Aula 26: Teorema Central do Limite

Disciplina: PGE950 - Probabilidade

Programa de Pós-graduação em Estatística da UFPE

Prof. Pablo M. Rodriguez

http://www.pablo-rodriguez.org

Universidade Federal de Pernambuco

CCEN Centro de Ciências Exatas e da Natureza

Conteúdo desta aula

- Propriedades de tipos de convergência.
- Teorema Central do Limite

Funções continuas e convergência

Proposição 26.1

Sejam $X, X_1, X_2, \ldots$ *variáveis aleatórias e* $g : \mathbb{R} \longrightarrow \mathbb{R}$ *contínua.*

1. *Se* $X_n \stackrel{q.c.}{\to} X$, *então* $g(X_n) \stackrel{q.c.}{\to} g(X)$.
2. *Se* $X_n \stackrel{P}{\to} X$, *então* $g(X_n) \stackrel{P}{\to} g(X)$.
3. *Se* $X_n \stackrel{D}{\to} X$, *então* $g(X_n) \stackrel{D}{\to} g(X)$.

Observação!

Para (3) usamos convergência das funções características. No que segue provamos (1), (2).

Prova da Proposição 26.1 (1). Se $X_n \overset{q.c.}{\to} X$ e

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}$$

então $P(A) = 1$. Seja $\omega \in A$ e note que, como

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

e g é uma função contínua então

$$\lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega)).$$

Isto é,

$$A \subset \left\{\omega \in \Omega : \lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega))\right\}.$$

Portanto,

$$1 = P(A) \leq P\left(\lim_{n\to\infty} g(X_n) = g(X)\right) \leq 1,$$

e assim $g(X_n) \overset{q.c.}{\to} g(X)$.

Prova da Proposição 26.1 (2). Se $X_n \xrightarrow{P} X$ então, para todo $\delta > 0$:

$$\lim_{n\to\infty} P(|X_n - X| < \delta) = 1. \tag{1}$$

Considere $\epsilon > 0$ tal que, para m suficientemente grande

$$P\left(-\frac{m}{2} \leq X \leq \frac{m}{2}\right) > 1 - \epsilon \tag{2}$$

o que é possível pois $\left\{-\frac{m}{2} \leq X \leq \frac{m}{2}\right\} \nearrow \Omega$. De (1) e (2) temos:

$$\lim_{n\to\infty} P\left(\{|X| \leq m/2\} \cap \{|X_n - X| < \delta\}\right) = P\left(|X| \leq m/2\right) > 1 - \epsilon \tag{3}$$

... continuação da prova da Proposição 26.1 (2). Como g é uniformemente continua em $[-m, m]$, existe $0 < \delta < m/2$ tal que se $\max\{|x|, |y|\} \leq m$ e $|x - y| < \delta$ então

$$|g(x) - g(y)| < \epsilon. \tag{4}$$

Note que, para $\delta < m/2$ temos que

$$\begin{aligned} \{|X| \leq m/2, |X_n - X| < \delta\} &\subset \{|X| \leq m, |X_n| \leq m, |X_n - X| < \delta\} \\ &\subset \{|g(X_n) - g(X)| \leq \epsilon\} \end{aligned}$$

De (3) concluímos que $P(|g(X_n) - g(X)| \leq \epsilon) > 1 - \epsilon$ para n grande. Agora, se $\xi \in (0, \epsilon)$ temos que, para n suficientemente grande:

$$P(|g(X_n) - g(X)| \leq \epsilon) > P(|g(X_n) - g(X)| \leq \xi) > 1 - \xi$$

e tomando $\xi \to 0$ concluímos que

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

Teorema Central do Limite

Teorema 26.1

Sejam $X_1, X_2, \ldots$ *i.i.d. com* $\mu = E(X_1)$ *e* $\sigma^2 = Var(X_1) < \infty$. *Então:*

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{D} Z \sim \mathcal{N}(0,1).$$

Em outras palavras:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \sim \mathcal{N}(0,1),$$

ou

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a),$$

ou

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \Phi(b) - \Phi(a), \quad \textit{para } n \textit{ suficientemente grande.}$$

Teorema Central do Limite de Lindeberg

Teorema 26.2

Sejam $X_1, X_2, \ldots$ *variáveis aleatórias independentes tais que* $E(X_n) = \mu_n$ *e* $Var(X_n) = \sigma)n^2$, *onde* $\sigma_n^2 < \infty$ *e pelo menos um* $\sigma_n^2 > 0$. *Sejam*

$$F_n := F_{X_n}, \quad S_n := \sum_{i=1}^{n} X_i, \quad s_n := \sqrt{Var(S_n)} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}.$$

Se para todo $\epsilon > 0$ *vale que*

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \epsilon s_n} (x - \mu_k)^2 \, dF_k(x) = 0 \tag{5}$$

então

$$\frac{S_n - E(S_n)}{s_n} \xrightarrow{D} Z \sim \mathcal{N}(0,1).$$

Condição de Lindeberg

A condição (1), chamada condição de Lindeberg, pode ser lida da seguinte forma para os casos discreto e continuo:

- Se X_k são discretas com $P(X_k = x_i) = p_k(x_i)$, então:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \left\{ \sum_{i:|x_i-\mu_k|>\epsilon s_n} (x_i - \mu_k)^2 p_k(x_i) \right\} = 0.$$

- Se X_k são continuas com densidade $f_k(x)$, então:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \left\{ \int_{-\infty}^{\mu_k-\epsilon s_n} (x-\mu_k)^2 f_k(x)\,dx + \int_{\mu_k+\epsilon s_n}^{+\infty} (x-\mu_k)^2 f_k(x)\,dx \right\} = 0.$$

Condição de Lindeberg: outra forma

A condição de Lindeberg pode ser escrita como: para todo $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|\leq \epsilon s_n} (x-\mu_k)^2 \, dF_k(x) = 1.$$

Prova. Resulta de observar que

$$\begin{aligned} \sigma_k^2 &= \int (x-\mu_k)^2 \, dF_k(x) \\ &= \int_{|x-\mu_k|\leq \epsilon s_n} (x-\mu_k)^2 \, dF_k(x) + \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 \, dF_k(x), \end{aligned}$$

isto é

$$\frac{\sigma_k^2}{s_n^2} = \frac{1}{s_n^2} \int_{|x-\mu_k|\leq \epsilon s_n} (x-\mu_k)^2 \, dF_k(x) + \frac{1}{s_n^2} \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 \, dF_k(x),$$

Condição de Lindeberg: outra forma

Como

$$\sum_{k=1}^{n} \frac{\sigma_k^2}{s_n^2} = 1,$$

temos que

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|\leq \epsilon s_n} (x-\mu_k)^2 \, dF_k(x) = 1$$

se, e somente se,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|> \epsilon s_n} (x-\mu_k)^2 \, dF_k(x) = 0.$$

Proposição 26.2

Sejam $X_1, X_2, \ldots$ variáveis aleatórias tais que $E(X_n) = \mu_n$ e $Var(X_n) = \sigma_n^2 < \infty$ para todo n. Se a sequência satisfaz a condição de Lindeberg, então:

$$\lim_{n\to\infty} \left\{ \max_{1\le k\le n} \frac{\sigma_k^2}{s_n^2} \right\} = 0.$$

Prova. Note que, para $k \in \{1, \ldots, n\}$

$$\begin{aligned}
\frac{\sigma_k^2}{s_n^2} &= \frac{1}{s_n^2}\int_{|x-\mu_k|\le\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \frac{1}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) \\
&\le \frac{1}{s_n^2}\int_{|x-\mu_k|\le\epsilon s_n} \epsilon^2 s_n^2 dF_k(x) + \frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\epsilon s_n} (x-\mu_i)^2 dF_i(x) \\
&\le \frac{1}{s_n^2}\int_{-\infty}^{\infty} \epsilon^2 s_n^2 dF_k(x) + \frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\epsilon s_n} (x-\mu_i)^2 dF_i(x)
\end{aligned}$$

... continuação da prova da Proposição 26.2. Logo,

$$\max_{1\le k\le n}\frac{\sigma_k^2}{s_n^2}\le \epsilon^2+\frac{1}{s_n^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\epsilon s_n}(x-\mu_i)^2\,dF_i(x).$$

Fazendo $n\to\infty$ temos pela condição de Lindeberg que

$$\lim_{n\to\infty}\max_{1\le k\le n}\frac{\sigma_k^2}{s_n^2}\le \epsilon^2$$

e fazendo $\epsilon\to 0$ concluímos o resultado.

Recíproca do TCL de Lindeberg

Teorema 26.3

Sejam $X_1, X_2, \ldots$ *variáveis aleatórias independentes tais que* $E(X_n) = \mu_n$ *e* $Var(X_n) = \sigma_n^2$*, onde* $\sigma_n^2 < \infty$ *e pelo menos um* $\sigma_n^2 > 0$*. Suponha que*

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2} = 0.$$

Se

$$\frac{S_n - E(S_n)}{s_n} \xrightarrow{D} Z \sim \mathcal{N}(0,1)$$

então, para todo $\epsilon > 0$ *vale que*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 \, dF_k(x) = 0.$$

TCL de i.i.d. como corolário

Podemos obter o Teorema 26.1 do Teorema 26.2: Vamos verificar que vale a condição de Lindeberg. De fato, como $s_n^2 = n\sigma^2$, para $\epsilon > 0$, temos que

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2\,dF_k(x)$$

é igual a

$$\frac{1}{\sigma^2}\int_{|x-\mu|\leq\epsilon\sigma\sqrt{n}}(x-\mu)^2\,dF(x) = \frac{1}{\sigma^2}\int_{\mu-\epsilon\sigma\sqrt{n}}^{\mu+\epsilon\sigma\sqrt{n}}(x-\mu)^2\,dF(x)$$

mas

$$\lim_{n\to\infty}\frac{1}{\sigma^2}\int_{\mu-\epsilon\sigma\sqrt{n}}^{\mu+\epsilon\sigma\sqrt{n}}(x-\mu)^2\,dF(x) = \frac{1}{\sigma^2}\int_{-\infty}^{+\infty}(x-\mu)^2\,dF(x) = \frac{\sigma^2}{\sigma^2} = 1.$$

Bom estudo!