Probabilidade e Estatística

Aula 3 – Convergência estocástica

Luis A. F. Alvarez

1 de fevereiro de 2025

Vetores aleatórios

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

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

Convergência quase-certa

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

$$\mathbb{P}[\{\omega : \boldsymbol{Y}_n(\omega) \nrightarrow \boldsymbol{Y}(\omega)\}] = 0\,.$$

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

Lema

$\boldsymbol{Y}_n \stackrel{q.c.}{\to} \boldsymbol{Y}$ se, e somente se, para todo $\epsilon > 0$:

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

Convergência em probabilidade

- Dizemos que $\boldsymbol{Y}_n$ converge em probabilidade para um vetor aleatório $\boldsymbol{Y}$, denotado por $\boldsymbol{Y}_n \stackrel{p}{\to} \boldsymbol{Y}$, se, para todo $\epsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left[\{\omega : \|\boldsymbol{Y}_n(\omega) - \boldsymbol{Y}(\omega)\| > \epsilon\}\right] = 0\,.$$

Lema

Se $\boldsymbol{Y}_n \stackrel{q.c.}{\to} \boldsymbol{Y}$, então $\boldsymbol{Y}_n \stackrel{p}{\to} \boldsymbol{Y}$.

Convergência em distribuição

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

Lema

Se $\boldsymbol{Y}_n \stackrel{p}{\rightarrow} \boldsymbol{Y}$, então $\boldsymbol{Y}_n \stackrel{d}{\rightarrow} \boldsymbol{Y}$.

Lema (Trecho do lema Portmanteau)

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

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

Convergência em L^p

- Dizemos que $\boldsymbol{Y}_n$ converge para um vetor aleatório $\boldsymbol{Y}$ na norma L^p, denotado por $\boldsymbol{Y}_n \stackrel{L_p}{\to} \boldsymbol{Y}$, se $\|\boldsymbol{Y}_n - \boldsymbol{Y}\|_p \to 0$.
- Pela desigualdade de Markov, convergência em L_p implica convergência em probabilidade.
    - Recíproca não é, no geral, verdadeira.

Teorema do mapa contínuo

Teorema

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

1. $\boldsymbol{X}_n \overset{q.c.}{\to} \boldsymbol{X} \implies f(\boldsymbol{X}_n) \overset{q.c.}{\to} f(\boldsymbol{X})$.
2. $\boldsymbol{X}_n \overset{p}{\to} \boldsymbol{X} \implies f(\boldsymbol{X}_n) \overset{p}{\to} f(\boldsymbol{X})$.
3. $\boldsymbol{X}_n \overset{d}{\to} \boldsymbol{X} \implies f(\boldsymbol{X}_n) \overset{d}{\to} f(\boldsymbol{X})$.

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

Resultados adicionais e lema de Slutsky

Lema

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

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 $\lim_{n\to\infty} \frac{x_n}{y_n} = 0$.
- 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* $h \to 0$, *então* $R(\boldsymbol{X_n}) = o_p(\|\boldsymbol{X_n}\|^p)$.
  - B *Se* $R(h) = O(\|h\|^p)$ *quando* $h \to 0$, *então* $R(\boldsymbol{X_n}) = O_p(\|\boldsymbol{X_n}\|^p)$.

Método Delta

- Seja $\boldsymbol{X}_n$ uma sequência de vetores aleatórios.
- Suponha que existe uma sequência de números reais $r_n \to \infty$, uma constante $\theta \in \mathbb{R}^k$ e um vetor aleatório $\boldsymbol{S}$ tais que: $r_n(\boldsymbol{X}_n - \theta) \xrightarrow{d} \boldsymbol{S}$.
  - Nesse caso, $\boldsymbol{X}_n \xrightarrow{p} \theta$ (por quê?).
- Será que podemos calcular $r_n(\psi(\boldsymbol{X}_n) - \psi(\theta))$ 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:

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

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 $\boldsymbol{X}$ um vetor aleatório de dimensão k. Sua função característica é dada por $\phi : \mathbb{R}^k \mapsto \mathbb{C}$, definida por:

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

- Função está sempre bem-definida, para todo $s \in \mathbb{R}^k$ (por quê?).
- Propriedade útil 1: se $\boldsymbol{X}$ é escalar ($k = 1$), e $\mathbb{E}[|\boldsymbol{X}|^j] < \infty$ 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 $\phi^{(j)}(0) = i^j\mathbb{E}[\boldsymbol{X}^j]$.
- Propriedade útil 2: se $\boldsymbol{X}$ e $\boldsymbol{Y}$ são vetores aleatórios independentes, $\mathbb{E}[\exp(is'(\boldsymbol{X} + \boldsymbol{Y}))] = \mathbb{E}[\exp(is'\boldsymbol{X})]\mathbb{E}[\exp(is'\boldsymbol{Y})]$.

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

Corolário (Dispositivo de Crámer-Wold)

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

Lei forte dos grandes números

- Dizemos que uma sequência de vetores aleatórios $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ 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 $F_{\boldsymbol{X}_j}$ coincidem.
  - Pelo lema do π-sistema, a lei induzida por cada $\boldsymbol{X}_j$ sobre $\mathcal{B}(\mathbb{R}^k)$ é igual.

Proposição (Lei forte de Kolmogorov)

Seja $X_1, X_2, \ldots$ 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 $n \to \infty$:

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

Lei fraca dos grandes números

Corolário (Lei fraca de Khinchine)

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

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

Proposição (Lei fraca de Markov)

Seja $X_1, X_2, \ldots$ 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 $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[X_j]$ existe em $\mathbb{R}$, então, denotando esse limite por μ, temos, quando $n \to \infty$:

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

Teorema Central do Limite

Proposição

Seja $X_1, X_2, \ldots$ 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 $n \to \infty$:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_1 - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)\,,$$

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