Politechnika Wrocławska

Warunki Lindeberga i Lapunowa

Jakub Koral

Wrocław, 18 listopada 2021

HR EXCELLENCE IN RESEARCH

Politechnika Wrocławska

Warunek Lindeberga

Weźmy ciąg **niezależnych** zmiennych losowych $\{X_k\}_{k=1}^{n}$
o skończonych wartościach oczekiwanych $\mathbb{E}X_k = \mu_k$
i wariancjach $\operatorname{Var} X_k = \sigma_k^2$. Zdefiniujmy teraz $s_n^2 := \sum_{k=1}^{n} \sigma_k^2$.
Taki ciąg spełnia **warunek Lindeberga**, gdy:

$$\forall \epsilon > 0 : \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[|X_k - \mu_k|^2 \cdot \mathbb{1}_{|X_k - \mu_k| > \epsilon s_n} \right] = 0. \quad (1)$$

Warunek Lindeberga - o co chodzi?

Intuicyjnie rzecz biorąc w warunku Lindeberga chodzi o to, żeby zmienne losowe były "równie małe". Dla dowolnie wybranego małego $\epsilon > 0$ wkład scentrowanych zmiennych losowych większych co do modułu od ϵs_n staje się w sumowaniu pomijalny. Mówiąc precyzyjniej, wkład wariancji zmiennych losowych staje się pomijalny.

Wnioski z warunku Lindeberga (1)

Uprośćmy warunek Lindeberga biorąc $\mu_k = 0$ i $\lim_{n\to\infty} s_n^2 = 1$. Wtedy otrzymujemy następujące wnioski:

$$\max_{k\in 1,\dots,n} \mathbb{E}X_k^2 \leqslant \epsilon^2 + \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right]. \tag{2}$$

Dowód (2).

$\mathbb{E}X_k^2 = \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|\leqslant\epsilon}\right] + \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right] \leqslant \epsilon^2 + \sum_{k=1}^n \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right]$ i (2) z niezależności X_k. $\square$

Wnioski z warunku Lindeberga (2)

Zatem:

$$\lim_{n\to\infty} \max_{k\in 1,\dots,n} \mathbb{E}X_k^2 = 0. \tag{3}$$

Z nierówności Czebyszewa dostajemy też:

$$P(|X_k| > \epsilon) \leqslant \frac{\mathbb{E}X_k^2}{\epsilon^2}. \tag{4}$$

Wynika z tego ważny wniosek:

$$\forall \epsilon > 0 : \lim_{n\to\infty} \max_{k\in 1,\dots,n} P(|X_k| > \epsilon) = 0. \tag{5}$$

Własność ta nazywa się **jednostajną asymptotyczną pomijalnością**. Ciąg z tą własnością oznaczna się najczęściej skrótem *u.a.n.* (ang. *uniformly asymptotically negligible*). Własność ta ma związek z nieskończoną podzielnością rozkładów.

CTG Lindeberga

Załóżmy, że ciąg $\{X_k\}_{k=1}^{n}$ spełnia warunek Lindeberga (1), wtedy zachodzi **Centralne Twierdzenie Graniczne Lindeberga**:

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow{d} \mathcal{N}(0,1). \tag{6}$$

Nierówność Berry'ego-Esseena

Oznaczmy dystrybuanty unormowanej sumy zmiennych losowych jako $F_n(x) = P\left(\frac{1}{s_n}\sum_{k=1}^{n}(X_k - \mu_k) \leqslant x\right)$, a zmiennej losowej z rozkładu normalnego jako $\Phi(x)$.
Carl-Gustav Esseen pokazał, że:

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leqslant C_0 \frac{1}{s_n^3} \sum_{k=1}^{n} \mathbb{E}|X_k - \mu_k|^3, \qquad (7)$$

gdzie C_0 to pewna stała. Najlepsze oszacowanie to obecnie:
$C_0 \in \left[\frac{\sqrt{10}+3}{6\sqrt{2\pi}}, 0.5600\right]$.

Kiedy (1) jest spełnione?

1. Ciąg $\{X_k\}_{k=1}^n$ jest *iid*:

Dowód.

$\lim_{n\to\infty} \frac{1}{n\sigma^2} \sum_{k=1}^n \mathbb{E}\left[|X_k - \mu|^2 \cdot \mathbb{1}_{|X_k-\mu|\leqslant\epsilon\sigma\sqrt{n}}\right]=0$
i dostajemy CTG Lévy'ego-Lindeberga. □

2. $\lim_{n\to\infty} s_n^2 = \lim_{n\to\infty} \sum_{k=1}^n \sigma_k^2 = \infty$:

Dowód.

$\lim_{n\to\infty} \frac{1}{s_n^2} = 0$ i warunek (1) natychmiast spełniony. □

3. Ciąg $\{X_k\}_{k=1}^n$ ma własności (5) i (6) (twierdzenie Fellera)

Dowód - szkic.

$\lim_{n\to\infty} \max_{k\in 1,\ldots,n} P(|X_k| > \epsilon) \leqslant \lim_{n\to\infty} \max_{k\in 1,\ldots,n} \frac{\mathbb{E}X_k^2}{\epsilon^2} \leqslant$
$\lim_{n\to\infty} 1 + \sum_{k=1}^n \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right]$. □

Warunek Lapunowa

Weźmy ciąg **niezależnych** zmiennych losowych $\{X_k\}_{k=1}^n$ o skończonych wartościach oczekiwanych $\mathbb{E}X_k = \mu_k$ i wariancjach $\operatorname{Var} X_k = \sigma_k^2$. Zdefiniujmy teraz $s_n^2 := \sum_{k=1}^n \sigma_k^2$. Taki ciąg spełnia **warunek Lapunowa**, gdy:

$$\exists\delta > 0 : \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k - \mu_k|^{2+\delta}\right] = 0. \qquad (8)$$

Oczywiście istnieje **Centralne Twierdzenie Graniczne Lapunowa**, która ma taką samą postać jak CTG Lindeberga, tylko zmienne losowe spełniają warunek (8).

Warunek Lapunowa⇒Lindeberga

Dowód.

Dowód dla $\mu_k = 0$ i $\lim_{n\to\infty} s_n^2 = 1$. Weźmy $\epsilon, \delta > 0$, wtedy dla każdej zmiennej losowej $|X_k| < \epsilon$:

$$X_k^2 = \frac{|X_k|^{2+\delta}}{|X_k|^\delta} \leqslant \frac{|X_k|^{2+\delta}}{\epsilon^\delta}; \tag{9}$$

$$\mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right] \leqslant \frac{\mathbb{E}\left[|X_k|^{2+\delta}\right]}{\epsilon^\delta}; \tag{10}$$

$$\sum_{k=1}^n \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right] \leqslant \frac{\sum_{k=1}^n \mathbb{E}\left[|X_k|^{2+\delta}\right]}{\epsilon^\delta}; \tag{11}$$

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right] \leqslant \frac{\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[|X_k|^{2+\delta}\right]}{\epsilon^\delta} = 0. \tag{12}$$

$\square$

Warunek Lapunowa $\not\Leftarrow$ Lindeberga

$$P(X = i) = \frac{c}{|i|^3 \log^2 |i|}, \quad i \in \mathbb{Z} \backslash \{-1, 0, 1\}. \tag{13}$$

Zauważmy, że z symetrii $\mathbb{E}X = 0$ oraz $\mathrm{Var}\, X = \mathbb{E}X^2 = \sigma^2 < \infty$. Jednak $\forall \delta > 0 : \mathbb{E}|X|^{2+\delta} = \infty$ i warunek Lapunowa **nie jest** spełniony. Mimo to spełniony jest warunek Lindeberga:

$$\begin{aligned} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^2 \cdot \mathbb{1}_{|X_k| > \epsilon s_n} \right] &= \frac{n}{n\sigma^2} \sum_{|i| > \epsilon\sqrt{n}\sigma} i^2 \frac{1}{|i|^3 \log^2 |i|} < \\ &< c \int_{c'\sqrt{n}}^{\infty} \frac{\mathrm{d}\, z}{z \log^2 z} = \int_{c'' \log n}^{\infty} \frac{\mathrm{d}\, u}{u^2} = \frac{1}{c'' \log n} \xrightarrow[n \to \infty]{} 0. \end{aligned} \tag{14}$$

Zatem warunek Lindeberga **jest** spełniony.

Dowód warunku Lindeberga - szkic

Dowód (1).

Weź Z_k z rozkładu normalnego z parametrami równymi X_k

$$\begin{aligned} S :=S_0 &:= X_1 + X_2 + \ldots + X_n, \\ S_1 &:= Z_1 + X_2 + \ldots + X_n, \\ &\vdots \\ T :=S_n &:= Z_1 + Z_2 + \ldots + Z_n. \end{aligned} \tag{15}$$

Wybierz funkcje f takie, że $|f^{(i)}| \leqslant K, i \in \{0,1,2,3\}$. Rozwiń przy pomocy szeregu Taylora $|\mathbb{E}f(S) - \mathbb{E}f(T)|$. Pokaż zbieżność do 0:

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leqslant K \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^2 \cdot \mathbb{1}_{|X_k|>\epsilon}\right] + \frac{K}{6}\epsilon\sigma^2 + \frac{cK}{6}\sum_{k=1}^{n} \mathbb{E}X_k^3. \tag{16}$$

□

Źródła

- Jim Pitman, *Lecture 10: Setup for the Central Limit Theorem* (Dowód warunku Lindeberga - link)
- Hongyi Liu, *Lecture 10: LLN, CLT and Local Linear Approximation* (link)
- Patrick Breheny, *Lindeberg-Feller central limit theorem* (link)
- Adam Osękowski, *Wykład z Rachunku Prawdopodobieństwa II: Centralne Twierdzenie Graniczne* (link)