Probability Theory

Chapter 4

Central limit theorem.

Lecturer: Nataliia Kruglova.

Лекція 14. Центральна гранична теорема.

Поняття про центральну граничну теорему

Центральними граничними теоремами (ЦГТ) називають групу граничних теорем, які встановлюють зв'язок між законом розподілу суми випадкових величин і нормальним законом розподілу.

Сформулюємо ЦГТ для випадку суми незалежних однаково розподілених випадкових величин (саме такий варіант ЦГТ є найпростішим і найчастіше використовується на практиці).

Теорема 14.1. *Нехай випадкові величини* $X_1, X_2, \ldots, X_n$ *незалежні, однаково розподілені, мають скінченне математичне сподівання* $E[X_i] = a$ *і дисперсію* $D(X_i) = \sigma^2, i = \overline{1,n}$. *Тоді функція розподілу центрованої і нормованої суми цих випадкових величин прямує при* $n \to \infty$ *до функції розподілу стандартної нормальної випадкової величини:*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]}{\sqrt{D\left(\sum_{i=1}^{n} X_i\right)}},$$

$$\lim_{n\to\infty} F_{Z_n}(x) = \lim_{n\to\infty} P\{Z_n < x\} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx. \tag{55}$$

З цієї теореми випливає, що при достатньо великому значенні n сума Z_n наближено розподілена за нормальним законом. А звідси випливає, що $\sum_{i=1}^{n} X_i$ наближено розподілена по нормальному закону з параметрами $(na, n\sigma^2)$. Говорять, що випадкова величина $\sum_{i=1}^{n} X_i$ асимптотично нормальна при $n \to \infty$.

Формула (55) дозволяє для достатньо великих n обчислювати ймовірності деяких подій, пов'язаних з сумами випадкових величин. Для цього потрібно перейти від суми до центрованої нормованої випадкової величини:

$$P\left\{\alpha \le \sum_{i=1}^{n} X_i \le \beta\right\} = P\left\{\frac{\alpha - na}{\sigma\sqrt{n}} \le \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}} \le \frac{\beta - na}{\sigma\sqrt{n}}\right\} \approx$$

$$\approx \Phi\left(\frac{\beta - na}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{\alpha - na}{\sigma\sqrt{n}}\right).$$

Приклад. Незалежні випадкові величини $X_1, X_2, \dots, X_n$ розподілені рівномірно на відрізку $[0, 2]$. Знайдіть закон розподілу випадкової величини $Y = \sum_{i=1}^{120} X_i$ і обчисліть ймовірність $P\{70 < Y < 110\}$.

Розв'язання. Обчислимо математичне сподівання і дисперсію випадкових величин X_i: $E[X_i] = \frac{2+0}{2} = 1$, $D(X_i) = \frac{(2-0)^2}{12} = \frac{1}{3}, i = \overline{1, n}$. Умови центральної граничної теореми виконуються, а тому:

$$E[Y] = E\left[\sum_{i=1}^{120} X_i\right] = 120 \cdot 1 = 120.$$

$$D(Y) = D\left(\sum_{i=1}^{120} X_i\right) = 120 \cdot \frac{1}{3} = 40.$$

$$f_Y(y) \approx \frac{e^{-\frac{(x-120)^2}{80}}}{\sqrt{80\pi}}$$

Тоді

$$P\{70 < Y < 110\} \approx \int_{70}^{110} \frac{e^{-\frac{(x-120)^2}{80}} dx}{\sqrt{80\pi}} = \Phi\left(\frac{110 - 120}{\sqrt{40}}\right) - \Phi\left(\frac{70 - 120}{\sqrt{40}}\right) \approx$$

$$= \Phi(-1,581) - \Phi(-7,9) = 0.057.$$

Приклад. Студент отримує на іспиті 5 з імовірністю 0,2; 4 – з імовірністю 0,4; 3 – з імовірністю 0,3 і 2 – з імовірністю 0,1. За час навчання він здає 40 іспитів. Знайти межі, в яких з ймовірністю 0,95 лежить середній бал студента.

Розв'язання. Позначимо через X_i, – бал отриманий на i-тому іспиті, $i = \overline{1, 40}$. Ці випадкові величини – незалежні і однаково розподілені. Зведемо розподіл цих випадкових величин у таблицю:

X_i	2	3	4	5
p	0.1	0.3	0.4	0.2

Обчислимо математичне сподівання і дисперсію випадкової величини X_1.

$$E[X_1] = 2 \cdot 0.1 + 3 \cdot 0.3 + 4 \cdot 0.4 + 5 \cdot 0.2 = 3.7,$$

$$D(X_1) = (2-3.7)^2 \cdot 0.1 + (3-3.7)^2 \cdot 0.3 + (4-3.7)^2 \cdot 0.4 + (5-3.7)^2 \cdot 0.2 = 0.81.$$

Позначимо через Y – середній бал студента після всіх іспитів. Тоді $Y = \frac{\sum_{i=1}^{40} X_i}{40}$. Знайдемо числові характеристики цієї випадкової величини, враховуючи властивості математичного сподівання і дисперсії.

$$E[Y] = \frac{E[X_1] + \ldots + E[X_n]}{40} = E[X_1] = 3.7,$$

$$D(Y) = D\left(\frac{X_1 + \ldots + X_n}{40}\right) = \frac{D(X_1)}{40} \approx 0.02.$$

Тоді

$$0.95 = P\left\{\left|\frac{Y - E[Y]}{\sqrt{D(y)}}\right| \le \varepsilon\right\} = 2\Phi(\varepsilon).$$

З таблиці для нормального розподілу:

$$\left|\frac{Y - 3.7}{\sqrt{0.02}}\right| \le 1.96.$$

$$3.42 \le Y \le 3.98.$$

Теорема Муавра-Лапласа.

Повернемось до інтегральної теореми Муавра-Лапласа, згідно з якою для кількості успіхів в n незалежних випробуваннях Бернуллі μ_n з імовірністю успіху p $(0 < p < 1)$ виконується:

$$\lim_{n\to\infty} P\left\{\frac{\mu_n - np}{\sqrt{npq}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

Тобто, функції розподілу випадкових величин $\frac{\mu_n - np}{\sqrt{npq}}$ прямують до функції розподілу стандартного нормального закону. Це можна сформулювати іншим чином.

Введемо випадкові величини X_k, які приймають значення 1, якщо у k-тому випробування випав успіх та 0, якщо - невдача з тими ж самими ймовірністями. Тоді $\mu_n = X_1 + X_2 + \ldots + X_n$. При цьому $E[X_k] = p$, $D(X_k) = pq$,

$$E[\mu_n] = np, \qquad D(\mu_n) = npq.$$

Тоді інтегральну теорему Муавра-Лапласа можна сформулювати наступним чином.

Теорема 14.2. *(Центральна гранична теорема) Якщо кожна з випадкових величин* $X_1, X_2, \ldots, X_n$ *може набувати лише двох значень 1 або 0 з імовірностями* p *та* q $(p+q=1,\ 0<p<1)$*, то при нескінченному зростанні* n *функція розподілу випадкової величини*

$$\frac{\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]}{\sqrt{D\left(\sum_{i=1}^{n}\right)}}$$

прямує до функції розподілу стандартного нормального закону і для $\mu_n = X_1 + X_2 + \ldots + X_n$*:*

$$P_n\{k_1 \le \mu_n \le k_2\} \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right).$$

Доведення.

$$E[\mu_n] = np, \qquad D(\mu_n) = npq.$$

$$Z_n = \frac{\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]}{\sqrt{D\left(\sum_{i=1}^{n}\right)}} = \frac{\mu_n - np}{\sqrt{npq}}.$$

Випадкова величина Z_n являється сумою n незалежних, однаково розподілених випадкових величин. Але $Z_n \sim N(0,1)$, оскільки

$$E[Z_n] = E\left[\frac{\mu_n - np}{\sqrt{npq}}\right] = \frac{1}{\sqrt{npq}} \cdot (np - np) = 0.$$

$$D(Z_n) = D\left(\frac{\mu_n - np}{\sqrt{npq}}\right) = \frac{1}{npq} \cdot D(\mu_n) = \frac{npq}{npq} = 1.$$

За ЦГТ випадкова величина Z_n при достатньо великих n наближено має нормальний розподіл. Оскільки

$$P\{\alpha \le Z_n \le \beta\} \approx \Phi(\beta) - \Phi(\alpha),$$

то, поклавши:

$$\beta = \frac{k_2 - np}{\sqrt{npq}}, \qquad \alpha = \frac{k_1 - np}{\sqrt{npq}}.$$

Нерівність $\frac{k_1 - np}{\sqrt{npq}} \le \frac{\mu_n - np}{\sqrt{npq}} \le \frac{k_2 - np}{\sqrt{npq}}$ можна переписати у вигляді: $k_1 \le \mu_n \le k_2$, а тому

$$P_n\{k_1 \le \mu_n \le k_2\} \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right).$$

Приклад. Імовірність того, що тест на COVID-19 дає неправильний результат, дорівнює 0.02. В лабораторії за день зробили 7000 тестів. Яка ймовірність того, що буде не більше 120 помилкових результатів?

Розв'язання. $p = 0.02$, $q = 1 - p = 1 - 0.02 = 0.98$, $n = 7000$. $np = 140$, $\frac{1}{\sqrt{npq}} = \frac{1}{\sqrt{140 \cdot 0.98}} \approx 0.085$, $\frac{m-np}{\sqrt{npq}} = \frac{120-140}{\sqrt{140 \cdot 0.98}} \approx -1.7$.

$$P\{\mu_n \leq 120\} \approx \Phi(-1.7) \approx 0.0446.$$

Приклад. Обчисліть ймовірність того, що число випадінь одиниці при 12000 підкиданнях гральної кістки знаходиться між 1900 і 2150.

Розв'язання. $p = \frac{1}{6}$, $q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$, $n = 12000$. $np = 2000$.

$$P\{1900 \leq \mu_n \leq 2150\} \approx \Phi\left(\frac{2150 - 2000}{\sqrt{2000 \cdot \frac{5}{6}}}\right) - \Phi\left(\frac{1900 - 2000}{\sqrt{2000 \cdot \frac{5}{6}}}\right) \approx 0.993.$$

Теорема Ляпунова.

Теорема 14.3. *Нехай взаємно незалежні випадкові величини $X_1, X_2, \ldots, X_n$ мають скінченний третій абсолютний момент. Покладемо*

$$a_n = E[X_n], \quad b_n^2 = D[X_n]), \quad c_n^3 = E[|X_n - a_n|^3],$$

$$A_n = \sum_{i=1}^{n} a_i, \quad B_n^2 = \sum_{i=1}^{n} b_i^2, \quad C_n^3 = \sum_{i=1}^{n} c_i^3.$$

Якщо

$$\lim_{n \to \infty} \frac{C_n}{B_n} = 0,$$

то при нескінченному зростанні n функція розподілу випадкової величини

$$\frac{\sum_{i=1}^{n} X_i - A_n}{B_n}$$

прямує до функції розподілу стандартного нормального розподілу.

Приклад. Нехай випадкові величини X_n мають рівномірний на $[a_n - 1; a_n + 1]$ розподіл, і $\sum_{i=1}^{\infty} |a_i| = a < \infty$. Обчисліть.

$$\lim_{n \to \infty} P\left\{0 < \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} < 1\right\}.$$

Розв'язання.

$$f_{X_n}(x) = \begin{cases} \frac{1}{2}, & x \in [a_n - 1; a_n + 1], \\ 0, & x \notin [a_n - 1; a_n + 1]. \end{cases}$$

$$E[X_n] = \frac{a_n + 1 + a_n - 1}{2} = a_n, \qquad D(X_n) = \frac{(a_n + 1 - a_n + 1)^2}{12} = \frac{1}{3} = b_n^2.$$

$$c_n^3 = \int_{a_n-1}^{a_n+1} |x - a_n|^3 dx = |x - a_n = t| = \int_{-1}^{1} |t|^3 dt = 2\int_0^1 t^3 dt = \frac{1}{2}.$$

$$A_n = \sum_{i=1}^{n} a_i, \quad B_n^2 = \frac{n}{3}, \quad C_n^3 = \frac{n}{2}$$

.

$$\lim_{n\to\infty} \frac{C_n}{B_n} = \lim_{n\to\infty} \frac{\sqrt[3]{n}\sqrt{3}}{\sqrt[3]{2}\sqrt{n}} = 0.$$

А тому за теоремою Ляпунова:

$$\lim_{n\to\infty} P\left\{0 < \frac{\sum_{i=1}^n X_i}{\sqrt{n}} < 1\right\} = \lim_{n\to\infty} P\left\{-\frac{-A_n\sqrt{3}}{\sqrt{n}} < \frac{\sum_{i=1}^n X_i - A_n}{\sqrt{\frac{n}{3}}} < \left(1 - \frac{A_n}{\sqrt{n}}\right)\sqrt{3}\right\} \approx$$

$$\approx \lim_{n\to\infty} \left(\Phi\left(\left(1 - \frac{A_n}{\sqrt{n}}\right)\sqrt{3}\right) - \Phi\left(-\frac{A_n}{\sqrt{n}}\sqrt{3}\right)\right) = \Phi(\sqrt{3}) - \Phi(0) \approx 0.458.$$

Література

1. Теорія ймовірностей : конспект лекцій, В.В. Булдигін, Ю.П. Буценко, О.О. Диховичний/ Київ, 1999, 97с.

2. Теорія ймовірностей та математична статистика: навч. посіб., Н. Д. Федоренко, О. І. Баліна, І. С. Безклубенко/К.: КНУБА, 2007, 104 с.

3. Теорія ймовірностей, математична статистика та імовірнісні процеси : навч. посіб., Ю. М. Слюсарчук, Й. Я. Хром'як, Л. Л. Джавала, В. М. Цимбал/ Львів : Вид-во Львів. політехніки, 2015, 364 с.

4. Курс теорії ймовірностей: навч. посіб., Б. В. Гнєденко/ К.: ВПЦ Київський університет, 2010, 464 с.

5. Теория вероятностей и математическая статистика. Изд.9. Учебное пособие для вузов, В. Е. Гмурман/М: Высшая школа, 2003, 479с.