

PROBABILIDADE E ESTATÍSTICA

AULA 3 – CONVERGÊNCIA ESTOCÁSTICA

Luis A. F. Alvarez

5 de fevereiro de 2025

VETORES ALEATÓRIOS

- Seja $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade.
- Um vetor aleatório $\mathbf{Y} : \Omega \mapsto \mathbb{R}^k$ é uma função tal que cada coordenada $\mathbf{Y}_l : \Omega \mapsto \mathbb{R}$, $l = 1, \dots, k$, é uma variável aleatória real.
- Um vetor aleatório induz uma distribuição de probabilidade sobre $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, dada por $\mathbb{P}_{\mathbf{Y}}[A] = \mathbb{P}[\mathbf{Y}^{-1}(A)]$, $A \in \mathcal{B}(\mathbb{R}^k)$.
- Pelo lema do π -sistema, essa distribuição de probabilidade é caracterizada pela função de distribuição $F_{\mathbf{Y}} : \mathbb{R}^k \mapsto [0, 1]$, dada por:

$$F_{\mathbf{Y}}(\mathbf{c}) := \mathbb{P}_{\mathbf{Y}} \left[\prod_{l=1}^k (-\infty, \mathbf{c}_l] \right], \quad \mathbf{c} \in \mathbb{R}^k.$$

CONVERGÊNCIA QUASE-CERTA

- Seja $(\Omega, \Sigma, \mathbb{P})$ um espaço de probabilidade, e $\mathbf{Y}_1, \mathbf{Y}_2, \dots$ uma sequência de vetores aleatórios.
- Dizemos que \mathbf{Y}_n converge quase-certamente para um vetor aleatório \mathbf{Y} , denotado por $\mathbf{Y}_n \xrightarrow{\text{q.c.}} \mathbf{Y}$, se:

$$\mathbb{P}[\{\omega : \mathbf{Y}_n(\omega) \not\rightarrow \mathbf{Y}(\omega)\}] = 0.$$

- Sequência de funções \mathbf{Y}_n convergem (ponto a ponto), a não ser num conjunto de pontos de probabilidade zero.

LEMA

$\mathbf{Y}_n \xrightarrow{\text{q.c.}} \mathbf{Y}$ se, e somente se, para todo $\epsilon > 0$:

$$\mathbb{P} \left[\limsup_n \{\omega : \|\mathbf{Y}_n(\omega) - \mathbf{Y}(\omega)\| > \epsilon\} \right] = 0.$$

CONVERGÊNCIA EM PROBABILIDADE

- Dizemos que \mathbf{Y}_n converge em probabilidade para um vetor aleatório \mathbf{Y} , denotado por $\mathbf{Y}_n \xrightarrow{P} \mathbf{Y}$, se, para todo $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} [\{\omega : \|\mathbf{Y}_n(\omega) - \mathbf{Y}(\omega)\| > \epsilon\}] = 0.$$

LEMA

Se $\mathbf{Y}_n \xrightarrow{q.c.} \mathbf{Y}$, então $\mathbf{Y}_n \xrightarrow{P} \mathbf{Y}$.

CONVERGÊNCIA EM DISTRIBUIÇÃO

- Dizemos que \mathbf{Y}_n converge em distribuição para um vetor aleatório \mathbf{Y} , denotado por $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$, se as funções de distribuição dos \mathbf{Y}_n , $\{F_{\mathbf{Y}_n}\}_{n \in \mathbb{N}}$, convergem para a função de distribuição $F_{\mathbf{Y}}$ nos pontos em que $F_{\mathbf{Y}}$ é contínua.
 - Isto é, $\lim_n F_{\mathbf{Y}_n}(\mathbf{c}) = F_{\mathbf{Y}}(\mathbf{c})$ para todo \mathbf{c} em que $F_{\mathbf{Y}}$ é contínua.
 - Conjunto de pontos em que uma função de distribuição é descontínua é enumerável \implies se há convergência em distribuição, então, para qualquer $\mathbf{c} \in \mathbb{R}^k$, é sempre possível encontrar um $\mathbf{c}' \geq \mathbf{c}$ em que a função de distribuição converge.

LEMA

Se $\mathbf{Y}_n \xrightarrow{p} \mathbf{Y}$, então $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$.

LEMA (TRECHO DO LEMA PORTMANTEAU)

$\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$ se, e somente se, para qualquer $f : \mathbb{R}^k \mapsto \mathbb{R}$ contínua e limitada.

$$\mathbb{E}[f(\mathbf{Y}_n)] \rightarrow \mathbb{E}[f(\mathbf{Y})].$$

CONVERGÊNCIA EM L^p

- Dizemos que \mathbf{Y}_n converge para um vetor aleatório \mathbf{Y} na norma L^p , denotado por $\mathbf{Y}_n \xrightarrow{L^p} \mathbf{Y}$, se $\|\mathbf{Y}_n - \mathbf{Y}\|_p \rightarrow 0$.
 - Para variáveis aleatórias, definimos a norma L_p na última aula como $\|X\|_p = (\mathbb{E}[|X|^p])^{\frac{1}{p}}$.
 - Para vetores aleatórios, estendemos a definição fazendo $\|\mathbf{X}\|_p = \max_{j=1,\dots,k} \|X_j\|_p$.
- Pela desigualdade de Markov, convergência em L_p implica convergência em probabilidade.
 - Recíproca não é, no geral, verdadeira.

TEOREMA DO MAPA CONTÍNUO

TEOREMA

Seja $(\mathbf{X}_n)_n$ uma sequência de vetores aleatórios, \mathbf{X} um vetor aleatório, e $f : \mathbb{R}^k \mapsto \mathbb{R}^l$ uma função contínua num conjunto C do domínio tal que: $\mathbb{P}[\{\omega : \mathbf{X}(\omega) \in C\}] = 1$. Então:

1. $\mathbf{X}_n \xrightarrow{q.c.} \mathbf{X} \implies f(\mathbf{X}_n) \xrightarrow{q.c.} f(\mathbf{X})$.
2. $\mathbf{X}_n \xrightarrow{p} \mathbf{X} \implies f(\mathbf{X}_n) \xrightarrow{p} f(\mathbf{X})$.
3. $\mathbf{X}_n \xrightarrow{d} \mathbf{X} \implies f(\mathbf{X}_n) \xrightarrow{d} f(\mathbf{X})$.

- Modos de convergência quase certa, em probabilidade e distribuição são preservados por transformações contínuas.

RESULTADOS ADICIONAIS E LEMA DE SLUTSKY

LEMA

1. Se $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ e $\|\mathbf{X}_n - \mathbf{Y}_n\| \xrightarrow{p} 0$, então $\mathbf{Y}_n \xrightarrow{d} \mathbf{X}$.
2. Se $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ e $\mathbf{Y}_n \xrightarrow{p} c$, onde $c \in \mathbb{R}^k$ é constante, então $(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{d} (\mathbf{X}, c)$.
3. Se $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ e $\mathbf{Y}_n \xrightarrow{p} \mathbf{Y}$, então $(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{p} (\mathbf{X}, \mathbf{Y})$.

COROLÁRIO (LEMA DE SLUTSKY)

Sejam $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{p} c$, duas sequências de variáveis aleatórias ou vetores aleatórios, onde c é constante, então:

- $X_n + Y_n \xrightarrow{d} X + c$.
- $X_n \cdot Y_n \xrightarrow{d} X \cdot c$.
- Se $c \neq 0$, $X_n/Y_n \xrightarrow{d} X/c$.

NOTAÇÃO O E o PARA SEQUÊNCIAS NÃO ESTOCÁSTICAS

- Sejam $(x_n)_n$ e $(y_n)_n$ duas sequências de **números reais**.
- Dizemos que $x_n = o(y_n)$ se $\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = 0$.
- Dizemos que $x_n = O(y_n)$ se existem $C > 0$ e $K \in \mathbb{N}$ tais que:
 $|x_n| \leq C|y_n|$, para todo $n \geq K$.
 - Nesse caso, a razão $\frac{x_n}{y_n}$ é limitada.

NOTAÇÃO O_p E o_p PARA SEQUÊNCIAS ESTOCÁSTICAS

- Sejam X_n e Y_n sequências de **variáveis aleatórias**.
- Dizemos que $X_n = o_p(Y_n)$ se $\frac{X_n}{Y_n} \xrightarrow{P} 0$.
- Dizemos que $X_n = O_p(Y_n)$ se, para todo $\epsilon > 0$, existem $M > 0$ e $K \in \mathbb{N}$ tais que:

$$\mathbb{P}[|X_n| > M|Y_n|] \leq \epsilon, \forall n \geq K.$$

- Nesse caso, dizemos que $\left| \frac{X_n}{Y_n} \right|$ é **limitada em probabilidade**.

NOTAÇÃO O_p E o_p : PROPRIEDADES

LEMA

- $o_p(1) + o_p(1) = o_p(1)$.
- $o_p(1) + O_p(1) = O_p(1)$.
- $O_p(1)o_p(1) = o_p(1)$.
- $\frac{1}{(1+o_p(1))} = O_p(1)$
- $o_p(Y_n) = Y_n o_p(1)$.
- $O_p(Y_n) = Y_n O_p(1)$.
- $o_p(O_p(1)) = o_p(1)$.
- Se X_n converge em distribuição, $X_n = O_p(1)$.
- Seja $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ uma função tal que $\phi(\mathbf{0}) = 0$. Se $X_n \xrightarrow{p} \mathbf{0}$, então, para qualquer $p > 0$:
 - A Se $R(h) = o(\|h\|^p)$ quando $h \rightarrow 0$, então $R(\mathbf{X}_n) = o_p(\|\mathbf{X}_n\|^p)$.
 - B Se $R(h) = O(\|h\|^p)$ quando $h \rightarrow 0$, então $R(\mathbf{X}_n) = O_p(\|\mathbf{X}_n\|^p)$.

MÉTODO DELTA

- Seja \mathbf{X}_n uma sequência de vetores aleatórios.
- Suponha que existe uma sequência de números reais $r_n \rightarrow \infty$, uma constante $\theta \in \mathbb{R}^k$ e um vetor aleatório \mathbf{S} tais que: $r_n(\mathbf{X}_n - \theta) \xrightarrow{d} \mathbf{S}$.
 - Nesse caso, $\mathbf{X}_n \xrightarrow{P} \theta$ (por quê?).
- Será que podemos calcular $r_n(\psi(\mathbf{X}_n) - \psi(\theta))$ para uma transformação $\psi : \mathbb{R}^k \mapsto \mathbb{R}^l$?
 - Interpretação estatística: derivar a distribuição em amostras grandes de uma estatística a partir de outra estatística.

LEMA (MÉTODO DELTA)

Se ψ é continuamente diferenciável numa vizinhança de θ , então:

$$r_n(\psi(\mathbf{X}_n) - \psi(\theta)) \xrightarrow{d} \nabla_x \psi(\theta) \mathbf{S},$$

onde $\nabla_x \psi(\theta)$ é o Jacobiano de ψ avaliado em θ

FUNÇÃO GERADORA DE MOMENTOS

- Seja X uma variável aleatória real. Definimos a função geradora de momentos, $M : \mathcal{S} \mapsto \mathbb{R}_+$, $\mathcal{S} \subseteq \mathbb{R}$, como a função dada por.

$$M(s) = \mathbb{E}[\exp(sX)], \quad s \in \mathcal{S},$$

onde \mathcal{S} é o conjunto de pontos em que a esperança $\mathbb{E}[\exp(sX)]$ é finita.

PROPOSIÇÃO

Se existe $\epsilon > 0$ tal que $(-\epsilon, \epsilon) \subseteq \mathcal{S}$, então X possui todos os momentos (i.e. $\mathbb{E}[|X|^j] < \infty$ para todo $j \in \mathbb{N}$), a expansão de Taylor infinita de $M(s)$ em torno de 0 é exata para $|s| < \epsilon$, i.e.

$$M(s) = \sum_{j=0}^{\infty} M^{(j)}(0) \frac{s^j}{j!},$$

e as derivadas em zero podem ser calculadas por diferenciação sob a integral, resultando em $M^{(j)}(0) = \mathbb{E}[X^j]$ para todo $j \in \mathbb{N}$.

FUNÇÃO CARACTERÍSTICA

- Dificuldade de utilizar função característica é que nem sempre ela está disponível numa vizinhança do zero.
- Por esse motivo, definimos uma função alternativa, conhecida como **função característica**.
- Seja \mathbf{X} um vetor aleatório de dimensão k . Sua função característica é dada por $\phi : \mathbb{R}^k \mapsto \mathbb{C}$, definida por:

$$\phi(s) = \mathbb{E}[e^{is'\mathbf{X}}] = \mathbb{E}[\cos(s'\mathbf{X})] + i\mathbb{E}[\sin(s'\mathbf{X})]$$

- Função está sempre bem-definida, para todo $s \in \mathbb{R}^k$ (por quê?).
- **Propriedade útil 1:** se \mathbf{X} é escalar ($k = 1$), e $\mathbb{E}[|\mathbf{X}|^j] < \infty$ para $j \in \mathbb{N}$ então ϕ é j -vezes diferenciável no zero, com j -ésima derivada dada por diferenciação sob a integral, resultando em $\phi^{(j)}(0) = i^j \mathbb{E}[\mathbf{X}^j]$.
- **Propriedade útil 2:** se \mathbf{X} e \mathbf{Y} são vetores aleatórios independentes, $\mathbb{E}[\exp(is'(\mathbf{X} + \mathbf{Y}))] = \mathbb{E}[\exp(is'\mathbf{X})]\mathbb{E}[\exp(is'\mathbf{Y})]$.

CARACTERIZAÇÕES COM BASE EM FUNÇÕES CARACTERÍSTICA

- Funções características apresentam propriedades poderosas de caracterização de distribuições, que coletamos (sem demonstrar) abaixo.

PROPOSIÇÃO

1. *Dois vetores aleatórios, \mathbf{X} e \mathbf{Y} , possuem funções distribuição iguais se, e somente se, suas funções características são idênticas.*
2. *Uma sequência de vetores aleatórios \mathbf{X}_n convergem em distribuição para um vetor aleatório \mathbf{Y} se, e somente se, as funções características de \mathbf{X}_n convergem para a função característica de \mathbf{Y} pontualmente em todo $s \in \mathbb{R}^k$, i.e. $\phi_{\mathbf{X}_n}(s) \rightarrow \phi_{\mathbf{Y}}(s)$ para todo $s \in \mathbb{R}^k$.*

COROLÁRIO (DISPOSITIVO DE CRÁMER-WOLD)

$\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ se, e somente se, $t' \mathbf{X}_n \xrightarrow{d} t' \mathbf{X}$ para todo $t \in \mathbb{R}^k$.

LEI FORTE DOS GRANDES NÚMEROS

- Dizemos que uma sequência de vetores aleatórios $\mathbf{X}_1, \mathbf{X}_2, \dots$ é independente e identicamente distribuída (iid) se os vetores aleatórios são independentes (i.e. suas σ -álgebras geradas são independentes) e se suas funções distribuição $F_{\mathbf{X}_j}$ coincidem.
 - Pelo lema do π -sistema, a lei induzida por cada \mathbf{X}_j sobre $\mathcal{B}(\mathbb{R}^k)$ é igual.

PROPOSIÇÃO (LEI FORTE DE KOLMOGOROV)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias iid em $L^1(\Omega, \Sigma, \mathbb{P})$. Então existe $\mu \in \mathbb{R}$ tal que $\mathbb{E}[X_j] = \mu$ para todo $j \in \mathbb{N}$ e, quando $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{q.c.} \mu.$$

LEI FRACA DOS GRANDES NÚMEROS

COROLÁRIO (LEI FRACA DE KHINCHINE)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias iid em $L^1(\Omega, \Sigma, \mathbb{P})$. Então temos quando $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} \mathbb{E}[X_1].$$

PROPOSIÇÃO (LEI FRACA DE MARKOV)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias em $L^2(\Omega, \Sigma, \mathbb{P})$, não correlacionadas e tais que $\sup_{j \in \mathbb{N}} \mathbb{V}[X_j] < \infty$. Se $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}[X_j]$ existe em \mathbb{R} , então, denotando esse limite por μ , temos, quando $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} \mu.$$

TEOREMA CENTRAL DO LIMITE

PROPOSIÇÃO

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias iid em $L^2(\Omega, \Sigma, \mathbb{P})$. Então, denotando por $\mu = \mathbb{E}[X_1]$ e $\sigma^2 = \mathbb{V}[X_1]$, temos, quando $n \rightarrow \infty$:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

onde $\mathcal{N}(0, \sigma^2)$ denota uma variável aleatória com distribuição normal com média zero e variância σ^2 .