrow{\text{qc}} X$, se para $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |X_n - X| \leq \epsilon) = 1, \tag{11}$$

ou equivalentemente

$$P(\lim_{n\to\infty} |X_n - X| > \epsilon) = 0. \tag{12}$$

□

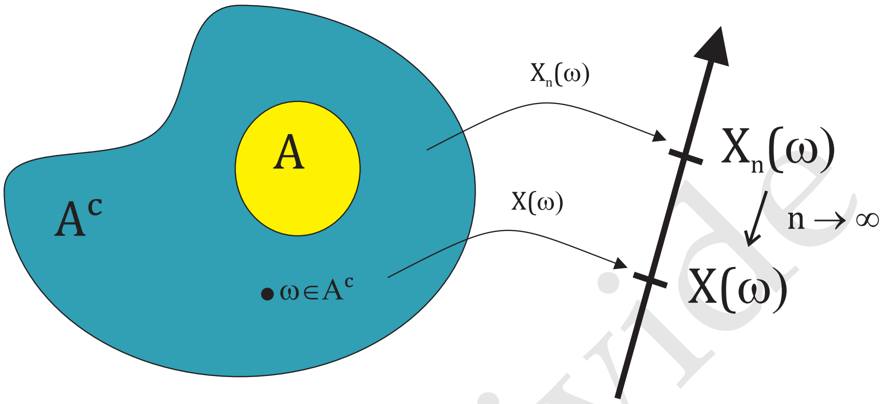
Para elucidar o próximo Teorema, Magalhães (2006, p. 301) define a convergência quase certa de uma outra forma.

Definição 4 (Convergência quase certa)**.** *Uma sequência de variáveis aleatórias* X_n*, definida no espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*, tem convergência quase certa, se existir um conjunto* $A \in \mathcal{F}$ *tal que* $P(A) = 0$ *e*

$$X_n \overset{qc}{\to} X \text{ em } A^c, \text{ quando } n \to \infty. \tag{13}$$

□

Isto implica que para $\omega \in A^c$, $X_n(\omega) \overset{qc}{\to} X(\omega)$ quando $n \to \infty$. Fora do conjunto A^c, nada se pode afirmar. Não quer dizer que para $\omega \in A$ $X_n \not\to X(\omega)$ quando $n \to \infty$. O que apenas se sabe é a convergência da sequência $X_n \overset{qc}{\to} X$ no conjunto A^c.



Teorema 3. *Seja uma sequência de variáveis aleatórias* X_n*, definida no espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*. Então* $X_n \overset{qc}{\to} X$ *se e somente se, para todo* $\epsilon > 0$*,*

$$P(|X_k - X| \leq \epsilon, \forall k \geq n) \underset{n\to\infty}{\to} 1. \tag{14}$$

□

Demonstração. Considere $A^c = \{w : X_n(w) \overset{qc}{\to} X(w)\}$ e $P(A^c) = 1$ e $B_n = \bigcap_{k\geq n}\{t : |X_n(t) - X(t)| \leq \epsilon\}$. A sequência de eventos $(B_n)_{n\in\mathbb{N}}$ é não-decrescente, uma vez que $X_n \overset{qc}{\to} X$. Se denotarmos $B = \bigcup_{n=1}^{\infty} B_n$ e que $B \subset A^c$, logo $P(B) = 1$. Como $B_n \to B$ quando $n \to \infty$, então $P(\lim_{n\to\infty} B_n) = P(B) = 1$, que é equivalente a expressão (14). □

Comparando (1) com (11) e (14) percebemos uma diferença crucial. Convergência em probabilidade é uma afirmação sobre um único X_n, isto é quando n é suficientemente grande, X_n é muito provável que se aproxime de X. Em contrapartida, a convergência quase certa é uma afirmação simultânea sobre uma sequência, quando n é suficientemente grande, é muito provável que todos os elementos da sequência $X_n, X_{n+1}, X_{n+2}, \ldots$, se aproximem de X. A convergência quase certa é uma convergência pontual. Já a convergência em probabilidade apenas afirma que para valores grandes de n, as variáveis X_n e X são aproximadamente iguais com alta probabilidade.

Evans e Rosenthal (2010) redefinem a convergência quase certa da seguinte forma: para algum ϵ, existirá um valor N_ϵ tal que $|X_n - X| < \epsilon$ para todo $n \geq N_\epsilon$. O valor de N_ϵ variará dependendo dos valores desencadeados da sequência $\{X_n - X\}$. Já a convergência em probabilidade apenas afirma que a distribuição de probabilidade $X_n - X$ se concentra sobre 0 quando n cresce, e não que os valores individuais de $X_n - X$ irão necessariamente todos estar próximo de zero.

A ilustração gráfica dessas duas convergências podem ser vistas em Micheaux e Liquet (2008).

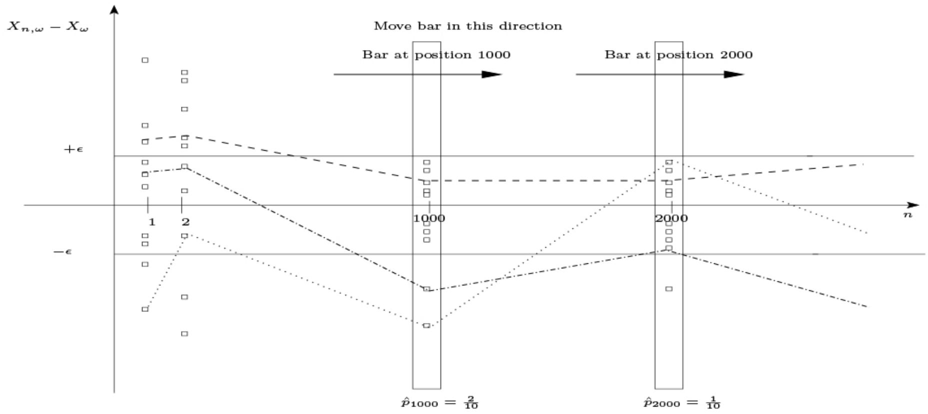


Figure 2: Seeing convergence in probability with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{p}_n = 2/10$ since we can see two sample paths lying outside the band $[-\epsilon, +\epsilon]$ in the bar at position 1000. For $n = 2000$, $\hat{p}_n = 1/10$ since we can see one sample path lying outside the band $[-\epsilon, +\epsilon]$ in the bar at position 2000.

Convergência em probabilidade

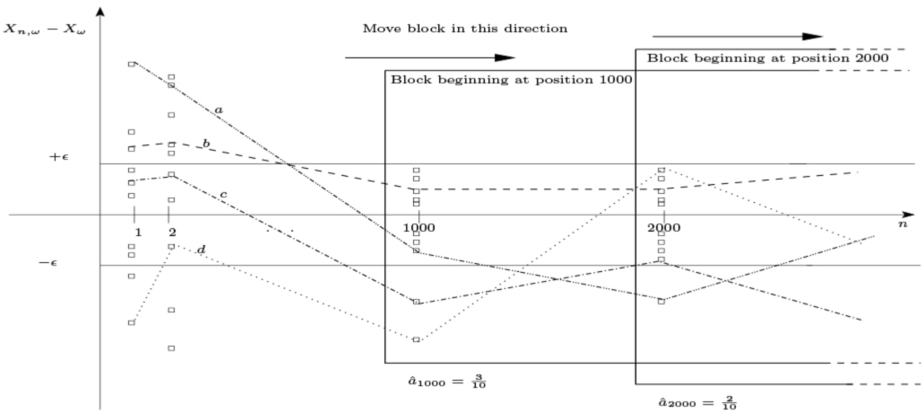


Figure 4: Seeing almost sure convergence with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{a}_n = 3/10$ since we can see 3 sample paths (a, c, d) lying outside the band $[-\epsilon, +\epsilon]$ in the block beginning at position 1000. For $n = 2000$, $\hat{a}_n = 2/10$ since we can see 2 sample paths (a, c) lying outside the band $[-\epsilon, +\epsilon]$ in the block beginning at position 2000.

Convergência quase certa

Definição 5 (Convergência em Distribuição)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias com função de distribuição $F_{X_n}(x)$. Considere ainda uma outra variável aleatória X com $F_X(x)$. Então X_n converge em distribuição para X, denotado por $X_n \xrightarrow{\text{d}} X$, se para todo ponto x em que F é contínua, tivermos*

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{15}$$

□

É importante notar que $X_n \xrightarrow{\text{d}} X$ implica em convergência em distribuição de função e não de variáveis aleatórias.

2.1 Relações entre os tipos de convergência

Teorema 4 ($X_n \xrightarrow{qc} X \Rightarrow X_n \xrightarrow{p} X$)**.** *Seja uma sequência de variáveis aleatórias X_n e X definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Se $X_n \xrightarrow{qc} X$ então $X_n \xrightarrow{p} X$.* □

Demonstração. □

Mostrar com um exemplo que o contrário não é verdade!

Teorema 5 ($X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$)**.** *Seja uma sequência de variáveis aleatórias X_n e X definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Se $X_n \xrightarrow{p} X$ então $X_n \xrightarrow{d} X$.* □

Demonstração. No material escrito! □

Teorema 6 ($X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{p} c$)**.** *Seja uma sequência de variáveis aleatórias X_n definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e c uma constante tal que $c \in \mathbb{R}$. Se $X_n \xrightarrow{d} c$ então $X_n \xrightarrow{p} c$.* □

Um importante Teorema que relaciona dois tipos de convergência é o Teorema de Slutsky.

Teorema 7 (Teorema de Slutsky)**.** *Se $X_n \xrightarrow{d} X$ em distribuição e $Y_n \xrightarrow{p} a$, uma constante, em probabilidade, então*

a) $Y_n X_n \xrightarrow{d} aX$ *em distribuição;*

b) $X_n + Y_n \xrightarrow{d} X + a$ *em distribuição.*

Demonstração. A prova está demonstrada na apostila [REGO,LC].Notas de aula do curso - probabilidade 4. 2010, pág. 51-53; MAGALHÃES (2006, p. 319-320) □

3 Função característica

Definição 6 (Função Característica)**.** *Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Então a função característica, denotada por φ_X, é definida por*

$$\varphi_X(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)], \tag{16}$$

em que $i = \sqrt{-1}$. □

A grande vantagem da função característica é que além de sempre existir, pois $\varphi_X(0) = 1$ e $|\varphi_X| \leq 1$, ela pode gerar os momentos de uma variável aleatória X, como também determina a função de distribuição da variável aleatória X. Por exemplo, se X tem esperança $E[X]$, os momentos de X podem ser expressos por $\varphi^d(0) = i^d E[X^d]$.

4 Lei dos Grandes Números

Teorema 8 (Lei Fraca dos Grandes Números (LFRGN)). *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias, independente e identicamente distribuídas (iid), definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, tal que $E[X] = \mu$ e $Var[X] = \sigma^2 < \infty$. Então, para cada $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \le \epsilon) = 1, \tag{17}$$

isto é, $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ converge em probabilidade para μ, denotada por $\bar{X}_n \xrightarrow{p} \mu$. □

Demonstração. Podemos afirmar que:

$$P(|\bar{X}_n - \mu| \le \epsilon) + P(|\bar{X}_n - \mu| > \epsilon) = 1. \tag{18}$$

Usando o Teorema 1, temos

$$\begin{aligned} P(|\bar{X}_n - \mu| \ge \epsilon) = P[(\bar{X} - \mu)^2 \ge \epsilon^2] &\le \frac{E[(\bar{X} - \mu)^2]}{\epsilon^2} \\ &\le \frac{\sigma^2}{n\epsilon^2}. \end{aligned} \tag{19}$$

Substituindo (18) em (19), temos

$$P(|\bar{X}_n - \mu| < \epsilon) \ge 1 - \frac{\sigma^2}{n\epsilon^2}. \tag{20}$$

Tomando $n \to \infty$ em (20), logo

$$\begin{aligned} \lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) &\ge \lim_{n\to\infty} \left(1 - \frac{\sigma^2}{n\epsilon^2}\right) \\ &= 1, \end{aligned}$$

provando o Teorema. □

Teorema 9 (Lei Forte dos Grandes Números (LFOGN)). *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias, independente e identicamente distribuídas (iid), definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, tal que $E[X] = \mu$ e $Var[X] = \sigma^2 < \infty$. Então, para cada $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| \le \epsilon) = 1, \tag{21}$$

isto é, $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ converge em quase certamente para μ, denotada por $\bar{X}_n \xrightarrow{\text{qc}} \mu$. □

5 Teorema Central do Limite

Teorema 10 (Teorema do Limite Central (TLC)). *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias, independente e identicamente distribuídas (iid),definidas no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, tal que $E[X] = \mu$ e $0 < Var[X] = \sigma^2 < \infty$. Então*

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\text{d}} Z \sim N(0, 1), \tag{22}$$

sendo $\bar{X}_n$ a média amostral. Assim, dizemos que Z_n converge em distribuição para Z. □

Demonstração. Seja,

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{n}{n}\mu)}{\sigma} = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i,$$

sendo $Y_i = \frac{\bar{X}-\mu}{\sigma}$. Como a função característica de uma soma de variáveis aleatórias independentes é igual ao produto das funções características das variáveis aleatórias, tem-se

$$\varphi_{Z_n}(t) = E[e^{itZ_n}] = E\left[e^{\frac{it\sum_{i=1}^{n} Y_i}{\sqrt{n}}}\right] = E\left[\prod_{i=1}^{n} e^{\frac{itY_i}{\sqrt{n}}}\right] = \prod_{i=1}^{n} E\left[e^{\frac{itY_i}{\sqrt{n}}}\right]$$

Usando ainda o fato que as variáveis aleatórias são identicamente distribuídas, temos

$$\varphi_{Z_n}(t) = \prod_{i=1}^{n} E\left[e^{\frac{itY_i}{\sqrt{n}}}\right] = \varphi_Y(t/\sqrt{n})^n.$$

Fazendo uma aproximação usando a expansão em série de Taylor de segunda ordem em torno de 0 e usando o fato que $\varphi^d(0) = i^d E[X^d]$, então,

$$\begin{aligned}
\varphi_Y(t) &\approx \varphi_Y(0) + \frac{t}{\sqrt{n}}\varphi_Y'(0) + \frac{t^2}{2n} + \varphi_Y''(0), \qquad \text{Casella, port. 215} \\
&\approx 1 + \frac{it\mu_Y}{\sqrt{n}} - \frac{\sigma_Y^2 t^2}{2n},
\end{aligned}$$

sendo $\mu_Y = 1$ e $\sigma_Y^2 = 1$. Assim,

$$\varphi_Y(t) \approx 1 - \frac{t^2}{2n}. \tag{23}$$

Tomando o limite $n \to \infty$ em (23), temos

$$\lim_{n\to\infty} \varphi_Y(t) \approx \lim_{n\to\infty}\left(1 - \frac{t^2}{2n}\right).$$

Fazendo $-\frac{t^2}{2n} = \frac{1}{k} \Rightarrow n = -\frac{kt^2}{2}$, logo

$$\lim_{k\to\infty}\left(1 + \frac{1}{k}\right)^{-k\frac{t^2}{2}} = \left[\lim_{k\to\infty}\left(1 + \frac{1}{k}\right)^k\right]^{-\frac{t^2}{2}} = e^{-\frac{t^2}{2}},$$

que é a função característica da distribuição normal padrão, o que prova o Teorema. □

6 Aplicações

6.1 Aplicações de LFRGN

Consistência...

6.2 Aplicações da LFOGN

Simulação Monte Carlo

6.3 Aplicações do TLC

Aproximação de distribuição, Intervalos de confiança e testes de hipóteses